ADA026695

ESD-TR-76-73

See 473

(12)

Technical Report

517

C. W. Therrien

The Application of Linear Prediction to Sequential Classification of Radar Target Signatures

DDC
RECEIVED
JUL 12 1976
A

25 March 1976

Prepared for the Ballistic Missile Defense Program Office,
Department of the Army,
under Electronic Systems Division Contract F19628-76-C-0002 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Approved for public release; distribution unlimited.

Best Available Copy

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This program is sponsored by the Ballistic Missile Defense Program Office, Department of the Army; it is supported by the Ballistic Missile Defense Advanced Technology Center under Air Force Contract F19628-76-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Eugene C. Raabe

Eugene C. Raabe, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document when it is no longer needed.

Best Available Copy

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

THE APPLICATION OF LINEAR PREDICTION TO SEQUENTIAL CLASSIFICATION OF RADAR TARGET SIGNATURES

C. W. THERRIEN

Group 92

TECHNICAL REPORT 517

25 MARCH 1976

Approved for public release; distribution unlimited.

Best Available Copy

LEXINGTON MASSACHUSETTS

ABSTRACT

A linear-predictive form of the quadratic classifier (the optimal decision rule for Gaussian random processes) is developed and applied to the discrimination and classification of radar target signatures. The classifier was devised to implement a sequential probability ratio test (SPRT); that is, consecutive radar returns are observed until the target can be classified with a prescribed probability of error. Because of the linear-predictive formulation, the computational and storage requirements for the classifier are related only to the number of returns necessary to predict the signature and not to the length of signature observed; a classifier with modest storage and computational requirements can be employed to classify signatures consisting of an arbitrarily large number of radar returns. The classifier is related to several results in mean-square filtering theory and has an interpretation in terms of the maximum entropy and maximum likelihood spectral estimates for the target signatures.

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

CONTENTS

FOREWORD

This report describes a decision rule structure for the discrimination of two discrete-time random processes. The structure is recursive and its discovery was motivated by a desire to perform sequential discrimination of targets tracked by a radar. For sequential discrimination, an object is illuminated with consecutive radar pulses until a classification of the object can be made to within a predetermined probability of error.

The approach taken brings some results from estimation theory (in particular, linear prediction) to bear on the discrimination problem. The structure of the decision rule can be related to estimator/correlator receiver realizations for random signal detection problems and both formulations have some common important characteristics. Because the linear prediction can often be performed with a relatively short history of the process, only modest amounts of storage and computational resources are required – even when the process is observed over a long time interval.

The sequential decision rule described here has been implemented in real time and operates in conjunction with the TRADEX radar at Kwajalein, M.I. The real-time implementation has been given the acronym LEAD for Linear Estimation And Discrimination.

The report is written in the context of the radar discrimination problem and is presented in a largely tutorial format. It should be noted, however, that the results are applicable to target discrimination involving other types of sensors (sonic, optical, etc.) and to other similar detection and identification problems as well. The reader with a basic knowledge of statistical decision procedures will find that the results are derived in a straightforward manner and that the report is self-contained. Several references are provided for the related material on mean-square filtering and spectral analysis.

THE APPLICATION OF LINEAR PREDICTION TO SEQUENTIAL CLASSIFICATION OF RADAR TARGET SIGNATURES

I. INTRODUCTION

This report describes a new form of quadratic classifier (the optimal decision rule for Gaussian random processes) that is being used for discrimination and classification of radar target signatures.* The classifier is sequential in nature and is well suited for real-time operation. The signatures are observed a few returns at a time and a classification may be made after any number of observations. Since most forms of the quadratic classifier can be applied only to signatures consisting of some fixed number of returns, this sequential form provides the following advantages:

(a) Signatures of targets that are easy to discriminate are classified with a relatively small number of returns. This saves time that would be wasted if it were required to observe the signature for a larger, fixed number of returns, and permits a greater number of such targets to be discriminated in a given real-time interval.

(b) Signatures of targets that are difficult to discriminate can be observed for an arbitrarily large number of returns before classification. This capability, which derives from our special formulation of the sequential decision rule, leads to a higher percentage of correct classifications of these targets than could be achieved if classifications were restricted to signatures with a fixed maximum number of returns.

The sequential classifier has no greater storage and computational requirements than a similarly configured non-sequential quadratic classifier, and when both are applied to signatures having the same number of returns, classification results are identical. However, because it permits targets to be discriminated with a varying number of observations, the sequential form of the classifier results in a more efficient use of radar resources and an overall improvement in discrimination performance.

The sequential decision rule on which the classifier is based is characterized by a set of recursive decision stages. Each stage consists of the observation of returns from one or more pulses illuminating the target, and an attempt to classify the target based on these and all previously observed returns. The process is repeated through successive stages until a classification can be made at a prescribed level of confidence.

This report shows that a realization of the sequential decision rule can be obtained that consists of two basic steps: linear prediction followed by an "incremental" classification. At each decision stage, a linear prediction, based on previous observations, is made of the radar returns to be observed during that stage. An error term, formed as the difference between the predicted and the observed values of the signature, is then applied to an "incremental" quadratic classifier. The output of the incremental classifier is accumulated in a sum of previous outputs and compared to a set of thresholds to make a classification decision.

---

* The term "signature" refers to the time-ordered returns from a set of consecutively transmitted radar pulses illuminating a given target.

The classifier structure suggests two modes of operation: a normal mode in which prediction of the radar signature is based on the returns observed in all previous decision stages, and an extended mode in which prediction of the signature is based on the returns observed in only some fixed maximum number of previous stages. It is important to note that while signature prediction in the extended mode is based on the radar returns observed in only a subset of the previous stages, signature classification is based on the returns observed in all previous stages.

The extended mode of operation leads to important advantages in computer storage and execution time. A sequential classifier with modest storage and computational requirements can achieve the performance of a non-sequential classifier with much larger storage and computational requirements. In addition, since the computation during each stage of extended operation is constant and proportional to the (fixed) number of returns used for prediction, computational requirements place no upper limit on the ultimate length of signature that can be observed.

The use of the sequential classifier in extended mode has produced some outstanding results. Two examples will illustrate these.

Table I shows the results of applying a sequential classifier to 500 simulated signatures each of a re-entry vehicle (RV) and a fragment from the associated missile final stage (tank).

TABLE I

CLASSIFICATION RESULTS FOR A SEQUENTIAL CLASSIFIER USED IN THE NORMAL MODE*

Number of Decision Stages	Total Number of Signatures Classified		Classification Errors (percent)		Average Number of Radar Returns	
	RV	Fragment	RV	Fragment	RV	Fragment
1	1	78	100.0	0.0	1.0	1.0
2	9	116	11.1	0.0	1.9	1.3
3	18	155	16.7	0.0	2.4	1.7
4	48	182	8.3	0.0	3.4	2.1
5	95	215	7.4	0.0	4.2	2.5
6	147	231	4.8	0.0	4.8	2.8
7	203	260	3.9	0.8	5.4	3.2
8	257	288	3.5	0.7	6.0	3.7
9	286	297	3.8	1.0	6.3	3.9
10	333	317	3.9	1.9	6.8	4.3
11	378	325	3.4	2.8	7.3	4.4
12	401	335	3.2	3.6	7.6	4.6
13	418	341	3.3	3.8	7.8	4.8
14	432	345	3.5	3.8	8.0	4.9
15	445	354	3.6	4.5	8.2	5.2
16	457	366	3.5	4.9	8.4	5.5
16†	500	500	5.8	20.4	9.1	8.3

* Dual-polarized simulated signatures at L-band.

† Remaining signatures classified according to a single default threshold.

The classifier was exercised in the normal mode (i.e., the returns from all previous decision stages were used to predict the signature) and a single radar return was observed at each decision stage. The table lists for each stage the number of signatures of each target classified, the cumulative error rates, and the average number of returns used to classify the signatures. For a given signature to be classified at a given stage, the classifier output for that signature must fall outside a pair of widely separated thresholds. After sixteen decision stages, all the unclassified signatures were classified according to a single default threshold. It was found that the final error rates of 5.8 and 20.4 percent were the same as those for a non-sequential quadratic classifier based on sixteen returns, while the average number of returns used for classification was only 9.1 for the RV and 8.3 for the fragment. In other words, the average classification time was approximately cut in half.

Table II shows the results of classifying the same signatures using a classifier with considerably reduced storage requirements. This classifier was used in the extended mode; prediction of the signature was based on a maximum of three previous returns. The final error rates and average number of returns for each target are almost identical to those in Table I. This classifier has storage and computational requirements that are about 1/16 of those for the

TABLE II

CLASSIFICATION RESULTS FOR A SEQUENTIAL CLASSIFIER WITH REDUCED STORAGE REQUIREMENTS USED IN THE EXTENDED MODE*

Number of Decision Stages	Total Number of Signatures Classified		Classification Errors (percent)		Average Number of Radar Returns	
	RV	Fragment	RV	Fragment	RV	Fragment
1	1	76	100.0	0.0	1.0	1.0
2	11	115	9.1	0.0	1.9	1.3
3	24	149	12.5	0.0	2.5	1.7
4	63	179	6.3	0.6	3.4	2.1
5	106	214	5.7	0.5	4.1	2.6
6	166	243	5.4	0.4	4.8	3.0
7	218	263	4.1	0.8	5.3	3.3
8	257	289	3.5	1.4	5.7	3.7
9	294	305	3.4	2.3	6.1	4.0
10	329	318	3.0	2.8	6.5	4.2
11	369	327	2.7	4.0	7.0	4.4
12	401	336	2.7	4.2	7.4	4.6
13	418	349	3.1	5.2	7.6	4.9
14	431	353	3.2	5.1	7.8	5.0
15	448	363	3.3	5.5	8.1	5.3
16	454	368	3.3	5.7	8.2	5.5
16†	500	500	6.0	20.0	8.9	8.2

* Dual-polarized simulated signatures at L-band.

† Remaining signatures classified according to a single default threshold.

previously cited non-sequential classifier; yet it obtained the same error rates and classified the targets with approximately half the number of returns.

