AD NUMBER: AD0869234
LIMITATION CHANGES
TO: Approved for public release; distribution is unlimited.
FROM: Distribution authorized to U.S. Gov't. agencies and their contractors; Export Control; 1 Apr 1970. Other requests shall be referred to the Air Force Avionics Lab (AVWW), Wright-Patterson AFB, OH 45433.
AUTHORITY
AFAL LTR, 23 JUN 1975

AFAL-TR-70-39

SIMULTANEOUS DETECTION AND ESTIMATION UNDER MULTIPLE HYPOTHESES

A. J. Fredriksen
The Johns Hopkins University
Carlyle Barton Laboratory

TECHNICAL REPORT AFAL-TR-70-39

April 1970

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Air Force Avionics Laboratory (AVWW).

Air Force Avionics Laboratory
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

65

NOTICES

SIMULTANEOUS DETECTION AND ESTIMATION UNDER MULTIPLE HYPOTHESES

A. J. Fredriksen

FOREWORD

This report was prepared by the Carlyle Barton Laboratory of The Johns Hopkins University, Baltimore, Maryland under United States Air Force Contract F33615-69-C-1317 administered by the Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio. Mr. Charles W. Ambuske (AVWW) is the Project Engineer. The research was carried out during the period February through September 1969.

The author gratefully acknowledges the assistance of Professor D. V. Vandelinde of the Electrical Engineering Department, D. Middleton consultant to Carlyle Barton Laboratory, and W. K. Sones formerly of Carlyle Barton and now at Comsat Laboratories. Also, thanks are due Dr. R. S. Bennett and Dr. P. J. Butterly for numerous enlightening discussions.

In addition the author wishes to thank Mr. C. M. Laporte for preparing the illustrations, Mrs. Mary Johnson and Mrs. Dolores Sullivan for typing the report.

The contractor's report number is AF-160.

This report was submitted by the Author on January 15, 1970.

This technical report has been reviewed and is approved.

Ronald G. Stimmel

for Joseph A. Dombrowski
Lt. Colonel, USAF
Chief, Electronics Warfare Division

ABSTRACT

This report treats the problem of simultaneous detection and estimation under multiple hypotheses when data from only one observation interval is available. The analysis is based on statistical decision theory.

In the past, parameter estimation was always performed under the assumption that the desired signal was present with probability one. Since detection is meaningful only when some uncertainty exists as to the presence or absence of the desired signal it is apparent that classical estimation theory must be modified if the two operations are to be performed simultaneously. In addition, this report considers the case when the operations of detection and estimation are coupled.

In this report specific detector and estimator structures are determined for the case where the two operations are strongly coupled and where the cost of estimation is given by a quadratic cost function. It is found that the detector structures are complex nonlinear functions of the received data, but nevertheless the one case considered in detail resulted in a type of correlation detector, where the basic operation is correlation of the received data with the various least squares estimators of the possible signals in the absence of uncertainty. The associated optimum estimator structures are determined for this case, and found to be weighted sums of the various least squares estimators in the absence of uncertainty. Finally, joint detection and estimation under multiple hypotheses is discussed for the case of a simple cost function assignment. It is shown that the estimators which result in this case may be interpreted as generalized maximum likelihood estimators.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS

Roman Symbols:

$A_o(X)$	Average a posteriori cost of deciding no signal is present defined by relation (3.9a).
$A_i^{(2)}$	A constant cost factor appearing in the expression for the simple cost function, relation (3.41a).
b_i	Bias factor.
$b_j(\underline{X})$	A function of the observations which determined the detector structure and is defined by relation (3.12a).
$b_j'(\underline{X})$	Similar to $B_j(\underline{X})$ and defined by (3.9b).
C	Cost coefficient.
$C(\underline{\theta}, \gamma)$	Cost of deciding γ when $\underline{\theta}$ is the true state of nature.
$D_j(\underline{X})$	A function of the observations which embodies the detector structure in the simple cost function case.
$G(\underline{X};\underline{m}, k)$	Multivariate Gaussian probability density function of the vector random variable $\underline{X}$, with mean $\underline{m}$ and covariance matrix k.
H_i	Hypothesis that a signal of class Ω_i is present.
$I_{\Omega_i}(\cdot)$	Indicator function for the set Ω_i.
i, j, k	Integer indexing variables.
$\underline{X}$	Received data vector.
K_n	Covariance matrix of additive Gaussian noise.
K, K_i, K_{ji}	Covariance matricies of Gaussian densities.
$L_{ij}(\underline{X})$	Modified likelihood ratio defined by relation (3.10).
M	Number of members in the signal alphabet or the message set.

m_j	Generic element of the message set.
$\underline{m}_i$	Mean value vector of a multivariate Gaussian probability density.
N	Dimensionality of our signal space and the observation space.
n(t)	Realization of an additive noise process.
$P(\cdot)$	Probability density function.
P_i	Prior probability of occurrence of hypothesis H_i.
R_{D+E}	Average risk associated with the operation of simultaneous detection and estimation.
$\underline{S}$	Signal waveform vector.
$\hat{\underline{S}}(\underline{X})$	Signal waveform estimator.
T	Duration of an observation interval. When it appears as a superscript it denotes transpose.
W	Signal bandwidth.

<u>Greek symbols</u>:

γ	Generic element of the decision space.
Γ	Observation space.
$\delta(\cdot)$	Dirac δ-function.
Δ	Decision space.
$\underline{\theta}$	Signal parameter vector.
Θ	Signal parameter space.
$\hat{\underline{\theta}}(\underline{X})$	Estimator of $\underline{\theta}$.
$\Lambda(\underline{X})$	Generalized likelihood ratio.
Ω	Signal space or signal parameter space.
Ω_j	Subset of Ω.

I. GENERAL INTRODUCTION

Optimum detection of a known signal corrupted by noise of known statistical characteristics has been treated by many authors from the viewpoint of statistical decision theory (1), (2). Many of the results may be applied to data processing procedures, and applications have been made to radar and communications (2), (5). A critical assumption in this existing theory has been that the estimation was performed using data known to contain the assigned signal with complete certainty. In many practical situations this is not the case, and a paper by Middleton and Esposito (4) considers the problem of joint detection and estimation, and the modifications to the estimation procedure that are necessary when the assumption that the signal is present during the observation interval with probability one is relaxed. Also, a recent paper by Nahi (15) investigates the problem of determining the form of optimum least mean square recursive linear estimators in the presence of uncertainty. In addition to these investigations, the past decade has produced many studies concerning the theory of learning machines both of the supervised (6) and unsupervised variety (7), (8).

The purpose of this report is to initiate a study of the data sorting problem which results when the reception situation consists of repetitive, pulsed signals whose waveforms are initially unknown at the receiver. It seems reasonable that optimum reception in such a situation will very likely involve adaptive methods, i.e., a receiver which is capable of modifying its own structure in accordance with what is learned of the signal parameters actually used in the transmitted signal set. Ideally an adaptive receiver would collect data during one observation interval and process it so as to make two optimal decisions; one a detection the other an estimation decision. The detection decision determines which signal class the observation is assigned, where the class consisting of the null signal accounts for the case of noise alone. If it is decided that something other than noise is present then an estimate is made of the relevant signal parameters. This estimate is then used to update the structure (future data processing procedure) of the receiver so as to improve its future performance. The receiver's adaptivity is to be associated with this updating procedure. The adaptive receiver just described makes binary decisions and necessarily has to estimate in the presence of uncertainty, i. e., it is not certain whether or not the signal of interest is present, also the two operations of detection and estimation are interrelated.

Any attempt to solve a realistic data sorting problem must treat the multiple hypotheses reception situation, i. e., where any one of several signals may be present during a given observation interval. Also, before a general method for treating the sequential operation of adaptive receivers can be developed the problem of joint detection and estimation under multiple hypotheses for one observation interval has to be investigated. It is this aspect of the general data sorting problem which is investigated in this report, namely, the problem of joint detection and estimation under multiple hypotheses, when data from only one observation interval is available. Extensions of the concept of simultaneous detection and estimation to the case when data is received sequentially will be investigated in a future report, however, some aspects of the extension to "multi-decision" simultaneous detection and estimation have been investigated in (16).

The results of this report show that in general nonlinear estimator and detector structures result when the two operations are related and estimation is carried out in the presence of uncertainty. However, as the capabilities of digital computers increase, e.g., as they become faster, more economical and have larger memory capacities the possibility arises of performing complex data processing operations as a matter of routine (17). Therefore it may not be necessary to exclude, on the grounds of impracticality, complex data processing procedures which are likely to result from the solution of any realistic data sorting problem.

A glance at the table of contents will give a good indication of the topics covered in this report. Section 3 is concerned with a treatment of simultaneous detection and estimation under multiple hypotheses. Examples are considered where the costs of estimation are given by the simple cost function (SCF) and the quadratic cost function (QCF). Also, special cases are considered and comparisons made with published results.

2. BACKGROUND AND STATEMENT OF THE PROBLEM

This section will provide some of the background and motivation for the topics studied in this report. As mentioned in the introduction, one of our principal objectives will be an investigation of adaptive receivers, and therefore we begin this section with a discussion of such devices.

An adaptive receiver is a receiver capable of modifying its structure so as to optimally receive either known or unknown signals in a changing or initially unknown environment. The limited amount of information at the receiver concerning the environment in which it operates is the result of either not knowing the characteristics of the signal source or a lack of information about the nature of the random disturbances introduced by the communication channel. For example, characteristics of the signal source which may not be known are:

(1) The waveform set used by the transmitter to encode the symbols of the message set.

(2) The prior probabilities of occurrence of the symbols of the message set, or equivalently the prior probabilities of the various signal waveforms used.

Continuing, we may list the following as examples of the kind of information that may be lacking concerning the communication channel:

(3) The noise introduced by the channel may be stationary but have unknown statistics, or the statistics may be known but the noise may be nonstationary and its temporal characteristics unknown.

(4) An adequate channel model may exist, but channel parameters such as the channel gain and delay time may not be known.

A solution to the general problem of optimum adaptive receiver design when all the above uncertainties are present simultaneously appears unlikely at present. It therefore seems reasonable to limit the type of uncertainty present and determine its effect on the structure of the optimum adaptive receiver. In this report we will only consider situations where the predominant uncertainties are of the form (1) and (4) because they are of general interest and both can be investigated by the same formalism. Accordingly, we shall assume that additive, stationary noise with known statistics is present and that the prior probabilities of occurrence of the various message symbols are given.

This paragraph is an elaboration on the comment made previously that the two types of uncertainties, (1) and (4), may be treated by the same formalism. If situation (1) prevails, the signal waveforms are not known precisely. This may involve a partial lack of information, perhaps only one parameter out of several needed to describe the signal waveform is uncertain, or possibly a total lack of information of the transmitted waveform, in which case 2WT sample values would suffice to describe a signal of bandwidth W which is essentially time limited to an interval of duration T. In the case of a total lack of information, which involves the largest number of unknown parameters, the parameter vector $\underline{\theta}_j$ might represent the sample values of the jth signal, $S_j(t;\underline{\theta}_j)$. Cases which involve fewer signal parameters are of course included in this formulation. For example, if the normalized waveforms, $s_j(t)$, used by the transmitter are known at the receiver and the only unknown parameter is an amplitude factor, then $\underline{\theta}_j$ reduces to a one component vector, namely the amplitude A_j of the signal $S_j(t;\underline{\theta}_j) = A_j s_j(t)$. The case of known signal waveforms but unknown channel parameters is handled by letting $\underline{\theta}_1 = \underline{\theta}_2 = \ldots \underline{\theta}_M = \underline{\theta}$, where $\underline{\theta}$ is a vector whose components are the unknown channel parameters. For example, if A is an unknown channel gain and τ the delay time introduced by the channel then $\underline{\theta}^T = (A, \tau)$, where superscript T indicates transpose.

In order to adapt to a changing or unknown environment, an adaptive receiver must attempt to learn certain features of the channel or the signal source and then incorporate this acquired knowledge into its own structure. This last statement suggests that all adaptive receivers share two important properties. First, some sort of learning feature must be present and second, a performance sensing device must exist which compares the present performance with the performance which is optimal in light of the newly acquired information made available by the learning feature. Presumably the output of the performance sensing device is used to appropriately adjust the structure of the adaptive receiver to bring its subsequent performance more in line with that of the optimum receiver in the absence of uncertainty.

If receiver adjustments are to be made so that optimum reception is approached, it is necessary to know the form of the optimum receiver in the absence of uncertainty. The structural form of optimum data processors in the absence of uncertainty can be determined from decision theory (2), (5). For example, it is well known that Bayes tests for the optimum detection of signals in additive noise whose statistics are known, is based upon the generalized likelihood ratio, which in the case of stationary, white, Gaussian noise with a known signal set reduces to tests based on cross-correlation or matched filtering, that is, the incoming data is cross correlated with stored replicas of the known signal waveforms and if the various costs of misclassification are all equal, then the signal corresponding to the correlator with the largest output is chosen to be the one which was transmitted. So if the same environmental conditions exist, except that the signal set being used is unknown, it appears reasonable that an adaptive receiver must attempt to learn the transmitted waveforms if it is to approach the performance of the optimum detector in the known signal case.

How does the adaptive receiver learn the signal waveform ? One rather obvious approach, known as decision-directed measurement, which may be used when no classified learning sequence exists, is to assume that the detector operation is error free, and that whenever the detector decides a signal is present an estimate of the unknown signal parameter is accepted and used to update the detector and estimator structure. Since we have assumed error free operation of the detector, the form of the estimator is that dictated by the usual estimation theory which assumes that the signal of interest is present with probability one. The assumption that the probability of the detector making an error is zero is reasonable provided the signal-to-noise ratio is high and the associated probability of error is small. However, in situations where the signal-to-noise ratio is high there is little need of the refinements in data processing provided by an optimum theory of reception. It is when the signal-to-noise ratio is small, and there is a fair degree of uncertainty as to the presence or absence of a signal, that an optimum theory of signal reception is most useful.

It therefore is apparent, that a theory which will provide the optimum structure of the learning feature part of an adaptive receiver even in the case of low signal-to-noise ratios, will involve several modifications in the existing theory of detection and estimation. First of all, estimation now has to be carried out in the presence of uncertainty, and the classical Bayes estimators have to be modified to reflect the fact that the signal may not be present with probability one. Second, if we picture a detector-directed estimator scheme as the basis of a learning device, as mentioned previously, then it is possible that the detector may decide that a signal is present when indeed one is not present and consequently false estimates of the signal may be accepted. Hence the two operations of detection and estimation should be coupled and the costs associated with the usual detection process modified to reflect the interaction between the two operations. How these modifications are to be achieved in the "single-shot" binary "on-off" case has been explored by Middleton and Esposito (4) and more recently extended to include multiple decision processes (16), their approach being based on a minimization of the Bayes average risk associated with the joint operations of detection and estimation. It is desirable to carry out a similar analysis for the multiple hypotheses reception situation and the consequences explored in some detail.

In the multishot reception situation there is a choice of two basic data processing arrangements depending on whether the updating is to take place after each observation interval, or whether the updating is to occur after the data of two or more observation intervals has accumulated. If the latter choice is made, and assuming that any of K distinct signals may occur during any observation interval of duration T, and we agree to update only after N such observation intervals, then there are a total of $M = K^N$ possible waveforms that may occur during the total time interval NT between updating. Consequently, such a situation may be regarded as a multiple hypotheses, "single-shot", joint detection and estimation problem. In this report we will treat by means of decision theory the problem of joint detection and estimation for the multiple hypotheses, "single-shot" reception situation, leaving for a subsequent report a consideration of sequential joint detection and estimation, where it is assumed that updating takes place after each observation interval.

