PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
ELECTRONIC SYSTEMS RESEARCH LABORATORY

ON THE OPTIMIZATION OF MIXTURE RESOLVING SIGNAL PROCESSING STRUCTURES

by

J. C. Hancock and W. D. Gregg

Technical Report No. TR-EE66-11
Supported by
National Science Foundation
and
National Aeronautics and Space Administration
Washington, D. C.

OCTOBER 1966
LAFAYETTE, INDIANA

RESEARCH GRANTS

No. NSF GP-2898, PRF 3955
and
No. NASA NsG-553, PRF 3823

ON THE OPTIMIZATION OF MIXTURE RESOLVING SIGNAL PROCESSING STRUCTURES

for

NATIONAL SCIENCE FOUNDATION

and

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

WASHINGTON, D.C.

by

J. C. Hancock and W. D. Gregg

School of Electrical Engineering
Purdue University
Lafayette, Indiana

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF FIGURES (cont'd)

ABSTRACT

This thesis is concerned with the development of the structural form of optimum linear detection operators when the pulse waveform and noise parameters are unknown and are to be obtained by mixture resolving estimation; and with the development of the mixture resolving estimators to learn or extract this parametric information from the noisy signal pattern mixture, in order to obtain the elements for the structure. The observation signal model consists of a discrete, multidimensional, binary (two category) gaussian mixture.

An eigenvalue approach is taken for the development of the structural form of the detection operator; the criterion of optimality being the minimum average conditional probability of error at the Nth stage, conditioned upon the mixture resolving estimating category. The mixture resolving categories developed consist of an optimized, time-varying-weighted decision-directed category and a moment method category. These categories differ from related work in that; (1) In the former, the initial reference is extracted from the signal mixture by correlating the first observation with the next and updating the result with successive-time-varying-weighted combinations of the separated time slots optimized to minimize a measure of distance and dispersion with a subsequent "maximization" of convergence rate and (2) in the latter, no apriori knowledge of either the pulse waveform <u>or</u> noise parameters is required.

Extensive experimental studies, via digital computer simulation, of the performance characteristics of these signal processing algorithms are carried out and compared with the conventional decision directed and Bayes matched filter algorithms under identical input conditions. A complete formulation of this approach, to include verification of the theory by digital computer simulation experimentation, however, is carried out only for the bi-polar case ($\underline{\theta}_2 = -\underline{\theta}_1$) in the equiprobable situation ($p_1 = p_2 = \frac{1}{2}$). An analysis for the general case is carried out and the difficulties encountered by the lack of specific apriori information are discussed.

Both algorithms developed in this work converge for negative db. values of SNR at a rate considerably higher than that of the conventional decision-directed algorithm and are bounded from above in performance by the conventional decision-directed algorithm over the entire SNR range investigated. In addition, the weighting in the optimized-weighted-decision directed ("rate maximized") algorithm, over the SNR range investigated, is dominated only by the observation stage, N, and the constraint coefficient, γ; and, in that sense, is non-parametric in the mixture pulse waveform and noise parameters. A signal processing interpretation of the digital computer simulation of the numerical experimentation implies that knowledge of the required signal dimensionality is available.

CHAPTER I

INTRODUCTION

1-1. Introduction to the Problem of Mixture Resolving Signal Processing

This thesis is concerned with the problem of developing and optimizing, in a specific manner, mixture resolving signal processing structures. The problem is formulated with the signal processing structure fundamentally process detection or classification oriented as opposed to a process extraction or recovery orientation.

The model of the observation signal consists of a formal mixture and thus assumes no pre-classification of time slots into isolated class or category ensembles. The work in this thesis is based upon a multidimensional, binary, gaussian mixture.

Strict mixture resolving signal processing implies detection or classification without any apriori knowledge of the signal and noise pattern properties. The formulation of the problem in this work assumes certain specific relationships between the mixing parameters (probabilities of the two classes in the mixture) and the location parameters (mean vectors of the two gaussian classes). These assumptions are pointed out in detail in Section 1-4.

In general, this problem in a detection context, requires some integrated operation of time slot classification and extraction of the information required for this time slot classification. In drawing an analogy with the signal detection problem, it is seen that the requirement is an integration of the operation of the classification of a noisy pulse waveform with the extraction of the pulse and noise parameters for this classification. This problem is common to all areas of active and passive radar, sonar, seismology; active communication systems and bioelectric signal analysis, in that, in these areas, rarely are there ever available complete apriori deterministic and statistical descriptions of the characteristics of the electro-magnetic, acoustic, and electrochemical channels, cross-sections, and information obscuring processes.

1-2. Summary of Literature Review

A literature review of detection oriented work in the adaptive and learning system areas and in mixture technology has resulted in the following findings. The problems have generally been defined in a discrete sense assuming a scalar or vector sequence of samples, $\{\underline{z}_k\}_N$, of a signal mixture, as a model for the observations. The approaches taken have been generally conventional minimum conditional risk formulated giving rise to a conditional likelihood ratio, $\ell(\underline{z}_{N+1}|\{\underline{z}_k\}_N)$, with the likelihood for the ith class or category, ω_i, given by

$$P(\underline{z}_{N+1}|\{\underline{z}_k\}_N;\ \omega_i) = \int_{R\{\Theta\}} P(\underline{z}_{N+1}|\{\Theta\};\ \omega_i)P(\{\Theta\}|\{\underline{z}_k\}_N;\ \omega_i)\ d\{\Theta\} \qquad (1.1)$$

thus introducing the requirement for an apriori probability density, $P(\{\Theta\})$, as a measure of the uncertainty about the parameter; $\{\Theta\}$, a sequence of parameters associated with the observations (mean vectors, covariance matrices, [pulse shapes, noise statistics]). The object is to develop an optimum (minimum probability of misclassification) decision equation (detector, classifier, etc.) structure for classifying an observation $\underline{z}_{N+1}$ conditioned upon a past sequence of observations, $\{\underline{z}_k\}_N = \underline{Z}$, with varying amounts of apriori information about the observations. Abramson, Braverman[1] and Keehn[2] formulated (1.1) for the scalar gaussian case with the mean unknown and for the vector gaussian case with the mean vector and covariance matrix unknown respectively. In both cases it was assumed that the apriori observations, $\underline{Z}$, were <u>classified</u> yielding an isolated ensemble for ω_i as

$$Z = Z^{(i)} \;;\; P(\{\Theta\}|\underline{Z};\,\omega_i) \sim P(\underline{Z}^{(i)}|\{\Theta\}^{(i)})P(\{\Theta\}^{(i)}) \qquad (1.2)$$

In addition, both assumed apriori knowledge of $P(\{\Theta\}^{(i)})$. Thus, a completion of the square in the exponents of the integral of (1.1) yields a conditional mean and associated covariance for the estimates $\{\hat{\Theta}\}$ in the form of <u>unweighted</u>, linear combinations of the estimates of the mean, $\underline{\theta}_i$, and covariance, Θ, conditioned upon, and linearly averaged with, apriori values reflected by $P(\{\Theta\}^{(i)})$. This is referred to as the supervised, classified or "with teacher" decision equation structure and is invariant since iterations occur only in the elements of the structure. The form of the structure is invariant, since Abramson, Braverman and Keehn assume a gaussian and wishart form for the apriori density of the unknown mean and covariance respectively which corresponds to a natural conjugate prior density for a guassian likelihood, $P(\underline{z}|\{\Theta\}^{(i)})$ (Raiffa[3]). Raiffa has shown that when a natural conjugate prior density on $\{\Theta\}^{(i)}$ exists, the posterior density, $P(\{\Theta\}^{(i)}|\underline{Z}^{(i)}, \omega_i)$, is of the same form (reproducibility or invariant structure). Spragins[4] has shown that when the observations are classified the prior natural conjugate exists for a number of cases including gaussian. Clearly the learning with teacher decision equation structure can be obtained by a linear substitution of the conditional maximum likelihood estimates, $\{\hat{\theta}\}$, from the classified sequence (isolated ensemble for the ith class) into the Bayes matched filter structure. In view of the classification, the estimator, $\{\hat{\theta}\}$, of the elements for the structure is classically consistent and unbiased and hence the structure is bounded and converges to the Bayes matched filter. Jakowatz[5], Shuey, and White consider a sub-optimum cross-correlation-detection-decision-directed approach for the case of one fixed waveform repeating at random in additive noise (multi-dimensional off-on case). Their approach is to take a given time slot, obtain a correlation with successive observations until the cross-correlation "detects" the presence of the waveform, and up-date the estimate of the waveform by a linear average of the active time slot with the previous estimate. The cross-correlation detection concept is also treated in Downing[6] for the detection of differentially coherent phase reversal keying where a previous symbol waveform is used as a reference against which to correlate the time slot currently under observation. Hinich[7], in performing a more formal analysis of the Jakowatz approach, showed the existence of an asymptotically stable (but not necessarily unbiased) estimator for the waveform and developed an expression for an asymptotically efficient estimator for the discrete autocorrelation function of the waveform in the Jakowatz model.

The first attempt to arrive at a minimum conditional risk formulation for the decision equation structure in the unsupervised or unclassified case was made by Daly[8], who considered the off-on multidimensional gaussian case and allowed the past sequence of observations, $\underline{Z}$, to be partitioned into all possible 2^N patterns, yielding

$$P(\underline{Z}|\{\theta\}) = \sum_{r=1}^{2^N} p_r \, P(\underline{Z}_r|\{\theta\}^{(r)}); \quad \underline{Z} = \{\underline{z}_1 \, \cdots \, \underline{z}_N\} \tag{1.3}$$

where $\underline{Z}_r$ is the format of the rth partition, $\{\theta\}^{(r)}$ is the sequence of parameters associated with the partition, and p_r is the probability of the partition. The implications of (1.3) are that for N past observations, 2^N possible signal pattern formats could have been present and consequently it is necessary to have the equivalent of 2^N decision equation structures containing the estimates of the elements based upon all possible 2^N patterns; thus, giving rise to an unbounded structure. Subsequently, Daly[9], showed for the off-on case that the partitioning of $\underline{Z}$, yields a decision equation structure that, in view of the formulation, is a minimum conditional risk structure at each stage, and by invoking martingale theory, proved convergence of the structure to that which is optimum for detecting known signals in known noise (Bayes matched filter). Daly either did not choose to explore or was unaware of the fact that for the binary gaussian case and independent observations, (1.3) could be written in an equivalent two category mixture representation as

$$P(\underline{Z}|\{\theta\}) = \prod_{k=1}^{N} [p_1 P(\underline{z}_k|\underline{\theta}_1, \theta) + p_2 P(\underline{z}_k|\underline{\theta}_2, \theta)] \tag{1.4}$$

and Fralick[10], by incorporating (1.4) iteratively for the off-on gaussian case in white noise, obtained a recursive form for $P(\underline{\theta}_1|\underline{Z})$. In addition, Fralick[11] also showed that the iterative conditional likelihood ratio structure is a bounded martingale, thus, proving bounded stable performance and an asymptotic limit for the off-on case when it is known that one of the means is zero. Fralick did not make further use of mixture concepts other than to utilize (1.4) and Hancock and Patrick[12] were the first to incorporate mixture representations such as (1.4) in conditional likelihood structures as given by (1.1) for the M-ary case; and showed the equivalence of (1.3) and (1.4) in the structure. Also, Hancock and Patrick[13] applied histogram concepts to the unsupervised case and showed that to estimate class conditional cumulative distribution functions (c.d.f.'s) results in a mixture of multinomial distributions; and established the conditions under which the parameters characterizing the multinomial distributions could be uniquely learned in the binary case. In the same work, they drew upon identifiability, a particular fundamental concept of mixtures which is a definition of mixture resolvability, in order to arrive at a means of determining the amount of apriori information sufficient for a Bayes solution to exist. Teicher [14,15,16] defined and established some very broad and formal conditions for the identifiability of mixtures although estimation of parameters in mixtures was treated first by Pearson[17] and subsequently by Rao[18], Rider[19], and Blischke[20]. All assumed independent observations and developed point estimators by the method of moments for the one-dimensional case. Pearson and Rao were concerned with the gaussian mixture, Rider treated the exponential mixture and Blischke worked with a bi-nomial mixture. Mix[21] developed computer simulations of the detector structures for the supervised, decision-directed, and unsupervised models and processed one-dimensional binary gaussian mixtures with the means unknown and noise variance known. The detector structures for the unsupervised case included the partitioned (Daly) and the iterative (Fralick) forms. Mix illustrated the contrasts in relative rates of convergence, and, in terms of digital-computer implementation of the data-processing structures, arrived at some measures of relative complexity in terms of memory and processing time required. He illustrated the equivalence in complexity of the partitioned and iterative forms.

The decision directed approach was treated by Scudder[22] who suggested a <u>linear</u> substitution of the mean vector estimates obtained by a linear, unweighted averaging of the decision-directed time slot with the class to which the observation was assigned, into the Bayes matched filter structure. Scudder assumed that the noise was white, that knowledge of this fact was available, and treated the on-off gaussian case. He demonstrated convergence[24] for signal to noise ratios considerably greater than 0 db and illustrated that asymptotic performance deviated from the performance of the Bayes matched filter inversely with the signal to noise ratio.

Chang [23], in developing the form of the minimum conditional risk detector structure for the unsupervised binary gaussian case with intersymbol interference between adjacent bands, invoked the mixture concept and arrived at a four-category multi-dimensional gaussian mixture structure, with the constraint that a given waveform and its overlap have the same sign. In addition, Chang , developed moment estimators for the means (samples of pulse waveforms with intersymbol interference between adjacent bauds), however, the estimators required apriori knowledge of the noise statistics. He illustrated convergence of the detector structures and consistency of the estimators by digital computer simulation.

1-3. Motivation for Detector Structure Optimization About Reduced Complexity and Convergence Time

It is clear that in the areas of radar, sonar, seismology, communications, and bio-electrics, the signals available are never classified into isolated ensembles and thus the with teacher model is not practical. Second, the complexity, in terms of computer time and memory required for instrumentation, of the parametric minimum conditional risk structure for the unsupervised case becomes prohibitive as it stands. In addition, some apriori parametric measure of the uncertainty associated with the unknown parameters, $\{\Theta\}$, is required. Third, attempts to reduce the structure complexity by a linear substitution of possible mixture resolving estimators for the elements into the Bayes matched filter structure; have not been justified as preserving optimality. In addition, the correlation-location, decision-directed structures simply possess poor reliability and performance characteristics for negative db signal to noise ratios. Finally, the moment mixture resolving elements perform somewhat more reliably, however, more apriori information about the mixture is required, and for the methods developed thus far[23], knowledge of the noise statistics is required.

In summary, it appears that the unsupervised structures possessing reliable Bayes convergence rates are prohibitively complex and that the less complex structures, such as the conventional-decision-directed system, possess poor reliability for the more practical signal to noise ratio levels.

The effort in this thesis research strives for a reduction in the complexity of an optimized adaptive or learning detector structure for processing noisy binary signal pattern mixtures. In addition, emphasis is placed upon maximizing convergence rates and arriving at reliable performance for low signal to noise ratio levels (0 db and less). The approach taken and the problem definitions in mathematical terms are outlined in the next section.

1-4. Summary of the Approach and Contributions

In this work, the observation signal model consists of a discrete, multi-dimensional, binary (two category) gaussian mixture. The vector means (pulse samples) and noise covariance (noise statistics) are assumed unknown. No apriori probability density on the parameters, $\{\Theta\}$, is assumed available, and in view of the non-bayesian formulation of the problem, none is required. A complete formulation of this approach, to include verification of the theory by digital computer simulation experimentation, is carried out only for the bi-polar case, ($\underline{\theta}_2 = -\underline{\theta}_1$), where the probabilities of each class occurring are equal ($p_1 = p_2$). An analysis for the general case is carried out and the difficulties encountered in this approach, by the lack of specific apriori information are discussed in Section 4-5.

A portion of the contribution of this thesis appears in Chapter II which contains a formal development of the average conditional probability of misclassification under linear discrimination weighting of the observation samples. The contribution lies in the fact that the error probability is developed in terms of the expectation of the mixture resolving estimators and the relationship of these expectations to the classification operator corresponding to the minimum of the error probability. In particular, it is shown that the unique operator structure for classification or detection which minimizes the average conditional error probability appears as a unique eigenvector of the maximized eigenvalue in the error probability expression.

Further contributions appear in Chapter III, in which, through the vehicle of variational calculus, is developed a mixture resolving category, J, which minimizes the expected distance, $\underset{*}{<} ||\hat{\underline{\theta}} - \underline{\theta}||^2 >$, while constraining the mean square estimation error, $\underset{*}{<} ||\hat{\underline{\theta}} - \underset{*}{<} \hat{\underline{\theta}} >||^2 >$. Further treatment shows that by varying the constraint, the convergence rate given by, $\frac{\partial}{\partial N} \underset{*}{<} ||\hat{\underline{\theta}} - \underline{\theta}||^2 >$, can be maximized. The results of Chapters II and III are applied in Chapter IV to the bi-polar case, $\{\underline{\theta}_1, -\underline{\theta}_1, \Theta\}$, and the more general case, $\{\underline{\theta}_1, G\underline{\theta}_1, \Theta\}$. As previously pointed out, results to include verification of the theory by digital computer simulation experimentation, are available only for the bi-polar case with $p_1 = p_2$. A general iterative expression for the recursion on $\hat{\underline{\theta}}_\ell(N)$ is obtained. It is verified experimentally that the convergence rate can be controlled by an adjustment of the La-Grange multiplier coefficient in the constraint.

Final contribution appears in Chapter V, which contains the development of a moment method mixture resolving estimator for the multi-dimensional case which does not require apriori knowledge of the noise statistics. Chapter VI contains a discussion of the convergence rates and performance characteristics. A computer simulation is developed and discussed in Chapter VI for the two signal processing algorithms developed in Chapters III, IV, and V, and for the conventional decision-directed algorithm and the Bayes matched filter. A comparison is made of the relative rates of convergence and the dynamic and asymptotic performance characteristics against the matched filter as a standard. Average performance is obtained over an ensemble of noisy signal pattern observations with a given signal and noise pattern applied simultaneously to all structures for an additional study of relative performance properties on a member basis. The superior convergence rate characteristics of the weighted-decision-directed ("rate maximized") structure are borne out in the digital computer simulation experimentation. In addition, the "rate-maximized" structure performs reliably and converges for negative db signal to noise ratios where the conventional decision directed method converges considerably slower and on a member run by member run, is less reliable.

Finally a numerical technique is illustrated to bound the probability of error.

CHAPTER II

EIGENVECTOR OPERATOR FORMULATION OF AN OPTIMAL SIGNAL MIXTURE PROCESSING STRUCTURE

2-1. Error Probability Under Linear Discrimination

Consider a discrete signal composed of two possible unknown pulses, $\Theta_1(mT_S)$, and $\Theta_2(mT_S)$, of duration $M_P T_S$ appearing independently every $M_D T_S$ sec. in additive, zero mean, gaussian noise with probabilities p_1 and p_2 respectively. Examples of an arbitrary and a bi-polar base-band observation appear in Fig. 2-1. The symbol T_S represents the sampling period and m ranges from one to M_{P_j} where j is associated with the active

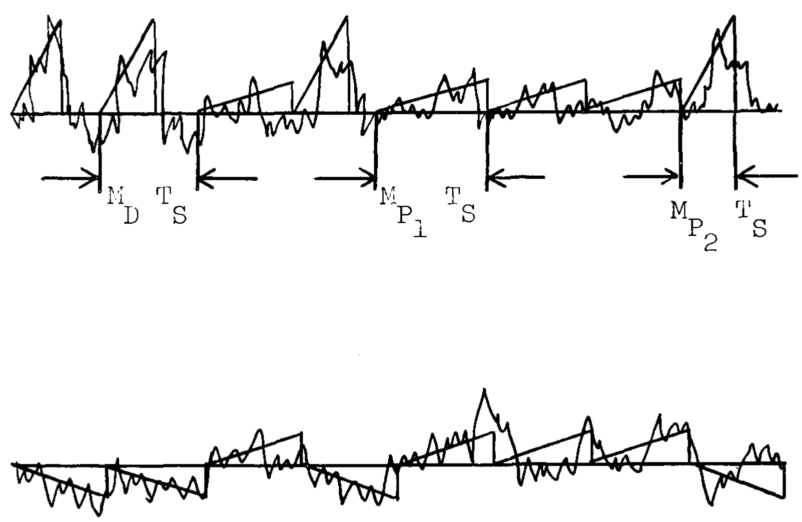

Fig. 2-1. Noisy Signal Pattern Models

pulse. The time samples, thus constitute a sequence of independent, multidimensional, gaussian observations, $\{\underline{z}_k\}_N$ from classes ω_1 and ω_2, with unknown parameters $(\underline{\theta}_1, \Theta)$, and $(\underline{\theta}_2, \Phi)$, occurring at random with probabilities p_1 and p_2 respectively. Since there is a non-zero "distance" between the population, $\underline{\theta}_1 \neq \underline{\theta}_2$, the mixture can be resolved (Teicher[14-16] [identifiability]).