As a second example, a sequential classifier was applied to the signatures of an experimental RV and one of its decoys. Both objects were contained in a cloud of chaff. Figure 1 shows the operating characteristic (plot of RV error rate vs decoy error rate) for sequential and non-sequential quadratic classifiers with identical storage requirements. The non-

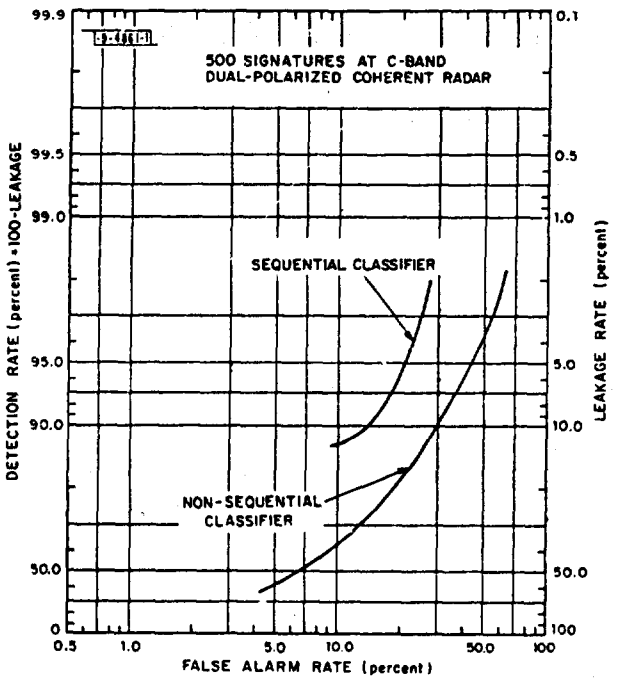


Fig. 1. Comparison of performance of sequential and non-sequential quadratic classifiers.

sequential classifier observed ten returns of each signature. The sequential classifier, applied in the extended mode, observed an average number of returns that varied between points on the operating characteristic, but which was generally in the range of 25 to 35. The performance improvement is clearly significant. For example, at the 10-percent decoy error (false alarm) rate, the non-sequential classifier had an RV error (leakage) rate of about 40 percent while the sequential classifier had a leakage rate of about 11 percent. Note that since the average number of returns observed by the sequential classifier was generally in the range of 25 to 35, some signatures were observed for a larger number of returns. Thus, the performance of the sequential classifier could probably not have been matched by even a (larger and computationally more expensive) non-sequential quadratic classifier designed for signatures with the average number of returns.

These examples illustrate the advantages that can be obtained by using the special sequential form of the quadratic classifier. The remainder of this report describes the classifier in detail and discusses its method of implementation. Section II develops the theory of the classifier while Sec. III deals with computational issues. Section IV relates the sequential classifier to the current literature in linear mean-square prediction theory and provides the justification for the use of the classifier in the extended mode. In addition, Sec. IV provides an interpretation of the classifier in terms of maximum likelihood and maximum entropy spectral estimation. Finally, Sec. V summarizes results.

II. DEVELOPMENT OF THE SEQUENTIAL QUADRATIC CLASSIFIER

A. Basic Form of the Sequential Decision Rule

Two-class sequential decision procedures have been employed in a number of applications.[1-4] For the application of these procedures to radar signature discrimination, an observation vector $\underline{x}_k$ is formed from samples of the radar signature at the k^{th} decision stage. In the so-called "fixed-boundary" sequential probability ratio test (SPRT), the likelihood ratio for the observation vector $\underline{x}_k$ is evaluated and compared to thresholds $T_A > T_B$. If the value of the likelihood ratio is greater than T_A or less than T_B, then Class 1 (ω_1) or Class 2 (ω_2) is decided, respectively.

$$\ell_k = \frac{p(\underline{x}_k|\omega_1)}{p(\underline{x}_k|\omega_2)} \quad \begin{cases} > T_A \rightarrow \omega_1 \\ < T_B \rightarrow \omega_2 \end{cases} . \tag{1}$$

Otherwise, the process continues for another stage. A fixed number of additional observations are made and appended to the observation vector; a new likelihood ratio is formed using the conditional multivariate densities for the larger observation vector, and the process continues until one of the boundaries (thresholds) is crossed. The probabilities of misclassification are related to the thresholds by the following equations.[1,2]

$$\text{Prob}[\text{Error}|\omega_1] = T_B(T_A - 1)/(T_A - T_B) \tag{2a}$$

$$\text{Prob}[\text{Error}|\omega_2] = (1 - T_B)/(T_A - T_B) . \tag{2b}$$

In the most general type of SPRT, the thresholds are not fixed at constants T_A and T_B but change at each decision stage.[2] For example, the thresholds or "boundaries" can be made to converge gradually to a single value as shown in Fig. 2(a). This forces the classifier to make a decision after some maximum number of stages k_o. For the case of general time-varying boundaries, Eqs. (2) do not apply.

A special case of time-varying boundaries is illustrated in Fig. 2(b). The boundaries remain fixed at T_A and T_B until the $(k_o)^{th}$ stage when both are changed to a common intermediate

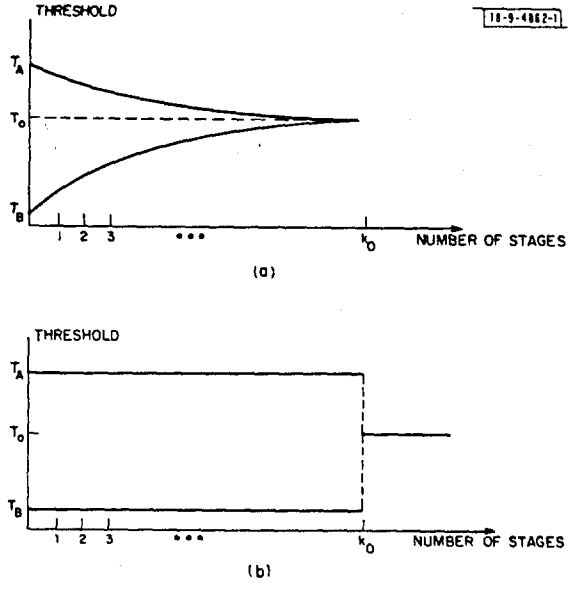


Fig. 2. Types of decision boundaries for sequential classification.

value T_o. This special case of time-varying boundaries differs from the fixed-boundary case only in that a decision is suddenly forced after k_o stages. If k_o is sufficiently large so that most observation vectors are classified before k_o stages, then Eqs. (2) provide a good approximation to the errors. Obviously in any practical implementation, sequential decision processes cannot be allowed to continue indefinitely so that if it is desired to guarantee classification of all observation vectors, decision boundaries of one of the forms of Fig. 2 must be employed.

B. Recursive Form of the Sequential Decision Rule

For the application of the SPRT to radar signature discrimination, the observation vector is partitioned as

$$\underline{x}_k = \begin{bmatrix} \underline{x}_{k-1} \\ \cdots\cdots \\ \Delta\underline{x}_k \end{bmatrix} \tag{3}$$

and constructed as shown in Fig. 3. Here $\underline{x}_{k-1}$ is the observation vector at the $(k-1)^{th}$ stage and $\Delta\underline{x}_k$ represents the new observations at the k^{th} stage. The dimension of $\Delta\underline{x}_k$ is equal to $n\Delta p$ where Δp is the number of new returns observed per stage (which is assumed to be the same for all stages) and n ; ; the number of scalar items of information available from each return.

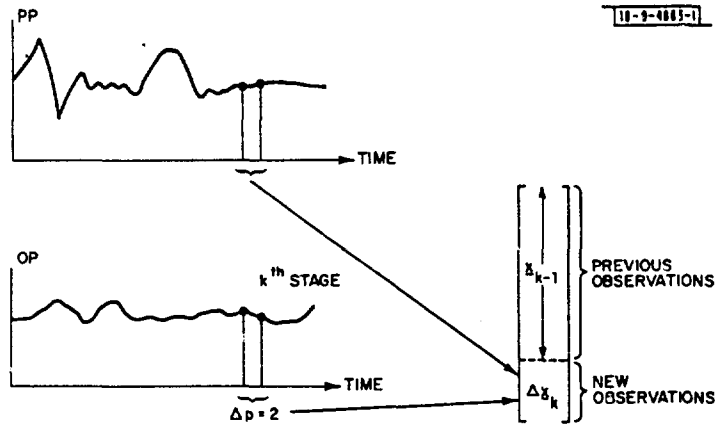


Fig. 3. Construction of observation vector for sequential classification of radar signatures.

In particular, if the radar is narrowband and noncoherent with only the principal polarization (PP), then n will be equal to 1. If the radar is coherent with only PP, or if the radar is noncoherent with both principal and orthogonal polarizations (OP), then n will be equal to 2. If the radar is coherent in both PP and OP, then n will be equal to 4.*

*Our formulation models the coherent signatures from each polarization as samples from two correlated real random processes (representing amplitude and phase or real and imaginary parts). An alternative formulation would treat the signatures as samples from a single complex process.

The likelihood ratio in Eq. (1) can be written in the recursive form

$$\ell_k = \frac{p(\Delta \underline{x}_k | \underline{x}_{k-1}, \omega_1)}{p(\Delta \underline{x}_k | \underline{x}_{k-1}, \omega_2)} \cdot \ell_{k-1} \quad . \tag{4}$$

Note that when the new observations $\Delta \underline{x}_k$ are independent of the previous observations $\underline{x}_{k-1}$, the first term in Eq. (4) depends only on $\Delta \underline{x}_k$ and so the decision rule is considerably simplified. In order to achieve independence for signatures described by Gaussian statistics one can, by the use of Karhunen-Loève analysis, transform the observations into a coordinate system where they are uncorrelated and therefore independent. This is the approach taken in Refs. 3 and 4. Our approach is somewhat different. It will be shown that when the observations are jointly Gaussian, but not necessarily independent, one can separate the decision rule into a step consisting of a linear estimation of the new observations and a step involving a likelihood ratio test based on the prediction errors. It can further be shown (see Sec. IV) that this is essentially a different path to achieving independence. The prediction errors used for classification are uncorrelated and represent only the new or "innovations" information present in the new observations.

C. Decision Rule for Jointly Gaussian Observations

If the observations $\underline{x}_k$ are jointly Gaussian, then the probability densities in Eq. (1) are given by

$$p(\underline{x}_k | \omega_i) = \frac{1}{(\sqrt{2\pi})^{kn\Delta p} |K_k^i|^{1/2}} \exp[-\tfrac{1}{2} Q_k^i] \quad ; \quad i = 1, 2 \tag{5a}$$

with

$$Q_k^i = (\underline{x}_k - \underline{m}_k^i)^T (K_k^i)^{-1} (\underline{x}_k - \underline{m}_k^i) \tag{5b}$$

where $\underline{m}_k^i$ is the mean vector and K_k^i is the covariance matrix for the observation vector $\underline{x}_k$ of the i^{th} class. Q_k^i is the squared Mahalanobis distance between the observation vector and the mean vector of the i^{th} class and is explicitly defined here because of its importance in deriving the ensuing results.