3. SIMULTANEOUS OPTIMUM DETECTION AND ESTIMATION UNDER MULTIPLE HYPOTHESES

3.1 Introduction

In this section we shall consider the multiple alternative joint detection and estimation problem of determining the structure of a receiver which produces at its output both optimum detector decisions and estimates of either unknown signal parameters or unknown channel parameters. The two basic reception situations investigated in this section are shown on the next page in Figures 1 and 2.

Throughout this section it will be assumed that data can be collected only during one observation interval of duration T, and it is this assumption to which the terms "single-shot" or "one-shot" refer. It will also be assumed that to each member of the finite message set, $\{m_j \mid j = 0, \ldots, M\}$, consisting of M + 1 elements, where $M \geqq 2$, there corresponds a distinct signal selected from the set $\{S_j(t) \mid j = 0, 1, \ldots, M\}$. The correspondence between a message symbol, m_j say, and the signal waveform $S_j(t)$, may involve an intermediate step consisting of the selection of a descriptive signal parameter vector $\underline{\theta}_j$ (vector quantities will be underlined), which in turn is associated with a signal waveform $S(t;\underline{\theta}_j)$, however, this waveform will also be abbreviated as $S_j(t)$. It should be mentioned that although the actual mapping of the message symbol, m_j say, onto the parameter vector $\underline{\theta}_j$ is perhaps in reality unique, because of the receivers assumed limited knowledge of the transmitted signal waveform there is a consequent uncertainty as to the precise value of the actual signal parameter vector chosen by the message source, and hence $\underline{\theta}_j$ can only be assigned to some subset Ω_j of the total signal parameter space Ω. This is indicated in Figure 1 by the mapping of a message symbol onto a set of signal waveforms.

When a message symbol, m_j, is selected by the message source a corresponding signal $S(t;\underline{\theta}_j)$ is inserted into the channel and the received waveform which emerges is $X(t;\underline{\theta}_j) = S(t;\underline{\theta}_j) + n(t)$,

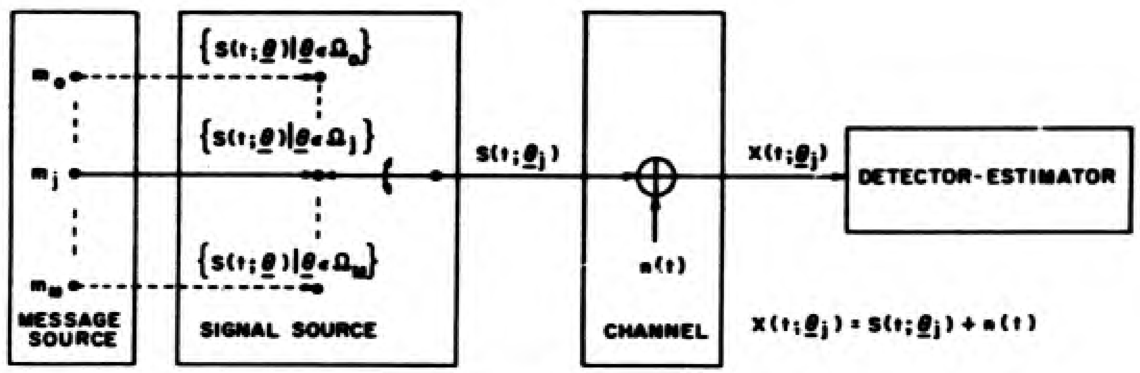

Fig. 1. Reception situation when the signal set is unknown at the receiver.

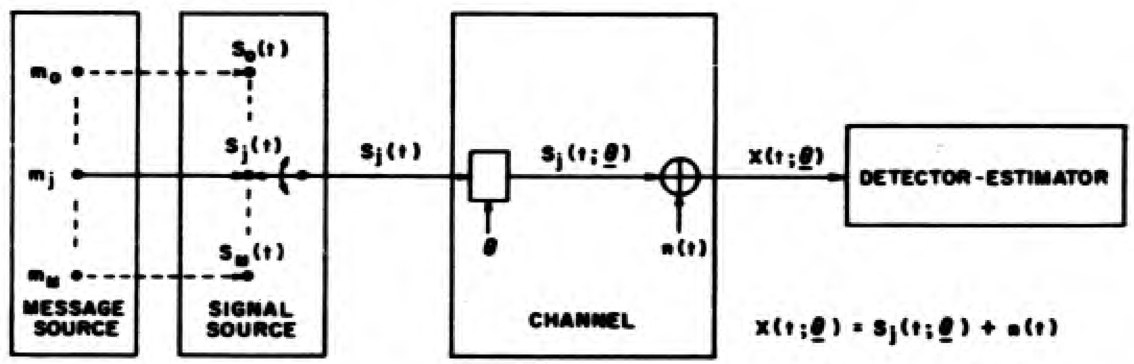

Fig. 2. Reception situation when the signal set is known at the receiver and the channel introduces an unknown parameter.

where n(t) is a realization of an additive noise process. The received waveform, $X(t;\underline{\theta}_j)$, will in general depend on an unknown parameter vector $\underline{\theta}_j$ which, as mentioned previously, results from either not knowing the signal waveform or because of unknown channel parameters.

A few remarks need to be made concerning the possibility that no signal is present during the observation interval (0, T) and that therefore the received data consists of noise alone. We can allow for this situation by letting one of the message symbols, m_0, correspond to the null signal, $S_0(t) = 0$, being emitted by the signal source. A problem arises when we attempt to identify subsets of an arbitrary signal parameter space with subsets of a signal space. Usually, a signal space Ω is considered to be a collection of waveforms, $\{S(t;\theta) \mid \theta \in \Theta\}$, indexed perhaps by some parameter θ whose values are contained in some signal parameter set Θ. What we would like to do, in order to keep notation to a minimum and make the analysis as general as possible, is identify the signal parameter set Θ with Ω. This identification would allow us to replace a signal space by a signal parameter space, the former being considered as a special case of the latter where $\underline{\theta} = \underline{S}$, where $\underline{S}$ is the vector of 2WT sample values. If no restrictions are imposed on the parameter space, then it is not true in general that $\theta = 0$ implies $\underline{S} = \underline{0}$, i.e., the null parameter doesn't correspond to the null signal. The phase angle ϕ of a sinusoidal signal $\sin(\omega t + \phi)$ is a good example to illustrate this point. Therefore, if we are going to identify signal spaces with signal parameter spaces, and allow for the possibility of noise alone being present, we have to restrict the signal parameter spaces. The restriction imposed will be that the parameters considered be non-nuisance or energy type parameters, for which $\theta = 0 \Rightarrow \underline{S} = \underline{0}$. The subset of the signal space or signal parameter space Ω corresponding to the null signal will be denoted by Ω_0.

In subsection 3.2 we will consider the case where the subsets $\Omega_0, \Omega_1, \ldots, \Omega_M$ of Ω assigned to the message symbols $m_0, m_1, \ldots, m_M$ are mutually disjoint or nonoverlapping. These subsets will be referred to as signal classes or hypothesis classes. Conceptually the nonoverlapping case is somewhat easier to deal with since it is always possible, having been given the actual parameter vector $\underline{\theta}$ selected by the transmitter and the decision γ made by the receiver, to determine whether a correct or an incorrect decision has been made. If a cost $C(\underline{\theta}, \gamma)$ is now associated with each parameter-decision pair $(\underline{\theta}, \gamma)$ it may be interpreted as the cost of either a correct or incorrect decision. Unfortunately the assumption that the signal spaces are nonoverlapping restricts the domain of definition of the prior densities used in Bayes decision theory to proper subsets of Ω, such as Ω_j, rather than to the total signal space, which excludes the use of certain convenient density functions. For example, if the signal parameters are the N = 2WT sample values then the signal parameter space is Euclidean N-space and the various nonoverlapping signal spaces $\Omega_0, \Omega_1, \ldots, \Omega_M$ are subsets of this space, which excludes the use of multivariate Guassian densities because their domain of definition is the entire space. We shall assume that such densities exist for the various nonoverlapping signal spaces involved in the problem of Bayes joint detection and estimation under multiple hypotheses.

In order to be able to use convenient densities, such as the Gaussian, we are forced to consider the case where the signal spaces $\Omega_0, \Omega_1, \ldots, \Omega_M$ overlap. This will be done in subsection 3.4. In the case of overlapping signal spaces or hypothesis classes it is impossible to decide when an error or a correct decision has been made for all elements of the signal space because the same signal may belong to two different hypothesis classes (see (3) page 22), and therefore the interpretation of $C(\underline{\theta}, \gamma)$ as the cost of a correct or incorrect decision is no longer valid. If we reinterpret $C(\underline{\theta}, \gamma_i)$ as the cost of assigning $\underline{\theta}$ to the hypothesis class Ω_i, then we can again define an average risk and attempt to determine decision rules which minimize it. A discussion of Bayes optimum detection for the case of overlapping hypothesis classes and the selection of a special cost function which results in detectors which are formally the same as those determined by

Middleton and Van Meter for the nonoverlapping case may be found in (9). In subsection 3.4 we shall follow a similar approach and show, for a special choice of cost function, that the detector and estimator structures for the overlapping case are formally the same as those found in subsections 3.2 and 3.3 for the nonoverlapping case.

The decision rules derived in subsections 3.2 and 3.3 are arrived at by a two stage minimization process, which consists of first determining a detection rule involving an unknown estimator of the signal or signal parameter, which if used results in the average risk assuming a relative minimum and secondly, determining that estimator which further minimizes the average risk. The final detection rule for this two stage process is the first detection rule with the optimum estimator substituted in place of the initially assumed estimator. This approach suggests that logically the estimation operation should precede the detection operation because the optimum estimator is used as part of the detection process. This is the reverse of the commonly thought of sequence of operations known as decision-directed measurement which was mentioned previously. This "reverse" strategy, for the case of detection, has been discussed originally by Esposito in (10) and (11), where it was found that the average likelihood ratio, which determines the detector structure, for known and Gaussian stochastic signals in Gaussian noise can be expressed as a function of a least squares or minimum variance estimate of the signal under the assumption that the signal is present with probability one. In a more recent paper, (12), Kailath shows that similar results are true for more general signal processes in additive Gaussian noise. The logic of performing the estimation process first, in connection with Bayes joint detection and estimation, is discussed in (16) where also a more satisfying approach to the minimization problem, involving only one step, is proposed. A similar approach to the multiple hypotheses joint detection and estimation problem will be discussed in a future report.

For a complete discussion of Bayes systems, it is always important to know whether or not nonrandom decision rules exist which achieve the Bayes risk. By again following an approach similar to that given in (9), it can be shown that our Bayes joint detection and estimation decision rules are non-random.

3.2 Nonoverlapping Hypothesis Classes

The notion of minimizing an average risk associated with the combined operation of detection and estimation provides not only the basis for analyzing the "single-shot", binary "on-off", but also the "single-shot", multiple alternative reception situation. Accordingly, we begin with the expression for the average risk for the combined operations of detection and estimation which will be denoted by R_{D+E} and is given by

$$R_{D+E} = \int_{\Gamma}\int_{\Delta}\int_{\Omega} C(\underline{\theta}, \gamma)P(\underline{\theta}, \underline{X}, \gamma)d\underline{\theta}d\underline{X}d\gamma \ . \tag{3.1}$$

In this equation $\underline{\theta}$ denotes the signal parameter vector, $\underline{X}$ is the sampled data vector of the received waveform $X(t;\underline{\theta})$, and γ denotes the decision made. The term $C(\underline{\theta}, \gamma)$ is the cost associated with the decision γ when $\underline{\theta}$ is the signal parameter actually present and $P(\underline{\theta}, \underline{X}, \gamma)$ is a joint probability density function. Since any decision, γ, can only depend on the observed data $\underline{X}$ and not on the unobservable true state of nature, $\underline{\theta}$, the risk, R_{D+E}, as given in (3.1), may be expressed in the following more familiar form:

$$R_{D+E} = \int_{\Gamma} d\underline{X} \int_{\Delta} d\gamma \int_{\Omega} d\underline{\theta} C(\underline{\theta}, \gamma)P(\gamma | \underline{X})P(\underline{X} | \underline{\theta})P(\underline{\theta}) \ . \tag{3.2}$$

Expression (3.2) follows from (3.1) by using Bayes rule twice to express $P(\underline{\theta}, \underline{X}, \gamma)$ in the following form:

$$P(\underline{\theta}, \underline{X}, \gamma) = P(\gamma | \underline{\theta}, \underline{X})P(\underline{X} | \underline{\theta})P(\underline{\theta}) \ .$$

Since the decision γ is independent of $\underline{\theta}$ but only depends on the received data $\underline{X}$, the last expression for $P(\underline{\theta}, \underline{X}, \gamma)$ can be written as

$$P(\underline{\theta}, \underline{X}, \gamma) = P(\gamma | \underline{X}) P(\underline{X} | \underline{\theta}) P(\underline{\theta}) . \qquad (3.3)$$

Substitution of (3.3) into (3.1) results in (3.2).

The space of all possible observed data vectors $\underline{X}$ is Γ, the space of all possible decisions γ is Δ, and the space of all possible signal parameters $\underline{\theta}$ is Ω. The spaces Γ, Δ, and Ω are in general quite arbitrary, and the probabilities are to be interpreted as appropriately defined measures on these spaces. In particular any of the three spaces may consist of discrete points in which case the integrals are to be interpreted as sums over these points.

Returning to our expression for the total average risk, we next observe that the a priori probability $P(\underline{\theta})$ may be decomposed in the following way:

$$P(\underline{\theta}) = \sum_{i=0}^{M} P(\underline{\theta} | H_i) P(H_i) \qquad (3.4)$$

where H_i denotes the hypothesis that a signal of class Ω_i is present. $P(H_i)$ is simply the probability of occurrence of signals of class Ω_i, and $P(\underline{\theta} | H_i)$ is a probability density function describing the probability that various values of $\underline{\theta}$ are assumed within the class Ω_i. Since $P(\underline{\theta} | H_i)$ has meaning only for values of $\underline{\theta}$ contained in Ω_i we may assume that $P(\underline{\theta} | H_i)$ is zero outside Ω_i and rewrite $P(\underline{\theta} | H_i)$ as $P_i(\underline{\theta}) I_{\Omega_i}(\underline{\theta})$, where I_{Ω_i} is the indicator function, namely

$$I_{\Omega_i}(\underline{\theta}) = \begin{cases} 1 \text{ of } \underline{\theta} \in \Omega_i \\ 0 \text{ otherwise} . \end{cases}$$

Also in what follows we shall use the abbreviated notation P_i for $P(H_i)$.