To arrive at some measure of distance between the classes of a mixture, consider a set of hypothetical hypersurfaces which would separate the populations with some associated set of proportionalities and risks. In particular, for the two-class gaussian mixture above, a collection of such surfaces consists of a sequence of hypothetical hyperplanes, $\{\beta_m(\underline{z})\}$, where a particular mth element is given by

$$\beta_m(\underline{z}) = \underline{\alpha}_m' \, \underline{z} - \ell_m \quad ; \quad m = 0, 1, 2, \ldots \qquad (2.1)$$

where the parameters, $\{\underline{\alpha}_m\}$, $\{\ell_m\}$, constitute sequences of associated weighting coefficients and thresholds respectively. Clearly, the optimal hyperplane, $\beta_o(\underline{z})$, Bayes matched filter, is an element of the set with parameters,

$$\underline{\alpha}_o = \Theta[\underline{\theta}_1 - \underline{\theta}_2] \quad ; \quad \ell_o = \frac{1}{2}\,\underline{\alpha}_o' \,[\underline{\theta}_1 + \underline{\theta}_2] + p_2C_1/p_1C_2 \quad ; \quad \Theta = \Sigma^{-1} \qquad (2.2)$$

where the C's are defined risks associated with mis-classification. By associating hypotheses H_1 and H_2 with the observation, $\underline{z}_k$, relative to the boundary; (2.1) becomes a decision equation.

$$\begin{aligned} \ell_m(\underline{z}_k) = \underline{\alpha}_m' \, \underline{z}_k &\geq \ell_m \qquad H_1 \equiv \underline{z}_k \in \omega_1 \\ &< \ell_m \qquad H_2 \equiv \underline{z}_k \in \omega_2 \end{aligned} \qquad (2.3)$$

Regardless of the element, $\underline{\alpha}_m$, chosen, a particular value of ℓ_m, ℓ_{o_m}, can always be found which corresponds to that region of classification or separation for that category m (that particular hypothetical set of parameters for the population, $(\Theta)_m$, out of the sequence $\{(\Theta)_m\}$) such that the probabilities of misclassification are equal. Consequently, the two associated error probabilities are

$$P\{H_2|\omega_1\} = \int_{\ell_m(\underline{z}) \geq \ell_{o_m}} P(\underline{z}|\omega_1)d\ell_m(\underline{z}) = \int_{\ell_m(\underline{z}) < \ell_{o_m}} P(\underline{z}|\omega_2)d\ell_m(\underline{z}) = P\{H_1|\omega_2\} \tag{2.4}$$

Taking

$$E[\ell_m(\underline{z})|\omega_{m_\ell}] = \underline{\alpha}_m' \, \underline{\theta}_{m_\ell} \quad ; \quad \ell = 1,2 \quad ; \quad m = 1,2, \ldots \tag{2.5}$$

and

$$E[(\ell_m(\underline{z}) - E[\ell_m(\underline{z})|\omega_{m_\ell}])^2|\omega_{m_\ell}] = \underline{\alpha}_m' \, \Theta_m^{-1} \, \underline{\alpha}_m \tag{2.6}$$

the identification, detection, or classification error probabilities become

$$P\{H_2|\omega_{m_1}\} = (2\pi Q_m)^{-\frac{1}{2}} \int_{\ell_m(\underline{z}) \geq \ell_{o_m}} \exp\,[-\tfrac{1}{2}\,(\ell_m(\underline{z}) - \underline{\alpha}_m' \, \underline{\theta}_{m_1})^2/Q_m]d\ell_m(\underline{z}) \tag{2.7}$$

$$= (2\pi Q_m)^{-\frac{1}{2}} \int_{\ell_m(\underline{z}) < \ell_{o_m}} \exp\,[-\tfrac{1}{2}\,(\ell_m(\underline{z}) - \underline{\alpha}_m' \, \underline{\theta}_{m_2})^2/Q_m]d\ell_m(z)$$

where

$$Q_m = \underline{\alpha}_m' \, \Theta_m^{-1} \, \underline{\alpha}_m \tag{2.8}$$

Since the gaussian family is symmetric and since Θ_m is the same for both populations,

$$\ell_{o_m} = (E[\ell_m(\underline{z})|\omega_{m_1}] + E[\ell_m(\underline{z})|\omega_{m_2}])/2 \tag{2.9}$$

The densities of the test statistics associated with the mth boundary, under the two hypotheses, are illustrated in Fig. 2-2 along with a hypothetical contrast to the densities of the Bayes matched-filter-weighted test statistics. The thresholds given by (2.9) are also illustrated for the mth group of elements by ℓ_{o_m}, and for the Bayes matched filter by ℓ_{o_o}. Since the terms in (2.7) are equal, equivalent results can be obtained for either, and the approach is simplified by considering one term.

Thus set

$$P_e(m) = P\{H_\ell | \omega_{m_k}\} \qquad k,\ell = 1, 2 \quad ; \quad k \neq \ell \tag{2.10}$$

and with the change of variables,

$$y_m = (\ell_m(\underline{z}) - \underline{\alpha}_m' \, \underline{\theta}_{m_2})/Q_m \tag{2.11}$$

using (2.9), (2.7) becomes

$$P_e(m) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{-(\lambda(m))^{\frac{1}{2}}} \exp\,[-\tfrac{1}{2}\, y_m^2]\, dy_m \tag{2.12}$$

From (2.12) it is seen that

$$\min_{\underline{\alpha}_m \in E_n} P_e(m) \equiv \max_{\underline{\alpha}_m \in E_n} \lambda(m) \tag{2.13}$$

$$E_n \overset{\Delta}{=} [\text{finite n dimensional space}]$$

$$\lambda(m) = \left[\frac{\underline{\alpha}_m' \, \underline{\delta}_m}{2Q_m^{\frac{1}{2}}}\right]^2 = \frac{\underline{\alpha}_m' \, \Delta_m \, \underline{\alpha}_m}{4Q_m}$$

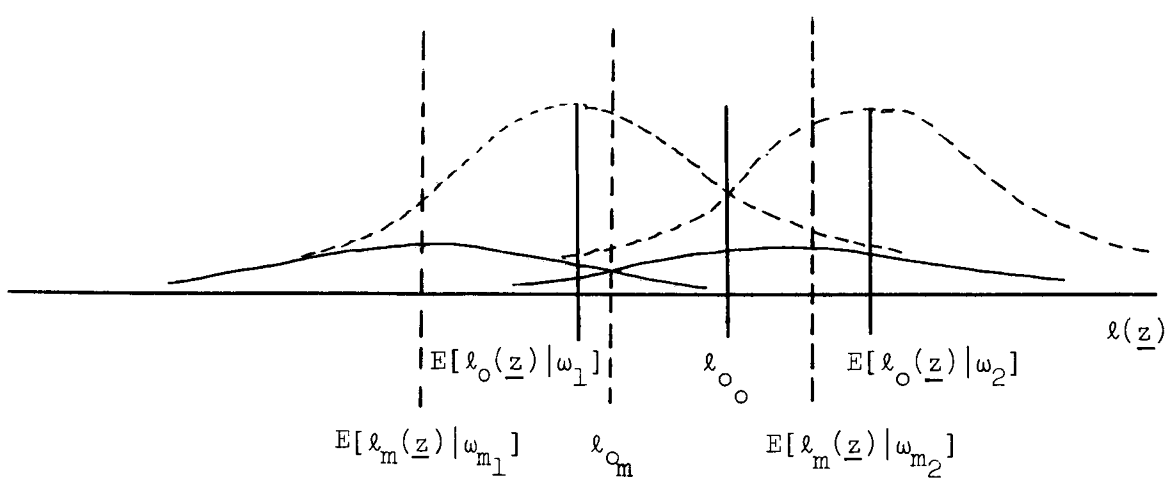

Fig. 2-2. Hypothetical Contrast in Test Statistic PDF's

where

$$\Delta_m = \underline{\delta}_m \underline{\delta}_m' \quad ; \quad \underline{\delta}_m \stackrel{\Delta}{=} (\underline{\theta}_{m_1} - \underline{\theta}_{m_2}) \tag{2.14}$$

and that the probability of error for the mth element, $P_e(m)$ is a monotonically decreasing function of the positive definite form, $\lambda(m)$; since Δ_m and Θ_m are both symmetric.

2-2. Structural Form of the mth Optimal Linear Detector

In view of (2.13), it is concluded that the minimal of error probability with respect to the mth element category corresponds to linear weighting of the signal samples by the elements of a vector operator which effects the maximum of $\lambda(m)$. Taking

$$\frac{\partial\lambda(m)}{\partial\underline{\alpha}_m} = \underline{0} \tag{2.15}$$

yields a homogeneous, eigenvalue equation

$$(\Theta_m \Delta_m - \lambda(m)I)\, \underline{\alpha}_m = \underline{0} \tag{2.16}$$

and Max $\lambda(m)$ corresponds to the largest eigenvalue of $\Theta_m \Delta_m$.

However, with R denoting rank,

$$R(\Theta_m \Delta_m) \leq \underset{R}{\text{Min}} \, [R(\Theta_m), R(\Delta_m)] \tag{2.17}$$

and in view of (2.14), since Θ_m is non-singular and $R(\Delta_m)$ is one, then the rank of the product is one. Consequently since,

$$\sum_{j=1}^{n} \lambda_j = \text{Trace } (M) \tag{2.18}$$

where the λ_j are the eigenvalues of M and since $R(\Theta_m \Delta_m)$ is one, there is only one eigenvalue. This eigenvalue is the maximum, for the mth group elements, and is given by

$$\lambda_o(m) = \text{Tr}(\Theta_m \Delta_m) = \text{Tr}(\Theta_m \underline{\delta}_m \underline{\delta}_m') = \underline{\delta}_m' \Theta_m \underline{\delta}_m \tag{2.19}$$

The desired observation-signal-sample weighting vector operator corresponds to the eigenvector corresponding to $\lambda_o(m)$. The substitution of (2.19) into (2.16) yields

$$\Theta_m \underline{\delta}_m \underline{\delta}_m' \underline{\alpha}_{o_m} = \underline{\delta}_m' \Theta_m \underline{\delta}_m \underline{\alpha}_{o_m} \tag{2.20}$$

and from (2.19), the desired vector operator is

$$\underline{\alpha}_{o_m} = \Theta_m \underline{\delta}_m \tag{2.21}$$

The Bayes matched filter vector operator would correspond to

$$\underline{\alpha}_{o_o} = \Theta\underline{\delta} \tag{2.22}$$

which is the eigenvector corresponding to the eigenvalue

$$\sup_m \{ \max_{\underline{\alpha}_m \varepsilon E_n} \lambda(m)\} = \sup_m \lambda_o(m) = \lambda_o(m_o) = \underline{\delta}' \Theta\underline{\delta} \tag{2.23}$$

where Θ, and $\underline{\delta}$ are true population parameters. This would have the effect of giving

$$\inf_m \{ \min_{\underline{\alpha}_m \varepsilon E_n} P_e(m)\} = P_e(m_o) \tag{2.24}$$

the error probability of the matched filter.

Thus the vector operator structure which operates upon a vector observation to yield $\min_{\underline{\alpha}_m} P_e(m)$, relative to the mth category elements in the sequence $\{(\Theta)_m\}$, is given by (2.21). The argument of the optimal operator structure relative to an element cateogry m out of a sequence, $\{(\Theta)_m\}$, is presented in order to introduce the concept of optimizing a structure within or about a category of elements, J, in particular an estimating category; which is carried out in the next section. The actual introduction of the elements from the mixture is treated in Chapter III.

2-3. Estimating Category Conditioned Error Probability

Consider a general vector, $\underline{h} \varepsilon E_n$, with all continuous possible values for the elements in (2.3) (in essence call the sequence of elements, $\{\underline{\alpha}_m\}$, the general variable $\underline{h}$ which can take on any continuom of values in the sequence). Consider the raw samples, $\{\underline{z}_k\}_N$. If any mixture resolving estimators, $\{\hat{\theta}(\{\underline{z}_k\}_N, N, J)\}$, for the parameters, $\{\theta\}$, from some estimating cateogry J, can be developed such that

$$\lim_{N\to\infty} P\{|\{\hat{\theta}(\{\underline{z}_k\}_N, N, J)\} - \{C\}| > \varepsilon\} = 0 \qquad (2.25)$$

or some other measure of consistency, exists, where $\{C\}$ is some sequence of constants, then one can take across an ensemble of error patterns

$$E[P_e(N, J)] \qquad (2.26)$$

That is, if some separation of an observation $\underline{z}_k$, from the mixture $\{\underline{z}_k\}_N$, can be achieved into spaces ω_1^*, ω_2^* by an operator with coefficients from some category, $J\varepsilon\{J_r\}_R$, where ω_1^*, ω_2^* need not initially coincide with ω_1 and ω_2, and J is not yet specified, then over a separation ensemble one can take the expectation, $< >$

$$< \ell(\underline{z}_{N+1} ; N, J)|\omega_\ell^* > = \underline{h}' < \hat{\underline{\theta}}_\ell(\{\underline{z}_k\}_N ; N, J)|\omega_\ell^* > ; \ \ell = 1, 2 \qquad (2.27)$$

and

$$<(\ell(\underline{z}_{N+1} ; N, J) - <\ell(\underline{z}_{N+1} ; N, J)|\omega_\ell^*>)^2|\omega_\ell^*>$$
$$= \underline{h}' <\hat{\Theta}^{-1}(\{\underline{z}_k\}_N ; N, J)> \underline{h} \qquad (2.28)$$

As a result, at any stage of observation and classification, N, for a particular estimating category, $J\varepsilon\{J_r\}_R$, the measure of distance or error probability would be given by

$$P_e\{N, J\} = (2\pi Q(N, J))^{-\frac{1}{2}} \int_{-\infty}^{(\underline{h}' E[\hat{\underline{\delta}} (N, J)]/2\, Q(N, J)^{\frac{1}{2}}} \exp\, [-\tfrac{1}{2}\, y_J^2] dy_J \qquad (2.29)$$

where

$$Q(N, J) = \underline{h}' \, E[\hat{\theta}^{-1}(N, J)] \, \underline{h} \tag{2.30}$$

and the change of variable

$$y_J = (\ell(\underline{z}_{N+1} \; ; N, J) - \underline{h}' \, \underline{\hat{\theta}}_2(N, J))/Q^{\frac{1}{2}}(N, J) \; ; \quad J = J_1, \ldots, J_r, \ldots J_R \tag{2.31}$$

has been introduced as in (2.11).

Thus from (2.29), it is seen that the error probability at the Nth stage is a monotonically decreasing function of the positive definite form

$$\lambda(N, J) = \frac{\underline{h}'\Delta(N, J)\underline{h}}{Q(N, J)} \tag{2.32}$$

where Q(N, J) is given in (2.30) and

$$\Delta(N, J) = < \underline{\hat{\delta}}(\{\underline{z}_k\}_N \; ; N, J) > < \underline{\hat{\delta}}(\{\underline{z}_k\}_N \; ; N, J) >' \tag{2.33}$$

2-4. Structural Form of the Estimating Category, $J \varepsilon \{J\}_R$, Conditional Optimal Linear Detector

The results of the previous section indicate that at the Nth stage within the estimating category, J, the minimum error probability arises when discrimination is carried out by a vector operator, $\underline{h}_o$, which is the vector operator maximizing the form $\lambda(N, J)$. Thus

$$\underset{\underline{h} \, \varepsilon \, E_n}{\text{Min}} \; P_e\{N, J\} \equiv \underset{\underline{h} \, \varepsilon \, E_n}{\text{Max}} \; \lambda(N, J) = \underset{\underline{h} \, \varepsilon \, E_n}{\text{Max}} \; \frac{\underline{h}'\Delta(N, J)\underline{h}}{Q(N, J)} \tag{2.34}$$

Taking, again

$$\frac{\partial \lambda(N, J)}{\partial \underline{h}} = \underline{0} \tag{2.35}$$

yields the corresponding, estimating category conditioned eigenvalue equation

$$(<\hat{\theta}(\{\underline{z}_k\}_N \; ; N, J)> \; \Delta(N, J) - \lambda(N, J)I)\underline{h} = \underline{0} \tag{2.36}$$

Assuming that a mixture resolving category, J, exists such that $\hat{\Theta}(\{\underline{z}_k\}_N, N, J)$ is non-singular (this is shown later), and in view of (2.33), the rank of $<\hat{\Theta}(\{\underline{z}_k\}_N, N, J)> \Delta(N, J)$ is one, hence there is one and only one eigenvalue of (2.36), which is given by

$$\max_{\underline{h} \in E_n} \lambda(N, J) = \lambda_o(N, J) = \mathrm{Tr} < \hat{\Theta}(\{\underline{z}_k\}_N; N, J) > \Delta(N, J)$$
$$= < \hat{\underline{\delta}}(N, J) >' < \hat{\Theta}(N, J) > < \hat{\underline{\delta}}(N, J) > \quad (2.37)$$

Now from (2.36) and (2.37) one obtains

$$< \hat{\Theta}(N, J) > \Delta(N, J)\underline{h} = \lambda_o(N, J)\underline{h} \quad (2.38)$$

and by inspection

$$\underline{h}_o = \underline{h}(N, J) = < \hat{\Theta}(N, J) > < \hat{\underline{\delta}}(N, J) > \quad (2.39)$$

Thus, (2.39) gives the structural form of the unique vector operator which minimizes the error probability at the Nth stage relative to or conditioned upon the estimating category, J, as the eigenvector of a rank one eigenvalue equation. By invoking properties of ergodicity, and martingale boundedness, it can be shown that for a given signal pattern, the range of the random variable error probability, $P_e(N, J)$, at the Nth stage, relative to the estimating category, J, is minimal when the operator structure is given by

$$\underline{h}(N, J) = \hat{\Theta}(\{\underline{z}_k\}_N, N, J)\ \hat{\underline{\delta}}(\{\underline{z}_k\}_N, N, J) \quad (2.40)$$

Thus, the form of the vector operator is fixed whereas the elements, $\{\hat{\Theta}(\{\underline{z}_k\}_N, N, J)\}$, are functions of the past observations, $\{\underline{z}_k\}_N$, the number of observations, N, and the estimating category, $J \varepsilon \{J_r\}_R$. Clearly, for an estimating category, $J \varepsilon \{J_r\}_R$, possessing consistency, one has

$$\lim_{N\to\infty} \overline{P}_e(N, J) \to P_e(J) \quad (2.41)$$

the asymptotic performance characteristic for the Jth category and

$$\langle \overline{P}_e(N, J) \rangle_m = P_e(N, J) \tag{2.42}$$

the true dynamic performance characteristic at the Nth stage. It remains to develop, the mixture resolving categories, $J \varepsilon \{J_r\}_R$, which either allow or approach

$$\operatorname*{Inf}_{J \varepsilon \{J_r\}_R} \left\{ \{ \operatorname*{Min}_{\underline{h} \varepsilon E_n} P_e(N, J) \} \right\} = \operatorname*{Inf}_{J \varepsilon \{J_r\}_R} P_{o_e}(N, J) \tag{2.43}$$

due to

$$\operatorname*{Sup}_{J \varepsilon \{J_r\}_R} \left\{ \operatorname*{Max}_{\underline{h} \varepsilon E_n} \frac{\underline{h}' \Delta(N, J) \underline{h}}{Q(N,J)} \right\} = \operatorname*{Sup}_{J \varepsilon \{J_r\}_R} \lambda_o(N, J) \tag{2.44}$$

Conditions (2.43) and (2.44) can be attained at each stage by the formal mixture approach, (Bayes partitioned or iterative formulation of the estimating category, J_B, for the elements of the structure), but will not be pursued due to the complexity required, as mentioned previously. The actual value of $P_e(J)$ (biasedness, if any, of asymptotic performance) as well as $\frac{\partial P_e(N,J)}{\partial N}$ (performance convergence rate) depends upon the specific mixture resolving characteristics of the element, parameter, signature, or feature, estimators; and would be expected to vary from category J_i to J_j. Consequently one is interested in that estimating category, J_o, which possesses some rapid convergence characteristic toward the asymptotic performance, $P_e(J_o)$, for that category while maintaining some minimal dispersion of performance about the optimal performance at that stage. The development of one estimating category, J_o, with the foregoing optimality criteria imposed, is carried out in the next chapter.

2-5. Optimality Preserved Under Linear Substitution of Mixture Resolving Estimators

Recalling (2.29) and (2.34)

$$\underset{\underline{h}\,\epsilon\,E_n}{\text{Min}} \; P_e(N, J) \equiv \underset{\underline{h}\,\epsilon\,E_n}{\text{Max}} \; \lambda(N, J) = \lambda_o(N, J) \tag{2.45}$$

It was seen that the average error probability at the Nth stage relative to the estimating category, $J \,\epsilon \{J_r\}_R$, is minimal when the vector operator applied to the observations consists of the unique eigenvector corresponding to $\lambda_o(N, J)$. Further it is seen from (2.40), that the elements of this operator are linear substitutions of the mixture resolving estimators, $\{\hat{\theta}(N, J)\}$, for the estimating category J. In addition, if an estimating category, J_r, exists which is statistically consistent for the mixture parameters, then

$$\underset{N\to\infty}{\text{Lim}} \; P\{|\{\hat{\theta}(N, J)\} - \{C\}| > \epsilon\} = 0 \tag{2.46}$$

and thus,

$$\underset{N\to\infty}{\text{Lim}} \; P_e(N, J_r) \to P_e(J_r) \tag{2.47}$$

If the category contains no bias, then

$$P_e(J_r) = P_e(J_B) \tag{2.48}$$

and the structure converges to the matched filter.

Thus, this research contribution includes a formal development of the vector operator structure which minimizes the estimating category conditioned error probability and justifies, in the development context, the linear substitution of the Jth category, mixture resolving estimates into the structure. For the gaussian case, the structure has the same form as the Bayes matched filter structure, however, the elements differ in that they possess the mixture resolving properties (convergence rate, consistency) of the particular Jth category contained therein.

CHAPTER III

VARIATIONAL DEVELOPMENT OF THE PARAMETER ELEMENT ESTIMATING OPERATOR

3-1. Initial Separation of an Observation from a Mixture

The work in this chapter is concerned with the development of an estimating category, $J \varepsilon \{J_r\}$, which is to resolve a mixture with a specific criteria of resolution imposed, in order to provide elements for the detector structure developed in the previous section. In particular, the criteria that is desired is some sort of bounded, time minimized consistency. The interpretation of boundedness is to be taken in the sense that one wishes to minimize the dispersion of the values of the parameters being extracted and thus minimize the dispersion of dynamic performance from the classically optimum. In addition, it is desired to achieve time minimized consistency in some sense in order to achieve asymptotic performance in minimum time, granting that the first criteria will insure learning or adaptation such that the finite time asymptotic performance reasonably approaches the classically optimum.

Consider the mixture of observations, $\{\underline{z}_k\}_N$, where each observation lies in the space given by

$$\Omega = \omega_1 \cup \omega_2 \tag{3.1}$$

as illustrated in Fig. 3-1.