The conditional probability densities appearing in Eq. (4) for the new observations $\Delta \underline{x}_k$ can be derived by first noting that

$$p(\Delta \underline{x}_k | \underline{x}_{k-1}, \omega_i) = \frac{p(\Delta \underline{x}_k, \underline{x}_{k-1} | \omega_i)}{p(\underline{x}_{k-1} | \omega_i)} = \frac{p(\underline{x}_k | \omega_i)}{p(\underline{x}_{k-1} | \omega_i)} \quad . \tag{6}$$

Thus, substituting Eq. (5) one obtains

$$p(\Delta \underline{x}_k | \underline{x}_{k-1}, \omega_i) = \frac{\frac{1}{(\sqrt{2\pi})^{kn\Delta p} |K_k^i|^{1/2}} \exp[-\frac{1}{2} Q_k^i]}{\frac{1}{(\sqrt{2\pi})^{(k-1)n\Delta p} |K_{k-1}^i|^{1/2}} \exp[-\frac{1}{2} Q_{k-1}^i]}$$

$$= \frac{1}{(\sqrt{2\pi})^{n\Delta p} |K_k^i|^{1/2} / |K_{k-1}^i|^{1/2}} \exp[-\tfrac{1}{2} \Delta Q_k^i] \tag{7}$$

where

$$\Delta Q_k^i = Q_k^i - Q_{k-1}^i \quad . \tag{8}$$

Let the mean vector and the covariance matrix be partitioned to correspond to the partitioning of the observation vector, i.e.,

$$\underline{m}_k = E\left[\begin{bmatrix} \underline{x}_{k-1} \\ \hline \Delta\underline{x}_k \end{bmatrix}\right] = \begin{bmatrix} \underline{m}_{k-1} \\ \hline \Delta\underline{m}_k \end{bmatrix} \tag{9a}*$$

$$K_k = E\left[\begin{bmatrix} \underline{x}_{k-1} \\ \hline \Delta\underline{x}_k \end{bmatrix} \begin{bmatrix} \underline{x}_{k-1}^T & \vdots & \Delta\underline{x}_k^T \end{bmatrix}\right] = \left[\begin{array}{c|c} K_{k-1} & B_k \\ \hline B_k^T & D_k \end{array}\right] \quad . \tag{9b}$$

It is shown in Appendix A that the corresponding inverse covariance matrix can be expressed in partitioned form as

$$K_k^{-1} = \left[\begin{array}{c|c} K_{k-1}^{-1} + G_k E_k^{-1} G_k^T & -G_k E_k^{-1} \\ \hline -E_k^{-1} G_k^T & E_k^{-1} \end{array}\right] \tag{10}$$

where

$$G_k = K_{k-1}^{-1} B_k \tag{11}$$

and

$$E_k = D_k - B_k^T K_{k-1}^{-1} B_k \quad . \tag{12}$$

It is further shown in Appendix A that ΔQ_k is of the form

$$\Delta Q_k = (\Delta\underline{x}_k - \underline{\mu}_k)^T E_k^{-1} (\Delta\underline{x}_k - \underline{\mu}_k) \tag{13}$$

where

$$\underline{\mu}_k = \Delta\underline{m}_k + G_k^T (\underline{x}_{k-1} - \underline{m}_{k-1}) \tag{14}$$

and that

$$|E_k| = \frac{|K_k|}{|K_{k-1}|} \quad . \tag{15}$$

---

* In order to reduce the complexity of notation, the class index superscripts (i) will be dropped in Eqs. (9) through (15). It should be understood that the results apply to both classes.

By substituting Eq. (15) into Eq. (7), one obtains

$$p(\Delta \underline{x}_k | \underline{x}_{k-1}, \omega_i) = \frac{1}{(\sqrt{2\pi})^{n\Delta p} |E_k^i|^{1/2}} \exp[-\tfrac{1}{2} \Delta Q_k^i] \tag{16}$$

where ΔQ_k^i is given by Eq. (13). This shows that the conditional density of $\Delta \underline{x}_k$ is a Gaussian density with mean vector $\underline{\mu}_k$ and covariance matrix E_k.

That the conditional density $p(\Delta \underline{x}_k | \underline{x}_{k-1})$ is Gaussian when $p(\underline{x}_k)$ is Gaussian is well-known (see Ref. 5, Appendix to Chap. 7). The mean of the conditional density $\underline{\mu}_k$, regarded as a function of the previous observations $\underline{x}_{k-1}$, is known in statistics as the regression of $\Delta \underline{x}_k$ with respect to $\underline{x}_{k-1}$ and it is the maximum a posteriori estimate of $\Delta \underline{x}_k$. What is most important here is that the mean of the conditional density for Gaussian processes is a linear function of the previous observations [Eq. (14)]. Therefore, the sequential classification algorithm can be decomposed into two simple steps; namely (1) a linear-predictive estimation of the new observations and (2) an evaluation of the likelihood ratio for the conditional densities [Eq. (4)]. The dimensionality of the vectors and matrices employed in this second step is $n\Delta p$, i.e., it is the dimension of just the new observations at each stage.

D. Interpretation as Sequential Quadratic Classifier

One can interpret the decision rule for non-sequential likelihood ratio tests when the probability densities are Gaussian as a quadratic classifier. That is, the likelihood ratio test is interpreted as forming a second-order polynomial boundary in the observation space between the two regions in which observation vectors are classified as Class 1 or as Class 2. As such, quadratic classifiers are employed to classify observation vectors even with non-Gaussian statistics; one considers the classifier as defining a second-order decision boundary which is fitted to the first and second moments of the observation vectors. The SPRT for Gaussian statistics can also be interpreted as a (sequential) quadratic classifier. In this situation, however, the observation space and the decision boundary increases in dimensionality with each successive stage.

Let h_k represent minus twice the log of the likelihood ratio at the k^{th} stage. Then, from Eqs. (1) and (5) the sequential decision rule is

$$h_k = Q_k^1 - Q_k^2 + \ln \frac{|K_k^1|}{|K_k^2|} \quad \begin{cases} < -2 \ln T_A \rightarrow \omega_1 \\ > -2 \ln T_B \rightarrow \omega_2 \end{cases} \tag{17}$$

where Q_k^1 and Q_k^2 are the squared Mahalanobis distances defined by Eq. (5b). The left side of Eq. (17) has the same form as a standard quadratic classifier for observation vectors at the k^{th} stage. The classifier can be expressed in a recursive form by applying Eqs. (8) and (15) to Eq. (17) or by applying Eqs. (4) and (16) to Eq. (1). The result is

$$h_k = h_{k-1} + \Delta h_k \quad \begin{cases} < -2 \ln T_A \rightarrow \omega_1 \\ > -2 \ln T_B \rightarrow \omega_2 \end{cases} \tag{18a}$$

$$\Delta h_k = \Delta Q_k^1 - \Delta Q_k^2 + \ln \frac{|E_k^1|}{|E_k^2|} \tag{18b}$$

where ΔQ_k^1 and ΔQ_k^2 are defined by Eq. (13) and where at the start of the recursion one has $E_1^i = K_1^i$ and $h_0 = 0$. Note that in the sequential form of the classifier we need to deal with only an incremental classifier [Eq. (18b)] at each stage. The dimension of vectors and matrices in the incremental classifier is equal to $n\Delta p$. The change in the squared Mahalanobis distances ΔQ_k^i at each stage involves a linear prediction of the new observations $\Delta \underline{x}_k$ from the previous observations $\underline{x}_{k-1}$ as discussed earlier. Equations (13) and (14), used to compute the ΔQ_k^i, can be written in an alternative form as

$$\Delta Q_k^i = \underline{\epsilon}_k^{iT} (E_k^i)^{-1} \underline{\epsilon}_k^i \tag{19a}$$

$$\underline{\epsilon}_k^i = \Delta \underline{x}_k^i - \Delta \hat{\underline{x}}_k^i \tag{19b}$$

$$\Delta \hat{\underline{x}}_k^i = G_k^{iT} \underline{x}_{k-1}^i \qquad i = 1, 2 \tag{19c}$$

where

$$\underline{x}_k^i = \begin{bmatrix} \underline{x}_{k-1}^i \\ \cdots\cdots \\ \Delta \underline{x}_k^i \end{bmatrix} = \begin{bmatrix} \underline{x}_{k-1} - \underline{m}_{k-1}^i \\ \cdots\cdots\cdots\cdots \\ \Delta \underline{x}_k - \Delta \underline{m}_k^i \end{bmatrix} \tag{20}$$

and where $\Delta \hat{\underline{x}}_1^i = \underline{0}$. The prediction/classification sequence is very clear from this formulation with the mean removed. At each stage in the classification sequence a prediction $\Delta \hat{\underline{x}}_k^i$ of the new observations is made by Eq. (19c). An error term $\underline{\epsilon}_k^i$ is then formed in Eq. (19b) and this is used in Eq. (19a) to evaluate the change in the squared Mahalanobis distance ΔQ_k^i. These changes are then used in the incremental classifier decision rule, Eqs. (18).

E. Extension of Results to Long Signatures

If the statistics of the radar signatures to be classified do not change over an extended period of time, then one would expect that there would be some number $k_s - 1$ of previous observations called the prediction order for which the estimate $\Delta \hat{\underline{x}}_k$ of the new observations would be no worse than it would be if the estimate were based on an infinite number of previous observations. That this is in fact true is shown in Sec. IV. Consequently, one can base the prediction of the radar signature on the observations acquired in only the previous $k_s - 1$ stages, but continue to exercise the classifier for any number of stages $k_o > k_s$. By operating the sequential classifier in this "extended mode," one can obtain the approximate performance of a classifier with much larger computational and storage requirements. In fact, one gains an ability to process long signatures that for reasons of excessive computation or storage requirements could not be processed with other classifiers. The benefits of operating a sequential classifier in the extended mode already have been demonstrated in Sec. I.

One can imagine that the sequential classifier views a signature through a sliding window that never exceeds $(k_s - 1)\,\Delta p$ radar returns in length. Since the classifier output $[h_k]$ at the k^{th} stage is an accumulation of results from all previous stages [Eqs. (18)], the classification

decision is based on all returns seen by the classifier up through the k^{th} stage, even if some of the earlier returns are currently outside the window. In this way the sequential quadratic classifier is able to obtain the effect of a much larger classifier at a fraction of the overhead. The next section discusses some of the practical issues of storage and computational requirements in detail.

III. COMPUTATIONAL CONSIDERATIONS

A. Computation and Storage of Classifier Parameters

The equations that implement the sequential classifier [Eqs. (18) through (20)] involve a set of (fixed) matrix parameters G_k^i and E_k^i which are derived from partitions of the covariance matrices for the two classes of signatures. The covariance matrices [as well as the mean vectors which appear in Eq. (20)] can be estimated from a set of training signatures for each of the two classes.