Substitution of these notational changes into (3.4) results in

$$P(\underline{\theta}) = \sum_{i=0}^{M} P_i P_i(\underline{\theta}) I_{\Omega_i}(\underline{\theta}) \quad , \tag{3.4'}$$

and substitution of (3.4') into the expression for the average risk (3.2) results in:

$$R_{D+E} = \int_{\Gamma} d\underline{X} \int_{\Delta} d\gamma \int_{\Omega} d\underline{\theta}\, C(\underline{\theta}, \gamma) P(\gamma | \underline{X}) P(\underline{X} | \underline{\theta}) \Big[\sum_{i=0}^{M} P_i P_i(\underline{\theta}) I_{\Omega_i}(\underline{\theta}) \Big]$$

$$= \int_{\Gamma} d\underline{X} \int_{\Delta} d\gamma\, P(\gamma | \underline{X}) \Big[\sum_{i=0}^{M} P_i \int_{\Omega} d\underline{\theta}\, C(\underline{\theta}, \gamma) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) I_{\Omega_i}(\underline{\theta}) \Big]$$

$$= \int_{\Gamma} d\underline{X} \int_{\Delta} d\gamma\, P(\gamma | \underline{X}) \Big[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta}\, C(\underline{\theta}, \gamma) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \Big] . \tag{3.5}$$

This last line follows from the fact that $I_{\Omega_i}(\underline{\theta})$ is zero outside of Ω_i. The next step is to decide on the structure of the decision space Δ. It will be taken to be the following discrete set.

$$\Delta \triangleq \{\gamma_0, \gamma_1, \ldots, \gamma_M\}$$

where γ_0 = decision that hypothesis H_0 is true, noise alone present, and no estimate is required.

γ_i = the decision that hypothesis H_i is true, signal of class Ω_i is present, and an estimate $\hat{\underline{\theta}}(\underline{X})$ of the signal parameter present is required.

Since the decision space is a finite set of discrete elements, the integral over the decision space in (3.5) becomes a sum. If we replace the integral by a sum, (3.5) becomes

$$R_{D+E} = \int_{\Gamma} d\underline{X} \Big\{ \sum_{j=0}^{M} P(\gamma_j | \underline{X}) \Big[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta}\, C(\underline{\theta}, \gamma_j) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \Big] \Big\} . \tag{3.6}$$

A definite decision is required at the end of the observation interval and therefore the following condition is satisfied

$$\sum_{j=0}^{M} P(\gamma_j | \underline{X}) = \sum_{j=0}^{M} P(H_j | \underline{X}) = 1 . \tag{3.7}$$

This condition allows us to express $P(\gamma_0 | \underline{X})$ as

$$P(\gamma_0 | \underline{X}) = 1 - \sum_{j=1}^{M} P(\gamma_j | \underline{X}) \tag{3.7'}$$

which will be substituted into the following rewritten form of (3.6)

$$R_{D+E} = \int_{\Gamma} d\underline{X}\, P(\gamma_0 | \underline{X}) \left[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta} C(\underline{\theta}, \gamma_0) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \right]$$

$$+ \int_{\Gamma} d\underline{X} \left\{ \sum_{j=1}^{M} P(\gamma_j | \underline{X}) \left[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta} C(\underline{\theta}, \gamma_j) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \right] \right\} . \tag{3.6'}$$

Substitution of (3.7') into (3.6') results in the following expression for the average risk:

$$R_{D+E} = \int_{\Gamma} d\underline{X} \left[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta} C(\underline{\theta}, \gamma_0) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \right]$$

$$+ \int_{\Gamma} d\underline{X} \left\{ \sum_{j=1}^{M} P(\gamma_j | \underline{X}) \left[\sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta} \left[C(\underline{\theta}, \gamma_j) - C(\underline{\theta}, \gamma_0) \right] P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \right\} . \tag{3.8}$$

The cost factor, $C(\underline{\theta}, \gamma_0)$, is assumed to be positive or equal to zero, which is reasonable because in the case of nonoverlapping signal classes it is either the cost of a correct decision if $\underline{\theta} = \underline{0}$, or an incorrect decision if $\underline{\theta} \neq \underline{0}$, provided that $\underline{\theta}$ is not a nuisance parameter. Furthermore, when γ_0 is decided, no estimate is required and hence $C(\underline{\theta}, \gamma_0)$ does not depend on $\underline{X}$ through an estimator $\hat{\underline{\theta}}(\underline{X})$. Equation (3.8) may now be rewritten in the following condensed form

$$R_{D+E} = \int_\Gamma d\underline{X}\, A_0(\underline{X}) + \int_\Gamma d\underline{X} \sum_{j=1}^{M} P(\gamma_j \mid \underline{X}) B_j'(\underline{X}) \tag{3.8'}$$

where

$$A_0(\underline{X}) \triangleq \sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta}\, C(\underline{\theta}, \gamma_0) P(\underline{X} \mid \underline{\theta}) P_i(\underline{\theta}) \tag{3.9a}$$

$$B_j'(\underline{X}) \triangleq \sum_{i=0}^{M} P_i \int_{\Omega_i} d\underline{\theta} \left[C(\underline{\theta}, \gamma_j) - C(\underline{\theta}, \gamma_0)\right] P(\underline{X} \mid \underline{\theta}) P_i(\underline{\theta})\ . \tag{3.9b}$$

From our previous comments about $C(\underline{\theta}, \gamma_0)$ it is clear that $A_0(\underline{X}) \geqq 0$ and therefore the first term in (3.8') is a positive constant independent of any decision rule, which means that it may be neglected in the minimization process which determines the optimum decision rule. In order to make this analysis as similar as possible to that of the multiple alternative pure detection situation we define the following modified average likelihood ratio

$$L_{ij}(\underline{X}) \triangleq \frac{\mu_i \int_{\Omega_i} d\underline{\theta} \left[C(\underline{\theta}, \gamma_0) - C(\underline{\theta}, \gamma_j)\right] P(\underline{X} \mid \underline{\theta}) P_i(\underline{\theta})}{\left[C(\underline{0}, \gamma_j) - C(\underline{0}, \gamma_0)\right] P(\underline{X} \mid \underline{0})} \qquad (i, j = 1, \dots, M) \tag{3.10}$$

where $\mu_i \triangleq \frac{P_i}{P_0}$. The average likelihood ratio $L_{11}(\underline{X})$ corresponds to Middleton and Esposito's modified likelihood ratio, Λ_g given by relation 2.10 of reference (16). Upon noticing that $P_0(\underline{\theta}) = \delta(\underline{\theta} - \underline{0})$, where $\delta(\cdot)$ is the Dirac δ-function, and using the definition of $L_{ij}(\underline{X})$, the definition of $B_j'(\underline{X})$ may be expressed in the following form:

$$B_j'(\underline{X}) = P_0\left[C(\underline{0},\gamma_j)-C(\underline{0},\gamma_0)\right]P(\underline{X}|\,\underline{0}) - \sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta}\left[C(\underline{\theta},\gamma_0)-C(\underline{\theta},\gamma_j)\right] P(\underline{X}|\,\underline{\theta})P_i(\underline{\theta})$$

$$= P_0\left[C(\underline{0},\gamma_j)-C(\underline{0},\gamma_0)\right]P(\underline{X}|\,\underline{0})\left[1-\sum_{i=1}^{M} L_{ij}(\underline{X})\right]$$

$$B_j'(\underline{X}) = P_0 P(\underline{X}|\,\underline{0})C_j(\underline{X})B_j(\underline{X}) \qquad (3.11)$$

where

$$B_j(\underline{X}) \triangleq 1 - \sum_{i=1}^{M} L_{ij}(\underline{X}) \qquad (i, j = 1, 2, \ldots, M) \qquad (3.12a)$$

$$C_j(\underline{X}) \triangleq C(\underline{0},\gamma_j) - C(\underline{0},\gamma_0) \quad (j = 1, \ldots, M)\,. \qquad (3.12b)$$

The average risk, (3.8'), with (3.11) substituted for $B_j'(\underline{X})$ becomes

$$R_{D+E} = \int_\Gamma d\underline{X}\, A_0(\underline{X}) + \int_\Gamma d\underline{X}\, P_0\, P(\underline{X}|\,\underline{0}) \sum_{j=1}^{M} P(\gamma_j|\,\underline{X})C_j(\underline{X})B_j(\underline{X})\,. \qquad (3.13a)$$

For the same state of nature the cost of an incorrect decision is taken to be greater than the cost of a correct decision which is usually assumed to be zero. Therefore $C_j(\underline{X}) \geqq 0$ for all X, and $j = 1, 2, \ldots, M$. Since the first term in (3.13a) is a constant and P_0 and $P(\underline{X}|\,\underline{0})$ are probabilities and therefore positive, we see that the following nonrandom decision rule minimizes the average risk.

Decision rule: Decide γ_j or $P(\gamma_j|\,\underline{X}) = 1$ if (a) $B_j(\underline{X}) < 0$ (3.13b)

(b) $C_j(\underline{X})B_j(\underline{X}) \leqq C_i(\underline{X})B_i(\underline{X})$ for all i.

It is clear from the definition of $B_j(\underline{X})$ that condition (a) in (13.b) may be replaced by $\sum_{i=1}^{M} L_{ij}(\underline{X}) > 1$, which is similar to the result obtained for the classical multiple alternative detection problem (see ref. (2), Chapter 23), in fact, if we had not allowed for the possibility of the cost assignment, $C(\underline{\theta}, \gamma_j)$, being dependent on $\underline{\theta}$ the results would be identical.

Since $C(\underline{\theta}, \gamma_j)$ is to reflect not only the cost of the detection or classification operation but also that of estimation we must allow for the possibility of $C(\underline{\theta}, \gamma_j)$ being dependent both on $\underline{\theta}$ and $\underline{X}$, and it is through this dependence that the coupling between the estimation and detection operations is achieved. For example, as an illustration of a possible choice of cost function which will be used later in sub-section 3.3, we suppose that $\underline{\theta} \in \Omega_i$ and that $C(\underline{\theta}, \gamma_j) = C_{ij}^{(1)} + C_{ij}^{(2)} \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]^T \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]$, where $C_{ij}^{(1)}$ is the usual constant cost assignments for detection and $C_{ij}^{(2)}\left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]^T \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]$ is a quadratic cost function encountered in least squares estimation.

The decision rule obtained above depends on $C(\underline{\theta}, \gamma_j)$ which, as we noted in the previous paragraph, may contain an unknown function of $\underline{X}$, namely the estimator, $\hat{\underline{\theta}}(\underline{X})$, of the actual signal parameter $\underline{\theta}$ present, and therefore only the first stage of a two stage minimization process has been completed. The strategy involved in the two stage minimization procedure followed here, is to first determine a decision rule which depends on an estimator $\hat{\underline{\theta}}(\underline{X})$ and which would be optimum if $\hat{\underline{\theta}}(\underline{X})$ were optimum. The second stage of the minimization process consists of finding the presumably unique function $\hat{\underline{\theta}}^*(\underline{X})$ which further minimizes the average risk (3.13a), which is rewritten below taking into account the decision rule (3.13b):

$$R_{D+E} = \int_{\Gamma} d\underline{X} A_0(\underline{X}) + \sum_{j=1}^{M} \int_{\Gamma_j} d\underline{X}\, P_0 P(\underline{X}|\,\underline{0}) C_j(\underline{X}) B_j(\underline{X}) \quad (3.14a)$$

where Γ_j denotes the region of the observation space for which γ_j is decided, namely,

$$\Gamma_j \triangleq \left\{ \underline{X} \,|\, B_j(\underline{X}) < 0 \text{ and } C_j(\underline{X}) B_j(\underline{X}) \leqq C_i(\underline{X}) B_i(\underline{X}) \text{ for all } i \right\} .$$

(3.14b)

The optimum estimator which results from the second minimization procedure, for the case where γ_j is decided after the first stage, is denoted by $\hat{\underline{\theta}}_j^*(\underline{X})$ and determined from the following condition:

$$\min_{\hat{\underline{\theta}}_j(\underline{X})} \int_{\Gamma_j} d\underline{X} P_0\, P(\underline{X}|\,\underline{0})\, C_j(\underline{X}) B_j(\underline{X}) \quad (j \neq 0) . \qquad (3.15)$$

We have included a subscript j on the optimum estimator to indicate that γ_j was decided, which is equivalent to deciding that $\theta \in \Omega_j$, $j \neq 0$, and therefore the resulting estimate is presumably the best obtainable estimate of the signal of class Ω_j. In order to obtain the final optimum detection rule, we merely substitute the optimum estimator back into the originally obtained detection rule, which in this case is (3.13b).

It should be noted, that the optimum decision rule for the detector depends on the optimum estimator and hence estimation must precede detection, which is the reverse of the commonly envisioned sequence of operations. The sequence of operations for the "single-shot", multiple alternative, joint detection and estimation device are shown schematically in Figures 3.

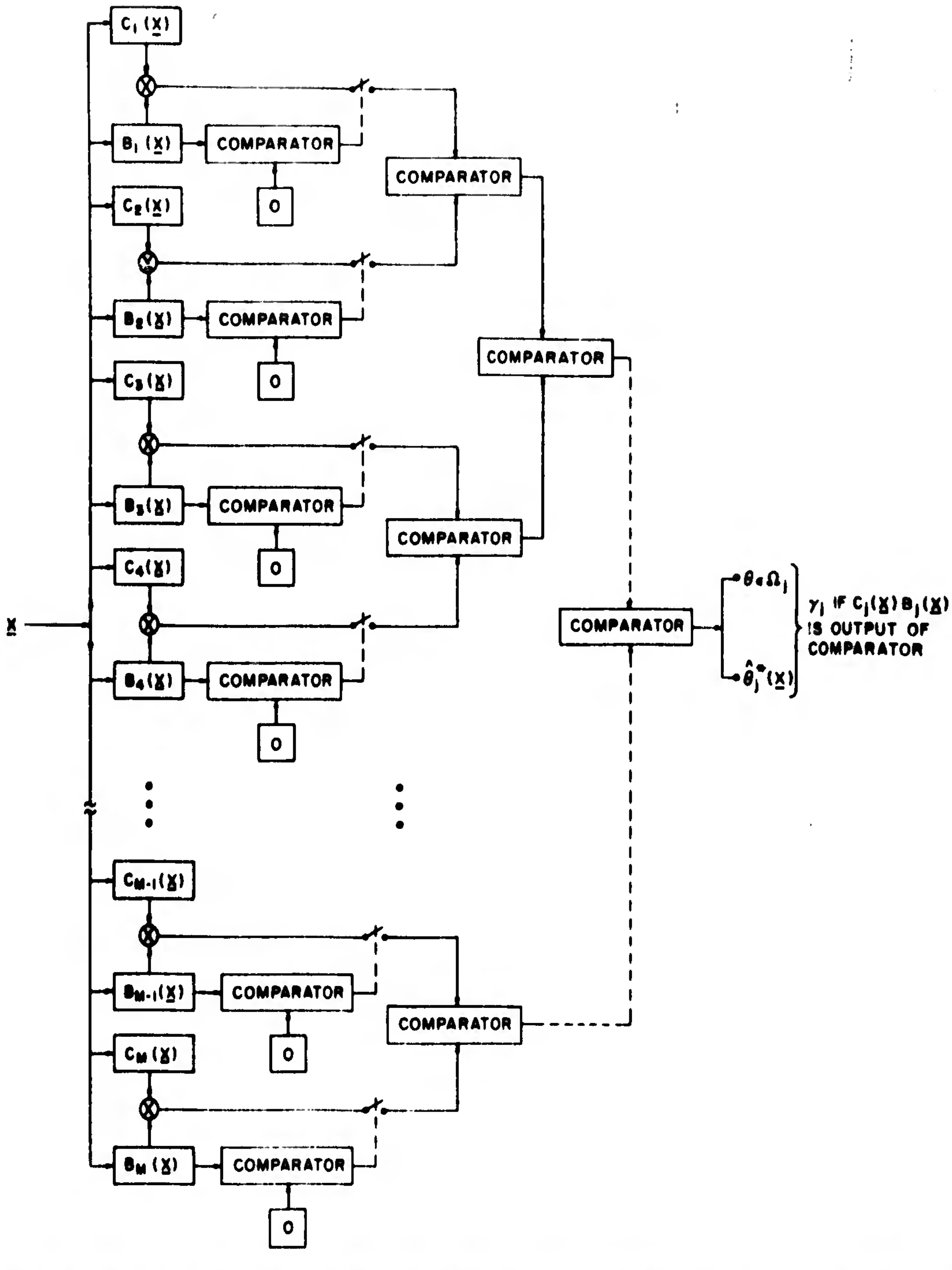

Figure 3a. Schematic diagram showing the sequence of operations resulting in the decision γ_j and given by relation 3.13b. The block labeled "comparator" has an output which is the smaller of its two inputs. A more detailed description of the block $B_j(\underline{X})$ is given in Figure 3b. A non-zero output of the comparator immediately following any of the B-blocks closes the switch to which it is connected.