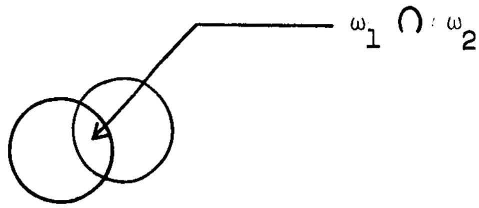

Fig. 3-1. Mixture Observation Space

In order to initiate a resolution of the mixture, $\{\underline{z}_k\}_N$, consider an initial separation of the jth observation, $\underline{z}_j$, from the mixture, by some process, into the space ω_i^*, i = 1, 2, where the subscript refers to the space of the reference for separation and the space is starred since it does not coincide with ω_i as a result of the true reference being unknown. Clearly, if the means, $\underline{\theta}_1$, $\underline{\theta}_2$, (samples of pulse waveforms), were known, they would serve as the reference (as elements in the detector structure), and the separation (classification, detection, and recognition in this case) would be achieved under the following total probability scheme.

$$
\begin{array}{ll}
P\{H_1|\underline{z}_j \ \varepsilon \ \omega_1\} = p_{11} & P\{H_2|\underline{z}_j \ \varepsilon \ \omega_1\} = p_{21} \\
P\{H_1|\underline{z}_j \ \varepsilon \ \omega_2\} = p_{12} & P\{H_2|\underline{z}_j \ \varepsilon \ \omega_2\} = p_{22}
\end{array}
\qquad (3.2)
$$

In the unsupervised case, the reference must be extracted from the mixture. Consider then a separation of the (j+1)th observation, $\underline{z}_{j+1}$, from the mixture, $\{\underline{z}_k\}_N$, by some process based upon the likeness or similarity of $\underline{z}_{j+1}$ with any other observation, $\underline{z}_k$, in particular with the immediately preceeding observation, $\underline{z}_j$. Consider the hypothetical decision rule,

$$\ell(\underline{z}_{j+1}, \underline{z}_j) \begin{array}{l} \geq \ell_j \equiv H_S \\ < \ell_j \equiv H_D \end{array} \tag{3.3}$$

Clearly, in order to preserve homogeneity of the system, the structure of (3.3) must be compatible with (2.3), in the sense that as the parameters are learned, no change in structure should be necessary with changes occurring only in the elements. This is seen to be so in Chapter IV where specific cases are studied. The significance of the subscripts, S and D, is that the hypotheses imply similarity and dissimilarity respectively, yielding the total starred probability scheme corresponding to the probabilities of the error of the first and second kind and the complementary probabilities of correct assignment for the similarity hypotheses.

$$\begin{aligned} P\{H_S \mid \underline{z}_j \in \omega_1 \cap \underline{z}_{j+1} \in \omega_1 \cup \underline{z}_j \in \omega_2 \cap \underline{z}_{j+1} \in \omega_2\} &= p_{11}(j) \\ P\{H_S \mid \underline{z}_j \in \omega_1 \cap \underline{z}_{j+1} \in \omega_2 \cup \underline{z}_j \in \omega_2 \cap \underline{z}_{j+1} \in \omega_1\} &= p_{12}(j) \\ P\{H_D \mid \underline{z}_j \in \omega_1 \cap \underline{z}_{j+1} \in \omega_1 \cup \underline{z}_j \in \omega_2 \cap \underline{z}_{j+1} \in \omega_2\} &= p_{21}(j) \\ P\{H_D \mid \underline{z}_j \in \omega_1 \cap \underline{z}_{j+1} \in \omega_2 \cup \underline{z}_j \in \omega_2 \cap \underline{z}_{j+1} \in \omega_1\} &= p_{22}(j) \end{aligned} \tag{3.4}$$

That is, the observation, $\underline{z}_{j+1}$, and the reference, $\underline{z}_j$, are either both from ω_1 or both from ω_2, or one from each in two possible ways. The starred space notation is used because the separation of an observation, $\underline{z}_{j+1}$, from the mixture, $\{\underline{z}_k\}_N$, by referencing with the observation, $\underline{z}_j$, separates the two into spaces such that if some combination of the separated observation were to be carried out to achieve an estimate, $\{\hat{\theta}\}^*$, the elements or moments achieved would not coincide with the elements of the true spaces, $\{\theta\}$, except possibly asymptotically.

It is appropriate to point out that the concept of one-shot separation on the basis of adjacent similarity is employed in the differentially coherent phase-shift-keyed system, however, no effort is made to achieve any improvement in separation or referencing, by any combination of the results of the referencing in order to effect an adaptation to or a learning of the true reference. Clearly a sequence of one-shot comparisons could be made, however an error followed by a sequence of correct comparisons of similarity would result in a severe error propagation format, which for low signal to noise ratios, is the most serious draw back of the DC-PSK system concept.

Thus the philosophy is to reinforce the structure, $\underline{h}$, given by (2.40), by combining the separated observations in a manner which imposes the desired criteria previously mentioned. The mathematical functionals imposing the desired criteria and the manner in which the observations are to be combined, are discussed in the sections which follow.

3-2. Time Dependent, Weighted Combination of the Separated Observations

Consider now, two or more observations, $\underline{z}_j$, $\underline{z}_{j+1}$, separated from $\{\underline{z}_k\}_N$ into ω_ℓ^*,

$$\{\underline{z}_j,\ \underline{z}_{j+1}\cdot\} = \underline{Z}^{(\omega_\ell^*)} \quad ; \quad \underline{z}_j = \begin{bmatrix} z_{1_j} \\ \vdots \\ z_{i_j} \\ \vdots \\ z_{n_j} \end{bmatrix} \qquad \ell = 1,\ 2 \tag{3.5}$$

Consider a time (observation) dependent, weighted combination of the two or more separated observations to form an estimate, $\hat{\underline{\theta}}(N_\ell,\ \underline{Z}^{(\omega_\ell^*)})$, of the mean associated with the space of that particular separation. The weighting process will always initiate with a combination of only one (the (j+1)th) observation with the similarity referenced observation (the jth as an example) and a discussion on the initial separation appears when specific cases are considered. The subscript ℓ is identified with the first similarity decision (say ω_1^*) and the next dis-similarly constitutes an observation separated into ω_2^*; or the reverse, since the order in which the space is assigned is of no consequence in the binary case.

Consider then, the time dependent, weighted average, given by

$$\underline{\hat{\theta}}(N_\ell, \underline{Z}^{(\omega_\ell^*)}) = N_\ell^{-1} H(N_\ell, \underline{Z}^{(\omega_\ell^*)}) \underline{Z}^{(\omega_\ell^*)} \tag{3.6}$$

where

$$H(N_\ell, \underline{Z}^{(\omega_\ell^*)}) = \begin{bmatrix} \underline{H}_1'(N_\ell) \\ \cdot \\ \cdot \\ \cdot \\ \underline{H}_n'(N_\ell) \end{bmatrix} \; ; \; \underline{H}_i(N_\ell) = \begin{bmatrix} H_{i_1}(N_\ell) \\ \cdot \\ \cdot \\ \cdot \\ H_{i_{N_\ell}}(N_\ell) \end{bmatrix} \tag{3.7}$$

and

$$\underline{Z}'^{(\omega_\ell^*)} = (\underline{z}_1, \ldots, \underline{z}_n) \; ; \; \underline{z}_i = \begin{bmatrix} z_{i_1} \\ \vdots \\ z_{i_{N_\ell}} \end{bmatrix} \tag{3.8}$$

For simplicity, since only one group of separations will be considered at a time, the subscripts ℓ, ω_ℓ^* can be dropped, and N_ℓ can be taken as N_1, the number of similar observations separated into one group prior to a dis-similar observation, under the condition that $N_1 \geq 2$. Thus (3.6) can be written for easier interpretation as

$$\underline{\hat{\theta}}(N_1, \underline{Z}) = N_1^{-1} H(N_1, \underline{Z}) \underline{Z} \tag{3.9}$$

and the estimate of the ith sample associated with the discrete unknown waveform of dimensionality n is given by

$$\hat{\Theta}_i = N_1^{-1} \underline{H}_i'(N_1) \underline{z}_i \tag{3.10}$$

where $\underline{H}_i(N_1)$ is the ith vector operator element of (3.7) and $\underline{z}_i$ is the vector of the ith samples in the separation sub-mixture, as

$$\underline{Z} = \{\underline{z}_j, \underline{z}_k, \cdot\} = \begin{bmatrix} z_{1_j} & \cdots & z_{i_j} & \cdots & z_{n_j} \\ \vdots & & & & \\ z_{1_k} & \cdots & z_{i_k} & \cdots & z_{n_k} \\ \vdots & & & & \\ z_{1_{N_1}} & \cdots & z_{i_{N_1}} & \cdots & z_{n_{N_1}} \end{bmatrix} ; \underline{z}_i = \begin{bmatrix} z_{i_j} \\ \vdots \\ z_{i_k} \\ \vdots \\ z_{i_{N_1}} \end{bmatrix} \tag{3.11}$$

The specific properties of the observation-time dependent, weighting, operator in (3.9) must take into account the criteria discussed in 3-1 and functionals for these criteria are developed in the next section.

3-3. Development of a Constrained Functional for Maximizing Convergence Rate and Minimizing Dispersion

Consider the following functionals for an estimating category J,

$$\Psi_R = \frac{\partial}{\partial N_1} \Psi_D \tag{3.12}$$

$$\Psi_D = \underset{*}{<} ||\hat{\underline{\theta}}(N_1, \underline{Z}) - \underline{\theta}||^2 > \tag{3.13}$$

$$\Psi_E = \underset{*}{<} ||\hat{\underline{\theta}}(N_1, \underline{Z}) - \underset{*}{<} \hat{\underline{\theta}}(N_1, \underline{Z}) >||^2 > \tag{3.14}$$

$$\lim_{N_1, N \to \infty} P\{|\hat{\underline{\theta}}(N_1, \underline{Z}) - \underline{C}| > \varepsilon\} \to 0 \tag{3.15}$$

where Ψ_D represents the normed measure of distance in n space, at state N_1, Ψ_R represents the rate of change of distance, Ψ_E is the mean-square combination error, and if $\underline{C}$ coinsides with the true pulse waveform samples, the estimating category is unbiased with probability one. For the ith element of the estimate associated with the ith element of the multi-dimensional observation,

$$\hat{\Theta}_i(N_1, \underline{Z}) = N_1^{-1} \underline{H}_i'(N_1, \underline{Z}) \underline{z}_i \tag{3.16}$$

consider the ith component in the functionals (3.12) through (3.14),

$$\psi_R \quad \frac{\partial}{\partial N_1} \psi_D \tag{3.17}$$

$$\psi_D = \underset{*}{<} (\frac{1}{N_1} \underline{H}_i^{'}(N_1, \underline{Z}) \, \underline{z}_i - \Theta_i)^2 > \tag{3.18}$$

$$\psi_E = \underset{*}{<} (\frac{1}{N_1} \underline{H}_i^{'}(N_1, \underline{Z}) \, \underline{z}_i - \underset{*}{<} \frac{1}{N_1} \underline{H}_i^{'}(N_1, \underline{Z}) \, \underline{z}_i >)^2 > \tag{3.19}$$

One then wishes a general functional, ψ, composed of constrained combinations of the functionals (3.17) through (3.18) in such a manner to effect an operator, $\underline{H}(N_1, \underline{Z})$, (henceforth referring to the ith element), corresponding to the estimating category J with the desired properties of maximizing the convergence rate and minimizing the dispersion of dynamic performance. The initial approach was to consider,

$$\psi = \psi_R + \gamma \, \psi_E \tag{3.20}$$

where γ is the so-called La-Grangian multiplier coefficient associated with the constraint, and to impose upon the variation the conditions

$$\underset{\underline{H} \, \varepsilon \, E_{N_1}}{\text{Max}} \quad \psi_R \quad ; \quad \psi_R < 0 \tag{3.21}$$

and

$$\underset{\underline{H} \, \varepsilon \, E_{N_1}}{\text{Min}} \quad \psi_E \tag{3.22}$$

In this manner it was hoped to directly maximize the negative rate of change of distance (convergence) and minimize the dispersion of the dynamic performance by constraining the mean square error of the Jth category. However from (3.16), it is seen that

$$\psi_R = \frac{\partial}{\partial N_1} \underset{*}{<} (\frac{1}{N_1} \underline{H}_i^{'}(N_1, \underline{Z}) \, \underline{z}_i - \Theta_i)^2 > \tag{3.23}$$

and with a variation still to be taken on $\underline{H}$, the expression for ψ_R is both explicit and implicit in N_1, and thus a derivative cannot be taken directly. However this problem can be circumvented to some extent by considering

$$\psi = \psi_D + \gamma\ \psi_E \tag{3.24}$$

as the functional upon which a variation on $\underline{H}$ is to be taken to minimize the normed measure of distance subject to a constraint on the mean square error. Then upon solution for $\underline{H}$, taking

$$\frac{\partial}{\partial\gamma} \psi_R(\underline{H}_o\ (N_1,\gamma)) = \frac{\partial}{\partial\gamma} \frac{\partial}{\partial N_1} \psi_D(\underline{H}_o(N_1,\ \gamma)) \tag{3.25}$$

and adjusting the normed rate of change of distance with the La-Grangian multiplier while maintaining some control over the mean-square error, it is hoped to effect

$$\underset{\gamma\ \varepsilon\{\gamma\}_m}{\text{Max}} \left\{ \frac{\partial}{\partial N_1} \underset{\underline{H}\ \varepsilon\ E_{N_1}}{\text{Min}} \psi_D \right\} = \underset{\gamma\ \varepsilon\{\gamma\}_m}{\text{Max}} \left\{ \frac{\partial}{\partial N_1} \psi_D(\underline{H}_o(N_1,\ \gamma)) \right\}$$

$$= \underset{\gamma\ \varepsilon\{\gamma\}_m}{\text{Max}} \psi_R(\underline{H}_o,\ \gamma) = \psi_R(\underline{H}_o,\ \gamma_o) \tag{3.26}$$

The validity of this approach is borne out in experimental studies of the control effected upon convergence rate and dispersion by the adjustment of the La-Grangian multiplier in $\underline{H}_o(N_1,\ \gamma)$. Thus the operator, $\underline{H}_o(N_1,\ \gamma)$ which satisfies the minimization of the functional, $\psi(\underline{H})$, as

$$\psi(\underline{H}_o(N_1,\ \underline{Z})) = \psi_D(\underline{H}_o(N_1,\ \gamma,\ \underline{Z})) + \gamma\ \psi_E(\underline{H}_o(N_1,\ \gamma,\ \underline{Z})) \tag{3.27}$$

followed by an adjustment of γ obtained by

$$\frac{\partial}{\partial\gamma} \frac{\partial}{\partial N_1} \psi_D\ (\underline{H}_o(N_1,\ \gamma,\ \underline{Z})) = 0 \tag{3.28}$$

to yield some $\underline{H}_o(N_1,\ \gamma_o,\ \underline{Z})$, is the operator desired to weight the combination of the separated observations, in effect yielding the optimum estimating category, J_o, desired. The extraction of the parameters, $\{\Theta\}$, in this manner is thus to reinforce the structure given by (2.40) in such a manner as to minimize convergence time and dispersion of dynamic performance. The mechanics associated with the minimization of ψ, and the solution for $\underline{H}$ are carried out in the following section.

3-4. Minimization of ψ and solution for $\underline{H}$.

Recalling the functional to be minimized,

$$\psi(\underline{H}) = \psi_D(\underline{H}) + \gamma\ \psi_E(\underline{H}) \tag{3.29}$$

by a variation on $\underline{H}$; it is seen that minimization can be achieved by partial differentiation as a result of the discrete, finite dimensionality of (3.29). Thus, specifically, recalling that (3.29) is associated with the ith element

$$\psi_D(\underline{H}) = \underset{*}{<} (N_1^{-1}\ \underline{H}_i'\ \underline{z}_i - \Theta_i)^2 > \tag{3.30}$$

which upon expansion becomes

$$\psi_D(\underline{H}) = \underset{*}{<} \Theta_i^2 - \frac{2}{N_1} \underline{H}_i'\ \underline{z}_i\ \Theta_i + \frac{1}{N_1{}^2} \underline{H}_i'\ \underline{z}_i\ \underline{z}_i'\ H_i > \tag{3.31}$$

and by a distribution of the expectation, yields

$$\psi_D(\underline{H}) = \Theta_i^2 - \frac{2}{N_1} \underline{H}_i' \underset{*}{<} \underline{z}_i\ \Theta_i > + \frac{1}{N_1{}^2} \underline{H}_i' \underset{*}{<} \underline{z}_i\ \underline{z}_i' > \underline{H}_i \tag{3.32}$$

The validity of (3.32) can be checked by expanding $(\frac{1}{N_1} \underline{H}_i'\ \underline{z}_i - \Theta_i)^2$ and writing the result as a series followed by the starred expectation on the finite series. Also

$$\psi_E(\underline{H}) = \underset{*}{<} (\frac{1}{N_1} \underline{H}_i'\ \underline{z}_i - \underset{*}{<} \frac{1}{N_1} \underline{H}_i'\ \underline{z}_i >)^2 > \tag{3.33}$$

which upon expansion becomes

$$\psi_E(\underline{H}) = \underset{*}{<} \frac{1}{N_1{}^2} \underline{H}_i'\ \underline{z}_i\ \underline{z}_i'\ \underline{H}_i - \frac{2}{N_1{}^2} \underset{*}{<} \underline{H}_i'\ \underline{z}_i > \underline{H}_i'\ \underline{z}_i + \frac{1}{N_1{}^2} \underset{*}{<} \underline{H}_i'\ \underline{z}_i >^2 > \tag{3.34}$$

and by a distribution of the expectation, yields

$$\psi_E(\underline{H}) = \frac{1}{N_1{}^2} (\underline{H}_i' \underset{*}{<} \underline{z}_i \; \underline{z}_i' > \underline{H}_i - 2\underline{H}_i' \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >' \underline{H}_i + \underline{H}_i' \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >' \underline{H}_i) \qquad (3.35)$$

Finally, by collecting terms

$$\psi_E(\underline{H}) = \frac{1}{N_1{}^2} \{\underline{H}_i' \; [\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >'] \; \underline{H}_i\} \qquad (3.36)$$

Now, combining (3.32) and (3.36)

$$\psi(\underline{H}) = \{\Theta_i{}^2 - \frac{2\Theta_i}{N_1} \underline{H}_i' \underset{*}{<} \underline{z}_i > + \frac{1}{N_1{}^2} \underline{H}_i' \underset{*}{<} \underline{z}_i \; \underline{z}_i' > \underline{H}_i\}$$
$$+ \gamma \frac{2}{N_1{}^2} \{\underline{H}_i' \; [\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >'] \; \underline{H}_i\} \qquad (3.37)$$

and taking

$$\frac{\partial}{\partial \underline{H}} \psi(\underline{H}) = \underline{0} \qquad (3.38)$$

yields

$$- \frac{2}{N_1} \Theta_i \underset{*}{<} \underline{z}_i > + \frac{2}{N_1{}^2} \underset{*}{<} \underline{z}_i \; \underline{z}_i' > \underline{H}_i + \gamma \frac{2}{N_1{}^2} [\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >'] \; \underline{H}_i = 0 \qquad (3.39)$$

as the minimized system of equations. Clearing (3.39) of N_1 and rewriting into the standard form, the system is

$$[\underset{*}{<} \underline{z}_i \; \underline{z}_i' > + \gamma \; (\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >')] \; \underline{H}_i = N_1 \; \Theta_i \underset{*}{<} \underline{z}_i > \qquad (3.40)$$

By the nature of matrices composed of first and second moment expectations,

$$R\{[\underset{*}{<} \underline{z}_i \; \underline{z}_i' > + \gamma \; (\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \; \underset{*}{<} \underline{z}_i >')]\} = N_1 \qquad (3.41)$$

and thus (3.40) is a full rank system of equations and consequently a unique solution, $\underline{H}_o$, for $\underline{H}$ exists, given by

$$\underline{H}_o(N_1, \gamma) = [\underset{*}{<} \underline{z}_i \; \underline{z}_i' > + \gamma \; (\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >')]^{-1} N_1 \Theta_i \underset{*}{<} \underline{z}_i > \qquad (3.42)$$

Thus the weighting operator, $\underline{H}_o$, which satisfies the conditions imposed by (3.27) is given in (3.42). In order to consider further, the application of (3.42) to specific cases, it is essential to examine the elements in the system matrix. This is carried out in the following section.

3-5. Elements of the Weighting Operator, $\underline{H}_o$

Recalling (3.42), the matrix of the system of equations is given by

$$C(N_1, \gamma) = [\underset{*}{<} \underline{z}_i \; \underline{z}_i' > + \gamma \, (\underset{*}{<} \underline{z}_i \; \underline{z}_i' > - \underset{*}{<} \underline{z}_i > \; \underset{*}{<} \underline{z}_i >')] \quad (3.43)$$

with the expectation taken with respect to the starred space defined by separation up to stage N_1. Also, recalling (3.4), it is seen that the separation initially produces two submixtures, $\underline{Z}^{(\omega_1^*)}$, $\underline{Z}^{(\omega_1^*)}$, with proportionality parameters given by (3.44) below. This idea is common to the case of detection with known parameters, where classification is not without error, and the classified observations constitute two mixtures of "zeros in ones" and "ones in zeros" with fixed parameters of mixture proportionality. In the adaptive and learning structures, the parameters of proportionality are observation dependent, reflecting the convergence properties, and constitute a Markov process with random transition probabilities. Consider an initial test of $\underline{z}_{j+1}$ against $\underline{z}_j$ as a reference, and the hypotheses given by (3.4). The following submixture generating scheme is observed, with mixing parameters,

$$\begin{aligned}
P\{H_S \cap A_{11} \cup H_D \cap A_{21}\} &= p_{11}(j)\, p_1^{\,2} + p_{22}(j)\, p_1 p_2 = p_{11}^*(j) \\
P\{H_S \cap A_{12} \cup H_D \cap A_{22}\} &= p_{12}(j)\, p_1 p_2 + p_{21}(j)\, p_2^{\,2} = p_{12}^*(j) \\
P\{H_S \cap A_{21} \cup H_D \cap A_{11}\} &= p_{12}(j)\, p_1 p_2 + p_{21}(j)\, p_1^{\,2} = p_{21}^*(j) \\
P\{H_S \cap A_{22} \cup H_D \cap A_{12}\} &= p_{11}(j)\, p_2^{\,2} + p_{22}(j)\, p_1 p_2 = p_{22}^*(j)
\end{aligned} \quad (3.44)$$

$$A_{k\ell} = \{\underline{z}_j \; \varepsilon \; \omega_k \quad \underline{z}_{j+1} \; \varepsilon \; \omega_\ell\}; \qquad k, \ell = 1, 2$$

In words, (3.44) means that referencing $\underline{z}_{j+1}$ against $\underline{z}_j$ results in a separation such that the observations are both similar and are separated as similar or dissimilar, or are dissimilar and are separated as similar or dissimilar respectively. The terms, $p_{11}(j)$ and $p_{22}(j)$ are the probabilities that the observations are statistically similar or dissimilar and are separated as such. Thus the values of the probabilities are constantly changing as the structure is re-inforced, characteristic of the learning or adaptation that is taking place, and this is represented by the subscript j. As such, the probabilities are random variables, in particular teranomial random variables (multi-nomial with four variables).

At any stage, j, consider the (k, ℓ)th element of

$$\underset{*}{<} \underline{z}_i \, \underline{z}_i' > = [r_{\underline{z}_i \underline{z}_i}(k, \ell)] \quad i = 1, 2, \ldots, n \quad k,\ell = 1, 2, \ldots, N_1 \tag{3.45}$$

given by

$$r_{k\ell} = \underset{*}{<} z_{i_k} \, z_{i_\ell} > = \underset{*}{<} (\Theta_{i_k} + N_{i_k})(\Theta_{i_\ell} + N_{i_\ell}) > \tag{3.46}$$

which for zero mean noise, upon expansion, becomes

$$r_{k\ell} = \underset{*}{<} (\Theta_{i_k} \Theta_{i_\ell} + N_{i_k} N_{i_\ell}) > \tag{3.47}$$

Taking the expectation over the starred space, with reference to (3.44), for the binary case, (3.47) yields

$$r_{k\ell} = (p_{11}^* \Theta_{1_i} + p_{12}^* \Theta_{2_i})^2 + (p_{11}^* + p_{12}^*)^2 < N_{i_k} N_{i_\ell} > \quad k \neq \ell \tag{3.48}$$

For k = ℓ, (3.46) is

$$r_{k\ell} = r_{kk} = \underset{*}{<} z_{i_k}^2 > = \underset{*}{<} (\Theta_{i_k} + N_{i_k})^2 > \tag{3.49}$$

which upon expansion, for the zero-mean noise case, and expectation $\underset{*}{<} \; >$ yields

$$r_{k\ell} = p_{11}^* \Theta_{1_i}^2 + p_{12}^* \Theta_{2_i}^2 + (p_{11}^* + p_{12}^*) < N_{i_k}^2 > \quad k = \ell \tag{3.50}$$

Since the expectation has been taken at the jth stage, the expectation of the random variable, $p^*(j)$, is replaced by the expected value p^*; at that stage and consequently no subscripts, j, appear in (3.48) and (3.50).