Computation of the G_k^i and the E_k^i is carried out prior to any real-time operation of the classifier. In fact, since the classifier operation involves not E_k^i itself but its inverse and its determinant, these latter quantities also are computed and stored prior to any real-time operation. The parameters of the classifier can be conveniently stored as a symmetric matrix

$$P^i = \begin{bmatrix} (E_1^i)^{-1} & G_2^i & G_3^i & \cdots & G_{k_s}^i \\ & (E_2^i)^{-1} & & & \\ & & (E_3^i)^{-1} & & \\ & & & & \left(E_{k_s}^i\right)^{-1} \end{bmatrix} \quad \begin{matrix} \leftarrow n\Delta p \rightarrow & \leftarrow n\Delta p \rightarrow & \leftarrow n\Delta p \rightarrow & & \leftarrow n\Delta p \rightarrow \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{matrix} \quad \left.\begin{matrix} \\ \\ \\ \\ \\ \end{matrix}\right\} k_s n\Delta p \qquad (21)$$

and a set of vectors

$$\underline{d} = \begin{bmatrix} d_1 \\ d_2 \\ . \\ . \\ . \\ d_{k_s} \end{bmatrix} \quad ; \quad d_k = \ln \frac{|E_k^1|}{|E_k^2|} \qquad (22)$$

and

$$\underline{m}^i_{k_s} = \left[\begin{array}{c} \underline{m}^i_k \\ \\ \\ \end{array}\right.\begin{array}{l} \\ km\,\Delta p \\ \\ \end{array}\left.\begin{array}{l} \\ \\ \\ \end{array}\right] \quad k_s n\,\Delta p \tag{23}$$

from which the mean vectors for any stage can be obtained as a partition (as shown). A special form of storage for the matrices P^i of Eq. (21) has been developed for use with FORTRAN so that the matrix partitions can be accessed directly and used with matrix manipulation programs from the IBM Scientific Subroutine Package.[6] This special form of storage is described in Appendix B.

The matrices E^i_1 are equal to the first-stage covariance matrices K^i_1. The matrices G^i_k and E^i_k for $k > 1$ are computed from the partitions of the k^{th} stage covariance matrices via Eqs. (11) and (12). If Eq. (10) is also employed, then a convenient recursive algorithm results for computation of the classifier parameters. One begins with the full-stage covariance matrices $K^i_{k_s}$ partitioned in a form corresponding to the partitioning of the P^i.

$$K^i_{k_s} = \left[\begin{array}{c|c|c|c|c} K^i_1 & B^i_2 & B^i_3 & \cdots & B^i_{k_s} \\ \hline & D^i_2 & & & \\ \hline & & D^i_3 & & \\ \hline & & & & \\ \hline & & & & D^i_{k_s} \end{array}\right] \tag{24}$$

At the first step of recursion, $K^i_1 = E^i_1$ is inverted (by any conventional procedure) and stored in P^i. The determinant $|E^i_1|$ which is usually obtained as a by-product of the inversion is retained and used to form the first element of the vector $\underline{d}$. At the second step of recursion, G^i_2 and E^i_2 are computed from Eqs. (11) and (12). The matrix E^i_2 is inverted and the results $[(E^i_2)^{-1}$ and $G^i_2]$ are stored in P^i. The determinant $|E^i_2|$ is retained and employed to form the second element of the vector $\underline{d}$. Equation (10) is then used to compute the inverted second-stage covariance matrix $(K^i_2)^{-1}$ which replaces the partitions K^i_1, B^i_2, and D^i_2 of Eq. (24) in

storage and the process continues. At the k^{th} step of recursion, one has stored the partially inverted covariance matrix

$$\begin{bmatrix} (K^i_{k-1})^{-1} & B^i_k & \cdots & B^i_{k_s} \\ & D^i_k & & \\ & & \ddots & D^i_{k_s} \end{bmatrix} \tag{25}$$

from which one can readily proceed to compute the partitions G^i_k and E^i_k of P^i and the inverse covariance matrix needed for the next step. The partitioned parameter matrices P^i, the full-stage mean vectors $\underline{m}^i_{k_s}$, and the determinant vector $\underline{d}$ completely define the classifier.

B. Real-Time Computation and Storage Requirements

Storage requirements for a sequential classifier that is designed to observe a maximum of $k_s \Delta p$ radar returns in the normal mode of operation are the same as those for a non-sequential quadratic classifier using that same number of returns. The primary storage allocation is for two symmetric matrices of dimension $k_s n\Delta p$ which requires $k_s n\Delta p \cdot (k_s n\Delta p + 1)$ words.

TABLE III
COMPUTATIONS TO EVALUATE ΔQ^i_k

Computed Item	Multiplications	Additions or Subtractions
$\Delta\hat{\underline{x}}^i_k = (G^i_k)^T \underline{x}^i_{k-1}$	$(k-1)(n\Delta p)^2$	$[(k-1)\, n\Delta p - 1]\, n\Delta p$
$\underline{\epsilon}^i_k = \Delta\underline{x}^i_k - \Delta\hat{\underline{x}}^i_k$	–	$n\Delta p$
$(\epsilon^i_k)^T (E^i_k)^{-1} \underline{\epsilon}^i_k$	$\frac{n\Delta p(n\Delta p + 3)}{2}$	$\frac{n\Delta p(n\Delta p + 1)}{2}$
Total	$(k - \frac{1}{2})(n\Delta p)^2 + \frac{3}{2} n\Delta p$	$(k - \frac{1}{2})(n\Delta p)^2 + \frac{1}{2} n\Delta p$

The central processor time required to execute the classifiers in real time can be approximated by the time required to compute the squared Mahalanobis distances Q^1_k and Q^2_k or their changes ΔQ^1_k and ΔQ^2_k. Table III lists the number of computations required to evaluate ΔQ^i_k at the k^{th} stage. Both the number of additions and the number of multiplications per stage are approximately equal to $k(n\Delta p)^2$ for $k \leqslant k_s$ and equal to $k_s(n\Delta p)^2$ thereafter. In other words, for the normal mode of operation [$k \leqslant k_s$], the computation is proportional to k and to the square of n

and Δp. If the number of arithmetic operations is summed over all stages up to the $(k_s)^{th}$, the total number of multiplications is $k_s n\Delta p(k_s n\Delta p + 3)/2$ and the number of additions is $k_s n\Delta p(k_s n\Delta p + 1)/2$, both of which are identical to the number of operations required to evaluate a non-sequential classifier of corresponding size. Thus, the total computational effort is the same for both types of classifiers when applied to signatures with the same number of returns. However, since the number of computations performed per stage by the sequential classifier is distributed so that it increases linearly with the number of stages, the real-time computational resources set aside for classifier evaluation in the normal mode of operation must be large enough to permit evaluation at the final most computationally expensive stage. This has the effect of doubling the total real-time computational resources that must be allocated to the classifier, although in the early decision stages this time is not used. While this may at first seem wasteful, more careful consideration of the application usually will show that it is possible to ration the unused time to lower priority computer tasks not related to classifier evaluation.

In the extended mode of operation, all the allocated time is used. Further, since the computational requirements at each stage are constant [proportional to $k_s(n\Delta p)^2$], there is no limit to the number of returns that may be observed in order to classify a target.

IV. RELATIONS TO MEAN-SQUARE FILTERING AND SPECTRAL ANALYSIS

The sequential classifier as formulated in this report is intimately related to certain problems in mean-square filtering and spectral estimation. It is in this context, in fact, that the classifier assumes its greatest significance and a better understanding of the classifier operation in both the normal and the extended mode emerges.

In order to relate the sequential classifier to the important literature in mean-square filtering and spectral analysis, we will restrict our attention for the most part to the case where the radar signature has only one component (i.e., PP) and the number of returns observed at each stage is one. Thus, the signatures can be modeled by single (univariate) discrete-time random processes which we will generally assume to be stationary. None of these special restrictions are absolutely necessary; the results can be formulated to include the more general case treated in the earlier portion of this report. However, generalization serves only to complicate the algebra and adds very little to the understanding of the basic relations.

It should be observed that whereas much of the literature referred to in this section is couched in terms of temporal averages of the random processes, we shall adhere to our use of ensemble averages. This should pose no particular problems as far as the results are concerned, since they can be formulated from either point of view. In addition, it will be assumed throughout that the mean of the signatures has been removed so that the random processes can be treated as zero-mean processes.

This section begins by showing that the predicted estimate $\Delta\hat{\underline{x}}_k$ of Eqs. (19) is the optimal linear mean-square estimate of the random process (the signature) regardless of the process statistics. For this result, none of the special restrictions cited earlier except the zero-mean condition are applied. Next, it is shown that when the processes are univariate and stationary, the set of recursive relations [Eqs. (10) through (12)] used to compute the classifier parameters reduce to the classical recursive relations originally formulated by Levinson.[7] The sequential classifier is then interpreted in a canonical form involving a pair of linear-predictive filters and an optimal classifier for two white Gaussian noise processes. Finally, the classifier is

interpreted in terms of the maximum likelihood and maximum entropy spectral estimates for the processes.

A. Linear Mean-Square Prediction of the Radar Signatures

The problem to be considered here is that of determining the best linear estimate $\Delta\hat{\underline{x}}_k$ of the observations $\Delta\underline{x}_k$ at the k^{th} stage from the set of previous observations $\underline{x}_{k-1}$ in the sense that

$$\mathcal{E}^2 = E[|\Delta\underline{x}_k - \Delta\hat{\underline{x}}_k|^2] \tag{26}$$

is minimum. The (arbitrary) linear estimate can be written in the form

$$\Delta\hat{\underline{x}}_k = G_k^T \underline{x}_{k-1} \quad . \tag{27}$$

It will be shown that the particular linear transformation G_k^T required to minimize Eq. (26) is given by Eq. (11).

In preparation for this result, we first define the estimation error $\underline{\epsilon}_k$ as

$$\underline{\epsilon}_k = \Delta\underline{x}_k - \Delta\hat{\underline{x}}_k \tag{28}$$

and state the following basic theorem of mean-square estimation which is known as the orthogonality principle. The proof of the theorem is given in Appendix C.