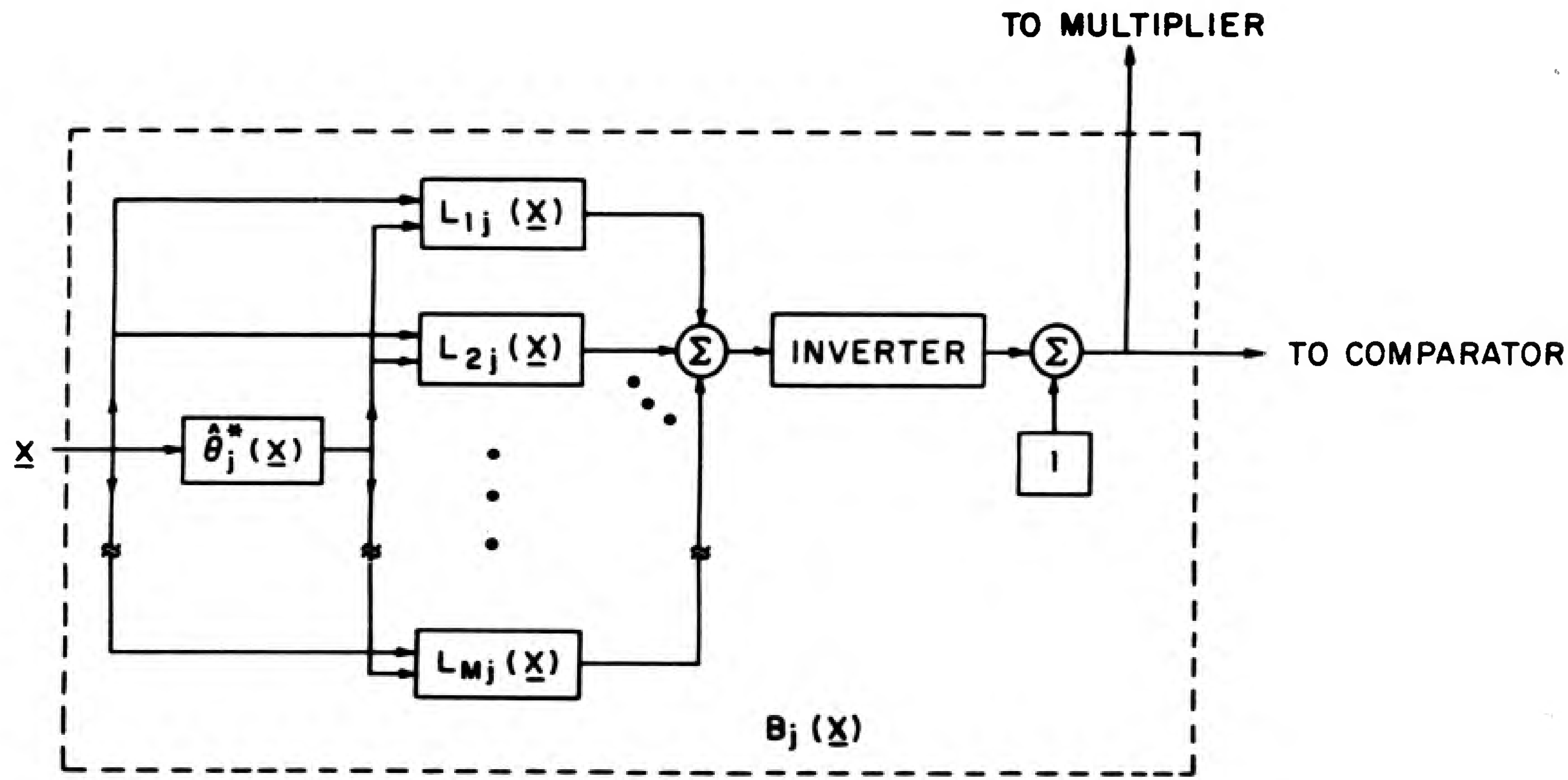

Figure 3. b. Schematic diagram of the operations performed by the blocks $B_j(\underline{X})$ ($j = 1, \ldots, M$) of Figure 3. a. The block labeled $\hat{\underline{\theta}}_j^*(\underline{X})$ is necessary because the likelihood ratios, $L_{ij}(\underline{X})$, depend on $\hat{\underline{\theta}}_j^*(\underline{X})$ through the cost function.

3.2a. Binary Special Cases

In this subsection some special cases of the results obtained so far will be considered, and wherever possible comparisons will be madewith previously published results. The cases considered in this subsection are: (1) pure detection, where no joint estimation operation is required, for the binary "on-off" case ($\underline{\theta} \neq \underline{0}$ versus $\underline{\theta} = \underline{0}$); (2) strong coupled joint detection and estimation problem for the binary "on-off" case ($\underline{\theta} \neq \underline{0}$ versus $\underline{\theta} = \underline{0}$); (3) the binary two signal case ($\underline{\theta}_1 \neq \underline{0}$ versus $\underline{\theta}_2 \neq \underline{0}$).

Case 1. Pure detection. Here we are only interested in the detection operation and therefore the cost functions reduce to the usual constant cost assignments, that is,

$$C(\underline{\theta}\epsilon\Omega_0, \gamma_0) = C_{00} \qquad C(\underline{\theta}\epsilon\Omega_1, \gamma_0) = C_{10}$$
$$C(\underline{\theta}\epsilon\Omega_0, \gamma_1) = C_{01} \qquad C(\underline{\theta}\epsilon\Omega_1, \gamma_1) = C_{11} \quad .$$

In the binary "on-off" case i, j = 0, 1 and therefore only one modified likelihood ratio $L_{11}(X)$, given by (3.10), exists. For the above cost assignments this likelihood ratio becomes

$$L_{11}(\underline{X}) = \frac{\mu_1 \int_{\Omega_1} (C_{10} - C_{11}) \, P(\underline{X} | \underline{\theta}) P_1(\underline{\theta}) d\underline{\theta}}{(C_{01} - C_{00}) P(\underline{X} | \underline{0})} = \frac{(C_{10} - C_{11})}{(C_{01} - C_{00})} \Lambda(\underline{X})$$

where $\Lambda(\underline{X})$ is the usual generalized likelihood ratio, viz.,

$$\Lambda(\underline{X}) = \frac{P_1 \int_{\Omega_1} d\underline{\theta} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta})}{P_0 P(\underline{X} | \underline{0})} \quad . \tag{3.16}$$

According to our decision rule (3.13b) we will decide γ_1, a signal is present, if $B_1(\underline{X}) < 0$. Since $B_1(\underline{X}) = 1 - L_{11}(\underline{X})$, it follows that γ_1 is decided if $L_{11}(\underline{X}) > 1$, or

$$\Lambda(\underline{X}) > \frac{C_{01} - C_{00}}{C_{10} - C_{11}} \triangleq K_{01} \quad .$$

This result is the well known detection rule (Ref. (3) page 24) where the generalized likelihood ratio $\Lambda(\underline{X})$ is compared with a threshold K_{01}, and the decision that a signal is present during the observation interval is made if $\Lambda(\underline{X}) > K_{01}$, otherwise it is decided that only noise is present.

Case 2. Joint detection and estimation; the strong coupled, binary "on-off" case. Here estimates are a part of our decision when it is decided that something other than noise is present, and therefore we have to include in our cost assignments a factor which accounts for the additional cost of estimation. The cost functions chosen will be of the form:

$$\begin{aligned}
C(\underline{\theta}\epsilon\Omega_0, \gamma_0) &= C(\underline{0}, \gamma_0) = C_{00}^{(1)} + C_{00}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}) = C_{1-\alpha} + C_{00} \\
C(\underline{\theta}\epsilon\Omega_0, \gamma_1) &= C(\underline{0}, \gamma_1) = C_{01}^{(1)} + C_{01}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}) = C_{\alpha} + f_{10}(\underline{\hat{\theta}}) \\
C(\underline{\theta}\epsilon\Omega_1, \gamma_0) &= C(\underline{\theta}, \gamma_0) = C_{10}^{(1)} + C_{10}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}) = C_{\beta} + f_{01}(\underline{\theta}) \qquad (3.17) \\
C(\underline{\theta}\epsilon\Omega_1, \gamma_1) &= C(\underline{\theta}, \gamma_1) = C_{11}^{(1)} + C_{11}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}) = C_{1-\beta} + f_{11}(\underline{\theta}, \underline{\hat{\theta}}) \; .
\end{aligned}$$

The cost assignments on the extreme right are those used by Middleton and Esposito (4). Again since we are dealing with the binary "on-off" case and i, j = 0, 1 there is only one likelihood ratio to deal with, namely $L_{11}(\underline{X})$ which is given by 3.10. Substitution of the cost assignments (3.17) with $\hat{\theta}(\underline{X})$ replaced by $\underline{\hat{\theta}}^*(\underline{X})$ results in:

$$L_{11}(\underline{X}) = \frac{(C_{10}^{(1)} - C_{11}^{(1)})\Lambda(\underline{X}) + \mu_1 \int_{\Omega_1} \left[C_{10}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}^*) - C_{11}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}^*)\right] \frac{P(\underline{X}|\underline{\theta})}{P(\underline{X}|\underline{0})} P_1(\underline{\theta}) d\underline{\theta}}{C_{01}^{(1)} - C_{00}^{(1)} + C_{01}(\underline{\theta}, \underline{\hat{\theta}}^*) - C_{00}(\underline{\theta}, \underline{\hat{\theta}}^*)}$$

or

$$L_{11}(\underline{X}) = \frac{(C_\beta - C_{1-\beta})\Lambda(\underline{X}) + \mu_1 \int_{\Omega_1} \left[f_{01}(\underline{\theta}) - f_{11}(\underline{\theta}, \underline{\hat{\theta}}^*)\right] \Lambda'(\underline{X}, \underline{\theta}) P_1(\underline{\theta}) d\underline{\theta}}{(C_\alpha - C_{1-\alpha}) + f_{10}(\underline{\hat{\theta}}^*) - C_{00}}$$

where $\Lambda(\underline{X})$ is the generalized likelihood ratio given by (3.16) and $\Lambda'(\underline{X}, \underline{\theta})$ is the "ordinary" likelihood ratio given by $\mu_1 \frac{P(\underline{X}|\underline{\theta})}{P(\underline{X}|\underline{0})}$. If we assume that $C_{1-\alpha} = C_{1-\beta} = C_{00} = f_{10}(\underline{\hat{\theta}}^*) = 0$, then

$$L_{11}(\underline{X}) = \frac{C_\beta}{C_\alpha} \Lambda(\underline{X}) + \frac{1}{C_\alpha} \int_{\Omega_1} \left[f_{01}(\underline{\theta}) - f_{11}(\underline{\theta}, \underline{\hat{\theta}}^*)\right] \Lambda'(\underline{X}, \underline{\theta}) P_1(\underline{\theta}) d\underline{\theta} .$$

This last expression is identical to relation (4.13) of reference (4). When $L_{11}(\underline{X}) > 1$, we would decide that a signal was present during the observation interval and the optimum estimate of $\underline{\theta}$ would be $\underline{\hat{\theta}}^*(\underline{X})$. If $L_{11}(\underline{X}) \leq 1$ we would decide that no signal was present and in this case no estimate is required.

Case 3. Joint detection and estimation; the strong coupled, binary case of $\underline{\theta}_1 \neq 0$ versus $\underline{\theta}_2 \neq 0$. Rather than try and modify the results of section 3.2 to fit the situation when the null hypothesis is excluded it is easier to derive the decision rule for the reception situation where one of M distinct and non-zero signals can be present during the observation interval.

When the null hypothesis is excluded we may assume that $P_0 = 0$ and $P(\gamma_0|\underline{X}) = 0$. The average risk, R_{D+E}, given by (3.6') then becomes

$$R_{D+E} = \int_\Gamma d\underline{X} \left\{ \sum_{j=1}^{M} P(\gamma_j|\underline{X}) \left[\sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta} C(\underline{\theta}, \gamma_j) P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) \right] \right\} .$$

The condition that a definite decision is required is expressed by

$$P(\gamma_k | \underline{X}) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{M} P(\gamma_j | \underline{X}) ,$$

and if this relation is substituted into the previous expression for R_{D+E} we see that

$$R_{D+E} = \int_{\Gamma} d\underline{X} \left\{ \sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta} C(\underline{\theta}, \gamma_k) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) + \sum_{\substack{j=1 \\ j \neq k}}^{M} P(\gamma_j | \underline{X}) B_{jk}(\underline{X}) \right\} \tag{3.18a}$$

where

$$B_{jk}(\underline{X}) \triangleq \sum_{i=1}^{M} P_i \int_{\Omega_i} (C(\underline{\theta}, \gamma_j) - C(\underline{\theta}, \gamma_k)) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} .$$

The first term on the right hand side is a positive constant independent of the probabilities, $P(\gamma_j | \underline{X})$, which embody the decision rule. Clearly the following nonrandom decision rule may be chosen.

$$\text{Decide } \gamma_j (j \neq k) \text{ or } P(\gamma_j | \underline{X}) = 1 \text{ if (a) } B_{jk}(\underline{X}) < 0$$

$$\text{(b) } B_{jk}(\underline{X}) \lessgtr B_{ik}(\underline{X}) \text{ for all } i \; (i \neq k) \tag{3.19}$$

otherwise decide γ_k or $P(\gamma_k | \underline{X}) = 1$.