The remaining matrix in (3.43) is

$$M_{\underline{z}_i \underline{z}_i}(k, \ell) = \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >' \tag{3.51}$$

where the (k, ℓ)th element is given by

$$m_{k\ell} = \underset{*}{<} z_{i_k} > \underset{*}{<} z_{i_\ell} > = \underset{*}{<} (\Theta_{i_k} + N_{i_k}) > \underset{*}{<} (\Theta_{i_\ell} + N_{i_\ell}) > \tag{3.52}$$

which for zero-mean noise becomes

$$m_{k\ell} = \underset{*}{<} \Theta_{i_k} > \underset{*}{<} \Theta_{i_\ell} > = (p_{11}^* \Theta_{1_i} + p_{12}^* \Theta_{2_i})^2 \quad k \neq \ell \tag{3.53}$$

For the case, $k = \ell$, in view of the zero mean noise,

$$m_{k\ell} = m_{kk} = \underset{*}{<} \Theta_{i_k} >^2 = (p_{11}^* \Theta_{1_i} + p_{12}^* \Theta_{2_i})^2 \quad k = \ell \tag{3.54}$$

The substitution of (3.48), (3.50), (3.53) and (3.54) into (3.43), yields the general expression of the matrix to be inverted, at the jth stage, in order to provide weighting coefficients for combining the separated observations in a manner such as to maximize convergence rate and minimize dispersion. With the results developed thus far, it is now possible to consider specific cases such as, $\{\underline{\Theta}_1 = -\underline{\Theta}_2\}$, $\{\underline{\Theta}_1, \underline{\Theta}_2\}$, etc.; and specific forms of $\underline{H}_o(N_1, \underline{Z}, \gamma)$, given by (3.42) are developed in the next chapter. In particular, an iterative format of the recursion for $\underline{H}_o(N_1, \underline{Z}, \gamma)$ is developed.

CHAPTER IV

THE PARAMETER ELEMENT ESTIMATING OPERATOR FOR THE BI-POLAR AND GENERAL TWO-CATEGORY SIGNAL PATTERN MODEL

4-1. The Integrated Signal Processing Structure

The work in the previous chapter resulted in the development of an operator,

$$\underline{H}_{o_i}(N_1, \gamma, \underline{Z}) = N_1 \, C^{-1}(N_1, \gamma) \, \Theta_i < \underline{z}_i > \tag{4.1}$$

as the weighting operator of the ith samples in a time slot of dimensionality n, where

$$C(N_1, \gamma) = [R + \gamma \, (R - M)] \tag{4.2}$$

and

$$< \underline{z}_i > = (p_{11}^* \, \Theta_{1_i} + p_{12}^* \, \Theta_{2_i}) \, \underline{1} \tag{4.3}$$

The vector operator was found to be the unique solution of a full rank system of equations obtained by extremizing a constrained functional ψ. The functional ψ was a combination of functionals arranged in a manner such as to allow the maximization of the convergence rate and minimization of the dispersion of dynamic performance, the contention being that maximization of the consistency of the time dependent, weighted estimates of the samples of the discrete unknown pulses with a constrained mean-square estimation error carries these properties over into the detection, recognition, and classification properties of the structure. These hypotheses are proven correct by a digital computer simulation of the signal processing structure, signal patterns in noise, and the actual processing of the observation signal by the structure. The results of the computer simulation experiments and a discussion of related performance appears in Chapter VI.

At this point it is appropriate to introduce a basic diagram of the data processing structure, Fig. 4-1, which can be interpreted as follows.

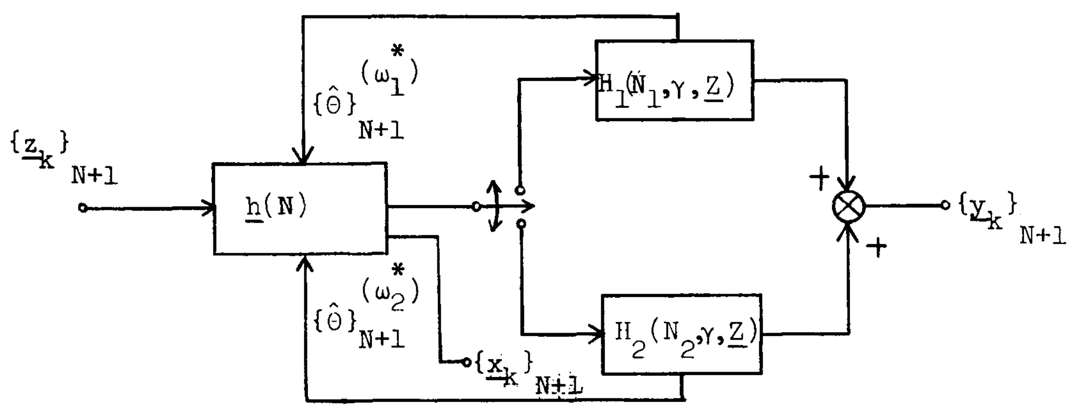

Fig. 4-1. Optimized Feedback-Mixture-Resolving Detector Structure

The input, $\{\underline{z}_k\}_{N+1}$, is an unclassified vector sequence of noisy discrete binary pulses (mixture) as described in the introduction, and neither the pulse waveforms nor the noise statistics are known. The structure $\underline{h}$, as given by (2.40), is the eigenvector operator detector structural form which minimizes the conditional average error probability as it initially isolates observations and, with the first re-inforcement of its elements in the structure, generates a sequence of binary decisions, $\{\underline{x}_k\}_{N+1}$. The structure $\underline{h}$, in its initial separation and subsequent decision, gates the time slots through to the operators $H_1(N_1, \gamma, \underline{Z}^{(\omega_1^*)})$

and $H_2(N_2, \gamma, \underline{Z}^{(\omega_2^*)})$, which combine, in a time-varying, weighted manner, the gated observations, to generate up-dated values of the elements (parameters) in the structure $\underline{h}$. The time varying operators, $H_1(N_1, \gamma, \underline{Z}^{(\omega_1^*)})$ and $H_2(N_2, \gamma, \underline{Z}^{(\omega_2^*)})$ are not both active simultaneously, however the sum of the outputs generates a continuous sequence, $\{y_k\}_{N+1}$, which is the extracted signal pattern as "recovered" from the observation mixture by the operators H_1 and H_2. As the elements converge, eg., as the outputs, $\{\hat{\theta}\}^{(\ell)}_{N_\ell+1}$, reinforce $\underline{h}$ toward more correct decisions, the estimates (extractions) of the discrete pulse waveforms improve and the sequence $\{y_k\}_{N+1}$ resembles more closely the true signal pattern. The sequence $\{\underline{x}_k\}_{N+1}$ is the decision sequence and agrees in format with $\{y_k\}_{N+1}$.

Thus it can be concluded that the above structure is a time-varying, weighted, decision-directed structure, which extracts the initial reference from the observations. The formulation and structure differs from the conventional decision-directed system in two fundamental ways. First, the conventional decision-directed system does not extract the initial reference from the observations, and second, unweighted linear averages are taken to provide reinforcement of the detector structure. As a result, the conventional decision-directed system is seen to be unreliable at signal-to-noise ratios from just above 0 db. down.

It can be seen that this is inherently a feedback structural reinforcement system and will be contrasted in Chapter V with a feed-forward mixture resolving structure. However, it will be seen that the feed-forward system requires more apriori information, in certain terms, regarding the observation mixture.

4-2. The Optimized Weighting Operator for the Bi-Polar Signal Pattern Model

For the mixture observation, $\{\underline{z}_k\}_{N+1}$, as described, the most general relationship between the two unknown discrete pulses is given by

$$\underline{\theta}_2 = G\ \underline{\theta}_1 \qquad (4.4)$$

where the elements differ by a matrix operator transformation G. For the case

$$G = k\ I \qquad (4.5)$$

the signal pattern model is said to consist of two signal categories differing only by a location parameter, and for the specific case; k = -1, the signal pattern is bi-polar with unknown pulse waveforms.

Consider the bi-polar case then and refer to the upper branch in Fig. 4-1 as ω_1^*. The weighting operator for the ith element of the discrete pulse $\underline{\theta}_1$, upon the ith elements of the sequence of n dimensional time slots, $\{\underline{Z}^{(\omega_1^*)}\}_{N_1}$, where $N_1 \geq 2$, depending upon the number of likeness decisions gated to ω_1^* or ω_2^* prior to at least two observations gated to the "other" branch, is given by

$$\underline{H}_{o_i}(N_1, \gamma, \underline{Z}^{(\omega_1^*)}) = N_1\ C^{-1}(N_1, \gamma)\ \Theta_{1_i} \underset{*}{<} \underline{z}_i > \qquad (4.6)$$

Now since $\underline{\theta}_1 = -\underline{\theta}_2$, (4.3) becomes

$$\underset{*}{<} \underline{z}_i > = (p_{11}^* - p_{12}^*)\ \Theta_{1_i}\ \underline{1} \qquad (4.7)$$

Also from (3.43) and (4.2),

$$C(N_1, \gamma) = [R + \gamma\ (R - M)] \qquad (4.8)$$

where R and M are N_1 dimensional matrices given by

$$R = [r_{\underline{z}_i\ \underline{z}_i}(k, \ell)] = \underset{*}{<} \underline{z}_i\ \underline{z}_i' > \qquad (4.9)$$

and

$$M = [m_{\underline{z}_i\ \underline{z}_i}(k, \ell)] = \underset{*}{<} \underline{z}_i > \underset{*}{<} \underline{z}_i >' \qquad (4.10)$$

For the bi-polar case, the (k, ℓ)th elements of R, from (3.48) and (3.50) become

$$r_{k\ell} = (p_{11}^* - p_{12}^*)^2\ \Theta_{1_i}^2 + (p_{11}^* + p_{12}^*)^2 < N_{i_k}\ N_{i_\ell} >; k \neq \ell \quad (4.11)$$

and

$$r_{k\ell} = (p_{11}^* + p_{12}^*)\ \Theta_{1_i}^2 + (p_{11}^* + p_{12}^*)\ \sigma_i^2 \quad k = \ell \quad (4.12)$$

where σ_i^2 is the variance of the zero-mean noise in the ith sample of each time slot. From (3.53) and (3.54), the (k, ℓ)th elements of M are

$$m_{k\ell} = (p_{11}^* - p_{12}^*)^2\ \Theta_{1_i}^2 \quad k,\ell = 1, 2, \ldots, N_1 \quad (4.13)$$

Consider for the time being, stationary, white noise, such that

$$\underset{*}{<} N_{i_k}\ N_{i_\ell} > = 0 \quad (4.14)$$

and

$$\sigma_i^2 = \sigma^2 \quad i = 1, 2, \ldots, n \quad (4.15)$$

Thus from (4.14), (4.15) and representing $(p_{11}^* - p_{12}^*)$ by Δp, the elements of R and M become

$$r_{k\ell} = \Delta p^2\ \Theta_{1_i}^2 \quad k \neq \ell \quad (4.16)$$

and

$$r_{k\ell} = (p_{11}^* + p_{12}^*)\ [\Theta_{1_i}^2 + \sigma^2] \quad k = \ell \quad (4.17)$$

Also

$$m_{k\ell} = \Delta p^2\ \Theta_{1_i}^2 \quad k,\ell = 1, 2, \ldots, N_1 \quad (4.18)$$

$$R - M = [r_{k\ell} - m_{k\ell}] \quad (4.19)$$

where

$$[r_{k\ell} - m_{k\ell}] = (p_{11}^* + p_{12}^*)\ [\Theta_{1_i}^2 + \sigma^2] - \Delta p^2\ \Theta_{1_i}^2 \quad k = \ell \quad (4.20)$$

$$= 0 \quad k \neq \ell$$

Thus from (4.16) through (4.20), (4.8) becomes

$$C(N_1, \gamma) = [R + \gamma\{(p_{11}^* + p_{12}^*)[\Theta_{1_i}^2 + \sigma^2] - \Delta p^2\, \Theta_{1_i}^2\}\, I]$$

$$= \begin{bmatrix} (p_{11}^* + p_{12}^*)[\Theta_{1_i}^2 + \sigma^2] & \Delta p^2\, \Theta_{1_i}^2 & \cdots & \Delta p^2\, \Theta_{1_i}^2 \\ \Delta p^2\, \Theta_{1_i}^2 & (p_{11}^* + p_{12}^*)[\Theta_{1_i}^2 + \sigma^2] & \Delta p^2\, \Theta_{1_i}^2 \cdots & \Delta p^2\, \Theta_{1_i}^2 \\ \vdots & & & \\ \Delta p^2\, \Theta_{1_i}^2 & \cdots \Delta p^2\, \Theta_{1_i}^2 & & (p_{11}^* + p_{12}^*)[\Theta_{1_i}^2 + \sigma^2] \end{bmatrix}$$

$$+ \gamma(r_{kk} - m_{kk})\, I \qquad (4.21)$$

where I is the identity matrix. Now by factoring $\Delta p^2\, \Theta_{1_i}^2$ from each term in (4.21) and combining, (4.21) becomes the N_1 dimensional, square, symmetric form

$$C(N_1, \gamma) = \Delta p^2\, \Theta_{1_i}^2 \begin{bmatrix} (a + \gamma b) & 1 & \cdots & 1 \\ 1 & (a + \gamma b) & 1 \cdots & 1 \\ \vdots & & & \\ 1 & \cdots & 1 & (a + \gamma b) \end{bmatrix} \qquad (4.22)$$

where

$$a \triangleq \frac{(p_{11}^* + p_{12}^*)\,[\Theta_{1_i}^2 + \sigma^2]}{\Delta p^2\, \Theta_{1_i}^2} \qquad (4.23)$$

and

$$b \triangleq \left(\frac{(p_{11}^* + p_{12}^*)[\Theta_{1_i}^2 + \sigma^2]}{\Delta p^2\, \Theta_{1_i}^2} - 1\right) \qquad (4.24)$$

For compactness, set

$$c = a + \gamma b \qquad (4.25)$$

yielding the system matrix representation

$$C(N_1, \gamma) = \begin{bmatrix} c & 1 & \cdots & \cdots & 1 \\ 1 & c & 1 & \cdots & 1 \\ \vdots & & & & \\ \vdots & & & & \\ 1 & \cdots & \cdots & 1 & c \end{bmatrix} \qquad (4.26)$$

The matrix $C(N_1, \gamma)$ is of full rank and a general inverse of (4.26) substituted into (4.6) would yield the specific form of the weighting operator $\underline{H}_{oi}$. However from (4.7) and (4.26), (4.6) can be rewritten and the equivalent of a general inverse can be more readily obtained in compact form. Thus (4.6) becomes

$$C(N_1, \gamma)\, \underline{H}_{oi}(N_1, \gamma, \underline{Z}) = N_1\, \Delta p^{-1}\, \underline{1} \qquad (4.27)$$

Now a full rank system of equations such as (4.27) has the form

$$C\, \underline{H} = \underline{g} \qquad (4.28)$$

with the alternate form of solution given by

$$\underline{H} = |C|^{-1}\, C^{*}\, \underline{g} \qquad (4.29)$$

where C^{*} is the adjoint of the cofactor matrix for C. Consequently, the kth element for $\underline{H}$ is given by

$$\underline{H}_k = |C|^{-1} \sum_{\ell=1}^{N_1} g_\ell\, C_{\ell k} \qquad (4.30)$$

where $\sum_{\ell=1}^{N_1} g_\ell\, C_{\ell k}$ is the expansion of the determinant formed from C by replacing its kth column with $\underline{g}$. However in (4.27)

$$\underline{g} = N_1\, \Delta p^{-1}\, \underline{1} = N_1\, \Delta p^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad (4.31)$$

and it can be shown that

$$\sum_{\ell=1}^{N_1} g_\ell\, C_{\ell k} = N_1\, \Delta p^{-1}(1) \sum_{\ell=1}^{N_1} C_{\ell k} \qquad k = 1, 2, \ldots, N_1 \qquad (4.32)$$

In particular, by the specific nature of the symmetry of $C(N_1, \gamma)$,

$$\sum_{\ell=1}^{N_1} C_{\ell k} = (c-1)^{(N_1-1)} \qquad k = 1, 2, \ldots, N_1 \tag{4.33}$$

and thus the elements of $\underline{H}_{o_i}(N_1, \gamma, \underline{Z})$ are given by

$$H_{o_{i_k}}(N_1, \gamma, \underline{Z}) = N_1 \, \Delta p^{-1} \, |C|^{-1}_{(N_1)} \, (c-1)^{(N_1-1)} \; ; \qquad k = 1, 2, \ldots, N_1 \tag{4.34}$$

where $|C|_{(N_1)}$ is an N_1th order determinant of the matrix $C(N_1, \gamma)$. In order to complete the development of $\underline{H}_{o_i}(N_1, \gamma, \underline{Z})$ into a compact closed form, it is necessary to obtain a compact closed form expression for $|C|_{(N_1)}$. Since C is of full rank, it has N_1 distinct eigenvalues, λ_k, and thus a similarity transformation, T, exists such that

$$T' \, CT = [\lambda_k] \tag{4.35}$$

where $[\lambda_k]$ is the diagonal matrix of eigenvalues of C and is similar to C. Furthermore, since eigenvalues are invariant under transformation,

$$|C|_{(N_1)} = |[\lambda_k]| = \prod_{k=1}^{N_1} \lambda_k \tag{4.36}$$

Also

$$\mathrm{Tr}\, C = \sum_{k=1}^{N_1} \lambda_k = \sum_{k=1}^{N_1} c_{kk} = N_1 \, c \tag{4.37}$$

Consequently, from (4.36), (4.37) and by the process of induction, the closed form expression for $|C|_{(N_1)}$ becomes

$$|C|_{(N_1)} = (c-1)^{(N_1-1)} \, (c + [N_1-1]) \tag{4.38}$$

Thus with (4.38) and (4.34), the operator for the ith element becomes, at the N_1th combination stage

$$\underline{H}_{o_i}(N_1, \gamma, \underline{Z}^{(\omega_1^*)}) = \frac{N_1 \, \Delta p^{-1}}{[N_1 + (c-1)]} \, \underline{1} \tag{4.39}$$

The iterative format of the recursion is not yet obvious, however, the apparent requirement of time dependent weighting can be deduced from the appearance of N_1.

4-3. The Recursive Operator Format for the Bi-Polar Signal Pattern Model

Recalling (3.16), the weighting estimate of the ith element of the discrete pulse, $\underline{\theta}_1$, is given by

$$\hat{\Theta}_{1_i}(N_1, \underline{Z}^{(\omega_1^*)}) = N_1^{-1} \, \underline{H}'_{o_i}(N_1, \gamma, \underline{Z}^{(\omega_1^*)})\underline{z}_i \tag{4.40}$$

Dropping the term $\underline{Z}^{(\omega_1^*)}$ for simplicity, and incorporating (4.39), (4.40) becomes

$$\hat{\Theta}_{1_i}(N_1) = \frac{1}{N_1} \frac{\Delta p^{-1} \, N_1}{[N_1 + (c-1)]} \, \underline{1}' \, \underline{z}_i \tag{4.41}$$

where, from (3.11), by subscripting, the vectors become

$$\underline{z}_i = \begin{bmatrix} z_{i_j} \\ z_{i_k} \\ \cdot \\ \cdot \\ \cdot \\ z_{i_{N_1}} \end{bmatrix} = \begin{bmatrix} z_i(j) \\ z_i(k) \\ \cdot \\ \cdot \\ \cdot \\ z_i(N_1) \end{bmatrix} \tag{4.42}$$

In particular, suppose j=1, the initial separation, then for ω_1^*, the combinations are as follows,

$$\hat{\Theta}_i(1) = \frac{1}{1} \left\{ \frac{\Delta p^{-1}}{[1 + (c-1)]} \; z_i(1) \right\} \tag{4.43}$$

$$\hat{\Theta}_i(2) = \frac{1}{2} \left\{ \frac{2 \, \Delta p^{-1}}{[2 + (c-1)]} \, z_i(1) + \frac{2 \, \Delta p^{-1}}{[2 + (c-1)]} \, z_i(2) \right\} \tag{4.44}$$

Now, by substituting (4.43) into (4.44), $\hat{\Theta}_i(2)$ becomes

$$\hat{\theta}_i(2) = \frac{1}{2}\left\{\frac{2[1-(c-1)]}{[2+(c-1)]}\hat{\theta}_i(1) + \frac{2\,\Delta p^{-1}}{[2+(c-1)]}z_i(2)\right\} \qquad (4.45)$$

Finally, following the substitution (4.43) through (4.45), (4.41) yields the iterative format of the recursion at the N_1th stage,

$$\hat{\theta}_i(N_1) = \left\{\frac{N_1(N_1-1+(c-1))}{(N_1-1)(N_1+(c-1))}\hat{\theta}_i(N_1-1) + \frac{N_1}{\Delta p(N_1+(c-1))}z_i(N_1)\right\}\frac{1}{N_1} \qquad (4.46)$$

It is immediately seen from (4.46) that the weighted combinations have a definite observation-time dependence. For any value of the coefficient (c-1), the effect of the time weighting is such that, in obtaining up-dated values of $\hat{\theta}_i(N_1)$ by a weighted combination of $z_i(N_1)$ with $\hat{\theta}_i(N_1-1)$, the weighting of the "accepted" observations incorporated in $\hat{\theta}_i(N_1-1)$ is larger, for positive values of c, than the weighting for the new observation to be "averaged" in. Also, it is seen that the time dependent weighting coefficients,

$$W_\Theta(N_1) = \frac{N_1(N_1-1+(c-1))}{(N_1-1)(N_1+(c-1))} \qquad (4.47)$$

and

$$W_z(N_1) = \frac{N_1}{(N_1+(c-1))} \qquad (4.48)$$

have

$$\lim_{N_1\to\infty} W_\Theta(N_1) = \lim_{N_1\to\infty} W_z(N_1) = 1 \qquad (4.49)$$

however the rates at which each approaches unity differ and depend upon the coefficient (c-1). From (4.23), (4.24), and (4.25), it is seen that c is a function of p_{11}^*, p_{12}^*, Θ_{1_i}, σ^2, and γ. It will be shown in the next section, that for the low-signal to noise ratios considered, an adjustment of the La-Grange multiplier coefficient γ for "maximizing" the convergence rate, renders the observation time variable, N_1, dominant in $w_\Theta(N_1)$ and $w_z(N_1)$ and hence the weighting operator and related weights become non-parametric in this sense.