> Theorem: Let the estimate $\Delta\hat{\underline{x}}_k$ be defined so that the error $\underline{\epsilon}_k$ is orthogonal to the observations $\underline{x}_{k-1}$; i.e., $E[\underline{\epsilon}_k \underline{x}_{k-1}^T]$ is a matrix of zeros. Then $\Delta\hat{\underline{x}}_k$ minimizes Eq. (26) and the minimum mean-square error is given by
>
> $$\mathcal{E}_{min}^2 = E[\underline{\epsilon}_k^T \Delta\underline{x}_k] \quad . \tag{29}$$

This theorem can be used to show that the optimal prediction matrix G_k is given by Eq. (11). The theorem requires that

$$E[\underline{\epsilon}_k \underline{x}_{k-1}^T] = \begin{bmatrix} 0 \end{bmatrix} \quad . \tag{30}$$

By substituting Eqs. (27) and (28) in Eq. (30) and employing the definitions in Eq. (9b), one has

$$E[(\Delta\underline{x}_k - G_k^T \underline{x}_{k-1})\, \underline{x}_{k-1}^T] = B_k^T - G_k^T K_{k-1} = \begin{bmatrix} 0 \end{bmatrix} \tag{31}^*$$

from which Eq. (11) follows directly. By substituting the same equations in Eq. (29), one can obtain an expression for the minimum mean-square error:

$$\begin{aligned} \mathcal{E}_{min}^2 &= E[\underline{\epsilon}_k^T \Delta\underline{x}_k] = E[\mathrm{tr}\, \underline{\epsilon}_k \Delta\underline{x}_k^T] \\ &= E[\mathrm{tr}(\Delta\underline{x}_k - G_k^T \underline{x}_{k-1})\, \Delta\underline{x}_k^T] \\ &= \mathrm{tr}(D_k - G_k^T B_k) \quad . \end{aligned} \tag{32}$$

* Equation (31) represents the Yule-Walker equations in statistical estimation[5,8] or the normal equations in regression analysis.[8]

Then, by virtue of Eqs. (11) and (12), Eq. (32) reduces to

$$\mathcal{E}^2_{\min} = \operatorname{tr} E_k \quad . \tag{33}$$

That is, the minimum mean-square error is the trace of the conditional covariance matrix.

The structure of the sequential classifier can be depicted as in Fig. 4. The classifier consists of a pair of (discrete-time) linear mean-square predictive filters each of which is designed to estimate one of the processes to be classified. The two resulting error processes $\underline{\epsilon}^1_k$ and $\underline{\epsilon}^2_k$ are then fed into a box that implements Eqs. (18) and which represents a quadratic classifier

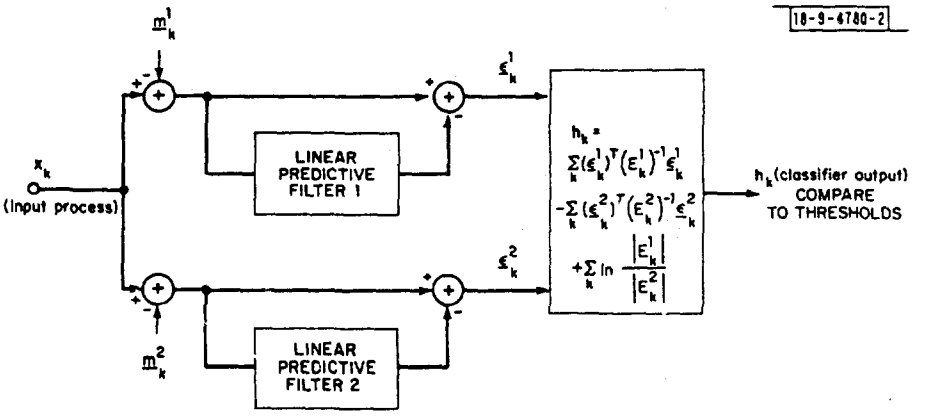


Fig. 4. Structure of sequential classifier.

for the two error processes that result when a signature of each class is applied to its corresponding filter. That the quadratic classifier for the error processes can be expressed as a sum of outputs from the incremental classifier as in Eq. (18a) is a result of the fact that the error $\underline{\epsilon}_k$ is uncorrelated with the previous errors $\underline{\epsilon}_{k-1}$, $\underline{\epsilon}_{k-2}$, etc. This fact in turn follows from the orthogonality principle expressed in the theorem. The classifier is optimal for Gaussian input processes and will be further interpreted for general random processes in Sec. IV-C. However, let us first consider how the matrix G_k that defines the linear-predictive filter can be more easily computed when the processes are univariate (i.e., when the signatures consist of the PP amplitude only) and stationary, and the number of pulses observed per stage is equal to one.

B. Recursion Relations for Classifier Parameters

A set of recursion relations for the computation of the classifier parameters was described in Sec. II. The recursion consists of Eqs. (10) through (12) with starting value $E_1 = K_1$. For the special case when the processes to be classified are univariate, and a single time sample (radar return) is observed at each stage, we have $n\Delta p = 1$. The matrices B_k and G_k defined in Sec. II then become column matrices or "vectors" and the matrices D_k and E_k degenerate to scalars. For this special case, the variables will be represented by lowercase symbols

$\underline{b}_k$, $\underline{g}_k$, d_k, and e_k. The recursion relations Eqs. (10) through (12) can be written in this notation as

$$K_k^{-1} = \left[\begin{array}{c|c} K_{k-1}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] + \frac{1}{e_k} \left[\begin{array}{c} -\underline{g}_k \\ \hline 1 \end{array}\right] \left[-\underline{g}_k^T \;\vdots\; 1\right] \tag{34}$$

$$\underline{g}_k = K_{k-1}^{-1} \underline{b}_k \tag{35}$$

$$e_k = d_k - \underline{b}_k^T K_{k-1}^{-1} \underline{b}_k \quad . \tag{36}$$

If, further, the random processes are stationary, the covariance matrix has the special Toeplitz form

$$K_k = \begin{bmatrix} R(0) & R(1) & \cdots & R(k-1) \\ R(1) & R(0) & & \\ & & \ddots & \\ & & & R(0) \; R(1) \\ R(k-1) & & & R(1) \; R(0) \end{bmatrix} \tag{37}$$

where R(i) represents the correlation function for the process evaluated at lag i.* If Eq. (37) is compared with Eq. (9b), one can observe that

$$\underline{b}_k = \begin{bmatrix} R(k-1) \\ \vdots \\ R(1) \end{bmatrix} \tag{38}$$

and that

$$d_k = R(0) \quad . \tag{39}$$

Let us now define the reversal of a vector as another vector whose components are the components of the original vector in reverse order. For the vector $\underline{b}_k$, one has

$$\tilde{\underline{b}}_k = \text{rev } \underline{b}_k = \begin{bmatrix} R(1) \\ \vdots \\ R(k-1) \end{bmatrix} \tag{40}$$

*Observe that since the process is zero-mean the covariance is equal to the correlation.

from which it follows that

$$\tilde{\underline{b}}_k = \begin{bmatrix} \tilde{\underline{b}}_{k-1} \\ \hline R(k-1) \end{bmatrix} \quad . \tag{41}$$

In order to derive a simplified set of recursion relations, Eq. (35) will now be written in a form involving the reversals of $\underline{g}_k$ and $\underline{b}_k$. Equation (35) can be written in the equivalent form

$$K_{k-1}\,\underline{g}_k = \underline{b}_k \quad . \tag{42}$$

If both vectors in Eq. (42) are replaced by their reversals, then in order to maintain a true relation, the matrix K_{k-1} must be replaced by one obtained from itself by reflecting it first about the main diagonal and then about the reverse diagonal. However, because K_{k-1} has the special form [Eq. (37)] (i.e., it is symmetric and Toeplitz), these two operations leave it unchanged. Therefore, one has the relation

$$K_{k-1}\,\tilde{\underline{g}}_k = \tilde{\underline{b}}_k \tag{43}$$

and thus

$$\tilde{\underline{g}}_k = K_{k-1}^{-1}\,\tilde{\underline{b}}_k \quad . \tag{44}$$

The desired simplified recursion now follows directly, because from Eqs. (44), (34), and (41) one can write

$$\begin{aligned} \tilde{\underline{g}}_k &= \left[\begin{bmatrix} K_{k-2}^{-1} & 0 \\ \hline 0 & 0 \end{bmatrix} + \frac{1}{e_{k-1}} \begin{bmatrix} -\underline{g}_{k-1} \\ \hline 1 \end{bmatrix} \begin{bmatrix} -g_{k-1}^T & 1 \end{bmatrix} \right] \cdot \begin{bmatrix} \tilde{\underline{b}}_{k-1} \\ \hline R(k-1) \end{bmatrix} \\ &= \begin{bmatrix} K_{k-2}^{-1} & 0 \\ \hline 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\underline{b}}_{k-1} \\ \hline R(k-1) \end{bmatrix} + \begin{bmatrix} -\underline{g}_{k-1} \\ \hline 1 \end{bmatrix} \cdot \frac{[R(k-1) - \underline{g}_{k-1}^T \tilde{\underline{b}}_{k-1}]}{e_{k-1}} \quad . \end{aligned} \tag{45}$$

By employing Eq. (44) once again and taking the reversal of the result, we obtain

$$\underline{g}_k = \begin{bmatrix} 0 \\ \hline \underline{g}_{k-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \hline -\tilde{\underline{g}}_{k-1} \end{bmatrix} \rho_{k-1} \tag{46}$$

where

$$\rho_k = [R(k) - \underline{g}_k^T \tilde{\underline{b}}_k]/e_k \tag{47}$$

and where from Eqs. (35), (36), and (39) we have

$$e_k = R(0) - \underline{g}_k^T \underline{b}_k \quad . \tag{48}$$

Equations (46) through (48) represent a concise form of the recursion relations to compute the filter coefficients for a univariate stationary process. These are identical to the recursion relations derived by Levinson and others.[7-14]

Most of the literature related to linear-predictive filtering and the associated statistical models defines the vector of filter coefficients as the reversal of $\underline{g}_k$. A comparison of our terminology to that which is most common in the literature is made in Table IV. Equations (46) through (48) can be written in the alternative terminology as

$$\underline{a}_{p+1} = \begin{bmatrix} \underline{a}_p \\ \hline 0 \end{bmatrix} + \begin{bmatrix} -\tilde{\underline{a}}_p \\ \hline 1 \end{bmatrix} \rho_{p+1} \tag{49}$$

$$\rho_{p+1} = [R(p+1) - \tilde{\underline{a}}_p^T \underline{r}_p]/e_{p+1} \tag{50}$$

$$e_{p+1} = R(0) - \underline{a}_p^T \underline{r}_p \quad . \tag{51}$$

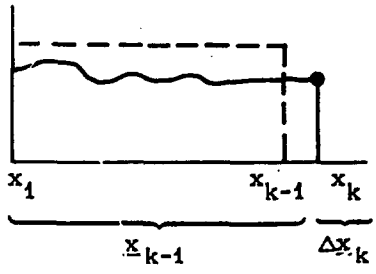
The variable ρ_k is the so-called reflection coefficient in linear-prediction theory or partial correlation coefficient in statistics. An engineering interpretation of the reflection coefficient is given in Refs. 12 and 14. By substituting Eq. (34) in Eq. (36), one can show that