Again let Γ_j be the region of the observation space for which γ_j is decided. The average risk, (3.18a), taking into account the decision rule (3.19) becomes:

$$R_{D+E} = \int_{\Gamma} d\underline{X} \left[\sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta} \, C(\underline{\theta}, \gamma_k) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) \right] + \sum_{\substack{j=1 \\ j \neq k}}^{M} \int_{\Gamma_j} d\underline{X} B_{jk}(\underline{X}) . \tag{3.18b}$$

The optimum estimator for the case when γ_j is decided, $\underline{\hat{\theta}}_j^*(\underline{X})_{P_j<1}$, is that function defined over Γ_j which minimizes the last expression for the average risk. Since the first term on the right-hand side of (3.18b) only involves the decision γ_k and therefore will be independent of $\underline{\hat{\theta}}_j^*(\underline{X})_{P_{j<1}}$, the condition which determines this estimator is:

$$\min_{\underline{\hat{\theta}}_j(\underline{X})} \int_{\Gamma_j} d\underline{X} \left[\sum_{i=1}^{M} P_i \int_{\Omega_i} (C(\underline{\theta}, \gamma_j) - C(\underline{\theta}, \gamma_k)) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \right]$$

and since $C(\underline{\theta}, \gamma_k)$ is independent of $\underline{\hat{\theta}}_j(\underline{X})$ this last relation becomes

$$\min_{\underline{\hat{\theta}}_j(\underline{X})} \int_{\Gamma_j} d\underline{X} \left[\sum_{i=1}^{M} P_i \int_{\Omega_i} C(\underline{\theta}, \gamma_j) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \right] . \qquad (3.20)$$

If $P(\gamma_j | \underline{X}) = 0$ for all j not equal to k, that is, γ_k is decided, then the average risk is given by just the first term in (3.18b) and optimum estimator, $\underline{\hat{\theta}}_k^*(\underline{X})_{P_k<1}$, for this case is determined from the following condition:

$$\min_{\underline{\hat{\theta}}_k(\underline{X})} \int_{\Gamma_k} d\underline{X} \left[\sum_{i=1}^{M} P_i \int_{\Omega_i} C(\underline{\theta}, \gamma_k) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \right] . \qquad (3.21)$$

This last relation is identical with (3.20) if we allow j = k. The optimum decision rule is now given by (3.19) with $\underline{\hat{\theta}}_j^*(\underline{X})_{P_j<1} (j=1, \ldots, M)$ substituted for the various unknown estimators involved in (3.19). If we now specialize to the binary two signal case we may take j = 1 and k = 2. Then

$$B_{12}(\underline{X}) = \sum_{i=1}^{2} P_i \int_{\Omega_i} (C(\theta, \gamma_1) - C(\theta, \gamma_2)) P(\underline{X} | \underline{\theta}) d\underline{\theta} \qquad (3.22)$$

and the estimators are determined from the following conditions:

$$\min_{\underline{\theta}_j(\underline{X})} \int_\Gamma d\underline{X} \left[\sum_{i=1}^{2} P_i \int_{\Omega_1} C(\underline{\theta}, \gamma_j) P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \right] (j = 1, 2) \quad . \tag{3.23}$$

Denote by $B^*_{12}(\underline{X})$ the factor $B_{12}(\underline{X})$ with $\hat{\underline{\theta}}^*_1(\underline{X})_{P_1<1}$ and $\hat{\underline{\theta}}^*_2(\underline{X})_{P_2<1}$ substituted in place of $\hat{\underline{\theta}}_1(\underline{X})$ and $\hat{\underline{\theta}}_2(\underline{X})$ respectively. The optimum decision rule for this case is given by the following:

$$P^*(\gamma_1 | \underline{X}) = 1 \text{ if } B^*_{12}(\underline{X}) < 0$$

$$P^*(\gamma_2 | \underline{X}) = 1 \text{ if } B^*_{12}(\underline{X}) \geqq 0 \quad .$$

3.3 Example: Quadratic Cost Functions - Nonoverlapping Hypothesis Classes

In order to make any further progress with the analysis, specific cost assignments have to be made and $\hat{\underline{\theta}}^*(\underline{X})$ determined in accordance with the minimization operation indicated in (3.15). An important example will now be discussed which will illustrate the general analysis as well as provide the functional form or structure of a specific estimator and detector.

As mentioned previously, the usual costs of detection decisions are modified to take into account the influence such decisions have on whether or not an estimate is presented at the output of the joint detection and estimation procedure. The specific cost function chosen for this example will be

$$C(\underline{\theta}, \gamma_j) = C^{(1)}_{ij} + C^{(2)}_{ij} \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]^T D \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right] (\underline{\theta} \in \Omega_i \text{ and } i, j = 0, 1, \ldots, M). \tag{3.24}$$

where D is a positive definite matrix.

The cost factors $C^{(1)}_{ij}$ are the usual constant cost assignments associated with the detection operation (see Middleton (2), Chapters 19 and 23), namely the cost of deciding that hypothesis H_j is true when actually H_i is true. The cost factor $C^{(1)}_{ii}$ is assoicated with a correct detection decision and we shall make the standard assumption that a correct detection decision costs nothing, i. e. $C^{(1)}_{ii} = 0$ for $i = 0, 1, \ldots, M$. The second term in (3. 24) is an additional cost which reflects the coupling between the detection and estimation operations and we shall now give a more detailed discussion of the $C^{(2)}_{ij}$ factors. Assuming that $\underline{\theta}$ is not a nuisance parameter, i. e., $\underline{0} = \underline{0}$ implies $\underline{S} = \underline{0}$, then $\underline{0} \in \Omega_0$ implies H_0 is true and if, in addition, γ_0 is decided then a correct decision has been made and no estimate is presented at the output. It is therefore reasonable to choose $C^{(2)}_{00} = 0$, and this combined with $C^{(1)}_{00} = 0$ results in $C(\underline{\theta}, \gamma_0) = 0$, if $\theta \in \Omega_0$. Now if γ_0 is decided and $\underline{\theta} \in \Omega_i$ $(i \neq 0)$, then an incorrect decision has been made, but nevertheless no estimate is presented at the output. If we visualize this joint detection and estimation procedure as one stage in an adaptive process, then no updating would take place as a result of this decision although its rate of "convergence" may be decreased. Such a mistake is not considered as serious as the use of erroneous estimates for updating the device, which would be the case if γ_j is decided and $\underline{\theta} \in \Omega_i$ $(i \neq j \neq 0)$. Accordingly, the following relations between the relative magnitudes of the cost factors $C^{(2)}_{ij}$ seems reasonable:

$$C^{(2)}_{0j} \geqq C^{(2)}_{jj} > C^{(2)}_{j0} \geqq C^{(2)}_{00} = 0 \quad (j \neq 0)$$

$$C^{(2)}_{ij} \geqq C^{(2)}_{jj} \quad (i, j = 1, 2, \ldots, M) \tag{3.25}$$

With these considerations in mind, the cost functions, $C(\underline{\theta}, \gamma_j)$, for various combinations of signal classes Ω_i and decisions γ_j, becomes

$$
\begin{aligned}
C(\underline{0}, \gamma_0) &= 0 \\
C(\underline{0}, \gamma_j) &= C^{(1)}_{0j} + C^{(2)}_{0j}\, \underline{\hat{\theta}}^{\,T}(\underline{X})\underline{\hat{\theta}}(\underline{X}) \qquad (j \neq 0) \qquad\qquad (3.26) \\
C(\underline{\theta}, \gamma_j) &= C^{(1)}_{ij} + C^{(2)}_{ij}\left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right]^T D\left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right] \; (\underline{\theta} \neq \underline{0} \text{ or } i \neq 0).
\end{aligned}
$$

Our immediate goal is to determine that $\underline{\hat{\theta}}^*(\underline{X})$ which minimizes the integral appearing in (3.15), the integrand of which involves the factor $C_j(\underline{X})B_j(\underline{X})$. From the definitions of $C_j(\underline{X})$ and $B_j(\underline{X})$, relations 3.12, we see that for $j \neq 0$

$$
C_j(\underline{X})B_j(\underline{X}) = C(\underline{0}, \gamma_j) - C(\underline{0}, \gamma_0) - \frac{1}{P_0 P(\underline{X}|\underline{0})} \sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta} \left[C(\underline{\theta}, \gamma_0) - C(\underline{\theta}, \gamma_j)\right] P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}).
$$

Multiplying this last relation through by $P_0 P(\underline{X}|\underline{0})$ and substitution of the relations (3.26) results in the following expression:

$$
\begin{aligned}
P_0 P(\underline{X}|\underline{0}) C_j(\underline{X}) B_j(\underline{X}) &= P_0 C^{(1)}_{0j} P(\underline{X}|\underline{0}) + P_0 C^{(2)}_{0j} \underline{\hat{\theta}}^T(\underline{X})\underline{\hat{\theta}}(\underline{X})\, P(\underline{X}|\underline{0}) \\
&- \sum_{i=1}^{M} P_i \int_{\Omega_1} \left[d\underline{\theta} \left\{ C^{(1)}_{i0} - C^{(1)}_{ij}\right] + \left[C^{(2)}_{i0} - C^{(2)}_{ij}\right]\left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right]^T D \left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right]\right\} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) .
\end{aligned}
$$

Integrating this last expression over the space Γ and noticing that $P_0(\underline{\theta}) = \delta(\underline{\theta} - \underline{0})$ and $C^{(2)}_{00} = 0$ we see that the integral in (3.15) becomes

$$
\begin{aligned}
\int_{\Gamma_j} d\underline{X}\, P_0 P(\underline{X}|\underline{0}) C_j(\underline{X}) B_j(\underline{X}) &= P_0 C^{(1)}_{0j} - \sum_{i=1}^{M} \left[C^{(1)}_{i0} - C^{(1)}_{ij}\right] P_i \int_{\Omega_i} d\underline{\theta} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) \\
&+ \int_{\Gamma_j} d\underline{X} \left\{ \sum_{i=0}^{M} \left[C^{(2)}_{ij} - C^{(2)}_{i0}\right] P_i \int_{\Omega_i} d\underline{\theta} \left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right]^T D \left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right] P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) \right\}
\end{aligned}
$$

where the minus one factor has been absorbed into the last term as indicated by the reverse order of the cost factors. The first two terms on the left-hand side of this last expression do not depend on $\underline{\hat{\theta}}(\underline{X})$ and therefore do not enter into the minimization process. Because of relations (3.25) it follows that $C_{ij}^{(2)} - C_{i0}^{(2)} > 0$ for $j \neq 0$, and therefore our problem reduces to finding that function $\underline{\hat{\theta}}^*(\underline{X})$ which minimizes the third term in the last expression. Let

$$k(\underline{X}, \underline{\hat{\theta}}(\underline{X})) \triangleq \sum_{i=0}^{M} \left[C_{ij}^{(2)} - C_{i0}^{(2)}\right] P_i \int_{\Omega_i} d\underline{\theta} \left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right]^T D \left[\underline{\theta} - \underline{\hat{\theta}}(\underline{X})\right] P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) . \tag{3.27}$$

The function $k(\underline{X}, \underline{\hat{\theta}}(\underline{X}))$ is positive for all $\underline{X}$, and therefore a function $\underline{\hat{\theta}}(\underline{X})$ which minimizes $k(\underline{X}, \underline{\hat{\theta}}(\underline{X}))$ will also minimize $\int_{\Gamma_j} d\underline{X}\ k(\underline{X}, \underline{\hat{\theta}}(\underline{X}))$. Therefore a necessary condition which a minimizing function, $\underline{\hat{\theta}}^*(\underline{X})$, must satisfy is

$$\nabla_{\underline{\hat{\theta}}}\, k(\underline{X}, \underline{\hat{\theta}}(\underline{X})) \Big|_{\underline{\hat{\theta}}(\underline{X}) = \underline{\hat{\theta}}^*(\underline{X})} = 0 . \tag{3.28a}$$

where $\nabla_{\hat{\theta}}$ is the gradient operator, that is, a column vector of first partial derivatives with respect to the components of $\underline{\hat{\theta}}(\underline{X})$. If the Hessian matrix, i.e., the matrix of second partial derivatives with respect to $\underline{\hat{\theta}}(\underline{X})$ evaluated at $\underline{\hat{\theta}}^*(\underline{X})$ is positive definite then the function, $\underline{\hat{\theta}}^*(\underline{X})$, determined from condition (3.28a) is the function which minimizes $k(\underline{X}, \underline{\hat{\theta}}(\underline{X}))$. Therefore, if the following matrix:

$$\nabla_{\underline{\hat{\theta}}} (\nabla_{\underline{\hat{\theta}}})^T\, k(\underline{X}, \underline{\hat{\theta}}(\underline{X})) \Big|_{\underline{\hat{\theta}} = \underline{\hat{\theta}}^*} \tag{3.28b}$$

is positive definite, then the function determined by (3.28a) is indeed a minimizing function. Substitution of (3.27) into (3.28b) results in

$$\nabla_{\underline{\hat{\theta}}} (\nabla_{\hat{\theta}})^T\, k(\underline{X}, \underline{\hat{\theta}}(\underline{X})) \Big|_{\underline{\hat{\theta}} = \underline{\hat{\theta}}^*} = 2 \left[\sum_{i=0}^{M} (C_{ij}^{(2)} - C_{i0}^{(2)}) P_i \int_{\Omega_i} d\underline{\theta}\ P(\underline{X} | \underline{\theta}) P_i(\underline{\theta})\right] D .$$

Since D is by definition positive definite and the scalar factor multiplying it is positive we see that the Hessian is positive definite, and therefore the solution of (3.28a) will minimize $\int_{\Gamma_j} d\underline{X}\, k(\underline{X}, \hat{\underline{\theta}}(\underline{X}))$.

Substitution of (3.27) into (3.28a) and noticing that

$$\nabla_{\underline{\theta}} \left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]^T D\left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right] = -2D\left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right] \qquad (3.29)$$

we find that the function $\hat{\underline{\theta}}_j^*(\underline{X})$ is a solution of

$$\sum_{i=0}^{M} \left[C_{ij}^{(2)} - C_{i0}^{(2)}\right] P_i \int_{\Omega_i} d\underline{\theta} D\left[\underline{\theta} - \hat{\underline{\theta}}_j^*(\underline{X})\right] P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) = 0 .$$

Solving for $\hat{\underline{\theta}}_j^*(\underline{X})$ we see that

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \frac{\sum_{i=0}^{M}\left[C_{ij}^{(2)} - C_{i0}^{(2)}\right] P_i \int_{\Omega_i} \underline{\theta} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}}{\sum_{i=0}^{M}\left[C_{ij}^{(2)} - C_{i0}^{(2)}\right] P_i \int_{\Omega_i} d\underline{\theta} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta})}$$

which, if we remember that $P_0(\underline{\theta}) = \delta(\underline{\theta} - \underline{0})$ and $C_{00}^{(2)} = 0$, can also be written as:

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \sum_{i=1}^{M} \frac{\left[\dfrac{C_{ij}^{(2)} - C_{i0}^{(2)}}{C_{0j}^{(2)}}\right]}{1 + \sum_{i=1}^{M}\left[\dfrac{C_{ij}^{(2)} - C_{i0}^{(2)}}{C_{0j}^{(2)}}\right] \Lambda_i(\underline{X})} \hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \qquad (3.30)$$

where $\Lambda_i(\underline{X})$ is the average likelihood ratio and $\hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1}$ is the least square or minimum variance estimator for a signal of class Ω_i in the presence of no uncertainty, i.e.,

$$\Lambda_i(\underline{X}) \triangleq \frac{P_i \int_{\Omega_i} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}}{P_0 P(\underline{X} | \underline{0})} \qquad (3.31a)$$

and

$$\hat{\theta}_i^*(\underline{X})_{P_i=1} \triangleq \frac{\int_{\Omega_i} \underline{\theta} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}}{\int_{\Omega_i} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}} \ . \qquad (3.31b)$$

If we now assume that $C_{i0}^{(2)} = 0$ and $C_{ij}^{(2)} = C_{0j}^{(2)} > 0$ for $i = 1, \ldots, M$ then (22) reduces to the simpler expression

$$\underline{\hat{\theta}}_j^*(\underline{X})_{P_j < 1} = \sum_{i=1}^{M} \frac{\Lambda_i(\underline{X})}{1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})} \, \underline{\hat{\theta}}_i^*(\underline{X})_{P_i = 1} \qquad (3.32)$$

which may also be expressed in the following more revealing form

$$\underline{\hat{\theta}}_j^*(\underline{X})_{P_j < 1} = \sum_{i=1}^{M} P(H_i | \underline{X}) \underline{\hat{\theta}}_i^*(\underline{X})_{P_i = 1} \qquad (3.33)$$

where $P(H_i | \underline{X})$ is the a posteriori probability of hypothesis H_i being true. It may easily be shown that $P(H_i | \underline{X}) = \Lambda_i(\underline{X}) \left[1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})\right]^{-1}$ by several applications of Bayes rule and the definition of a marginal probability density. From Bayes rule and our notational convention that $P_i = P(H_i)$ we have

$$P(\underline{X}, \underline{\theta}, H_i) = P(\underline{X}, \underline{\theta} | H_i) P_i = P(\underline{X} | \underline{\theta}, H_i) P(\underline{\theta} | H_i) P_i \ . \qquad (3.34a)$$

The definition of a marginal density results in the first equality in

(3. 34b) below and the second follows from (3. 34a) and the convention that $P(\underline{\theta}|H_i) = P_i(\underline{\theta})$. If H_i is given, then $\theta \epsilon \Omega_i$ and we may restrict the range of integration to the Ω_i subset of Ω and hence the last equality in (3. 34b).

$$P_i P(\underline{X}|H_i) = P_i \int_{\Omega} P(\underline{X}, \underline{\theta}|H_i) d\underline{\theta} = P_i \int_{\Omega} P(\underline{X}|\underline{\theta}, H_i) P_i(\underline{\theta}) = P_i \int_{\Omega_i} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \tag{3. 34b}$$

The relations below follow from (3. 34b) and noticing that $P_o(\underline{\theta}) = \delta(\underline{\theta} - \underline{0})$.

$$P(\underline{X}) = \sum_{i=0}^{M} P(\underline{X}|H_i) P(H_i) = \sum_{i=0}^{M} P_i \int_{\Omega_i} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} = P_0 P(\underline{X}|\underline{0}) + \sum_{i=1}^{M} P_i \int_{\Omega_i} d\underline{\theta} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) \; . \tag{3. 34c}$$

Again from Bayes rule we see that the first equality in (3. 34d) holds and the second follows by substitution from (3. 34b) and (3. 34c).

$$P(H_i|\underline{X}) = \frac{P(\underline{X}|H_i) P(H_i)}{P(\underline{X})} = \frac{P_i \int_{\Omega_i} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}}{P_0 P(\underline{X}|\underline{0}) + \sum_{i=1}^{M} P_i \int_{\Omega_i} P(\underline{X}|\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}} = \frac{\Lambda_i(\underline{X})}{1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})} \tag{3. 34d}$$

The last equality is clear from the definition of $\Lambda_i(\underline{X})$.