4-4. Adjustment of the La-Grange Multiplier for Maximizing Convergence Rate

The analysis for the adjustment of the constraining coefficient proceeds as follows. From (3.32), the ith element of the normed distance measure functional is

$$\psi_D(\underline{H}_o(N_1, \gamma)) = \Theta_{1_i}^{\ 2} - \frac{2}{N_1} \underline{H}_o'(N_1, \gamma)\Theta_{1_i}^{\ 2} \Delta p \, \underline{1} + \frac{1}{N_1^{\ 2}} \underline{H}_o'(N_1, \gamma) \underset{*}{<} \underline{z}_i \, \underline{z}_i' > \underline{H}_o(N_1, \gamma) \quad (4.50)$$

and upon the substitution of (4.9), the rate of change of normed distance, closure or convergence, immediately follows as

$$\psi_R(\underline{H}_o(N_1, \gamma)) = \frac{\partial}{\partial N_1} \psi_D(\underline{H}_o(N_1, \gamma)) = \frac{-2\Theta_{1_i}^{\ 2} \Delta p}{(N_1 + b[\gamma+1])} + \frac{(b_1 N_1 + b_2)}{(N_1 + b[\gamma+1])^2} - \frac{2\Delta p^2 \Theta_{1_i}^{\ 2} N_1 (N_1 + b)}{(N_1 + b[\gamma+1])^3} \quad (4.51)$$

where

$$b_1 = 2\Theta_{1_i}^{\ 2} \Delta p(1 + \Delta p) \qquad b_2 = b \, \Delta p^2 \, \Theta_{1_i}^{\ 2}$$

$$A = \begin{bmatrix} a & 1 & \cdots\cdots\cdots\cdots & 1 \\ 1 & a & \cdots\cdots\cdots\cdots & 1 \\ \vdots & & & \\ 1 & \cdots\cdots\cdots\cdots & 1 & a \end{bmatrix} \quad (4.52)$$

and a is given by (4.23). Since

$$\underline{1}' A \, \underline{1} = N_1(a + N_1 - 1)$$
$$\underline{1}' \, \underline{1} = N_1 \quad (4.53)$$

from (4.39) and (4.53), ψ_R can be normalized and becomes

$$\psi_{D'} = \frac{\psi_R(\underline{H}_o(N_1, \gamma))}{2\Theta_{1_i}^2 \Delta p} = \frac{-1}{(N_1 + b\,[\gamma+1])} + \frac{(1+\Delta p)N_1 + b\,\Delta p}{(N_1 + b[\gamma+1])^2} - \frac{\Delta p\, N_1(N_1 + b)}{(N_1 + b[\gamma+1])^3} \qquad (4.54)$$

eg., $\psi_{D'}$ is the normalized closure rate for the ith element of the discrete pulse waveform,

$$\psi_{D'}(\underline{H}_o(N_1, \gamma)) \stackrel{\Delta}{=} \frac{\partial}{\partial N_1} \{||\underline{\theta}_1 - \hat{\underline{\theta}}_1(N_1, \gamma)||^2\}_i \qquad (4.55)$$

The associated normalized convergence or closure acceleration follows from (4.54) as

$$\psi_{D''} = \frac{\frac{\partial}{\partial N_1}\psi_R}{2\Theta_{1_i}^2 \Delta p} = \frac{(2 + \Delta p)}{(N_1 + b[\gamma+1])^2} - \frac{(\Delta p\, \Theta_{1_i}^2)^{-1} N_1(1+b_1) + \frac{b}{2} + \frac{1}{2\Delta p^2\, \Theta_{1_i}^2}}{(N_1 + b[\gamma+1])^3} + \frac{3\Delta p\, N_1(N_1 + b)}{(N_1 + b[\gamma+1])^4} \qquad (4.56)$$

The remaining measure, ψ_E, is associated with dispersion and from (3.36)

$$\psi_E(\underline{H}_o(N_1, \gamma)) = \frac{\Delta p^2\, \Theta_{1_i}^2\, b}{N_1^2}\; \underline{H}_o'(N_1, \gamma)\, I\, \underline{H}_o(N_1, \gamma) \qquad (4.57)$$

From (4.39) and (4.53),

$$\underline{H}_o'(N_1, \gamma) I\, \underline{H}_o(N_1, \gamma) = \frac{N_1^3\ \ \Delta p^{-2}}{[N_1 + (a-1) + \gamma b]^2} \qquad (4.58)$$

and thus ψ_E becomes

$$\psi_E(\underline{H}_o(N_1,\gamma)) = \frac{\Theta_{1_i}^2\, N_1}{[(N_1 + a-1)/b + \gamma]^2} \qquad (4.59)$$

Consequently, from (4.59), it is seen that for any stage N_1,

$$\psi_E \simeq \frac{1}{\gamma^2} \tag{4.60}$$

and the dispersion of performance with ψ_E as a measure is reduced by large positive values of γ. At the same time, from (4.54), the convergence or closure rate is enhanced when the negative term is dominant and yields

$$\psi_{D'}(\underline{H}_o(N_1, \gamma)) < 0 \tag{4.61}$$

In particular, the closure rate is increased by the assignment of small positive and possibly negative values to γ since, including the operation

$$\frac{\partial}{\partial\gamma} \psi_{D'}(\underline{H}_o(N_1, \gamma)) = 0 \tag{4.62}$$

would yield some γ_o as

$$\alpha_1 \gamma_o^2(N_1) + \alpha_2 \gamma_o(N_1) + \alpha_3 = 0 \tag{4.63}$$

Recalling (4.23), and for small signal to noise ratios,

$$||\underline{\theta}||^2/\sigma^2 << 1 \tag{4.64}$$

produces a value of γ, under the above operation, for "maximized" convergence without invoking any constraint upon the dispersion, ψ_E, as

$$\gamma_o \simeq K_N^2 \frac{\Delta p}{b} + \frac{K_N(K_N - b\Delta p)^{\frac{1}{2}}}{b} - 1 \tag{4.65}$$

From (4.56), it is seen that the measure of closure acceleration, $\psi_{D''}$, is largest when the positive term is dominant. This condition is compatible with (4.54) in form, in that the <u>same range</u> of values for γ results in a negative rate of change of distance (closure). At this point, from (4.54), (4.56), (4.59), and the ensuing discussion, it is obvious that a trade-off between dispersion of dynamic performance and closure or convergence rate and acceleration is required. The required trade-off is not obvious analytically and results are arrived at experimentally by digital computer simulation of the processing with various constraint coefficients. These results appear in Chapter VI where this method is compared with a feed-forward mixture resolving method, the conventional decision-directed method, and the matched filter. In particular, it is seen from (4.23) and (4.64), that for the low signal to noise ratio range, the coefficient of $\psi_{D'}$, $\psi_{D''}$, and ψ_E is

$$\frac{1}{N_1 + b(\gamma+1)} \tag{4.66}$$

and hence the weighting is primarily dominated by the time (stage of observation) and not as much upon the parameters of the mixture, and hence, in this sense is "non-parametric".

4-5. The Optimized Weighting Operator for the General Two-Category Signal Pattern Model

For the more general two-category signal pattern model given by

$$\underline{\theta}_2 = G\,\underline{\theta}_1 \tag{4.67}$$

where

$$G = [\underline{g}_1, \ldots, \underline{g}_i, \ldots, \underline{g}_n] \tag{4.68}$$

The operator elements in (4.3) take on the form

$$\underset{*}{<} \underline{z}_i > = (p_{11}^* \, \theta_{1_i} + \underline{g}_i{}' \, \underline{\theta}_1 \, p_{12}^*) \, \underline{1} \tag{4.69}$$

Also the elements of

$$C(N_1, \gamma) = [R + \gamma(R - M)] \tag{4.70}$$

become, from (3.48) and (3.50),

$$r_{k\ell} = (p_{11}^* \, \theta_{1_i} + p_{12}^* \, \underline{g}_i' \, \underline{\theta}_1)^2 + (p_{11}^* + p_{12}^*)^2 < N_{i_k} \, N_{i_\ell} > \quad k \neq \ell \tag{4.71}$$

and

$$r_{k\ell} = r_{kk} = (p_{11}^* \, \theta_{1_i}^2 + p_{12}^* \, (\underline{g}_i' \, \underline{\theta}_1)^2) + (p_{11}^* + p_{12}^*) \, \sigma_i^2 \qquad k = \ell \tag{4.72}$$

Likewise, from (3.53) and (3.54)

$$m_{k\ell} = (p_{11}^* \Theta_{1_i} + p_{12}^* \underline{g}_i' \underline{\theta}_1)^2 \qquad k,\ell = 1, 2, \ldots, N_1 \tag{4.73}$$

Finally for stationary, white noise, from (4.18) and the above,

$$(r_{k\ell} - m_{k\ell}) = \{(p_{11}^* \Theta_{1_i}^2 + p_{12}^* (\underline{g}_i' \underline{\theta}_1)^2) + (p_{11}^* + p_{12}^*)\sigma^2 - (p_{11}^* \Theta_{1_i} + p_{12}^* \underline{g}_i' \underline{\theta}_1)^2\} \quad k = \ell$$

$$= 0 \qquad k \neq \ell \tag{4.74}$$

Thus the general model matrix given by (4.70) becomes

$$C(N_1, \gamma) = [[r_{k\ell}] + \gamma\,([r_{k\ell} - m_{k\ell}])] \tag{4.75}$$

and by factoring and combining terms, can be reduced to

$$C(N_1, \gamma) = r_{k\ell} \begin{bmatrix} c_{kk} & 1 & \cdots & 1 \\ 1 & c_{kk} & 1 \cdots & 1 \\ \vdots & & & \\ 1 & \cdots & 1 & c_{kk} \end{bmatrix} \tag{4.76}$$

where the diagonal terms are

$$c_{kk} = \frac{r_{kk}}{r_{k\ell}} + \gamma\left(\frac{r_{kk} - m_{kk}}{r_{k\ell}}\right) \tag{4.77}$$

The equivalent of (4.27) for the general case now becomes

$$C(N_1, \gamma)\, \underline{H}_{o_i}(N_1, \gamma, \underline{Z}^{(\omega_1^*)}) = N_1 \Theta_{1_i} (p_{11}^* \Theta_{1_i} + \underline{g}_i' \underline{\theta}_1 p_{12}^*)\, \underline{1} \tag{4.78}$$

and following the techniques from (4.28) through (4.41), the form of the general weighting operator for the ith element is

$$\underline{H}_{o_i}(N_1, \gamma) = \frac{N_1 r_{k\ell}^{-1/2}}{[N_1 - 1 + (r_{kk} + \gamma(r_{kk} - m_{kk}))/r_{k\ell}]} \tag{4.79}$$

Likewise, following the procedure from (4.40) through (4.46), the iterative format of the recursion takes on the form

$$\hat{\theta}_i(N_1) = \{w_\Theta(N_1)\ \hat{\theta}_i(N_1-1) + r_{k\ell}^{-\frac{1}{2}}\ w_z(N_1)\ z_i(N_1)\}\ \frac{1}{N_1} \tag{4.80}$$

and is similar in appearance to the form for the bi-polar signal pattern model given by (4.46). However in this case, the weighting elements

$$w_\Theta(N_1) = \frac{N_1(N_1-1 + (c_{kk}-1))}{(N_1-1)(N_1 + (c_{kk}-1))} \tag{4.81}$$

and

$$w_z(N_1) = \frac{N_1}{(N_1 + (c_{kk}-1))} \tag{4.82}$$

contain the coefficient c_{kk}, which from (4.77), (4.71), (4.72), and (4.74) is seen to depend upon terms such as $(p_{11}^{*}\ \Theta_{1_i} + p_{12}^{*}\ \underline{g}_i'\ \underline{\theta}_1)$. This presents a completely different problem, in that, the previous factoring and adjusting of the weights for low signal to noise ratio cannot be carried out. Specifically, since G is unknown, it is not readily established how the constraining coefficient, γ, might be adjusted for low signal to noise ratio conditions to render the weighting primarily observation-time dependent. Thus, for this more general case, the required operator characteristics, for this formulation, are developed to this point and no simulation studies are carried out.

CHAPTER V

DEVELOPMENT OF A FEEDFORWARD MIXTURE RESOLVING DETECTOR STRUCTURE

5-1. The Feedforward Signal Processing Structure

It is recalled that the work in previous chapters contains the development of a signal processing structure which computes a decision statistic, $\underline{h}'(N+1)\ \underline{z}(N+1)$, and gates the observations, $\{\underline{z}_k\}_{N+1}$, through to operators, $H_o(N_1, \gamma, \underline{Z}^{(\omega_1^*)})$, $H_o(N_2, \gamma, \underline{Z}^{(\omega_2^*)})$, for optimized, weighted combination and reinforcement of the structure $\underline{h}$. The fundamental character of the reinforcement, in that case, is of a feedback nature, and the data processing structure is illustrated in Fig. 4-1.

In Chapter II it was shown that, relative to an estimating category, $J \in \{J_r\}_R$, the average conditional error probability is minimized when the detector structure takes on the form given by (2.40), where the elements of the structure are linear substitutions of the estimating category J. The actual values of the dynamic and asymptotic, average conditional error probability depend upon the properties of the category, J, in question. The specific category, J_o, developed in the previous chapters was based upon the minimization of distance and dispersion functionals, ψ_D and ψ_E, respectively; with a constraint coefficient, γ, adjustment for maximization of convergence or closure rate and acceleration, $\psi_{D'}$, and $\psi_{D''}$.

It is the object of this section to consider a feed-forward detector structure, where the raw mixture observations, $\{\underline{z}_k\}_{N+1}$, are processed by a mixture resolving structure in order to generate the parameters, $\{\hat{\theta}\}_{N+1}^{(\omega_1)}$, $\{\hat{\theta}\}_{N+1}^{(\omega_2)}$, for reinforcement of the structure, $\underline{h}(N)$, and classification of the observation, $\underline{z}(N+1)$. The fundamental difference between the feedback and feedforward data-processing structure is the following. In the feedback structure, $\underline{h}(N)$, carries out a classification operation upon $\underline{z}(N+1)$ prior to extraction operations upon $\underline{z}(N+1)$ to reinforce $\underline{h}(N)$ to $\underline{h}(N+1)$. In the feedforward structure, $\underline{z}(N+1)$ is operated upon by the mixture resolving operator to extract information for the up-dating or reinforcement of $\underline{h}(N)$ to $\underline{h}(N+1)$ prior to a classification operation of $\underline{z}(N+1)$ by $\underline{h}(N+1)$. A basic block diagram of this operation appears in Fig. 5-1, where $\{\underline{x}_k\}_{N+1}$ is a sequence of binary decisions as in Fig. 4-1.

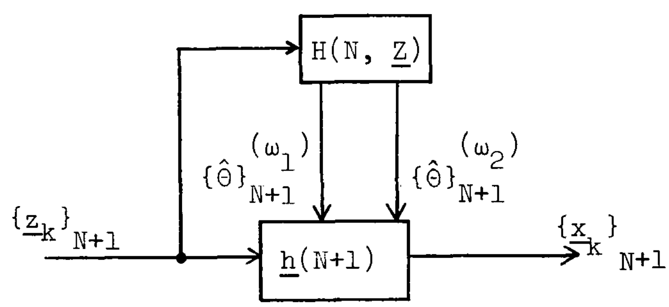

Fig. 5-1. Feedforward Structure

5-2. <u>The Moment Mixture Resolving Category, J_M</u>

Thus it is of interest to develop a mixture resolving category, $J \varepsilon \{J_r\}_R$, consistent with the feedforward concept which inherently requires operation on the raw, un-pre-processed observations, $\{\underline{z}_k\}_{N+1}$.

There is no specific optimization of the category in this case, and the category selected is the moment mixture resolving category, J_M. It was pointed out in Chapter I that the moment method was first applied to the mixture resolution problem by Pearson[17] and Rao,[18] and subsequently by Rider,[19] and Blischke.[20] All developed point estimators for the parameters of the mixture by the method of moments for the one dimensional case. Pearson and Rao were concerned with the gaussian mixture, Rider treated the exponential mixture, and Blischke worked with a bi-nomial mixture. Chang[23] developed moment estimators for the means of a four-category multi-dimensional gaussian mixture, however his moment method required apriori knowledge of the noise statistics. This chapter is concerned with the development of moment estimators for the two-category multi-dimensional signal pattern model which does not require apriori knowledge of the noise statistics, and the development proceeds as follows.

There is no formal, systematic method for developing a system of equations, such as in the previous chapters, to resolve a mixture of observations by the moment concept into some parametric decomposition. Rather, the approach is one of analytical trial and error, the object being to obtain a desired system of equations for the parameters in the mixture. The specific desired system of equations depends upon the signal pattern model assumed and the apriori information available. If the noise statistics were known, for instance, then the desired system of equations could be parametrically independent of the noise statistics, however should a specific system of equations be parametrically dependent upon the noise parameters, such dependence, in that case, would not prevent resolution. If the noise statistics, along with the discrete pulse waveform, are unknown, as assumed in this work, then the "desired" system of equations must be independent of the pulse and noise parameters. The moment method of mixture resolution consists of first equating raw or central mixture population moments to raw or central mixture sample moments and subsequently combining the raw or central moments, linearly, non-linearly, weighted, or unweighted, in order to arrive at the "desired" system of equations. In this development, a number of first moment, auto-correlation, cross-correlation, and higher moment equations were examined prior to arriving upon the appropriate combination of moments to satisfy the apriori information constraints; thus yielding the desired system of equations.

Proceeding, consider then the raw mixture of observations, $\{\underline{z}_k\}_N$, where,

$$\underline{z}_k = \underline{\theta}_k + \underline{N}_k \quad ; \quad \underline{z}_k = \begin{bmatrix} z_{1_k} \\ \vdots \\ z_{i_k} \\ \vdots \\ z_{j_k} \\ \vdots \\ z_{n_k} \end{bmatrix} \quad ; \quad k = 1, 2, \ldots, N \tag{5.1}$$

and the auto-correlation matrix of the kth observation,

$$R_{\underline{z}_k \underline{z}_k} = [r_{\underline{z}_k \underline{z}_k}(i,j)] = [r_{ij}] \tag{5.2}$$

where

$$r_{ij} = < z_{i_k} z_{j_k} > - < z_{i_k} >< z_{j_k} > \qquad i,j = 1, 2, \ldots, n \tag{5.3}$$

and $< >$ denotes the expectation over the mixture. From (5.1),

$$< z_{i_k} z_{j_k} > = \{< \theta_{i_k} \theta_{j_k} > + < N_{i_k} \theta_{j_k} > + < \theta_{i_k} N_{j_k} > + < N_{i_k} N_{j_k} >\} \tag{5.4}$$

and

$$< z_{i_k} > = < \theta_{i_k} > + < N_{i_k} > \tag{5.5}$$

In general, over the mixture,

$$< \theta_{i_j} \theta_{j_k} > = p_1(\theta_{1_{i_k}} \theta_{1_{j_k}}) + p_2(\theta_{2_{i_k}} \theta_{2_{j_k}}) \tag{5.6}$$

and

$$< \theta_{i_k} > = p_1\theta_{1_{i_k}} + p_2\theta_{2_{i_k}} \tag{5.7}$$

since $\theta_{1_{i_k}}$ and $\theta_{1_{j_k}}$ are both elements of the discrete pulse waveform within the same time slot and, where, for the two-category mixture,

$$p_1 = P\{\underline{z}_k \, \varepsilon \, \omega_1\} \quad ; \quad p_2 = P\{\underline{z}_k \, \varepsilon \, \omega_2\} \tag{5.8}$$

As in the previous chapter, for zero-mean, white noise,

$$< N_{i_k} N_{j_k} > = < N_{i_k} > = < N_{j_k} > = 0 \tag{5.9}$$

and using (5.7) and (5.8), the elements of $R_{\underline{z}_k \underline{z}_k}$ become

$$< z_{i_k} z_{j_k} > = p_1(\theta_{1_{i_k}} \theta_{1_{j_k}}) + p_2(\theta_{2_{i_k}} \theta_{2_{j_k}}) \quad i \neq j \tag{5.10}$$

and

$$< z_{i_k} z_{j_k} > = < z_{i_k}^2 > = p_1 \, \theta_{1_{i_k}}^{\;2} + p_2 \, \theta_{2_{i_k}}^{\;2} + \sigma_i^2 \quad i = j \tag{5.11}$$

Also

$$< z_{i_k} > = p_1 \, \theta_{1_{i_k}} + p_2 \, \theta_{2_{i_k}} \tag{5.12}$$

and in general,

$$\underline{\theta}_2 = G\underline{\theta}_1 \tag{5.13}$$

yielding

$$\theta_{2_i} = \underline{g}_i' \, \underline{\theta}_1 \qquad i = 1, 2, \ldots, j, \ldots, n \tag{5.14}$$

It is then necessary to combine equations (5.10) through (5.12), using what apriori information is available from (5.13) in such a manner that the solution for the $\underline{\theta}$'s is independent of the noise.

5-3. Multidimensional Moment Estimators Requiring no Apriori Noise or Pulse Waveform Information.