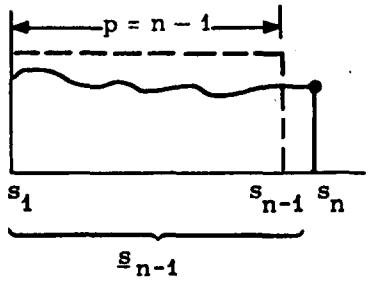
$$e_k = e_{k-1}(1 - \rho_{k-1}^2) \tag{52}$$

which can be used as an alternative to Eq. (48) or Eq. (51) in the recursion. Since, by virtue of Eq. (33), e_k represents the mean-square error in prediction of the process using $k-1$ previous observations, e_k will always be greater than or equal to zero. Equation (52) thus implies that $|\rho_k|$ must be less than or equal to one, which in turn implies that the mean-square error is a decreasing function of k. This is the key justification for the use of the sequential classifier in the extended mode. When a number of decision stages k_s or a prediction order $k_s - 1$ is reached such that the mean-square error "levels off" and/or approaches zero, then there is no need to consider any higher orders of prediction. In terms of classical statistical models one can say that the signatures are adequately modeled by an autoregressive process of order $k_s - 1$. Autoregressive processes are discussed in various references in the statistical literature (e.g., Refs. 5 and 8); such processes are represented exactly by a weighted sum of some number p of their past values plus white noise. Equation (52) implies that since the mean-square error of prediction of a process is a decreasing function of k, one can model an arbitrary discrete-time random process to any desired degree of accuracy by an autoregressive process of finite order p. This implication can be further interpreted when the results are presented from a spectral analysis point of view.

TABLE IV

COMPARISON OR TERMINOLOGY FOR SEQUENTIAL CLASSIFIER TO THAT USED FOR LINEAR-PREDICTIVE OR AUTOREGRESSIVE MODELS





$$\Delta\hat{\underline{x}}_k = \sum_{j=1}^{k-1} (\underline{g}_k)_j x_j = \underline{g}_k^T \underline{x}_{k-1}$$

$$\hat{s}_n = \sum_{i=1}^{p} (\underline{a}_p)_i s_{n-i} = \underline{a}_p^T \tilde{\underline{s}}_{n-1} = \tilde{\underline{a}}_p^T \underline{s}_{n-1}$$

$$K_k = \begin{bmatrix} R(0) & R(1) & \cdots & R(k-1) \\ R(1) & R(0) & \cdots & \\ & & & \\ R(k-1) & & \cdots & R(0) \end{bmatrix}$$

$$R_p = \begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(0) & \cdots & \\ & & & \\ R(p-1) & & \cdots & R(0) \end{bmatrix}$$

$$\underline{b}_k = \begin{bmatrix} R(k-1) \\ R(k-2) \\ \vdots \\ R(1) \end{bmatrix}$$

$$\underline{r}_p = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}$$

$$\underline{b}_k = K_{k-1} \underline{g}_k$$

$$\underline{r}_p = R_p \underline{a}_p$$

$$\underline{g}_k = K_{k-1}^{-1} \underline{b}_k$$

$$\underline{a}_p = R_p^{-1} \underline{r}_p$$

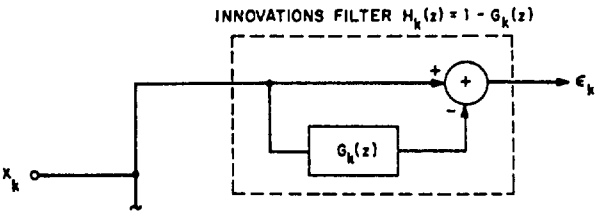


Fig. 5. Prediction branch of sequential classifier.

C. Relations to Spectral Analysis

Figure 5 depicts one branch of the sequential classifier. The linear-predictive filter has been represented in the frequency domain by its z transform

$$G_k(z) = \underline{g}_k^T \begin{bmatrix} z^{-k+1} \\ z^{-k+2} \\ \cdot \\ \cdot \\ \cdot \\ z^{-1} \end{bmatrix} \tag{53}$$

where z is the discrete-time frequency variable equal to $e^{j2\pi fT}$ and T is the time between samples of the process (the radar pulse repetition interval). Note that when the input x_k is the process which the filter was designed to predict, the output process ϵ_k is uncorrelated, i.e.,

$$E[\epsilon_k \epsilon_{k-j}] = \begin{cases} e_k \delta_{k,k-j} & ; \quad j \leq k \leq k_s \\ e_{k_s} \delta_{k,j} & ; \quad k > k_s \end{cases} . \tag{54}$$

This fact has been discussed extensively in the literature and follows directly from the theorem in Sec. IV-A. During the period when the classifier is used in the normal mode, the filter is time varying (i.e., the coefficients $\underline{g}_k$ depend on the time index k) and the error has a flat spectrum whose level is e_k. When the classifier is used in the extended mode, the filter becomes time invariant with prediction coefficients $\underline{g}_{k_s}$ and the error process ϵ_k becomes (to any desired degree of approximation) a white process with spectral level e_{k_s}. If the input process x_k is Gaussian, then the error process is also Gaussian and is referred to as the innovations process of x_k. The filter $H(z) = 1 - G(z)$ (shown within dotted lines in Fig. 5) that generates the error from the input is called the innovations filter. The innovations process is defined as a white Gaussian process obtained through a causal and causally invertible filter[15] [this is clearly the case for H(z)]. The importance of the innovations filter in general is that it can be used in the inverted form to generate the input process from white Gaussian noise. This technique has been used extensively for the artificial generation of speech,[12-14] and has recently been studied in connection with the generation of ei nbles of radar signatures to approximate live flight test data.[16]

The innovations filter provides an approach to obtaining an estimate of the spectrum of the input process. Since the output process is white noise with spectral level e_k, one has

$$\hat{S}_x^k(f) = \frac{e_k}{|1 - G_k(e^{j2\pi fT})|^2} \tag{55}$$

which implies that the process x_k can be represented to any desired degree of approximation by a system function with k poles. Equation (55) is identical to the so-called maximum entropy spectral estimate.[17-19] The maximum entropy spectral estimate is the spectrum of the "most random" Gaussian process (i.e., that with maximum entropy) whose correlation function matches the correlation function of the process x_k on the lag interval $[-k + 1, k - 1]$ (Refs. 11 and 19).

Another spectral estimate that will be of interest is the maximum likelihood spectral estimate $\hat{\hat{S}}_x^k(f)$ (Refs. 20 and 21) which can be interpreted as the power output of a narrowband causal filter of duration k centered at frequency f that minimizes the power due to frequencies other than f (Ref. 21). It can be shown that the maximum likelihood and the maximum entropy spectral estimates are related[11,22] as

$$\frac{1}{\hat{\hat{S}}_x^k(f)} = \sum_{j=1}^{k} \frac{1}{\hat{S}_x^j(f)} \quad . \tag{56}$$

The spectral analysis methods provide an important interpretation of the classifier. In particular, from the definition of the power spectral density for discrete time signals,[23] one can write

$$E[(\epsilon_k)^2] = T\int_{-1/2T}^{1/2T} S_\epsilon^k(f)\, df = T\int_{-1/2T}^{1/2T} |1 - G_k(e^{j2\pi fT})|^2\, S_x(f)\, df \tag{57}$$

where $S_x(f)$ is the true spectrum of the input. Observe that Eq. (57) holds regardless of the input process; it is not necessary to assume that the input is chosen to make ϵ_k a white process. The classifier output h_k, given recursively by Eqs. (18), can be expressed using Eqs. (19a) and (15) as

$$h_k = \sum_{j=1}^{k} \frac{(\epsilon_j^1)^2}{e_j^1} - \sum_{j=1}^{k} \frac{(\epsilon_j^2)^2}{e_j^2} + \ln \frac{|K_k^1|}{|K_k^2|} \quad ; \quad k \leqslant k_s \quad . \tag{58}$$

When the classifier is used in the extended mode, its output is given by

$$h_k = h_{k_s} + \frac{1}{e_{k_s}^1} \cdot \sum_{j=k_s+1}^{k} (\epsilon_j^1)^2 - \frac{1}{e_{k_s}^2} \cdot \sum_{j=k_s+1}^{k} (\epsilon_j^2)^2 + (k - k_s) \ln\left(\frac{e_{k_s}^1}{e_{k_s}^2}\right) \quad ; \quad k > k_s \quad . \tag{59}$$

By applying Eqs. (55) through (57) to Eqs. (58) and (59), one can express the mean value of the classifier output as

$$E[h_k] = T\int_{1/2T}^{1/2T} \frac{S_x(f)}{\hat{\hat{S}}_1^k(f)}\, df - T\int_{1/2T}^{1/2T} \frac{S_x(f)}{\hat{\hat{S}}_2^k(f)}\, df + \ln \frac{|K_k^1|}{|K_k^2|} \quad ; \quad k \leqslant k_s \tag{60}$$

and

$$E[h_k] = E[h_{k_s}] + (k - k_s)\left[T\int_{-1/2T}^{1/2T} \frac{S_x(f)}{\hat{S}_1^{k_s}(f)}\,df - T\int_{-1/2T}^{1/2T} \frac{S_x(f)}{\hat{S}_2^{k_s}(f)}\,df + \ln\left(\frac{e_{k_s}^1}{e_{k_s}^2}\right)\right] \quad ; \quad k > k_s \tag{61}$$

where $\hat{\hat{S}}_1^k(f)$ and $\hat{\hat{S}}_2^k(f)$ are the maximum likelihood spectral estimates and $\hat{S}_1^k(f)$ and $\hat{S}_2^k(f)$ are the maximum entropy spectral estimates for the two processes to be identified. Equations (60) and (61) show how the spectra of the processes are employed by the sequential classifier to perform the classification. When the classifier is used in the normal mode [Eq. (60)], the mean classifier output involves a comparison of the spectrum of the input process to the maximum likelihood spectral estimates of each of the two classes of processes to be identified. When the classifier is used in the extended mode [Eq. (61)], the mean output involves, in addition, a comparison of the input process spectrum to the maximum entropy spectral estimates of the processes. Clearly, if the classifier is used in the extended mode for a large number of decision stages $k >> k_s$, the mean classifier output will depend more on the maximum entropy than on the maximum likelihood spectral estimates. Since, in general, the maximum entropy estimate provides a higher degree of resolution than the maximum likelihood estimate, use of the classifier in the extended mode should be more sensitive to the fine structure of the spectra. On the other hand, since the maximum likelihood method tends to produce a smoother spectral estimate whose amplitude more accurately represents the true power level, use of the classifier in the normal mode should be more sensitive to the overall power level of the spectra. In addition, if the random processes are ergodic, then for $k >> k_s$, the actual value of the classifier output becomes proportional to the difference of the integrated spectral quotients expressed in Eq. (61). That is

$$h_k \approx (k - k_s)\; T\int_{-1/2T}^{1/2T} \frac{S_x(f)}{\hat{S}_1^{k_s}(f)}\,df - T\int_{-1/2T}^{1/2T} \frac{S_x(f)}{\hat{S}_2^{k_s}(f)}\,df + \ln\left(\frac{e_{k_s}^1}{e_{k_s}^2}\right) \quad ; \quad k >> k_s \quad . \tag{62}$$

To see this, observe that if the processes are ergodic, then for $k >> k_s$

$$\sum_{j=k_s+1}^{k} (\epsilon_j^i)^2 \rightarrow (k - k_s)\, E[(\epsilon_j^i)^2] \quad . \tag{63}$$

By using Eq. (63) in Eq. (59) and applying Eqs. (55) and (57), one obtains the desired result [Eq. (62)].