So far in subsection 3. 3 we have determined two estimator structures, namely (3. 30) and (3. 32), for the case where the additional terms in the modified costs of classification are quadratic in the difference $(\underline{\theta}^T - \hat{\underline{\theta}}(\underline{X}))$. The estimator given in (3. 32) is not as complicated as that given in (3. 30), and has the reasonable interpretation (3. 33) of being a weighted sum of least squares estimators for the case of no uncertainty with respect to the various hypotheses, with the weights being the a posteriori probability of the respective hypothesis. From its symmetry, the estimator given by (3. 32) is seen to be independent of j, and hence whatever decision $\gamma_j (j \neq 0)$ is made the same estimate is presented at the output and therefore for this case we could have dropped the subscript j on $\hat{\underline{\theta}}_j^*(\underline{X})$. This is not the case with (3. 30) derived under less restrictive assumptions on the cost assignments. It is readily verified, for example, that if $C_{0j}^{(2)} \neq C_{0k}^{(2)}$, then $\hat{\underline{\theta}}_j^*(\underline{X})_{P_j < 1}$ will differ from $\hat{\underline{\theta}}_k^*(\underline{X})_{P_k < 1}$ if $j \neq k$ and $j, k \neq 0$. Therefore, an important advantage of choosing costs that lead to an estimator as given by (3. 32) is that a joint detection and estimation device need only be constructed with one estimator structure instead of M.

After the determination of the optimum estimator structure, which in this example is either (3. 30) or (3. 32), the last step in the determination of the final optimum decision rule is the substitution of these estimators into the originally obtained decision rule (3. 13b), which for convenience is rewritten below in a slightly altered form after noticing that $C_j(\underline{X}) > 0$ for all $\underline{X}$ and $j \neq 0$, and including the dependence of B_j and C_j on $\hat{\underline{\theta}}_j(X)$:

$$\text{Decide } \gamma_j \text{ or } P(\gamma_j | \underline{X}) = 1 \text{ if (a) } C_j(\underline{X}, \hat{\underline{\theta}}_j(X))B_j(\underline{X}, \hat{\underline{\theta}}_j(\underline{X})) < 0$$

$$\text{(b) } C_j(\underline{X}, \hat{\underline{\theta}}_j(\underline{X}))B_j(\underline{X}, \hat{\underline{\theta}}_j(\underline{X}) \leqq C_i(\underline{X}, \hat{\underline{\theta}}_i(\underline{X}))\cdot B_i(\underline{X}, \hat{\underline{\theta}}_i(\underline{X}))$$

$$\text{for all } i.$$

With $\hat{\underline{\theta}}_j^*(X)$ substituted for $\hat{\underline{\theta}}_j(\underline{X})$ the final optimum detection rule becomes

$$\text{Decide } \gamma_j \text{ or } P^*(\gamma_j | \underline{X}) = 1 \text{ if (a) } C_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X}))B_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X})) < 0$$

$$\text{(b) } C_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X}))B_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X})) \leqq C_i(\underline{X}, \hat{\underline{\theta}}_i^*(\underline{X})) \cdot B_i(\underline{X}, \hat{\underline{\theta}}_i^*(\underline{X}))$$

$$\text{for all } i. \qquad (3.\ 35)$$

It is seen from (3.35) that the important factor necessary for the explicit determination of the optimum detector structure is $C_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X})B_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X}))$. This factor is given below in (3.36a) for the case when the cost assignments are such that they lead to the estimator given by (3.32) plus the assumption that $C_{ij}^{(2)} = 1$, and can be derived in a straightforward manner after some tedious algebra. There are two reasons for presenting this explicit result at this point: (1) it illustrates the typical complex nonlinear nature of the detector structures which result from joint detection and estimation schemes, and (2) it will be useful in subsection 3.6 where we shall investigate this same example for the case of overlapping hypothesis classes and consequently where it is possible to employ Gaussian prior densities.

$$C_j(\underline{X})B_j(\underline{X}, \hat{\underline{\theta}}_j^*(\underline{X})) = C_{0j}^{(1)} + \sum_{i=1}^{M}\left[C_{ij}^{(1)} - C_{i0}^{(1)}\right]\Lambda_i(\underline{X}) + \sum_{i=1}^{M}\Lambda_i(\underline{X})\widehat{(\underline{\theta}^T\underline{\theta})}_i^*(\underline{X})_{P_i=1}$$
$$- \left(1 + \sum_{i=1}^{M}\Lambda_i(\underline{X})\right)\sum_{i=1}^{M}\sum_{k=1}^{M} P(H_i \mid \underline{X})P(H_k \mid \underline{X})\left[\hat{\underline{\theta}}_k^*(\underline{X})_{P_k=1}\right]^T\hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \tag{3.36a}$$

where

$$\widehat{(\underline{\theta}^T\underline{\theta})}_i^*(\underline{X})_{P_i=1} \triangleq \frac{\int_{\Omega_i}\underline{\theta}^T\underline{\theta}P(\underline{X} \mid \underline{\theta})P_i(\underline{\theta})d\underline{\theta}}{\int_{\Omega_i}P(\underline{X} \mid \underline{\theta})P_i(\underline{\theta})d\underline{\theta}} \tag{3.36b}$$

3.3a Binary Special Cases

Case 1. Pure estimation. This special case considers the problem of uncoupled estimation and detection where we are only interested in estimation operation. The costs of detection, the $C_{ij}^{(1)}$ factors in (3.17), are therefore assumed to be equal to zero. Again $i, j = 0, 1$ since we are interested in the binary "on-off" case, i.e., the observation interval contains noise alone or signal plus noise. If we assume the cost function has the following form:

$$C(\underline{\theta}, \hat{\underline{\theta}}(\underline{X})) = C_{ij}^{(2)}(\underline{\theta}-\hat{\underline{\theta}})^T D(\underline{\theta}-\hat{\underline{\theta}})$$

and that $C_{00}^{(2)} = 0$ then (3.30) yields the optimum estimator structure

$$\hat{\underline{\theta}}_1^*(\underline{X})_{P_1<1} = \frac{\left[\dfrac{C_{11}^{(2)} - C_{10}^{(2)}}{C_{01}^{(2)}}\right]\Lambda(\underline{X})}{1 + \left[\dfrac{C_{11}^{(2)} - C_{10}^{(2)}}{C_{01}^{(2)}}\right]\Lambda(\underline{X})} \hat{\underline{\theta}}_1^*(\underline{X})_{P_1 = 1} \quad (3.37a)$$

Furthermore, if we assume that $C_{10}^{(2)} = 0$ and $C_{11}^{(2)} = C_{01}^{(2)}$ then (3.37a) reduces to

$$\hat{\underline{\theta}}_1^*(\underline{X})_{P_1<1} = \frac{\Lambda(X)}{1+\Lambda(\underline{X})} \hat{\underline{\theta}}_1^*(\underline{X})_{P_1=1} = P(H_1 | \underline{X})\hat{\underline{\theta}}_1^*(\underline{X})_{P_1=1} \quad (3.37b)$$

which is identical to result 3.7 in reference (4).

Case 2. Strong coupling, binary "on-off". In this case we shall assume strong coupling between the detection and estimation operation. The cost of deciding γ_j where $\underline{\theta}\epsilon\Omega_i (i, j = 0, 1)$ is assumed to have the following form

$$C(\underline{\theta}, \gamma_j) = C_{ij}^{(1)} + C_{ij}^{(2)}\left[\underline{\theta}-\hat{\underline{\theta}}(\underline{X})\right]^T D\left[\underline{\theta}-\hat{\underline{\theta}}(\underline{X})\right]$$

and that $C(\underline{\theta}=\underline{0}, \gamma_0) = 0$. Then, depending on what assumptions are made concerning the cost factors, $C_{ij}^{(2)}$, we obtain either (3.37a) or (3.37b) as estimator structures. Furthermore if we assume that $C_{10}^{(2)} = 0$ and that $C_{11}^{(2)} = C_{01}^{(2)} = C$, then the estimator structure is given by (3.37b), and the detector structure is given by (3.36a), which for this case becomes:

$$C_1(\underline{X}, \underline{\hat{\theta}}_1^*(\underline{X})_{P_1<1})B_1(\underline{X}, \underline{\hat{\theta}}_1^*(\underline{X})_{P_1<1}) = C - C\Lambda(\underline{X}) + \widehat{(\underline{\theta}^T\underline{\theta})}^*(\underline{X})_{P_1<1}\Lambda(\underline{X}) - \frac{\Lambda^2(X)}{1+\Lambda(\underline{X})}(\underline{\hat{\theta}}_1^*(\underline{X}))^T\underline{\hat{\theta}}_1^*(\underline{X}) \qquad (3.38)$$

Case 3. Strong coupling, binary: $\underline{\theta}_1 \neq \underline{0}$ versus $\underline{\theta}_2 \neq \underline{0}$. In this case the occurrence of the null signal is excluded and therefore P_0 has to be set equal to zero in our previous results, e.g., the estimator relations (3.30) or (3.32). Also because the null hypothesis is excluded $C_{i0}^{(2)} = 0$ for all i. Finally, the binary assumption means that i, j = 1, 2. The cost function for this example will be of the form

$$C(\underline{\theta}, \gamma_j) = C_{ij}^{(1)} + C_{ij}^{(2)}\left[\underline{\theta}-\underline{\hat{\theta}}_j(\underline{X})\right]^T D\left[\underline{\theta}-\underline{\hat{\theta}}_j(\underline{X})\right] \qquad (\underline{\theta}\in\Omega_i).$$

The estimator structure for this example follows directly upon substituting the assumed cost function into (3.23), or by applying the above restrictions to (3.30). In either case the estimator structure which results is given by:

$$\underline{\hat{\theta}}_j(\underline{X})_{P_j<1} = \sum_{i=1}^{2} \frac{C_{ij}^{(2)}}{\sum_{k=1}^{2} C_{kj}^{(2)}\Lambda_{ki}(\underline{X})}\underline{\hat{\theta}}_i^*(\underline{X})_{P_i=1} = \sum_{i=1}^{2} \frac{1}{\sum_{i=1}^{2} \frac{C_{kj}^{(2)}}{C_{ij}^{(2)}}\Lambda_{ki}(\underline{X})}\underline{\hat{\theta}}_i^*(\underline{X})_{P_i=1}$$

(3.40a)

where $\hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1}$ is defined by (3.31b) and $\Lambda_{ki}(\underline{X})$ is given by

$$\Lambda_{ki}(\underline{X}) \triangleq \frac{P_k \int_{\Omega_k} P(\underline{X} | \underline{\theta}) P_k(\underline{\theta}) d\underline{\theta}}{P_i \int_{\Omega_i} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}} \quad . \tag{3.41}$$

It can easily be verified that: $\Lambda_{ii}(\underline{X}) = 1$ and $\Lambda_{ik}(\underline{X})\Lambda_{ki}(\underline{X}) = 1$. If it can be assumed that $C_{ij}^{(2)} = C_{kj}^{(2)}$ for all i and k then (3.40a) reduces to

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \sum_{i=1}^{2} \frac{1}{\sum_{ki}^{2} \Lambda_{ki}(\underline{X})} \hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \tag{3.40b}$$

which may also be expressed as

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \sum_{i=1}^{2} P(H_i | \underline{X}) \hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \tag{3.40c}$$

because $P(H_i | \underline{X}) = (\sum_{k=1}^{2} \Lambda_{ki}(\underline{X}))^{-1}$, as is easily verified by setting $P_0 = 0$ in the set of relations (3.34d). The detector structure in this case is determined by the factor $B_{12}^*(\underline{X})$, defined immediately after Equation (3.23). The function $B_{12}^*(\underline{X})$ will be determined explicitly in subsection 3.6 for the case where $\underline{\theta} = \underline{S}$.

3.4 Example: Simple Cost Function — Nonoverlapping Hypothesis Classes.

In subsection 3.3 we used a quadratic cost function to illustrate the general theory of joint detection and estimation under multiple hypotheses. Another important cost assignment which is often used in estimation theory is the simple cost function (SCF), and the reason for its importance is that it leads to unconditional maximum likelihood estimators. This subsection is devoted to a discussion of joint detection and estimation under multiple hypotheses when a simple cost function assignment is assumed. The simple cost function used throughout this subsection is given by:

$$C(\underline{\theta}, \gamma_j) = C_{ij}^{(1)} + C_i^{(2)} (A_i^{(2)} - \delta(\hat{\underline{\theta}} - \underline{\theta})) \qquad (\underline{\theta} \epsilon \Omega_i) \tag{3.41a}$$

where

$$\delta(\hat{\underline{\theta}}(\underline{X}) - \underline{\theta}) = \prod_{k=1}^{n} \delta(\hat{\theta}_k(\underline{X}) - \theta_k) \ . \tag{3.41b}$$

In this last relation, $\delta(\cdot)$ is the Dirac δ-function, $\hat{\theta}_k(\underline{X})$ is the k^{th} component of $\hat{\underline{\theta}}(\underline{X})$, and $\hat{\underline{\theta}}(\underline{X})$ and $\underline{\theta}$ are n-component vectors. The factors $C_{ij}^{(1)}$ are the usual constant cost assignments, i. e., the cost of deciding that a signal of class Ω_j is present when really one of class Ω_i is present. The factors $C_i^{(2)}$ and $A_i^{(2)}$ are related to the cost of correct and incorrect estimation decisions and a discussion of them is found in reference (2), chapter 23.