Consider the bi-polar case again, where apriori knowledge of the bi-polarity provides apriori information about G in the form

$$G = kI \quad ; \quad k = -1 \tag{5.15}$$

Then

$$\underline{g}_i = \begin{bmatrix} 0_1 \\ \vdots \\ 0_{i-1} \\ g_{i_i} \\ 0_{i+1} \\ \vdots \\ 0_n \end{bmatrix} \qquad g_{i_i} = -1 \tag{5.16}$$

and

$$\Theta_{2_i} = - \Theta_{1_i} \quad ; \quad i = 1, 2, \ldots, j, \ldots, n \tag{5.17}$$

yielding

$$< z_{i_k} \, z_{j_k} > = \Theta_{1_{i_k}} \Theta_{1_{j_k}} \qquad i \neq j \tag{5.18}$$

and

$$< z_{i_k} \, z_{j_k} > = < z_{i_k}^2 > = \Theta_{1_{i_k}}^2 + \sigma_i^2 \qquad i = j \tag{5.19}$$

$$< z_{i_k} > = (p_1 - p_2) \, \Theta_{1_{i_k}} \tag{5.20}$$

Since all subsequent discussion will refer to the kth observation, and since the noise is stationary, the subscripts, k, associated with the moments and, i, associated with the variance of the noise samples can be dropped. Recalling (5.3), the above development, and assuming that the two pulses have equal probability of occurrence,

$$r_{ij} = < z_i \, z_j > - < z_i > \; < z_j > = \Theta_{1_j} \Theta_{1_i} \qquad i \neq j \tag{5.21}$$

$$r_{ii} = < z_i^2 > - < z_i >^2 = \Theta_{1_i}^2 + \sigma^2 \qquad i = j \tag{5.22}$$

$$r_{jj} = \langle z_j^2 \rangle - \langle z_j \rangle^2 = \Theta_{1_j}^2 + \sigma^2 \qquad j = i \tag{5.23}$$

Now considering

$$r_{ii} - r_{jj} = \Theta_{1_i}^2 + \sigma^2 - \Theta_{1_j}^2 - \sigma^2 \tag{5.24}$$

it is seen that the noise variance is eliminated. In addition, from (5.21)

$$\Theta_{1_i}^2 = r_{ij}^2/\Theta_{1_j}^2 \tag{5.25}$$

In view of the bi-polar apriori information from (5.15), the subscript, 1, can be dropped, since whatever the value of a specific $\hat{\theta}_i(N)$, the specific value of the other category is the negative of $\hat{\theta}_i(N)$. Now with (5.24) and (5.25), a quadratic equation in θ_j^2 can be developed as

$$\Theta_j^4 - (r_{jj} - r_{ii})\,\Theta_j^2 - r_{ij}^2 = 0 \tag{5.26}$$

Also, by rewriting (5.25) as

$$\Theta_{1_j}^2 = r_{ij}^2/\,\Theta_{1_i}^2 \tag{5.27}$$

and taking the negative of (5.24), a similar equation for the ith element of the discrete pulse waveform is obtained,

$$\Theta_i^4 - (r_{ii} - r_{jj})\,\Theta_i^2 - r_{ij}^2 = 0 \tag{5.28}$$

It is appropriate to point out, at this time, that as long as there are only two categories, and auto-correlation or cross-correlation matrix elements are used to combine and develop a system of equations, the resulting equations will always be quadratic in the unknown elements of the discrete pulse waveform. From (5.26) and (5.28), the form of the estimates of the category J_M for the moment method of mixture resolution becomes,

$$\hat{\Theta}_i = \left\{\tfrac{1}{2}(r_{ii} - r_{jj}) + \left(\tfrac{1}{4}(r_{jj} - r_{ii})^2 + r_{ij}^2\right)^{\frac{1}{2}}\right\}^{\frac{1}{2}} \tag{5.29}$$

$$i = 1,\ldots,j,\ldots,n$$

and

$$\hat{\theta}_j = \{\frac{1}{2}(r_{jj} - r_{ii}) + (\frac{1}{4}(r_{ii} - r_{jj})^2 + r_{ij}^2)^{\frac{1}{2}}\}^{\frac{1}{2}} \qquad (5.30)$$

$$j = 1,\ldots,i,\ldots,n$$

Recalling the previous discussion, the moment technique consists of equating the population moments r_{ij} to the <u>mixture</u> sample moments of the corresponding population moments. For the population moments given by (5.21) through (5.23), for the bi-polar signal pattern model, the corresponding mixture sample moments are given by the norms and inner products respectively

$$r_{ii}(N) = N^{-1}\ ||\underline{z}_i||^2 \qquad i = 1, 2, \ldots, j, \ldots, n \qquad (5.31)$$

$$r_{jj}(N) = N^{-1}\ ||\underline{z}_j||^2 \qquad (5.32)$$

$$r_{ij}(N) = N^{-1}\ (\underline{z}_i, \underline{z}_j) \qquad (5.33)$$

where

$$\underline{z}_i = \begin{bmatrix} z_{i_1} \\ \vdots \\ z_{i_k} \\ \vdots \\ z_{i_n} \end{bmatrix} \qquad i = 1, 2, \ldots, j, \ldots, n \qquad (5.34)$$

is the vector of samples at the ith instant of each of the N time slots. Hence a substitution of the mixture sample moments given by (5.31) through (5.33) into (5.29) and (5.30) yields for the mixture resolving estimators of the ith and jth elements of the discrete pulse waveform at the Nth stage,

$$\hat{\theta}_i(N) = \left\{\frac{(||\underline{z}_i||^2 - ||\underline{z}_j||^2)}{2N} + \left(\frac{(||\underline{z}_j||^2 - ||\underline{z}_i||^2)^2}{4N^2} + \frac{(\underline{z}_i, \underline{z}_j)^2}{N^2}\right)^{\frac{1}{2}}\right\}^{\frac{1}{2}} \qquad (5.35)$$

$$\hat{\theta}_j(N) = \left\{\frac{(||\underline{z}_j||^2 - ||\underline{z}_i||^2)}{2N} + \left[\frac{(||\underline{z}_i||^2 - ||\underline{z}_j||^2)^2}{4N^2} + \frac{(\underline{z}_i, \underline{z}_j)^2}{N^2}\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}} \tag{5.36}$$

5-4. Convergence of the Mixture Sample Moments

The convergence or consistency of the estimating category, J_M, the moment method mixture resolving category, follows from (5.18) through (5.20). In particular, since the mixture of observations, $\{\underline{z}_k\}_N$, consists of a sequence of N, independent, n dimensional random vectors, then the ith element of each vector constitutes a sequence, $\{z_i\}_N$, of independent, identically distributed random variables. In addition, the elements of $||\underline{z}_i||^2$, $||\underline{z}_j||^2$, and $(\underline{z}_i, \underline{z}_j)$, given by $z_{i_1}^2, \ldots, z_{i_N}^2$; $z_{j_1}^2, \ldots, z_{j_N}^2$; $z_{i_1} z_{j_1}, \ldots, z_{i_N} z_{j_N}$, constitute three sequences of independent, identically distributed random variables within each sequence, since one ith square, one jth square, and one ijth product is taken from each independent vector. With the above, and from (5.18) and (5.19), since the population moments over the mixture, $\langle z_i^2 \rangle$, $\langle z_j^2 \rangle$, and $\langle z_i z_j \rangle$ exist, then by the Kolmogorov theorem for the strong law of large number, these are necessary and sufficient conditions for

$$P\{\lim_{N\to\infty} \frac{1}{N} ||\underline{z}_i||^2 = r_{ii} = \Theta_i^2 + \sigma^2\} = 1 \quad i = 1,2, \ldots, j, \ldots, n \tag{5.37}$$

and

$$P\{\lim_{N\to\infty} \frac{1}{N} (\underline{z}_i, \underline{z}_j) = r_{ij} = \Theta_i \Theta_j\} = 1 \tag{5.38}$$

and hence the mixture sample moments of the elements of the auto-correlation matrix, $R_{\underline{z}\,\underline{z}}$, are said to converge with probability one. Consequently, sequences of random variables, $\{\hat{\theta}_i(N)\}_N$, formed from conbinations of $||\underline{z}_i||^2$ and $(\underline{z}_i, \underline{z}_j)$, as in (5.35) and (5.36) obey the law of large numbers and thus the mixture resolving moment estimators for the elements of the pulse waveform, $\hat{\underline{\theta}}_i(N)$, converge with probability one.

The mixture resolving characteristics of the category J_M, given by (5.35) and (5.36) are established by digital computer simulation of noisy signal patterns, the moment mixture resolving operations, and signal pattern classification of the mixture of observations by the feedforward structure discussed in section 5-1. The results of the simulation are discussed in Chapter VI and compared with the optimally weighted feedback structure, the conventional decision-directed structure, and the Bayes matched filter.

CHAPTER VI

DIGITAL COMPUTER SIMULATION OF THE SIGNAL PROCESSING ALGORITHMS

6-1. Conditions of the Digital Computer Simulation

In order to test the performance of the two "mixture resolving" detector structures established in the previous chapters, digital computer simulations were developed for noisy binary signal pattern models, the data processing algorithms, and the actual signal processing. The data processing algorithms studied experimentally, via digital computer simulation, were;

(1) The "Convergence Rate Optimized" Feedback Detector Structure (Chapters III and IV)

(2) The Moment-Mixture-Resolving Feedforward Detector Structure (Chapter V)

(3) The Conventional Decision Directed Detector Structure.

A simulation of the classical Bayes matched filter detector structure was included, under identical signal observation conditions, in order to establish some absolute standard of comparison.

The noisy signal pattern model consisted of a sequence of discrete, independent, n-dimensional, binary, pulse vectors, $\{\underline{\theta}_k\}_N$, with parameters of proportionality, p_1 and p_2, subjected to additive, white, zero-mean, gaussian noise; giving rise to a two-category, n-dimensional mixture of observations, $\{\underline{z}_k\}_N$, as discussed in Chapter I. The probability density function of any observation, $\underline{z}_k$, is thus

$$p(\underline{z}_k) = p_1 \, N(\underline{\theta}_1, \Theta) + p_2 \, N(\underline{\theta}_2, \Theta) \qquad (6.1)$$

where Θ is the inverse of the noise covariance matrix. The binary signal pattern sequence, $\{\underline{\theta}_k\}_N$, was simulated by selecting fixed vectors, $\underline{\theta}_1$ or $\underline{\theta}_2$, at random, successively, with probabilities p_1 and p_2 dependent upon the outcome of a uniform random-number-generator subroutine. The additive noise pattern was simulated by an uncorrelated gaussian random number generator subroutine approximating N(0,1). The generation of the noisy signal pattern model is illustrated in Fig. 6-1(a)

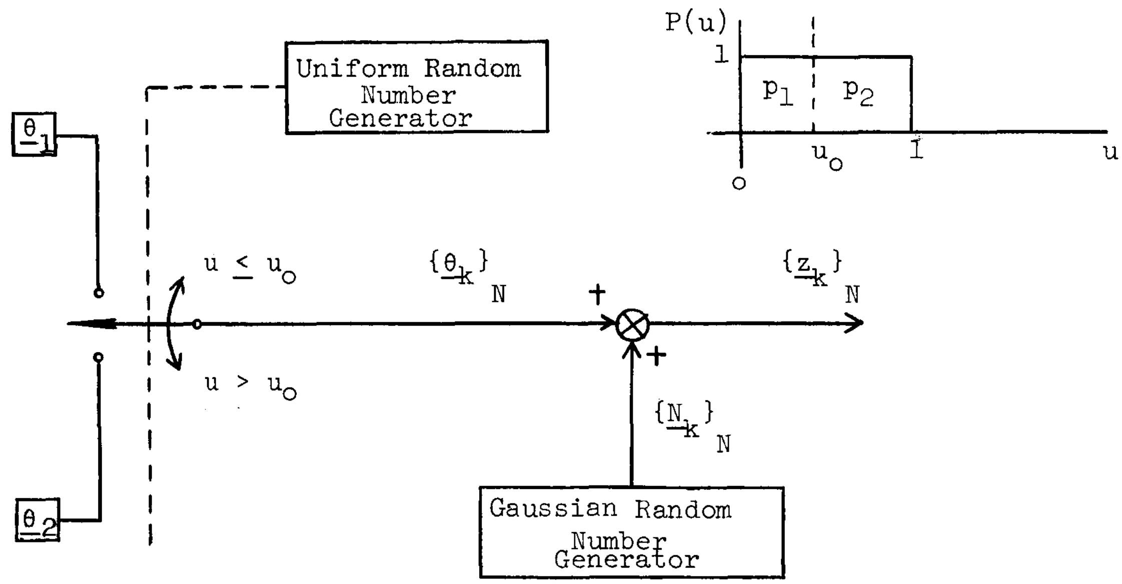

Fig. 6-1(a). Generation of Noisy Signal Pattern

The noise variance was maintained at one and the signal to noise ratio, given by

$$\text{SNR} \triangleq \ ||\underline{\theta}_1 - \underline{\theta}_2||^2/\sigma^2 \qquad (6.2)$$

was adjusted by adjusting the energy of the discrete pulses; e.g.,

$||\underline{\theta}_1 - \underline{\theta}_2||^2$. Various discrete waveforms were considered, including the on-off case. In view of the extent of the analysis and development for the bi-polar case, this was chosen as the case, to be treated for all of the detector structures considered. Thus, with

$$\underline{\theta}_2 = G\,\underline{\theta}_1 \;;\quad G = kI \;;\quad k = -1 \tag{6.3}$$

the particular shape of the discrete pulse was selected to follow

$$\theta_1(t) = \mu t \qquad 0 \leq t \leq M_P T_S \tag{6.4}$$

yielding the discrete pulse

$$\underline{\theta}_1 = \mu \underline{i} \qquad 0 \leq i \leq n \tag{6.5}$$

where i is the discrete time increment. The slope, norm, and thus SNR could be adjusted by the scalar μ. For each SNR case, $M = 50$ independent sequences (members), were generated and processed, where each sequence $\{\underline{z}_k\}_N$, consisted of $N = 300$, n-dimensional vector observations. In addition, the simulation was developed such that, for a given SNR case, the processing of a given member sequence, $\{\underline{z}_k\}$, of the 50, was carried out by all of the detector structures under identical input conditions in order to allow a study of relative performance characteristics on a given sequence. The range of SNR selected for experimentation, under the conditions imposed by (5.2), was

$$-9.48 \text{ db} \leq \text{SNR} \leq +\ .96 \text{ db} \tag{6.6}$$

For the simulation, the dimensionality of the pulses was selected to be $n = 2$. The pulse vectors corresponding to the lower and upper ends of the SNR range were thus

$$\begin{bmatrix} .15 \\ .30 \end{bmatrix} \leq \underline{\theta}_1 \leq \begin{bmatrix} .50 \\ 1.0 \end{bmatrix} \tag{6.7}$$

which were added to noise with a variance of one. A typical sequence, $\{\underline{z}_k\}_N$, for the "low" end of the SNR range is illustrated in Fig. 6-2.

6-2. Simulation of the Convergence Rate Optimized Feedback Detector Structure

A block diagram of the data processing by this algorithm is illustrated in Fig. 4-1. The decision equation, for the feedback structure, from (2.3) is

$$\ell(\underline{z}(N)) = \underline{h}'(N)\, \underline{z}(N+1) \begin{array}{c} \geq \ell_o \\ < \ell_o \end{array} \tag{6.8}$$

with the discrimination, or detection vector operator, from (2.40) given by

$$\underline{h}(N) = \hat{\theta}(N, \underline{Z}_N)\, \hat{\underline{\delta}}(N, \underline{Z}_N) \tag{6.9}$$

The elements of the vector operator structure are

$$\hat{\underline{\delta}}(N, \underline{Z}_N) = \hat{\underline{\theta}}_1(N_1, \underline{Z}^{(\omega_1^*)}) - \hat{\underline{\theta}}_2(N_2, \underline{Z}^{(\omega_2^*)}) \tag{6.10}$$

where

$$\hat{\underline{\theta}}_\ell(N, \underline{Z}^{(\omega_\ell^*)}) = \begin{bmatrix} \hat{\theta}_{1_\ell}(N_\ell) \\ \vdots \\ \hat{\theta}_{i_\ell}(N_\ell) \\ \vdots \\ \hat{\theta}_{n_\ell}(N_\ell) \end{bmatrix} \quad ; \quad \ell = 1, 2 \tag{6.11}$$

with the "optimally" weighted estimate of the ith element, from (4.40) and (4.41), given by

$$\hat{\theta}_{i_\ell}(N_\ell) = N_\ell^{-1}\, \underline{H}'_{o_i}(N_\ell, \gamma)\, \underline{z}_i = N_\ell^{-1} \left\{ \frac{N_\ell}{[N_\ell + c-1]} \right\} \underline{1}'\, \underline{z}_i \tag{6.12}$$

From (4.46), the recursion, normalized with respect to Δp, is iteratively expressed as

$$\hat{\theta}_{\ell_i}(N_\ell) = \{w_\Theta(N_\ell)\ \hat{\theta}_{\ell_i}(N_\ell-1) + w_z(N_\ell)\ z_i(N_\ell)\}\ \frac{1}{N_\ell} \tag{6.13}$$

Since the mixing parameters, p_{ij}^*, are unknown, actual processing requires the above normalization, which has the effect of producing an asymptotic, SNR dependent bias on $\hat{\theta}_{\ell_i}(N_\ell)$; a result that can be seen by taking $\lim_{N\to\infty}$ of the expectation of $\hat{\theta}_{\ell_i}(N_\ell)$, as given by (4.46), over the submixture ω_i^*. From (4.47) and (4.48), for the low SNR case, the coefficients are

$$w_\Theta(N_\ell) = \frac{N_\ell(N_\ell-1+\gamma)}{(N_\ell-1)(N_\ell+\gamma)} \tag{6.14}$$

and

$$w_z(N_\ell) = \frac{N_\ell}{(N_\ell+\gamma)} \tag{6.15}$$

From (6.13), it is seen that, initially what corresponds to $N_1 = 1$ has the requirements, that

$$\hat{\theta}_{\ell_i}(1) = w_\Theta(1)\ \hat{\theta}_{\ell_i}(0) + w_z(1)\ z_i(1) \tag{6.16}$$

in essence, that $z_i(1)$ be combined with $\hat{\theta}_{\ell_i}(0)$, some apriori reference. In order that this "apriori reference" reflect information about the population ω_ℓ, and still preserve the structural form given by (6.9), consider

$$\underline{h}'(0)\ \underline{z}(1) \begin{array}{l} \geq \ell_o \\ < \ell_o \end{array} \tag{6.17}$$

Now since

$$\underline{h}(0) = \hat{\theta}(0)\ \underline{\delta}(0) \tag{6.18}$$

in the absence of a difference, $\underline{\delta}(0)$, the simplest "pure" measure of information about the population value of $\underline{\delta}$ is

$$\hat{\underline{\delta}}(0) = \underline{z}(0) \quad ; \quad \hat{\theta}(0) = K\ I \tag{6.19}$$

such that (6.17) becomes

$$\underline{h}'(0)\ \underline{z}(1) = \underline{z}'(0)\ \underline{z}(1) \begin{array}{ll} \geq K^{-1}\ell_o & H_S \\ < K^{-1}\ell_o & H_D \end{array} \tag{6.20}$$

These assumptions are discussed in the section which follows (section 6-3).

Now from (2.9),

$$\ell_o = (E[\underline{h}'(0)\ \underline{z}(1) |\ \omega_S] + E[\underline{h}'(0)\ \underline{z}(1) |\ \omega_D])/2 \tag{6.21}$$

where, from (6.20) and the bi-polar case,

$$E[\underline{h}'(0)\ \underline{z}(1) | \omega_S] = (p_1{}^2 + p_2{}^2)\quad ||\underline{\theta}_1||^2 \tag{6.22}$$

and

$$E[\underline{h}'(0)\ \underline{z}(1) | \omega_D] = -2p_1p_2\quad ||\underline{\theta}_1||^2 \tag{6.23}$$

and where ω_S and ω_D are the spaces of two adjacent similar and dis-similar vectors respectively. For $p_1 = p_2 = 1/2$, $\ell_o = 0$, and regardless of the noise variance, the above threshold is zero and is minimax for this case. Thus the initial reference, $\hat{\theta}_{\ell_i}(0)$, is extracted from the observations by correlating the "first" observation, $\underline{z}(1)$, with the initial observation, $\underline{z}(0)$, and isolating both to ω_1^*, both to ω_2^*, or separated, as discussed in Section 3-1. The classification space for the initial separation is illustrated in Fig. 6-1(b) where $p_{12}(j)$ and $p_{21}(j)$ are the error probabilities in similarity and dis-similarity as given in (3.44).

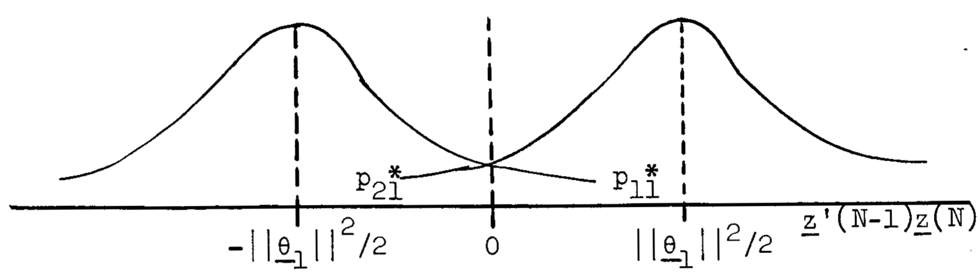

Fig. 6-1(b). Decision Space for Similarity Tests

Subsequent tests follow (6-8), and the elements of $\underline{\hat{\delta}}(N, \underline{Z}_N)$ are up-dated in the manner illustrated by (6.10) through (6.15). The matrix, $\hat{\Theta}(N, \underline{Z}_N)$, is given by

$$\hat{\Theta}(N, \underline{Z}_N) = \hat{\Sigma}^{-1}(N, \underline{Z}_N) = [A_{ij}(N)]' \, [|A(N)|]^{-1} \qquad (6.24)$$

where, in view of the stationary noise assumption (equal noise covariance), the $A_{ij}(N)$'s are cofactors of the elements, $\hat{S}_{ij}$, in $\hat{\Sigma}(N, \underline{Z}_N)$. The elements in $\hat{\Sigma}(N, \underline{Z}_N)$ are given by

$$\hat{S}_{ij} = \tfrac{1}{2}[\hat{S}_{ij}^{(\omega_1^*)} + \hat{S}_{ij}^{(\omega_2^*)}] \qquad i,j = 1, 2, \ldots, n \qquad (6.25)$$

where

$$\hat{S}_{ij}^{(\omega_\ell^*)} = \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} (z_{i_k}^{(\omega_\ell^*)} - \hat{\Theta}_i(N_\ell))(z_{j_k}^{(\omega_\ell^*)} - \hat{\Theta}_j(N_\ell)) \quad \ell = 1, 2 \qquad (6.26)$$

The pooling of the individual elements, $\hat{S}_{ij}^{(\omega_1^*)}$ with corresponding individual elements $\hat{S}_{ij}^{(\omega_2^*)}$, in order to obtain the elements $\hat{S}_{ij}$ of $\hat{\Sigma}(N, \underline{Z}_N)$ is carried out in order to allow for a comparison with the conventional decision directed system under identical conditions for the $\hat{\Sigma}(N, \underline{Z}_N)$. In the absence of a required form in (6.25), a computation of the pooled covariance elements would be more correctly carried out by using the within-between method[26].

The ith element of $\underline{h}(N)$ then becomes

$$h_i(N) = \underline{\hat{\theta}}_i'(N, \underline{Z}_N) \, \underline{\hat{\delta}}(N, \underline{Z}_N) \qquad (6.27)$$

where $\underline{\hat{\theta}}_i(N, \underline{Z}_N)$ is the ith row vector of the matrix in (6.24). For the case n = 2, this yields

$$\begin{aligned} h_1(N) &= (A_{22}(N)\,\hat{\delta}_1(N) - A_{12}(N)\,\hat{\delta}_2(N))\,|A(N)|^{-1} \\ h_2(N) &= (-A_{21}(N)\,\hat{\delta}_1(N) + A_{11}(N)\,\hat{\delta}_2(N))\,|A(N)|^{-1} \end{aligned} \qquad (6.28)$$

It was illustrated in section 4-4, that for the signal to noise ratio range defined by (6.6), the weighting coefficients given by (6.14) and (6.15) are dominated by the observation time, N_ℓ, and the constraint coefficient, γ, e.g.,

$$w_\Theta(N_\ell) \approx \frac{N_\ell(N_\ell-1+\gamma)}{(N_\ell-1)(N_\ell+\gamma)} \tag{6.29}$$

and

$$w_z(N_\ell) \approx \frac{N_\ell}{(N_\ell+\gamma)} \tag{6.30}$$

Numerous data processing simulations were carried out with γ ranging over a 10:1 interval

$$1 \leq \gamma \leq 10 \tag{6.31}$$

The detection performance characteristics, represented by $\overline{P_e}(N)$, averaged over M = 50 runs of N = 300 time slots each, are illustrated in Figs. 6-3 through 6-6 for 4 cases of SNR and the two extreme values of the constraint coefficient γ, given by γ_1 and γ_2. The analytical and computer simulated values of $\overline{P_e}(N)$ for the matched filter appear as a solid and a dotted line respectively. The effect of γ upon the convergence rate, as experimentally observed, supports the analytical conclusions arrived at in section 4-4. A discussion of relative performance characteristics is contained in the final section of this chapter for all detector structures considered.