V. SUMMARY

This report describes a sequential form of the quadratic classifier (the optimal decision rule for Gaussian random processes) and its application to the discrimination and classification of radar signatures. A special formulation of the classifier shows that each decision stage consists of two steps; namely (a) linear prediction of observations to be made at the given stage and

(b) application of the error between the actual and predicted observations to an incremental classifier to make a classification decision.

The storage and computational requirements for the sequential form of classifier are the same as those for a non-sequential quadratic classifier when both are applied to signatures with the same number of returns, and classification results are identical. However, the sequential classifier has the ability to classify signatures by observing only a relatively small number of returns and so can provide a definite real-time advantage.

In addition, the sequential classifier can be operated in an "extended mode" to classify signatures with more returns than those used to design the classifier. As a result, the sequential quadratic classifier can approximate the performance of a non-sequential classifier at a fraction of the storage and computational requirements. Since the sequential classifier rapidly classifies easily identifiable signatures and extends its observation time on more difficult signatures, it makes better use of the radar resources and can achieve a higher overall level of performance.

The sequential classifier as formulated here is intimately related to estimation-theoretic results in mean-square filtering, autoregressive time series analysis, and innovations process concepts. In particular, the linear prediction step produces the optimum linear mean-square estimate of the signature based on the previous observations regardless of the process statistics; when the input to the classifier is a signature from one of the classes that it was designed to recognize, then the prediction error process for that signature is white noise. The incremental classifier is the basic recursive component in a classifier designed to optimally discriminate the two white Gaussian noise processes.

The classifier can be related to some recent results in spectral estimation. These relations provide an interpretation of the classifier in the frequency domain.

APPENDIX A
COMPUTATION OF THE CONDITIONAL DENSITY FOR GAUSSIAN OBSERVATIONS

The purpose of this appendix is to show that the density for $\Delta\underline{x}_{k-1}$ conditioned on $\underline{x}_{k-1}$ is Gaussian and given by Eq. (16) whenever the observation vector $\underline{x}_k$ is Gaussian. To show this, it is necessary to prove Eqs. (13) and (15).

To prove Eq. (13), one first obtains the inverse of the covariance matrix in partitioned form by an adoption of the well-known Gauss-Jordan technique[24] to partitioned matrices. One starts with the relation

$$K_k K_k^{-1} = I \tag{A-1}$$

and performs certain "elementary operations" on both sides of the equation in order to reduce K_k to the identity matrix. At the j^{th} step in the process one has the relation

$$R^{(j)} K_k^{-1} = A^{(j)} \tag{A-2}$$

where $R^{(j)}$ is the reduced covariance matrix after performing some number of elementary operations, and $A^{(j)}$ is the matrix resulting from performing those same operations on the identity matrix. When $R^{(j)}$ has been reduced to the identity matrix, $A^{(j)}$ is the desired inverse.

Table A-I lists the steps involved in deriving the partitioned inverse covariance matrix. First, the upper row of partitions is multiplied by K_{k-1}^{-1}, then the upper row is multiplied by B_k^T and subtracted from the lower row, and so on. The resulting partitioned inverse is shown in the last step and repeated here using the matrices defined by Eqs. (11) and (12).

$$K_k^{-1} = \left[\begin{array}{c|c} K_{k-1}^{-1} + G_k E_k^{-1} G_k^T & -G_k E_k^{-1} \\ \hline -E_k^{-1} G_k^T & E_k^{-1} \end{array}\right] . \tag{A-3}$$

Equation (A-3) can be written in the equivalent form

$$K_k^{-1} = \left[\begin{array}{c|c} K_{k-1}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] + \left[\begin{array}{c} -G_k \\ \hline I \end{array}\right] E_k^{-1} \left[\begin{array}{c|c} -G_k^T & I \end{array}\right] . \tag{A-4}$$

By applying Eq. (5b) to Eq. (8) and employing Eqs. (A-4), (3), and (9a), one arrives at the result

TABLE A-I

OPERATIONS TO DERIVE PARTITIONED INVERSE OF COVARIANCE MATRIX

Step j	$R^{(j)}$	$A^{(j)}$
I	$\begin{bmatrix} K_{k-1} & B_k \\ B_k^T & D_k \end{bmatrix}$	$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$
II	$\begin{bmatrix} I & K_{k-1}^{-1} B_k \\ B_k^T & D_k \end{bmatrix}$	$\begin{bmatrix} K_{k-1}^{-1} & 0 \\ 0 & I \end{bmatrix}$
III	$\begin{bmatrix} I & K_{k-1}^{-1} B_k \\ 0 & E_k \end{bmatrix}$	$\begin{bmatrix} K_{k-1}^{-1} & 0 \\ -B_k^T K_{k-1}^{-1} & I \end{bmatrix}$

where $E_k = D_k - B_k^T K_{k-1}^{-1} B_k$

IV	$\begin{bmatrix} I & K_{k-1}^{-1} B_k \\ 0 & I \end{bmatrix}$	$\begin{bmatrix} K_{k-1}^{-1} & 0 \\ -E_k^{-1} B_k^T K_{k-1}^{-1} & E_k^{-1} \end{bmatrix}$
V	$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$	$\begin{bmatrix} K_{k-1}^{-1} + K_{k-1}^{-1} B_k E_k^{-1} B_k^T K_{k-1}^{-1} & -K_{k-1}^{-1} B_k E_k^{-1} \\ -E_k^{-1} B_k^T K_{k-1}^{-1} & E_k^{-1} \end{bmatrix}$

$$\Delta Q_k = (\underline{x}_k - \underline{m}_k)^T K_k^{-1}(\underline{x}_k - \underline{m}_k) - (\underline{x}_{k-1} - \underline{m}_{k-1})^T K_{k-1}^{-1}(\underline{x}_{k-1} - \underline{m}_{k-1})$$

$$= \left[(\underline{x}_{k-1} - \underline{m}_{k-1})^T \;\vdots\; (\Delta\underline{x}_k - \Delta\underline{m}_k)^T\right] \begin{bmatrix} -G_k \\ \cdots\cdots \\ I \end{bmatrix} E_k^{-1} \left[-G_k^T \;\vdots\; I\right] \begin{bmatrix} \underline{x}_{k-1} - \underline{m}_{k-1} \\ \cdots\cdots\cdots\cdots \\ \Delta\underline{x}_k - \Delta\underline{m}_k \end{bmatrix}$$

$$= [-G_k^T(\underline{x}_{k-1} - \underline{m}_{k-1}) + (\Delta\underline{x}_k - \Delta\underline{m}_k)]^T E_k^{-1} [-G_k^T(\underline{x}_{k-1} - \underline{m}_{k-1}) + (\Delta\underline{x}_k - \Delta\underline{m}_k)] \tag{A-5}$$

which proves Eq. (13).

To prove Eq. (15), note from Eq. (A-2) that at any step j the deierminants satisfy

$$|R^{(j)}| \cdot |K_k^{-1}| = |A^{(j)}| \quad . \tag{A-6}$$

If one evaluates this at the third or fourth step in Table A-I, one can obtain the desired result. In particular, for j = III one has

$$|E_k| \cdot |K_k^{-1}| = |K_{k-1}^{-1}| \tag{A-7}$$

which proves Eq. (15). Equation (15) when substituted in Eq. (7) leads to the conditional density given by Eq. (16). The additional result, Eq. (13), shows that this density is Gaussian.

APPENDIX B
PARTITIONED SYMMETRIC STORAGE

The parameter matrices P^1 and P^2 of Eq. (21) are stored in a special form that will be called "partitioned symmetric" form. The partitioned symmetric storage form permits the partitions of a symmetric matrix to be conveniently accessed for processing in FORTRAN. The storage mode is in the form of a large linear array and is completely compatible with the storage modes used for subroutines in the IBM Scientific Subroutine Package.[6]

To access a partition, one needs to have a pointer (index) to the first word in the partition. Partitions along the main diagonal are stored in symmetric form; that is, only the elements above and including the main diagonal are stored (by column). Partitions off the main diagonal are stored in general form; that is, all their elements are stored (by column). This is illustrated in Fig. B-1. Figure B-2 illustrates the difference between symmetric storage and partitioned symmetric storage for a 6 × 6 matrix with the indicated elements and partitions.

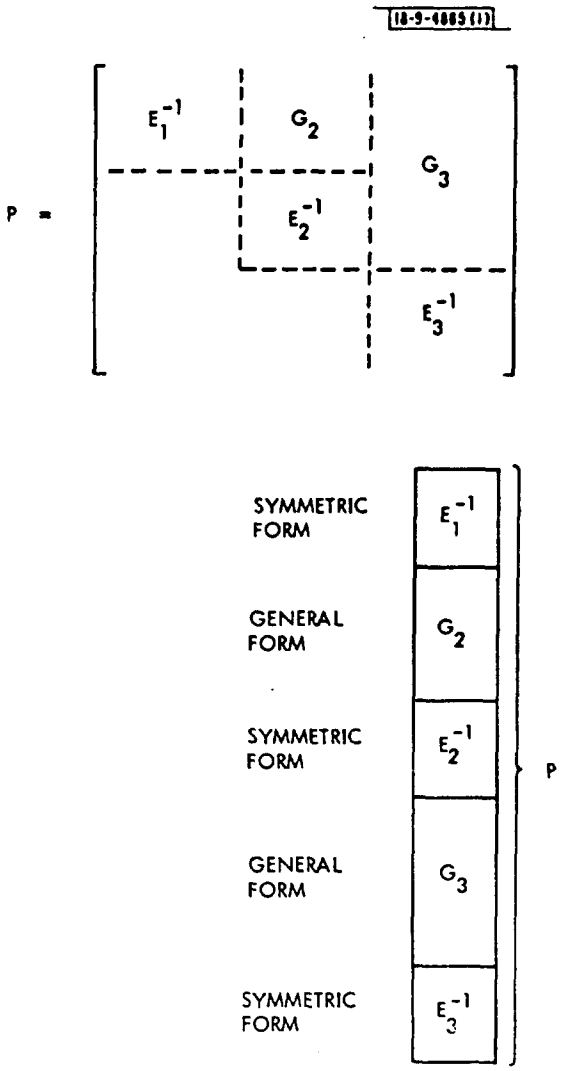


Fig. B-1. Partitioned symmetric storage.