Substitution of (3.41a) into relation (3.8) for the average risk R_{D+E} results in:

$$\begin{aligned} R_{D+E} = & \int_\Gamma d\underline{X} \Big[\sum_{i=0}^{M} (C_{i0}^{(1)} + C_i^{(2)} A_i^{(2)}) P_i \int_{\Omega_i} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \Big] \\ & + \int d\underline{X} \Big[\sum_{j=1}^{M} P(\gamma_j | \underline{X}) \Big\{ \sum_{i=0}^{M} (C_{ij}^{(1)} - C_{i0}^{(1)}) P_i \int_{\Omega_i} P(\underline{X} | \underline{\theta}) P_i(\underline{\theta}) d\underline{\theta} \Big\} \Big] \\ & - \int_\Gamma d\underline{X} \Big[C_0^{(2)} P_0 P(\underline{X} | \underline{0}) \, \delta(\hat{\underline{\theta}}) + \sum_{i=1}^{M} C_i^{(2)} P_i P(\underline{X} | \hat{\underline{\theta}}) P_i(\hat{\underline{\theta}}) \Big] \end{aligned} \tag{3.42}$$

The first term on the right side of (3.42) is a positive constant independent of the decision rule or the estimator. The second term depends only on the decision rule and not on the estimator, whereas the third term depends on the estimator, $\underline{\hat{\theta}}(\underline{X})$, and not on the decision rule. The integrand of the third term is always positive and therefore the integral will also be positive. The integral is preceded by a minus sign and therefore that part of the average risk will be minimized if we determine $\underline{\hat{\theta}}^*(\underline{X})$ in the following way:

$$\max_{\underline{\hat{\theta}}(\underline{X})} \int_\Gamma d\underline{X}\left[C_0^{(2)} P_0 P(\underline{X}|\underline{0})\delta(\underline{\hat{\theta}}) + \sum_{i=1}^{M} C_i^{(2)} P_i P(\underline{X}|\underline{\hat{\theta}})P_i(\underline{\hat{\theta}})\right] \quad (3.43)$$

The estimator, $\underline{\hat{\theta}}^*(\underline{X})$, is not the usual unconditional maximum-likelihood estimator but it may be thought of as a generalized version of maximum likelihood estimation (4). If the factors $C_i^{(2)}$ are all equal to C_0 then condition (3.43) becomes:

$$\max_{\underline{\hat{\theta}}(\underline{X})} \int_\Gamma d\underline{X}\left[P_0 P(\underline{X}|\underline{0})\,\delta(\underline{\hat{\theta}}) + \sum_{i=1}^{M} P_i P(\underline{X}|\underline{\hat{\theta}})P_i(\underline{\hat{\theta}})\right]. \quad (3.44)$$

In the binary "on-off" case where either signal plus noise or noise alone is present during the observation interval and M = 1, (3.44) reduces to

$$\max_{\underline{\hat{\theta}}(\underline{X})} \int_\Gamma d\underline{X}\left[P_0 P(\underline{X}|\underline{0})\,\delta(\underline{\hat{\theta}}) + P_1 P(\underline{X}|\underline{\hat{\theta}})P_1(\underline{\hat{\theta}})\right]$$

which is equivalent to relation (3.15) of reference (4).
The optimum decision rule for this case is given by:

decide γ_j or $P(\gamma_j|\underline{X}) = 1$ if (a) $D_j(\underline{X}) < 0$

(b) $D_j(\underline{X}) \leqq D_k(\underline{X})$ for all $k = 1, \ldots, M$

(3.45)

otherwise decide γ_0, where

$$D_j(\underline{X}) \triangleq \sum_{i=0}^{M}(C_{ij}^{(1)} - C_{i0}^{(1)})P_i\int_{\Omega_i} P(\underline{X}|\underline{\theta})P_i(\underline{\theta})d\underline{\theta} .$$

3.5 Overlapping Hypothesis Classes

In subsection 3.2 we investigated the "one-shot" joint detection and estimation problem for the case of non-overlapping hypothesis classes. The requirement that the hypothesis classes be nonoverlapping will now be relaxed and the corresponding problem studied. The only significant difference between the two cases is the interpretation of the cost function and one will be chosen which has reasonable properties and whose form is such, that the resulting optimum detector and estimator structures are identical with those oı the previous subsections.

In the case of overlapping hypothesis classes the same signal or descriptive signal parameter may belong to two or more hypothesis classes, and therefore it is no longer true that an incorrect decision has been made when $\underline{\theta} \varepsilon \Omega_i$ and it is decided that $H_j(i \neq j)$ is true. Therefore $C(\underline{\theta}, \gamma_j)$ for $\underline{\theta} \varepsilon \Omega_i$ can no longer be interpreted as the cost of an incorrect decision, however it can be interpreted as the cost of assigning $\underline{\theta} \varepsilon \Omega_i$ to hypothesis class Ω_j.

A reasonable cost function suggested by Ogg (10), and discussed by Middleton (5) p. 32, for the overlapping signal class case in Bayes detection theory is as follows:

$$C(\underline{\theta}, \gamma_j) = \frac{\sum_{i=0}^{M} C_{ij}^{(1)} P_i P_i(\underline{\theta})}{\sum_{i=0}^{M} P_i P_i(\underline{\theta})} . \tag{3.46}$$

Clearly this cost function reduces to the usual constant cost of misclassification when the hypothesis classes are nonoverlapping. Also, $C(\underline{\theta}, \gamma_j)$ has the property of being continuous on the prior densities. When the hypothesis classes overlap, the cost function (3.46) has the desirable property, which on the average seems plausible, that a less probable decision costs more than a more probable one, i.e., if $i \neq j$, $\underline{\theta}$ is contained in both Ω_i and Ω_j, and $P_i P_i(\underline{\theta}) < P_j P_j(\underline{\theta})$ then $C(\underline{\theta}, \gamma_j) < C(\underline{\theta}, \gamma_i)$.

The cost function which will be used in this subsection for the case of overlapping hypothesis classes is a direct analog of (3.46) and is given by:

$$C(\underline{\theta}, \gamma_j) = \frac{\sum_{i=0}^{M} \left[C_{ij}^{(1)} + C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))\right] P_i P_i(\underline{\theta})}{\sum_{i=0}^{M} P_i P_i(\underline{\theta})} \tag{3.47}$$

where $C_{ij}^{(1)}$ is the cost of classifying $\underline{\theta}$ in Ω_j when $\underline{\theta} \in \Omega_i$, and $C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))$ is the additional cost of presenting $\underline{\hat{\theta}}(\underline{X})$ as an estimate of a parameter of class Ω_j when $\underline{\theta} \in \Omega_i$.

Again an average risk may be defined and it is the same expression as that given previously in (3.2) and (3.5). If the structure of the decision space Δ is the same as in subsection 3.2 and the cost assignment is assumed to be of the form (3.47), then (3.5) becomes:

$$R_{D+E} = \int_{\Gamma} d\underline{X} \left\{ \sum_{j=0}^{M} P(\gamma_j | \underline{X}) \int_{\Omega} d\underline{\theta} P(\underline{X} | \underline{\theta}) \left[\frac{\sum_{i=0}^{M} \left[C_{ij}^{(1)} + C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))\right] P_i P_i(\underline{\theta})}{\sum_{i=0}^{M} P_i P_i(\underline{\theta})} \right] \sum_{i=0}^{M} P_i P_i(\underline{\theta}) \right\}$$

or

$$R_{D+E} = \int_\Gamma d\underline{X}\left\{\sum_{j=0}^{M}{}' P(\gamma_j \mid \underline{X})\left[\sum_{i=0}^{M}{}' P_i \int_\Omega d\underline{\theta}\left[C_{ij}^{(1)}+C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))\right] P(\underline{X} \mid \underline{\theta}) P_i(\underline{\theta})\right]\right\}. \tag{3.48}$$

The last expression for the average risk is identical with (3.6) except for $C(\underline{\theta}, \gamma_j)$ being expressed in more detail and the integral extended over the entire signal parameter space Ω, rather than just over the subspaces Ω_i. Extending the integral to Ω takes into account the possibility that the subspaces, Ω_i, may overlap and perhaps all be identical with Ω. Therefore the special choice of cost function given by (3.47) leads in the overlapping hypothesis class case to an average risk identical with that for the non-overlapping case, provided $C(\underline{\theta}, \gamma_j)$ in (3.6) is identified with $C_{ij}^{(1)} + C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))$ when $\underline{\theta} \epsilon \Omega_i$. Let

$$C'(\underline{\theta}, \gamma_j) \triangleq C_{ij}^{(1)} + C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X})) \qquad (\underline{\theta} \epsilon \Omega_i)$$

and if we remember that the objective is again to find the decision rule which minimizes the average risk, it is clear that the optimum decision rule obtained in the overlapping case is identical with that obtained in the nonoverlapping case provided we replace $C(\underline{\theta}, \gamma_j)$ by $C'(\underline{\theta}, \gamma_j)$. Assuming that $C'(\underline{\theta}, \gamma_0) = 0$ when $\underline{\theta} \epsilon \Omega_0$ we see that

$$C_j(\underline{X}, \underline{\hat{\theta}}(X)) = (C_{0j}^{(1)} - C_{00}^{(1)}) + C_{0j}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(X))$$

$$B_j(\underline{X}, \underline{\hat{\theta}}(\underline{X})) = 1 - \sum_{i=1}^{M} \frac{\mu_i \int_\Omega d\underline{\theta}\left[C'(\underline{\theta}, \gamma_0) - C'(\underline{\theta}, \gamma_j)\right] P(\underline{X} \mid \underline{\theta}) P_i(\underline{\theta})}{C_j(\underline{X}, \underline{\hat{\theta}}(\underline{X})) P(\underline{X} \mid \underline{0})}$$

where $C_{ij}^{(1)}$ and $C_{ij}^{(2)}(\underline{\theta}, \underline{\hat{\theta}}(\underline{X}))$ are determined by (3.47) and the optimum estimator is determined from the following condition:

$$\min_{\hat{\underline{\theta}}(\underline{X})} \int_{\Gamma_j} d\underline{X}\, P_0 P(\underline{X}|\,\underline{0})\, C_j(\underline{X}, \hat{\underline{\theta}}(\underline{X})) B_j(\underline{X}, \hat{\underline{\theta}}(\underline{X})) \qquad (3.49a)$$

and the optimum detector structure is given by:

$$\text{Decide } \gamma_j \text{ or } P^*(\gamma_j|\,\underline{X}) = 1 \text{ if (a) } C_j(\underline{X}, \hat{\underline{\theta}}^*_j(\underline{X}))B_j(\underline{X}, \hat{\underline{\theta}}^*_j(\underline{X})) < 0$$

$$\text{(b) } C_j(\underline{X}, \hat{\underline{\theta}}^*_j(\underline{X}))B_j(\underline{X}, \hat{\underline{\theta}}^*_j(\underline{X}) \leqq C_i(\underline{X}, \hat{\underline{\theta}}^*_i(\underline{X}))\cdot B_i(\underline{X}, \hat{\underline{\theta}}^*_i(\underline{X}))$$

$$\text{for all } i. \qquad (3.49b)$$

3.6 Example: Quadratic Cost Function - Overlapping Hypothesis Classes.

The results of the previous subsection showed that when the hypothesis classes overlap it is possible to choose a reasonable cost function so that the resulting detector and estimator structures are identical with those obtained in the nonoverlapping case. As mentioned previously this allows us to use convenient prior probability functions, such as the Gaussian, in the analysis of Bayes detection and estimation systems.

With the above comments in mind it is clear that if we choose the factor $C^{(2)}_{ij}(\underline{\theta}, \hat{\underline{\theta}})$ in (3.47), to have the following quadratic dependence on $\underline{\theta}$ and $\hat{\underline{\theta}}(\underline{X})$:

$$C^{(2)}_{ij}(\underline{\theta}, \hat{\underline{\theta}}(\underline{X})) \triangleq c^{(2)}_{ij}\left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right]^T D\left[\underline{\theta} - \hat{\underline{\theta}}(\underline{X})\right], \qquad (3.50)$$

then the resulting estimator and detector structures for the overlapping hypothesis class case are identical with the specific forms obtained in the quadratic cost function example discussed in subsection 3.3. Therefore the following estimator structures are obtained in the overlapping signal class case when the cost assignment is given by (3.47) and $C_{ij}^{(2)}(\underline{\theta}, \hat{\underline{\theta}}(\underline{X}))$ is given by (3.50):

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \sum_{i=1}^{M} \frac{\left[\frac{C_{ij}^{(2)} - C_{i0}^{(2)}}{C_{0j}^{(2)}}\right]\Lambda_i(\underline{X})}{1 + \sum_{i=1}^{M}\left[\frac{C_{ij}^{(2)} - C_{i0}^{(2)}}{C_{0j}^{(2)}}\right]\Lambda_i(\underline{X})} \hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \qquad (3.51a)$$

where $C_{0j}^{(2)} \geqq C_{jj}^{(2)} > C_{j0}^{(2)} \geqq C_{00}^{(2)} = 0$ $(j \neq 0)$ and $C_{ij}^{(2)} \geqq C_{jj}^{(2)}$ $(i, j = 1, 2, \ldots, M)$. If in addition, $C_{i0}^{(2)} = 0$ and $C_{ij}^{(2)} = C_{0j}^{(2)}$ for $i, j \neq 0$, then

$$\hat{\underline{\theta}}_j^*(\underline{X})_{P_j<1} = \sum_{i=1}^{M} \frac{\Lambda_i(\underline{X})}{1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})} \hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} \qquad (3.51b)$$

The generalized likelihood ratio, $\Lambda_i(\underline{X})$, and the estimators in the absence of uncertainty, $\hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1}$, are given by

$$\Lambda_i(\underline{X}) = \frac{P_i \int_{\Omega} d\underline{\theta} P(\underline{X}|\,\underline{\theta}) P_i(\underline{\theta})}{P_0 P(\underline{X}|\,\underline{0})} \qquad (3.52a)$$

$$\hat{\underline{\theta}}_i^*(\underline{X})_{P_i=1} = \frac{\int_{\Omega} \underline{\theta} P(\underline{X}|\,\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}}{\int_{\Omega} P(\underline{X}|\,\underline{\theta}) P_i(\underline{\theta}) d\underline{\theta}} \qquad (3.52b)$$

where now the range of integration has been extended to the whole signal parameter space. Furthermore, if we assume that $C_{ij}^{(2)} = 1$, then the estimator structure is still given by (3.51b) and the detector structure is given by (3.36a) which is rewritten below for the case where the parameter vector is the waveform vector, i.e., $\underline{\theta} = \underline{S}$.

$$C_j(X,\hat{\underline{S}}_j^*(\underline{X}))B_j(\underline{X},\hat{\underline{S}}_j^*(\underline{X})) = C_{0j}^{(1)} + \sum_{i=1}^{M}\left[C_{ij}^{(1)} - C_{i0}^{(1)}\right]\Lambda_i(\underline{X}) + \sum_{i=1}^{M}\Lambda_i(\underline{X})\widehat{(\underline{S}^T\underline{S})}_i^*(\underline{X})_{P_i=1}$$

$$-\left[1+\sum_{i=1}^{M}\Lambda_i(\underline{X})\right]\sum_{i=1}^{M}\sum_{k=1}^{M}P(H_i\,|\,\underline{X})P(H_k\,|\,\underline{X})(\hat{\underline{S}}_k^*(\underline{X})_{P_k=1})^T\hat{\underline{S}}_i^*(\underline{X})_{P_i=1} \tag{3.53a}$$

where

$$\widehat{(\underline{S}^T\underline{S})}_i^*(\underline{X})_{P_{i=1}} \triangleq \frac{\int_\Omega \underline{S}^T\underline{S}\, P(\underline{X}\,|\,\underline{S})P_i(\underline{S})d\underline{S}}{\int_\Omega P(\underline{X}\,|\,\underline{S})\, P_i(\underline{S})d\underline{S}} \,. \tag{3.53b}$$