6-3. Discussion of Initial Assumptions on $\underline{h}$ and Alternate Procedures of Classification

In view of the assumptions given by (6.19), since $\underline{\theta}_2 = k\underline{\theta}_1$; k = -1; and $\Sigma = \Theta^{-1} = \sigma_N{}^2 I$; $\underline{h}(0)$ becomes

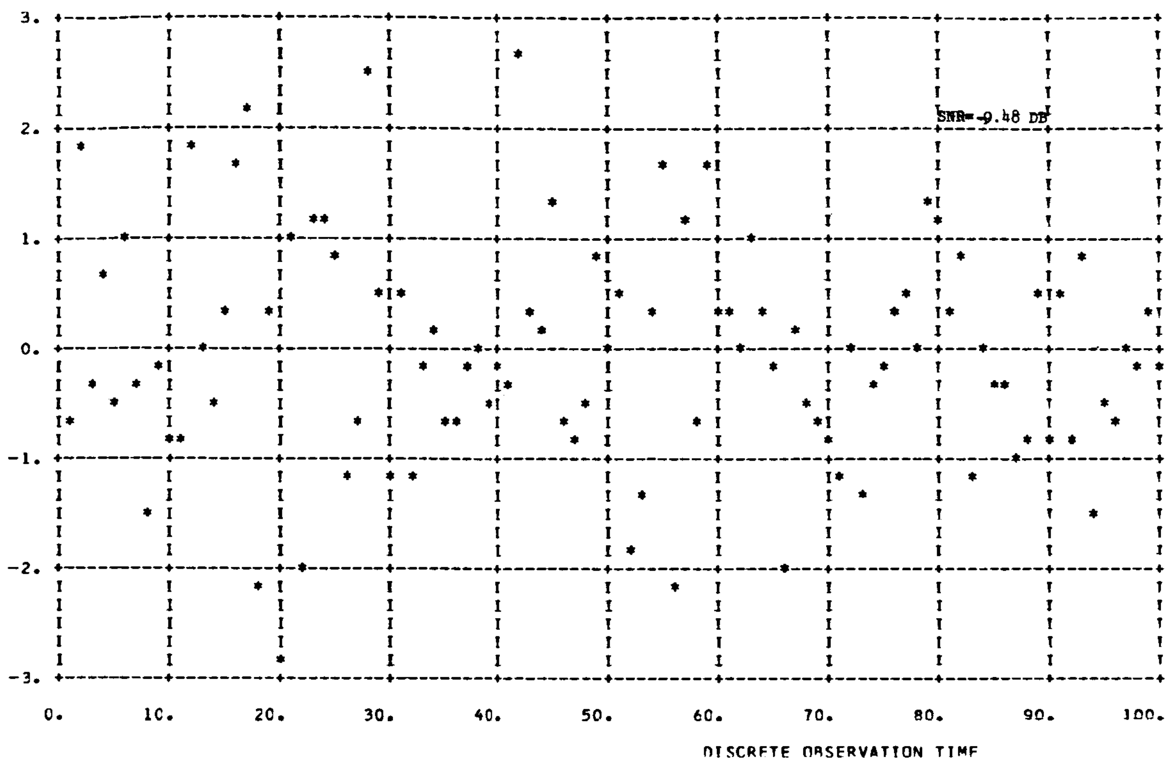

Fig. 6-2. Typical Observance Sequence

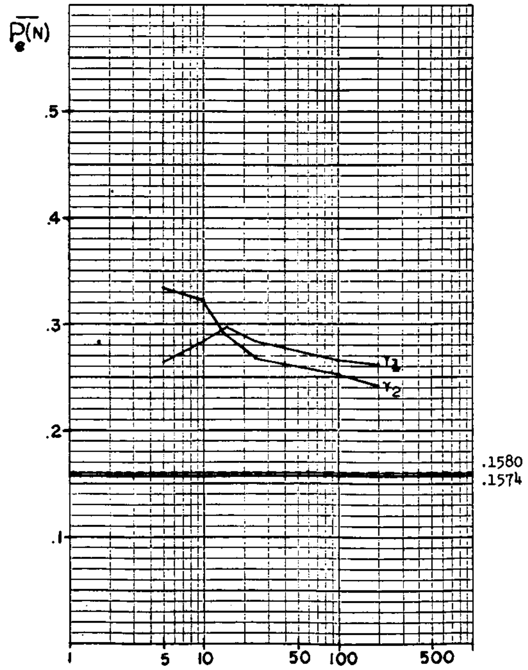

Fig. 6-3. Weighted Decision Directed $\overline{P_e(N)}$
(SNR = +.96 DB)

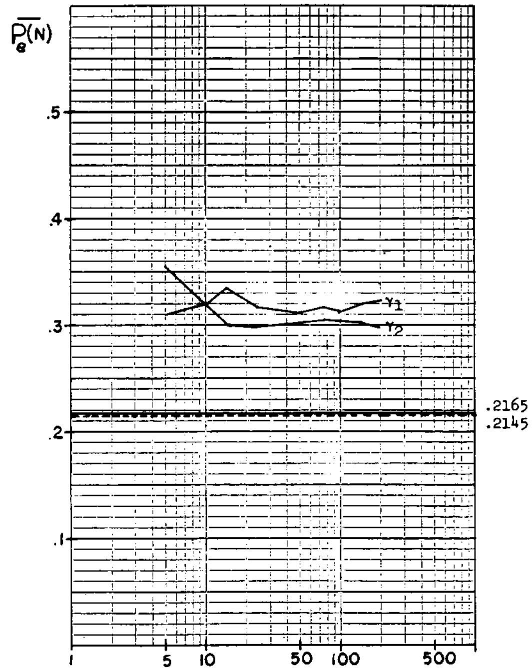

Fig. 6-4. Weighted Decision Directed $\overline{P_e(N)}$
(SNR = 1.06 DB)

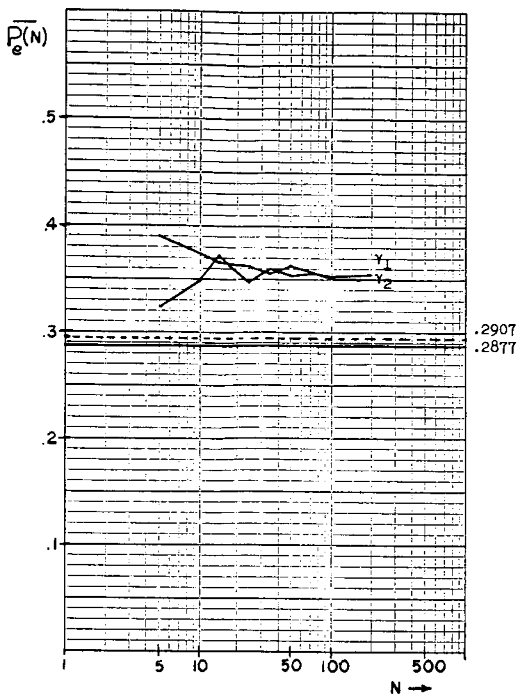

Fig. 6-5. Weighted Decision Directed $\overline{P_e(N)}$
(SNR = -2.53 DB)

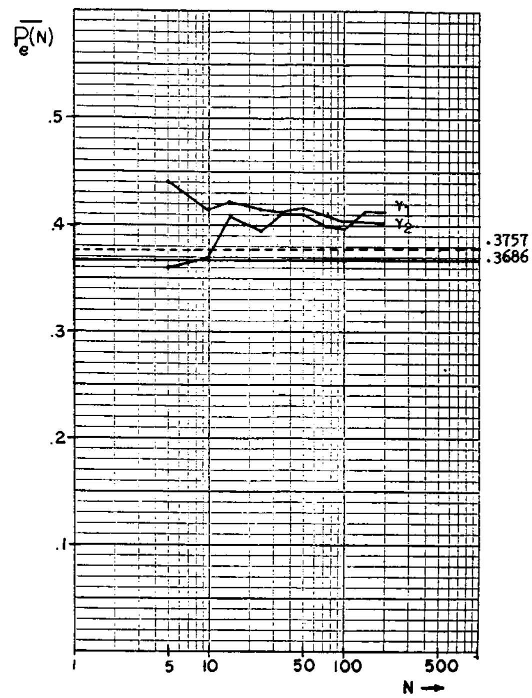

Fig. 6-6. Weighted Decision Directed $\overline{P_e(N)}$
(SNR = -9.48 DB)

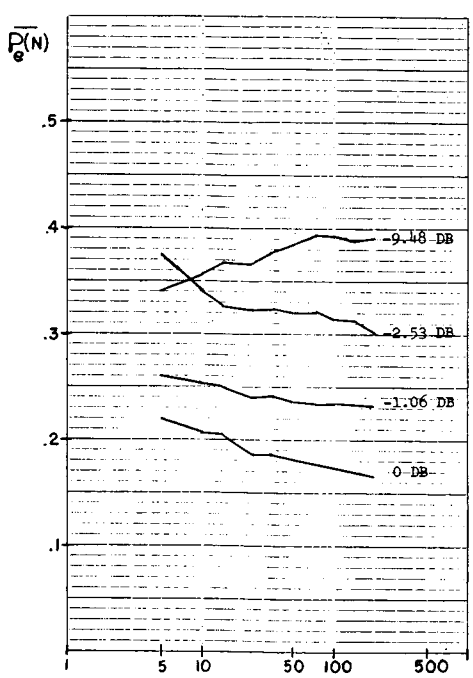

Fig. 6-7. Moment Resolution $\overline{P_e(N)}$

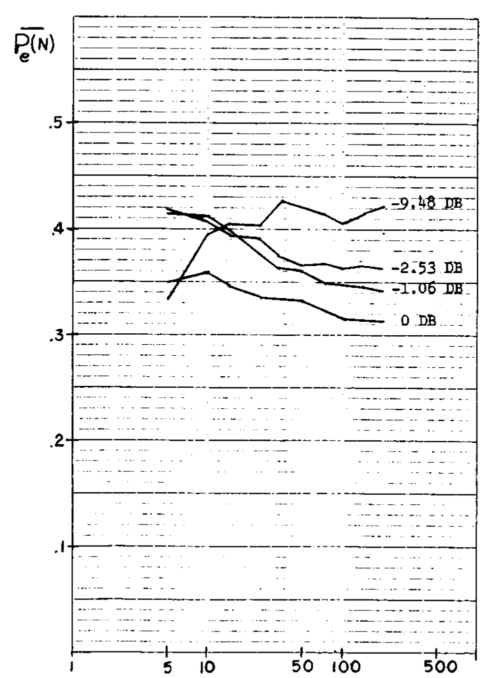

Fig. 6-8. Conventional Decision Directed $\overline{P_e(N)}$

$$\underline{h}(0) = K\, I\underline{z}(0) \tag{6.32}$$

The assumption, $\hat{\theta}(0) = K\, I$, implies that knowledge is available that the noise is white, however, it is not known what the noise power (variance) happens to be. Since it is further assumed that $p_1 = p_2 = 1/2$, the threshold, ℓ_o, given by (6.21) is zero, and the decision equation, (6.20) becomes

$$\underline{h}'(0)\, \underline{z}(1) = K\, \underline{z}'(0)\, \underline{z}(1) \geq \ell_o = 0 \quad H_S \tag{6.33}$$
$$< \ell_o = 0 \quad H_D$$

which is equivalent to

$$\underline{z}'(0)\, \underline{z}(1) \geq 0 \quad H_S \tag{6.34}$$
$$< 0 \quad H_D$$

In view of the equi-probable situation, knowledge of the actual value of the noise power (variance), given by K, is not required since ℓ_o is zero. The nature of the conditions established by the equi-probable, bi-polar situation, for the decision equation, (6.20), utilized for initial separation in this work, actually constitutes the minimax solution for testing the null hypothesis, $H_S(\underline{z}(0)$ and $\underline{z}(1)$ both from ω_1 or both from $\omega_2)$, against the alternative hypothesis, $H_D(\underline{z}(0)$ and $\underline{z}(1)$ from ω_1 and ω_2 or ω_2 and $\omega_1)$. It is emphasized that the inner-product statistic achieves this minimax initial separation for the equi-probable, bi-polar case only.

There is actually a deeper significance* to the assumption $\hat{\underline{\delta}}(0)=\underline{z}(0)$. For the bi-polar case, $\underline{\delta} = 2\underline{\theta}$, and thus $\underline{z}(0)$ is taken for $2\hat{\theta}(0)$. Thus, this interpretation yields, for (6.34)

* Personal discussion with Prof. K. C. S. Pillai, Dept. of Mathematical Statistics, Purdue University, West Lafayette, Indiana, July, 1966.

$$\underline{z}'(0)\ \underline{z}(1) \equiv \hat{\underline{\theta}}'(0)\ \underline{z}(1) \geq 0 \quad H_S \qquad (6.35)$$
$$< 0 \quad H_D$$

in essence, a referencing of the first observation with the above initial estimate for one of the means. Under the above interpretation, it must be recalled that the classical, minimum risk, decision equation, for two gaussian categories is given by[25]

$$[\Theta\underline{\delta}]'[\underline{z} - \tfrac{1}{2}\underline{s}] \geq \log \eta \qquad H_1 : \underline{z} \ \varepsilon\ \omega_1 \qquad (6.36)$$
$$< \log \eta \qquad H_2 : \underline{z} \ \varepsilon\ \omega_2$$

where $\underline{\delta} = \underline{\theta}_2 - \underline{\theta}_1$, $s = \underline{\theta}_2 + \underline{\theta}_1$, and $\Theta = \Sigma^{-1}$. For the equal cost, equi-probable, bi-polar situation in white noise, (6.36) becomes

$$\underline{\theta}'\ \underline{z} \geq 0 \qquad H_1 \qquad (6.37)$$
$$< 0 \qquad H_2$$

Thus it might be concluded, under the above interpretation, that the assumption, $\hat{\underline{\delta}}(0) = \underline{z}(0)$, in essence assumes $\hat{\underline{\theta}}(0) = \tfrac{1}{2}\underline{z}(0)$.

It was suggested* later in this work, that initial separation might be alternatively achieved by a use of the statistic

$$\left|\left|\frac{1}{\sqrt{2\sigma_N^2}}\ [\underline{z}(j) + \underline{z}(j+1)]\right|\right|^2 \geq \chi_n^2(\alpha) \quad H_S \quad j = 0,1,2,\ldots \qquad (6.38)$$
$$< \chi_n^2(\alpha) \quad H_D$$

for some desired significance level α, where $\chi_n^2(\alpha)$ is the associated central chi-square variable for the null hypothesis, H_D, in this case. The degree of freedom, n, corresponds to the vector dimensionality. Each element in the norm of (6.38) given by

$$\frac{z_i(j) + z_i(j+1)}{\sqrt{2\sigma_N^2}} \qquad (6.39)$$

* Pillai, op. cit.

is N(0, 1) under the null hypothesis for the equi-probable, bi-polar case. It is seen that a use of the statistic given by (6.39) would require no assumption in regard to the relationship of $\underline{z}(0)$ to $\underline{\theta}$. Use of this statistic would require a knowledge of the noise power (variance).

A second alternative for initial separation might be an application of Hotelling's[25] T^2 statistic, applied under the following modification.

$$T^2(N) = N_1 \left[\underline{\bar{z}}(N_1) - \underline{\theta}_o^{(\omega_1^*)}\right]' \hat{\Sigma}^{-1}(M) \left[\underline{\bar{z}}(N_1) - \underline{\theta}_o^{(\omega_1^*)}\right] \begin{array}{ll} \geq T_o^2 & H_D \\ < T_o^2 & H_S \end{array} \qquad (6.40)$$

where

$$T_o^2 = \frac{Mn}{(M - n+1)} F_{n,M-n}(\alpha) \qquad (6.41)$$

and M is the number of degrees of freedom associated with $\hat{\Sigma}(M)$; n is the vector dimensionality. Also N_1 is the number of vector observations that have been pooled in $\underline{\bar{z}}(N_1)$ and $\underline{\theta}_o^{(\omega_1^*)}$, as indicated below, by the assignment as similar under H_S, up to stage N. These vectors in $T^2(N)$ are given as

$$\underline{\bar{z}}(N_1) = \left(N_1 \underline{\theta}_o^{(\omega_1^*)} + \underline{z}(N)\right) \frac{1}{N_1+1} \; ; \; \underline{\theta}_o^{(\omega_1^*)} = \frac{1}{N_1} \sum_{\ell=1}^{N_1} \underline{z}^{(\omega_1^*)}(\ell) \qquad (6.42)$$

It is seen that $\underline{\theta}_o^{(\omega_1^*)}$ is the sample mean of the vectors assigned to ω_1^* (grouped together as similar) by N_1 outcomes of $T^2(N) < T_o^2$ prior to stage N. During the test, at each stage, $\underline{\theta}_o^{(\omega_1^*)}$ is to be treated as a constant.

Further insight into the relationship of $T^2(N)$ to (6.20), can be developed if (6.40) is modified to consist only of the vector, $\underline{z}(N)$, under observation, yielding

$$T^2(N) = N_1[\underline{z}(N) - \underline{\theta}_o^{(\omega_1^*)}]'\hat{\Sigma}^{-1}(M)\ [\underline{z}(N) - \underline{\theta}_o^{(\omega_1^*)}] \begin{matrix} \geq T_o^2 \\ < T_o^2 \end{matrix} \qquad (6.43)$$

The number of degrees of freedom associated with $\hat{\Sigma}(M)$ in T_o^2 remains as M. For Σ known or M large, such that

$$\hat{\Sigma}(M) \approx \sigma_N^2\ I \qquad (6.44)$$

for the white noise case, (6.44) becomes a chi-square statistic,

$$Q^2(N) = \frac{1}{\sigma_N^2}\ ||\underline{z}(N) - \underline{\theta}_o^{(\omega_1^*)}||^2 \begin{matrix} \geq \chi_n^2(\alpha) & H_D \\ < \chi_n^2(\alpha) & H_S \end{matrix} \qquad (6.45)$$

In either case, use of (6.45) would require either a knowledge of the noise statistics or M classified vectors to obtain $\hat{\Sigma}(M)$, and thus $\hat{\sigma}_N^2(M)$. For the initial stage, N = 1 and $\underline{\theta}_o^{(\omega_1^*)} = \underline{z}(0)$. Thus (6.45) becomes

$$Q^2(N) = Q^2(1) = ||\underline{z}(1) - \underline{z}(0)||^2 \begin{matrix} \geq \sigma_N^2\ \chi_n^2(\alpha) & H_D \\ < \sigma_N^2\ \chi_n^2(\alpha) & H_S \end{matrix} \qquad (6.46)$$

which, upon expansion, yields the equivalent test,

$$\underline{z}'(0)\ \underline{z}(1) \begin{matrix} \leq \frac{1}{2}\ [\{||\underline{z}(0)||^2+||\underline{z}(1)||^2\} - \sigma_N^2\ \chi_n^2(\alpha)] & H_D \\ > \frac{1}{2}\ [\{||\underline{z}(0)||^2+||\underline{z}(1)||^2\} - \sigma_N^2\ \chi_n^2(\alpha)] & H_S \end{matrix} \qquad (6.47)$$

This can be compared with (6.20), the statistic used for initial separation in this thesis, given by

$$\underline{z}'(0)\ \underline{z}(1) \begin{matrix} \geq 0 & H_S \\ < 0 & H_D \end{matrix} \qquad (6.48)$$

As it turns out, for the equi-probable, bi-polar case, the zero threshold for the statistic $\underline{z}'(0)\ \underline{z}(1)$ coincides with the minimax solution for initial separation on the basis of similarity or dis-similarity of vector means, and is related to the interpretation of $\underline{z}(0)$ as $\hat{\theta}(0)$. It must be pointed out that (6.48) applied successively simply carries out one-shot minimax separation. At the same time, successive application of (6.47) in the form (6.45) would result in an increase of information about $\underline{\theta}$ which would be reflected in $\underline{\theta}_o^{(\omega_1^*)}$ given by (6.42). No actual computer simulation experimentation was carried out for the alternative cases since a knowledge of the noise statistics would have been required. The tests for initial separation, as carried out with the use of (6.48),(6.20) required no knowledge of the noise statistics.

6-4. Simulation of the Moment Mixture Resolving Feedforward Detector Structure

The decision equation for this structure is given by

$$\ell(\underline{z}(N)) = \underline{h}'(N)\ \underline{z}(N) \begin{array}{l} \geq \ell_o \\ < \ell_o \end{array} \tag{6.49}$$

as discussed in section 5-1 and $\underline{h}(N)$ has the structural form given by (6.9). The elements of the structure differ, however, and recalling (5.34), the elements in (6.11) given by (6.12), for this case, are

$$\hat{\theta}_i(N) = \left\{ \frac{(||\underline{z}_i||^2 - ||\underline{z}_j||^2)}{2N} + \left[\frac{(||\underline{z}_j||^2 - ||\underline{z}_i||^2)^2}{4N^2} + \frac{(\underline{z}_i,\ \underline{z}_j)^2}{N^2} \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}} \tag{6.50}$$

The covariance, $\hat{\Theta}(N, \underline{Z}_N)$, has the same form as (6.24), however, under the assumption of apriori information necessary to develop the moment category of Chapter V, the elements of $\hat{\Theta}(N, \underline{Z}_N)$ are

$$\begin{aligned} A_{ij}(N) &= 0 \qquad i \neq j \\ &= S_{ii}(N) \quad i = j \quad ; \quad i = 1, 2, \ldots, n \end{aligned} \tag{6.51}$$

with

$$S_{ii}(N) = N^{-1} \sum_{k=1}^{N} (z_{i_k} - \hat{\theta}_i(N))^2 \quad i = 1, 2, \ldots, n \tag{6.52}$$

The remaining elements follow (6.27) and (6.28), except for the above differences in $\hat{\theta}_i(N)$, $A_{ij}(N)$, and where, in view of the bi-polar model,

$$\hat{\delta}_i(N) = (\hat{\theta}_i(N) - (-\hat{\theta}_i(N)) = 2\hat{\theta}_i(N) \qquad (6.53)$$

The performance characteristics of this structure, as indicated by $\overline{P_e}(N)$ for values of SNR identical to those used for the previous detector structure, are illustrated in Fig. 6-7.