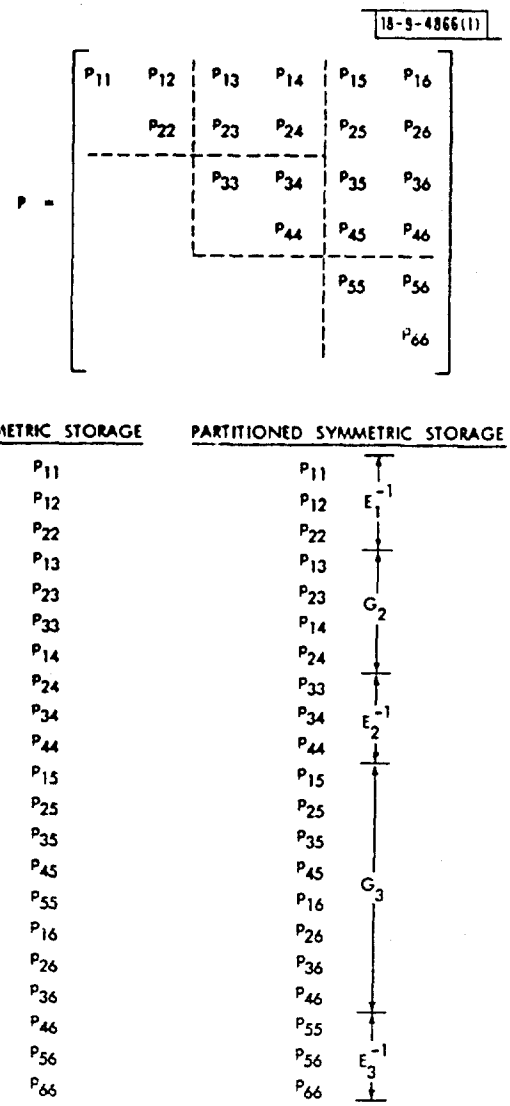


Fig. B-2. Comparison of symmetric and partitioned symmetric storage forms.

APPENDIX C
PROOF OF THE ORTHOGONALITY PRINCIPLE

The proof of the theorem stated in Sec. IV-A is given here. The proof follows that of Papoulis.[25]

Theorem: Let the estimate $\Delta\hat{\underline{x}}_k$ be defined so that the error $\underline{\epsilon}_k$ is orthogonal to the observations $\underline{x}_{k-1}$, i.e., $E[\underline{\epsilon}_k \underline{x}_{k-1}^T]$ is a matrix of zeros. Then $\Delta\hat{\underline{x}}_k$ minimizes Eq. (26) and the minimum mean-square error is given by

$$\delta_{min}^2 = E[\underline{\epsilon}_k^T \Delta\underline{x}_k] \quad .$$

Proof: First note that since $\underline{\epsilon}_k$ is orthogonal to $\underline{x}_{k-1}$, $\underline{\epsilon}_k$ is also orthogonal to any linear transformation of $\underline{x}_{k-1}$. That is, since $E[\underline{\epsilon}_k \underline{x}_{k-1}^T] = \begin{bmatrix} 0 \end{bmatrix}$, then $E[\underline{\epsilon}_k (A\underline{x}_{k-1})^T] = \begin{bmatrix} 0 \end{bmatrix}$ for any matrix A. Now let H be any matrix of predictor coefficients and let G be the matrix that results in $\underline{\epsilon}_k$ orthogonal to $\underline{x}_{k-1}$. The prediction error using H is thus given by

$$\Delta\underline{x}_k - H^T\underline{x}_{k-1} = \underline{\epsilon}_k + (G - H)^T\underline{x}_{k-1}$$

where $\underline{\epsilon}_k$ is the error that results when G is used. Since $\underline{\epsilon}_k$ is orthogonal to any linear transformation of $\underline{x}_{k-1}$ [and, in particular, the transformation $(G - H)^T$], the mean-square error is given by

$$E[|\Delta\underline{x}_k - H^T\underline{x}_{k-1}|^2] = E[|\underline{\epsilon}_k|^2] + E[|(G - H)^T\underline{x}_{k-1}|^2]$$

which is minimized for H = G. Further, since $\underline{\epsilon}_k$ is orthogonal to any linear transformation of $\underline{x}_{k-1}$, one has

$$\delta_{min}^2 = E[\underline{\epsilon}_k^T \underline{\epsilon}_k] = E[\underline{\epsilon}_k^T(\Delta\underline{x}_k - G^T\underline{x}_{k-1})] = E[\underline{\epsilon}_k^T \Delta\underline{x}_k]$$

Q.E.D.

THIS

PAGE

IS

MISSING

IN

ORIGINAL

DOCUMENT

REFERENCES

1. A. Wald, Sequential Analysis (Wiley, New York, 1947).
2. K. S. Fu, Sequential Methods in Pattern Recognition and Machine Learning (Academic Press, New York, 1968).
3. I. Selin, "The Sequential Estimation and Detection of Signals in Normal Noise," Part I, Information and Control 7, 512-534 (1964).
4. I. Selin, "The Sequential Estimation and Detection of Signals in Normal Noise," Part II, Information and Control 8, 1-35 (1965).
5. G. E. P. Box and G. M. Jenkins, Time Series Analysis Forecasting and Control (Holden-Day, San Francisco, 1970).
6. System/360 Scientific Subroutine Package Version III Programmer's Manual, IBM Document GH20-0205-4 (International Business Machines Corporation, White Plains, New York, 1970).
7. N. Levinson, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," J. Math. Phys. 25, 261-278 (1947); also Appendix B in N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series (The Technology Press and Wiley, New York, 1957).
8. T. W. Anderson, The Statistical Analysis of Time Series (Wiley, New York, 1958).
9. E. A. Robinson, Statistical Communication and Detection (Hafner Publishing Company, New York, 1967), pp. 274-279.
10. J. Durbin, "The Fitting of Time-Series Models," Int. Stat. Inst., Rev. 28, 233-243 (1960).
11. L. C. Pusey, "High Resolution Spectral Estimates," Technical Note 1975-7, Lincoln Laboratory, M.I.T. (21 January 1975), DDC AD-A006406/3.
12. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE 63, 561-580 (1975).
13. J. D. Markel and A. H. Gray, Jr., "On Autocorrelation Equations as Applied to Speech Analysis," IEEE Trans. Audio Electroacoust. AU-21, 69-79 (1973).
14. E. M. Hofstetter, "An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis," Technical Note 1973-36 (Rev. 1), Lincoln Laboratory, M.I.T. (12 April 1974), DDC AD-777579/4.
15. T. Kailath, "The Innovations Approach to Detection and Estimation Theory," Proc. IEEE 58, No. 5, 680-695 (May 1970).
16. W. H. Schoendorf (private communication).
17. J. P. Burg, "Maximum Entropy Spectral Estimation," 37th Annual International SEG Meeting, Oklahoma City, 31 December 1967.
18. J. P. Burg, "New Concepts in Power Spectra Estimation," 40th Annual International SEG Meeting, New Orleans, 11 November 1970.
19. A. Van Den Bos, "Alternative Interpretation of Maximum Entropy Spectral Analysis," IEEE Trans. Inf. Theory (July 1971).
20. J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," Proc. IEEE 57, 1408-1418 (1969), DDC AD-696880.

21. R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics 36, 661-675 (1971), DDC AD-734104.

22. J. P. Burg, "The Relationship Between Maximum Entropy and Maximum Likelihood Spectra," Geophysics 37, 375-376 (1972).

23. L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing (Prentice-Hall, Englewood Cliffs, New Jersey, 1975).

24. F. B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, Englewood Cliffs, New Jersey, 1965).

25. A. Papoulis, Probability, Random Variables and Stochastic Processes (McGraw-Hill, New York, 1965).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

(19) **REPORT DOCUMENTATION PAGE**

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) ESD-TR-76-73	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)* (6) The Application of Linear Prediction to Sequential Classification of Radar Target Signatures	5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report
6. PERFORMING ORG. REPORT NUMBER Technical Report 517	
7. AUTHOR(s) (10) Charles W. Therrien	8. CONTRACT OR GRANT NUMBER(s) (15) F19628-76-C-0002
(14) TR-517	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Project No. 8X363304D215
11. CONTROLLING OFFICE NAME AND ADDRESS Ballistic Missile Defense Program Office Department of the Army 1320 Wilson Boulevard Arlington, VA 22209	12. REPORT DATE (11) 25 Mar 1976
	13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Electronic Systems Division Hanscom AFB Bedford, MA 01731 (12) 39p.	15. SECURITY CLASS. *(of this report)* Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

(16) DA-8-X-363304-D-215

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

18. SUPPLEMENTARY NOTES

None

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

radar target discrimination	maximum entropy spectral estimate	sensor target discrimination
mean-square filtering	linear prediction	stochastic process discrimination
sequential discrimination	Gaussian classifier	sequential detection
innovation filter	quadratic classifier	discrete-time Wiener filtering
innovations process	sequential probability ratio test	autoregressive time series
maximum likelihood spectral estimate		general binary detection

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

A linear-predictive form of the quadratic classifier (the optimal decision rule for Gaussian random processes) is developed and applied to the discrimination and classification of radar target signatures. The classifier was devised to implement a sequential probability ratio test (SPRT); that is, consecutive radar returns are observed until the target can be classified with a prescribed probability of error. Because of the linear-predictive formulation, the computational and storage requirements for the classifier are related only to the number of returns necessary to predict the signature and not to the length of signature observed; a classifier with modest storage and computational requirements can be employed to classify signatures consisting of an arbitrarily large number of radar returns. The classifier is related to several results in mean-square filtering theory and has an interpretation in terms of the maximum entropy and maximum likelihood spectral estimates for the target signatures.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

207 650