Before, in subsection 3.3, we stopped with the analysis of the detector structure with equations (3.36) because at that point specific assumptions concerning the prior probability density functions were required. Now that the prior probability densities, $P_i(\underline{\theta})$ or $P_i(\underline{S})$, are no longer restricted to a subspace of the signal parameter space, Ω, we may carry the analysis a step further. For example, if Gaussian zero mean noise is assumed to be present and that $P_i(\underline{S})$ are also Gaussian, i.e.,

$$P_i(\underline{S}) = G(\underline{S};\underline{m}_i, K_i) \triangleq (2\pi)^{-\frac{N}{2}}(\det K_i)^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}(\underline{S}-\underline{m}_i)^T K_i^{-1}(\underline{S}-\underline{m}_i)\right] \tag{3.54a}$$

$$P(\underline{X}\,|\,\underline{S}) = G(\underline{X};\underline{S}, K_n) \tag{3.54b}$$

then we may evaluate $\widehat{(\underline{S}^T\underline{S})}_i^*(\underline{S})_{P_i=1}$ and thereby completely determine the detectors explicit dependence on $\underline{X}$, the received data. Both the numerator and the denominator of (3.53b) contain the factor $P(\underline{X}\,|\,\underline{S})P_i(\underline{S})$ which under our Gaussian assumptions (3.54) becomes $G(\underline{X};\underline{S}, K_n)G(\underline{S};\underline{m}_i, K_i)$. This product of Gaussian densities can also be expressed as in (3.55a) below, which can be derived in a straightforward manner.

$$G(\underline{X};\underline{S},K_n)G(\underline{S};\underline{m}_i,K_i) = \frac{G(\underline{X};\underline{0},K_n)G(\underline{m}_i;\underline{0},K_i)}{G(\underline{\hat{S}}_i^*,\underline{0},K_{1i})} G(\underline{S};\underline{\hat{S}}_i^*,K_{1i}) \quad (3.55a)$$

where

$$K_{1i} \triangleq (K_n^{-1} + K_i)^{-1} \quad (3.55b)$$

$$\underline{\hat{S}}_i^* \triangleq K_{1i}(K_n^{-1}\underline{X} + K_i^{-1}\underline{m}_i) \ . \quad (3.55c)$$

The asterisk in (3.55c) indicates that this is an optimum least squares estimate of the waveform vector in the absence of uncertainty and is equivalent to $\underline{\hat{S}}_i^*(\underline{X})_{P_i=1}$ in the quadratic cost function case. Since $G(\underline{S};\underline{\hat{S}}_i^*,K_{1i})$ is a probability density function, its integral over the signal space Ω is one. Therefore, after substitution of (3.55a) into (3.53b), we obtain the result that

$$\widehat{(\underline{S}^T\underline{S})}_i^*(\underline{X})_{P_i=1} = \int_{-\infty}^{\infty}\!\cdots\!\int \underline{S}^T\underline{S}\,G(\underline{S};\underline{\hat{S}}_i^*,K_{1i})\,d\underline{S} \ .$$

From the way in which the elements of the covariance matrix of a Gaussian probability density are defined, it is easy to see that the integral on the left hand side of this last relation is given by:

$$\int_{-\infty}^{\infty}\!\cdots\!\int \underline{S}^T\underline{S}\,G(\underline{S};\underline{S}_i^*,K_{1i})d\underline{S} = Tr(K_{1i}) + (\underline{\hat{S}}_i^*)^T\,\underline{\hat{S}}_i^*$$

where $Tr(K_{1i})$ denotes the trace of the matrix K_{1i}. Therefore

$$\widehat{(\underline{S}^T\underline{S})}_i^*(\underline{X})_{P_i=1} = Tr(K_{1i}) + (\underline{\hat{S}}_i^*)^T\underline{\hat{S}}_i^* \quad (3.56)$$

and if we substitute this expression into (3.53a), the following expression for the detector structure, when $\underline{\theta} = \underline{S}$ is obtained:

$$C_j(\underline{X}, \hat{\underline{S}}_j^*(\underline{X}))B_j(\underline{X}, \hat{\underline{S}}_j^*(\underline{X})) = C_{0j}^{(1)} + \sum_{i=1}^{M} \Big[(C_{ij}^{(1)} - C_{i0}^{(1)}) + Tr(K_{1i}) + (\hat{\underline{S}}_i^*(\underline{X})_{P_i=1})^T \hat{\underline{S}}_i^*(\underline{X})_{P_i=1} \Big] \Lambda_i(\underline{X})$$

$$- \left(1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})\right) \sum_{i=1}^{M} \sum_{k=1}^{M} P(H_i \mid \underline{X}) P(H_k \mid \underline{X}) (\hat{\underline{S}}_k^*(\underline{X})_{P_{k=1}})^T \hat{\underline{S}}_i^*(\underline{X})_{P_i=1} \qquad (3.57)$$

where we have equated $\hat{\underline{S}}_i^*$ with $\hat{\underline{S}}_i^*(\underline{X})_{P_{i=1}}$. Since $P(H_i \mid \underline{X}) = \Lambda_i(\underline{X}) \cdot \left[1 + \sum_{i=1}^{M} \Lambda_i(\underline{X})\right]^{-1}$, we see that the detector structure can be expressed as a function of the generalized likelihood ratios and the least squares estimators of the various signals in the absence of uncertainty. Furthermore, the left hand side of (3.57) can be expressed solely as a function of the generalized likelihood ratios, $\Lambda_i(\underline{X})$ for $i = 1, \ldots, M$, or as a function of the least squares estimators, $\hat{\underline{S}}_i^*(\underline{X})_{P_i=1}$ for $i = 1, \ldots, M$. This last observation follows from some relations discussed by Esposito in (10) and (11). In particular it is shown in (11) that the least squares or minimum variance estimator $\hat{\underline{S}}_i^*(\underline{X})_{P_i=1}$ can be expressed as the following function of $\Lambda_i(\underline{X})$:

$$\hat{\underline{S}}_i^*(\underline{X})_{P_i=1} = K_n \nabla \log \Lambda_i(\underline{X}) \qquad (3.58)$$

where $\nabla \log \Lambda_i(\underline{X})$ is a column vector whose components are $\frac{1}{\Lambda_i} \frac{\partial \Lambda_i}{\partial X_1}, \ldots, \frac{1}{\Lambda_i} \frac{\partial \Lambda_i}{\partial X_N}$. If relation (3.58) is substituted into (3.57) then we have expressed the detector structure solely in terms of the generalized likelihood ratios. Conversely, the generalized likelihood ratios can be expressed as a function of the various least squares estimators, $\hat{\underline{S}}_i^*(\underline{X})_{P_i=1}$, namely

$$\Lambda_i(\underline{X}) = e^{(\underline{X}^T K_n^{-1} \hat{\underline{S}}_i^*(\underline{X})_{P_i=1} + b_i)} \tag{3.59}$$

where $b_i \triangleq C - \int \underline{X}^T K_n^{-1} d\underline{S}_i$ is a bias term (see Eq. 8 in (11)). Substitution into (3.57) of the various generalized likelihood ratios, expressed as functions of the received data and the least squares estimators, as in (3.59), results in a nonlinear detector structure which is only a function of the least squares estimators and the received data. Furthermore, the detector is seen to be a type of estimator - correlator, the basic operations being correlation of the received data with the various least squares estimators in the absence of uncertainty, and cross correlation of these same estimators with themselves.

Although the multiple alternative joint detection and estimation theory discussed in this section assumes a "one-shot" reception situation it could have been considered as one stage in a sequential decision system. It therefore should be noted at this point that estimator - correlator structures, such as the one under discussion, have a built-in adaptive feature and a discussion of this point is to be found in Kailath (18).

4. CONCLUSIONS

This report is primarily concerned with the development of a general formulation of joint detection and estimation under multiple hypotheses based on statistical decision theory. The problem of joint detection and estimation under multiple hypotheses arose as one aspect of the analysis of adaptive communication receivers and pattern recognizers. The problem was investigated under the assumption that data from one observation interval are available, however, the results obtained may be applied to the case when data are received sequentially and this case will be explored in a subsequent report.

If reasonable cost assignments and simple coupling strategies between the detection and estimation processes are assumed, then optimum detector and estimator structures may be derived. In this report we were concerned with detection directed estimation and in the examples the estimation cost of the total cost assignment was either given by a quadratic cost function or a simple cost function. In the case of a quadratic cost assignment reasonable estimators were derived and it was shown that the estimator in the presence of uncertainty was a weighted sum of least squares estimators in the absence of uncertainty. The weighting coefficients were functions of all the generalized likelihood ratios and cost-of-estimation coefficients, $C_{ij}^{(2)}$, and in the case when $C_{ij}^{(2)} = C_{0j}^{(2)}$ $(i=1,\ldots,M)$ the weighting coefficients reduced to the posterior probabilities of the various hypotheses.

Since it has been shown in (16) that Sherman's theorem does not hold for estimators derived under the assumption of a quadratic cost function and in the presence of uncertainty as to whether or not the signal of interest is present, these estimators are rather specialized and hence it is important to consider other cost assignments, the most important of which is the simple cost function. For a simple cost function assignment for the cost of estimation, we found that the optimum estimator could be interpreted as a generalized maximum likelihood estimator. Both the quadratic cost function and the simple cost function resulted in estimators which when specialized to the binary "on-off" case reduced to estimators given in (4).

It is readily evident from a glance at the forms which had to be compared for classification decisions, that the detectors which resulted from coupled estimation and detection where complex nonlinear functions of the data. However, in the important case when $C_{ij}^{(2)} = C_{0j}^{(2)}$ and both the noise and prior densities of the signal parameters are assumed to be Gaussian, we were able to determine the detector structure explicitly. We found that it was a form of correlation detector, consisting of cross correlation between the various minimum variance estimators in the absence of uncertainty and cross correlation of the received data and these same estimators. Conversely, since the minimum variance estimators in the absence of uncertainty can be expressed as the gradient of the relevant generalized likelihood ratios, the detector structure may be expressed completely as a function of of the received data and the various generalized likelihood ratios. Therefore, the possibility exists that in certain cases the complicated detector structures which result from coupled detection and estimation may consist of a reasonable number of operations of the same type, and that perhaps in limiting cases, e.g., low signal-to-noise ratio, they may have implementable forms. Finally, it should be noted that the optimum detector structure was a function of the optimum estimator and therefore, even though we were concerned with detection directed estimation, the estimation operation should logically precede the detection operation.

REFERENCES

1. Middleton, D. and D. Van Meter: Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory, J. Soc. Ind. Appl. Math., 3: 192 (1955), 4: 86 (1956).

2. Middleton, D.: "An Introduction to Statistical Communication Theory," McGraw-Hill Book Company, New York, 1960.

3. Middleton, D. : "Topics in Communication Theory," McGraw-Hill Book Company, New York, 1965.

4. Middleton, D. and Esposito, R., "Simultaneous Optimum Detection and Estimation of Signals in Noise," IEEE Transactions on Information Theory, Vol. IT-14, Number 3, May 1968, pp. 434-444.

5. Helstrom, C. : "Statistical Theory of Signal Detection," Pergamon Press, New York, 1960.

6. Abramson, N., and Braverman, D., "Learning to Recognize Patterns in a Random Environment," IRE Transactions on Information Theory, Vol. IT-8, pp. 58-63, September 1962.

7. Spragins, J., "Learning Without a Teacher," IEEE Transactions on Information Theory, Vol. IT-12, Number 2, April 1966, pp. 223-230.

8. Hancock, J. C., and Wintz, P. A., "Signal Theory Detection," McGraw-Hill Book Company, New York 1966.

9. Ogg, F. C., Jr. : A Note on Bayes Detection of Signals, IRE Transactions on Information Theory, IT-10, Number 1, January 1964, pp. 57-60.

10. Esposito, R., "A Class of Estimators for Optimum Adaptive Detection," Information and Control, Vol. 10, pp. 137-148, February 1967.

11. Esposito, R., "On a Relation Between Detection and Estimation in Decision Theory," Information and Control, Vol. 12, Number 2, pp. 116-120, February 1968.

12. Kailath, T., "A General Likelihood - Ratio Formula for Random Signals in Gaussian Noise," IEEE Transactions on Information Theory, Vol. IT-15, Number 3, pp. 350-361, May 1969.

13. Kailath, T., "A Note on Least-Squares Estimates from Likelihood Ratios," Information and Control, Vol. 13, pp. 535-540, December 1968.

14. Fralick, S. C., "Learning to Recognize Patterns Without a Teacher," Report SI-SEL-65-011 (TR No. 6103-10), Stanford Electronics Laboratories, Stanford, California, March 1965.

15. Nahi, N. E., "Optimal Recursive Estimation with Uncertain Observation," IEEE Transactions on Information Theory, Vol. IT-15, Number 4, pp. 457-462, July 1969.

16. Middleton, D. and Esposito, R., forthcoming Raytheon technical report, 1969.

17. Balakrishnan, A. V., "Communication Theory," Inter-University Electronics Series, Vol. 6, p. 329, McGraw-Hill Book Company, New York, 1968.

18. Kailath, T., Adaptive Matched Filters, in "Mathematical Optimization Techniques," (R. Bellman, ed.), Chapter 6, University of California Press, Berkeley (1963).

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) The Johns Hopkins University, Carlyle Barton Lab. Charles and 34th Streets Baltimore, Maryland 21218	2a REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP ---

3 REPORT TITLE

Simultaneous Detection and Estimation Under Multiple Hypotheses

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5 AUTHOR(S) (Last name, first name, initial)

A. J. Fredriksen

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1970	55	18

8a. CONTRACT OR GRANT NO. F33615-69-C-1317 b. PROJECT NO. 4036, Task No. 403601 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) AF-160 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFAL-TR-70-39

10. AVAILABILITY/LIMITATION NOTICES This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Air Force Avionics Laboratory (AVWW), Wright-Patterson Air Force Base, Ohio

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Force Avionics Laboratoru Wright-Patterson AFB, Ohio 45433

13. ABSTRACT This report treats the problem of simultaneous detection and estimation under multiple hypotheses when data from only one observation interval is available. The analysis is based on statistical decision theory. In the past, parameter estimation was always performed under the assumption that the desired signal was present with probability one. Since detection is meaningful only when some uncertainty exists as to the presence or absence of the desired signal it is apparent that classical estimation theory must be modified if the two operations are to be performed simultaneously. In addition, this report considers the case when the operations of detection and estimation are coupled. In this report specific detection and estimator structures are determined for the case where the two operations are strongly coupled and where the cost of estimation is given by a quadratic cost function. It is found that the detector structures are complex nonlinear functions of the received data, but nevertheless the one case considered in detail resulted in a type of correlation detector, where the basic operation is correlation of the received data with the various least squares estimators of the possible signals in the absence of uncertainty. The associated optimum estimator structures are determined for this case, and found to be weighted sums of the various least squares estimators in the absence of uncertainty. Finally, joint detection and estimation under multiple hypotheses is discussed for the case of a simple cost function assignment. It is shown that the estimators which result in this case may be interpreted as generalized maximum likelihood estimators.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Simultaneous Detection and Estimation						
Optimum Decision Rules						
Statistical Estimation						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.