6-5. Simulation of the Conventional-Decision-Directed and Matched-Filter Detector Structures

A simulation of these two detector structures was carried out under input conditions exactly identical to those of the systems developed in this thesis, in order to establish a standard for comparison of performance on a relative and an absolute basis respectively. The decision equation for the conventional decision directed structure is given by

$$\ell(\underline{z}(N)) = \underline{h}'(N)\ \underline{z}(N+1) \begin{array}{l} \geq \ell_o \\ < \ell_o \end{array} \qquad (6.54)$$

where the structure of $\underline{h}(N)$ is given by (6.9), however, where the elements differ in a very fundamental manner. In the conventional decision-directed structure, there are no specific, time-varying or otherwise, weights and the estimate of the ith pulse sample corresponding to (6.11) is given by

$$\hat{\theta}_{\ell_i}(N_\ell) = N_\ell^{-1}\ \underline{1}'\ \underline{z}_i \quad ; \underline{1} = \begin{bmatrix} 1_1 \\ \vdots \\ 1_j \\ \vdots \\ 1_{N_\ell} \end{bmatrix} \quad ; \ \ell = 1, 2\ ; \quad i = 1, 2, \ldots, n \qquad (6.55)$$

with the iterative form

$$\hat{\theta}_{\ell_i}(N_\ell) = \{(N_\ell - 1)\, \hat{\theta}_{\ell_i}(N_\ell - 1) + (1)\, z_i(N_\ell)\} \frac{1}{N_\ell} \qquad (6.56)$$

which is fundamentally different from (6.13). The remaining elements of the structure follow (6.24) through (6.28) with $\hat{\theta}_{\ell_i}(N_\ell)$ inserted as given by (6.56). The decision equation of the Bayes matched filter, for comparison purposes is

$$\ell(\underline{z}(N)) = \underline{h}_o'\, \underline{z}(N) \begin{array}{l} \geq \ell_o \\ < \ell_o \end{array} \qquad (6.57)$$

where, for the bi-polar case

$$\underline{h}_o = 2\Theta\, \underline{\theta}_1 \quad ; \quad \Theta = \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{bmatrix} \qquad (6.58)$$

and where the elements of $\underline{h}_o$, for the conditions of the simulation are

$$h_{o_1} = \frac{2}{\sigma^2}\theta_{11} \quad ; \quad h_{o_2} = \frac{2}{\sigma^2}\theta_{12} \qquad (6.59)$$

The characteristics of $\overline{P_e}(N)$ for input conditions identical to the previous cases, are illustrated in Fig. 6-8. The relatively slower convergence rate of the conventional decision-directed structure is evidenced by comparing the results in Fig. 6-8 with those of Figs. 6-3 through 6-6. In addition, for a typical noisy signal pattern with a SNR of +.96 db, the convergence characteristics of the mixture resolving estimates of the elements of $\underline{\theta}_1$, for the weighted and conventional decision directed algorithms, are illustrated in Figs. 6-9 and 6-10, for that particular noisy signal pattern, respectively. It is seen that the weighted algorithm achieves quite accurate stable values within 100 vector observations, whereas the elements obtained by the conventional decision directed algorithm are totally in error. This is but one illustration in numerous cases, where the weighted algorithm converged accurately for a given

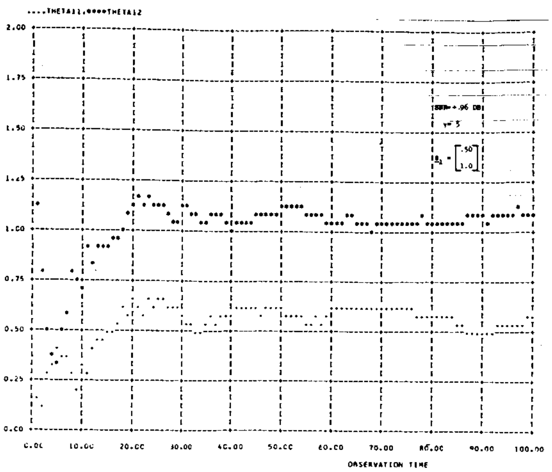

Fig. 6-9. Typical Convergence of $\hat{\underline{\theta}}_1(N_1)$ for a Weighted Decision Directed Estimate

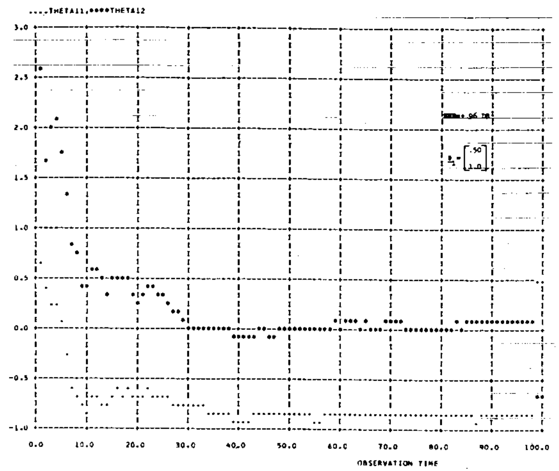

Fig. 6-10. Typical Convergence of $\hat{\underline{\theta}}_1(N_1)$ for a Conventional Decision Directed Estimate

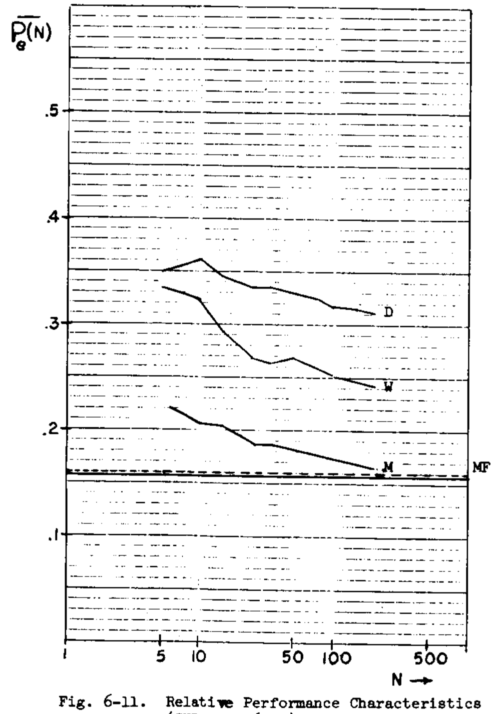

Fig. 6-11. Relative Performance Characteristics (SNR = +.96 DB)

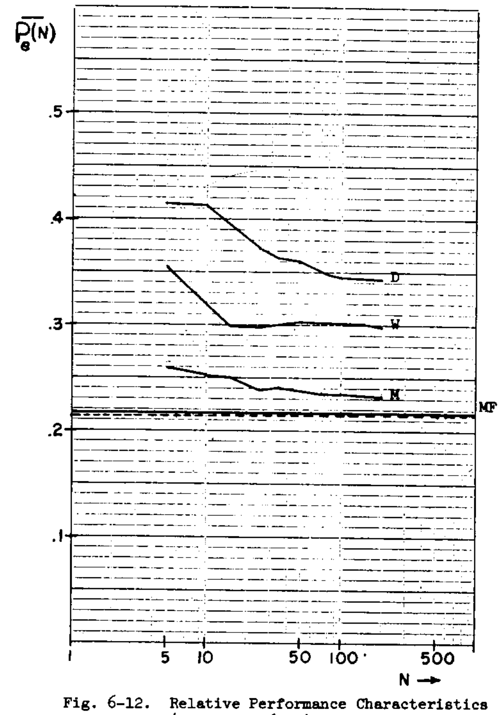

Fig. 6-12. Relative Performance Characteristics (SNR = -1.06 DB)

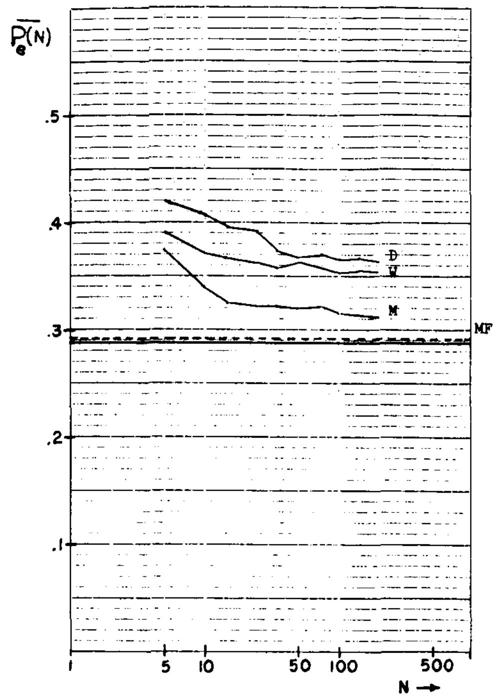

Fig. 6-13. Relative Performance Characteristics (SNR = -2.53 DB)

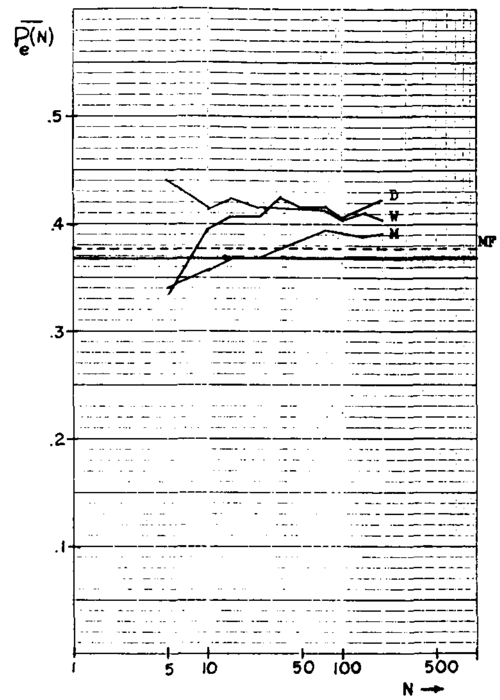

Fig. 6-14. Relative Performance Characteristics (SNR = -9.48 DB)

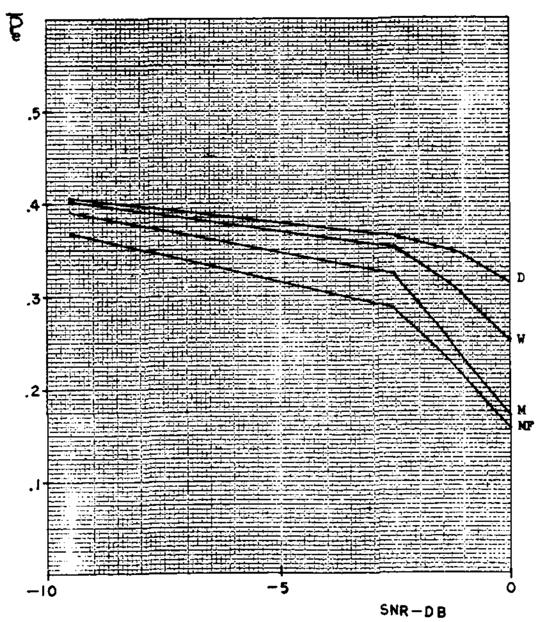

Fig. 6-15. Relative Performance After 100 Observations

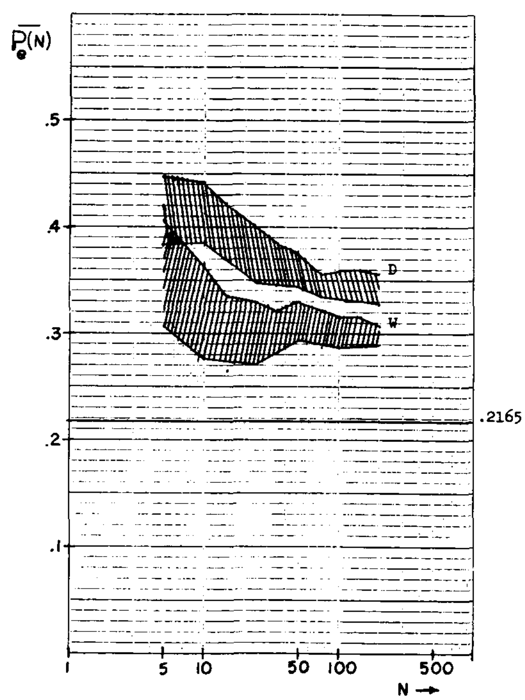

Fig. 6-16. Relative Performance Characteristics with Experimental Variance Bounds (SNR = -1.06 DB)

noisy signal pattern, whereas the conventional decision directed algorithm converged in error for that particular pattern.

The performance of the Bayes matched filter detector, as obtained by the simulation, appears as a dotted line in Figs. 6-3 through Fig. 6-14. This value differs slightly from the average probability of mis-classification computed, for the bi-polar and minimax conditions, analytically as

$$P_{e_o} = (2\pi)^{-\frac{1}{2}} \int_{\frac{1}{2}(\underline{\delta}\ \Theta\ \underline{\delta})^{\frac{1}{2}}}^{\infty} \exp\ [-\tfrac{1}{2}x^2]\ dx = (2\pi)^{-\frac{1}{2}} \int_{||\underline{\theta}_1|| / \sigma}^{\infty} \exp\ [-\tfrac{1}{2}x^2]\ dx \qquad (6.60)$$

which is to be expected in view of the finite sample size, M = 50, N = 300.

A comparison of the convergence characteristics for the moment, weighted, and decision-directed structures is presented in Figs. 6-11 through 6-14 for values of SNR ranging from .96db. to -9.48 db., and indicated by the letters M, W, and D respectively. In addition, the average relative performance after N = 100 vector observations, as a function of SNR, is illustrated in Fig. 6-15. Finally, a relative comparison of the convergence characteristics and dynamic performance properties of the weighted and conventional decision-directed structures, with simulation computed variance bounds, is illustrated in Fig. 6-16.

CHAPTER VII

CONCLUSIONS

7-1. Summary and Conclusions

This research has been concerned with;

(1) A formal development of the structural form of an optimum linear detection operator when the classification parameters are unknown,

(2) A formal development of mixture resolving estimating categories to learn or extract parametric information from the noisy signal pattern (mixture) in order to obtain the elements for the structure.

The criterion of optimality for the detector structure has been the average conditional probability of detection-error (misclassification). A formal development of this criteria has shown that the minimum average conditional probability of error at the Nth stage, conditioned upon the mixture resolving estimating category, corresponds to the maximum of a basically related, normalized, quadratic form. The maximization of this form leads to a homogeneous eigenvalue equation where the eigenvalues *are* the quadratic forms. Thus the largest eigenvalue of this equation is the maximum of the quadratic form. By the nature of the binary gaussian signal pattern model, the equation has rank one and the maximum eigenvalue is established. The associated eigenvector, $\underline{h}_o(N,J)$, is the unique optimum linear detection operator. It is a function of N, the observation time, and J, the mixture resolving estimating category reinforcing its elements. It is thus seen that the expectation of the mixture resolving estimators, relative to their estimating category, under linear substitution into the detector operator structure, $\underline{h}_o$, at the Nth stage, minimizes the average conditional probability of error at that stage, relative to the category of the estimators.

Two techniques, feedback and feedforward, have been developed to reinforce the detector structure as the noise characteristics and discrete pulse waveforms are jointly learned through mixture resolving estimation. Characteristics common to both categories are;

(1) Neither category requires apriori knowledge of the noise parameters or discrete pulse waveforms.

(2) Neither category requires apriori classification of the time slots of the noisy signal pattern observation (mixture).

The mixture resolving estimating categories developed, consist of;

(1) An optimized, time-varying-weighted decision-directed category (feedback reinforcement).

(2) A moment method cateogry (feedforward reinforcement).

The categories differ fundamentally, in that in the first, the decision is made and then the structure is re-inforced, and in the latter, the structure is re-inforced and then the decision is made. The first category differs from the conventional decision directed category in that;

(1) The initial reference is extracted from the observations by correlating the first observation with the next adjacent observation.

(2) Time-varying-weighted combination of the separated time slots is carried out under optimization to minimize distance and dispersion,

with a subsequent adjustment in the constraint to "maximize" the convergence rate. An iterative format for the weighting recursion in estimating the pulse waveform is developed.

The second category differs from previous work using the method of moments, in that apriori knowledge of neither the pulse nor the noise parameters is required.

The integration of the mixture resolving estimators into the detection operator structure constitutes an adaptive or learning decision equation and signal processing algorithm. Extensive experimental studies, via digital computer simulation, of the performance characteristics of these signal processing algorithms have been carried out for the equiprobable bi-polar case only. The dynamic and asymptotic performance characteristics of these algorithms have been compared with simulation established performance characteristics of the conventional decision-directed and Bayes matched filter algorithms.

The principal results, observations, and conclusions are as follows;

(1) Both the optimally weighted decision-directed and the moment method categories converge for negative db. values of SNR at a rate considerably higher than that of the conventional decision-directed category.

(2) The performance of the algorithms developed in this work is bounded from above by the conventional decision directed algorithm over the entire SNR range investigated and reasonably approaches the performance of the Bayes matched filter without the computational complexity of the formal Bayes mixture approach.

(3) For the optimally weighted, decision directed algorithm, over the entire SNR range investigated; the weighting is dominated only by

the observation stage, N, and the constraint coefficient, γ, and is thus non-parametric in the mixture pulse waveform and noise parameters.

7-2. Recommendations for Future Investigation

It has been encouraging, in this relatively non-Bayesian approach, to observe significant improvement over the conventional decision-directed algorithm by processing with a formally developed weighting iteration without the computational complexity of the formal Bayes mixture requirement. In addition, it appears promising, in the areas of digital data processing and estimation, to demand more specifics of the output by placing the additional and particular conditioned, extremized, and constrained properties upon the operators.

In work related to this area, it would be interesting to carry out some formal study of threshold effects for a given weighting system and to investigate specific tradeoff effects between dispersion, convergence rate, acceleration, and perhaps other criteria.

An inspection of the underlying principles of both the conventional and weighted decision directed philosophies, indicates that after classification, the classified variables are never reviewed. It appears that a lowering of the threshold, in terms of SNR, for reliable performance; as well as a further increase in the rate of convergence with a reduction in asymptotic bias, might be achieved by some systematic, combinatorial, re-classification of the early data. One possibility would be to reclassify all observations up to stage N_A, which is that time at which the estimates of the elements for the structures (mean vectors, covariance matrix) had stabilized, having attained their "asymptotic" values. In that case, the reclassification of the early data, $\{\underline{z}_k\}_{N_A}$, using "asymptotic" values for the elements of the structure would correspond to classification with a filter mis-matched by the amount of the "asymptotic" bias in the elements.

In as much as adaptive and learning algorithms can be formulated in a Markov process context, it would appear to be worthwhile to consider the development of a learning, adaptive, and mixture resolving algorithm by working with the transition probability matrix equations and perhaps constraining and extremizing the transition probabilities.

Finally, the development of mixture resolving estimating categories in this thesis as well as in all work to date, insofar as a rather intensive area investigation has revealed, assumes some form of time slot synchronization. In general, this assumption is valid, from a practical standpoint, only in digital communication systems employing a transmitted reference, or other scheme, for binary digit synchronization. Much more effort is required in investigating techniques and developing references for locating, isolating, and separating "unknown events" in more general mixtures such as discrimination radar, sonar, and seismic echoes; and bio-electric signals.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Abramson, N., and D. Braverman, "Learning to Recognize Patterns in a Random Environment," IRE International Symposium on Information Theory, Vol. IT-8, pp. 558-563, July, 1962.

2. Keehn, D. G., "Learning the Mean Vector and Covariance Matrix of Gaussian Signals in Pattern Recognition," Stanford Electronics Laboratories, Tech. Report No. 2003-6, February, 1963.

3. Raiffa, H. and R. Schlaifer, Applied Statistical Decision Theory, Division of Research - Harvard Business School, Cambridge, Mass., 1961.

4. Spragins, J. D., "Reproducing Distributions for Machine Learning," Stanford Electronics Laboratories, Tech. Report No. 6103-7, November, 1963.

5. Jakowatz, C. V., R. L. Shuey, and G. M. White, "Adaptive Waveform Recognition," G. E. Research Lab., Tech. Report No. 60-RL-2 353E, September, 1960.

6. Downing, J. J., Modulation Systems and Noise, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

7. Hinich, M., "Large-Sample Estimation of an Unknown Discrete Waveform Which is Randomly Repeating in Gaussian Noise," Dept. of Statistics, Stanford University, Tech. Report No. 93, December, 1963.

8. Daly, R. F., "Adaptive Binary Detectors," Stanford Electronics Laboratories, Tech. Report No. 2003-2, March, 1961.

9. -----------, "The Adaptive Binary Detection Problem on the Real Line," Stanford Electronics Laboratories, Tech. Report No. 2003-3, February, 1962.

10. Fralick, S. C., "The Synthesis of Machines Which Learn Without a Teacher," Stanford Electronics Laboratories, Tech. Report No. 6103-8, April, 1964.

11. -----------, "Learning to Recognize Patterns Without a Teacher, Stanford Electronics Laboratories, Tech. Report No. 6103-10, March, 1965.

12. Hancock, J. C. and E. A. Patrick, "The Non-supervised Learning of Probability Spaces and Recognition of Patterns," IEEE International Convention Record - Information II, New York, N. Y., March 1965.

13. --------, "Learning Probability Spaces for Classification and Recognition of Patterns With or Without Supervision," School of Electrical Engineering - Purdue University, Tech. Report No. TR-EE 65-21, November, 1965.

14. Teicher, Henry, "On the Mixture of Distributions," Annals of Math. Stat., Vol. 31, pp. 55-73, March, 1960.

15. -------, "Identifiability of Finite Mixtures," Annals of Math. Stat., Vol. 32, pp. 244-248, March, 1961.

16. -------, "Identifiability of Finite Mixtures," Annals of Math. Stat., Vol. 34, pp. 1265-1269, December, 1963.

17. Pearson, K. P., "Contributions to the Mathematical Theory of Evolution," Phil. Trans. Royal Soc. 185A, 71-110, 1894.

18. Rao, C. R., Advanced Statistical Methods in Biometric Research, John Wiley and Sons, New York, 1952.

19. Rider, P. R., "The Method of Moments Applied to a Mixture of Two Exponential Distributions," Annals of Math. Stat., Vol. 32, pp. 143-147, March, 1961.

20. Blischke, W. R., "Estimation of Parameters of Mixtures of Binomial Distributions," Journal of the American Stat. Assoc., Vol. 59, pp. 510-528, 1964.

21. Hancock, J. C., and D. F. Mix, "Learning Theory Applied to Communications," School of Electrical Engineering, Purdue University, Tech. Report No. TR-EE 65-20, October, 1965.

22. Scudder, H. J., "Adaptive Communication Receivers," Electronics Research Lab., University of California, Tech. Report No. 64-3, April, 1964.

23. Hancock, J. C. and R. W. Chang, "On Reception of Signals for Channels Having Memory," School of Electrical Engineering, Purdue University, Tech. Report No. TR-EE 65-19, October, 1965.

24. Scudder, H. J., "Probabilitiy of Error of Some Adaptive Pattern-Recognition Machines," IEEE Transactions on Information Theory, Vol. IT-11, No. 3, July, 1965.

25. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York, 1958.

26. Kenney, J. F., and E. S. Keeping, Mathematics of Statistics, D. Van Nostrand Co., New York, 1953.

27. Bodewig, E., Matrix Calculus, North Holland Publishing Co., Amsterdam, 1959.

28. Indritz, Jack, Methods in Analysis, The MacMillan Co., New York, 1963.

29. Wilks, S. S., Mathematical Statistics, John Wiley and Sons, New York, 1962.