

Index Terms—Incomplete gamma function, interleaved memory, random mapping.

In an n way interleaved memory, the effective bandwidth (BW) depends on the average number of concurrently active modules. Hellerman [2] has proposed a theoretical model in order to estimate the effect of n way interleaving for large n . We shall show in this correspondence that this model is equivalent to another which has occurred in studies of epidemics, and of random number generators, and in other statistical applications. We also present a simple "algebra" for dealing with sums of the type Hellerman encountered.

The theoretical model for interleaved memory assumes that a sequence of memory requests $a_1, a_2, a_3, \dots$ is made at high speed, where each a_i is a uniformly distributed random integer between 1 and n , independent of the other a_j . We find the smallest m such that $a_m = a_j$ for some $j < m$, and then the first $m - 1$ memory requests are serviced. The process continues on the next cycle by considering the sequence $a_m, a_{m+1}, a_{m+2}, \dots$ in the same way. Thus the model assumes the following. 1) The memory never idles from lack of work. 2) The time to inspect the requests can be overlapped with the accessing. 3) There is no queuing of requests on busy modules.

Hellerman proved that the average number of active modules is

$$\sum_{k=1}^n \frac{k^2(n-1)!}{(n-k)!n^k}. \quad (1)$$

He tabulated this quantity for $1 \leq n \leq 45$ and suggested that $n^{0.56}$ would be a good approximation based on these data. We shall show that the preceding quantity equals $Q(n)$, a well-studied function whose value is

$$\left(\frac{\pi n}{2}\right)^{1/2} - \frac{1}{3} + \frac{1}{12} \left(\frac{\pi}{2n}\right)^{1/2} - \frac{4}{135} n^{-1} + O(n^{-3/2}). \quad (2)$$

The sum in (1) takes the form

$$\frac{1^2}{n} + \frac{2^2}{n} \frac{n-1}{n} + \frac{3^2}{n} \frac{n-1}{n} \frac{n-2}{n} + \dots + \frac{n^2}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{1}{n} \quad (3)$$

when written out term by term. In dealing with such sums it is convenient to define the notation

$$\langle a_0, a_1, a_2, \dots \rangle = a_0 + a_1 \frac{n-1}{n} + a_2 \frac{n-1}{n} \frac{n-2}{n} + \dots \quad (4)$$

regarding n as a fixed positive integer. We will use this notation for infinite sequences $a_0, a_1, a_2, \dots$; but the quantity $\langle a_0, a_1, a_2, \dots \rangle$ is actually independent of $a_n, a_{n+1}, \dots$, since the terms on the right-hand side eventually are all multiplied by zero. Thus there is no problem of convergence, nor do we need to worry about the upper limit of summation.

The bracketed function in (4) is clearly linear in all components

$$\langle a_0, a_1, a_2, \dots \rangle + \langle b_0, b_1, b_2, \dots \rangle = \langle a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots \rangle. \quad (5)$$

Furthermore we have the identity

$$\langle a_0, a_1, a_2, \dots \rangle = a_0 + \langle a_1, a_2, a_3, \dots \rangle - \frac{1}{n} \langle 1a_1, 2a_2, 3a_3, \dots \rangle \quad (6)$$

as is readily verified from the definition. Equation (6) is a useful relation, since it proves, for example, that $\langle 1, 2, 3, 4, \dots \rangle = n$ if we set all the a_i to 1.

Hellerman's sum (3) is

$$\frac{1}{n} \langle 1^2, 2^2, 3^2, \dots \rangle$$

and by setting $a_i = j$ for all j , we have

$$\frac{1}{n} \langle 1^2, 2^2, 3^2, \dots \rangle = \langle 1, 2, 3, 4, \dots \rangle - \langle 0, 1, 2, 3, \dots \rangle = \langle 1, 1, 1, 1, \dots \rangle$$

by (6) and (5). Therefore the average number of memory modules active per cycle is

$$\langle 1, 1, 1, \dots \rangle = 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \dots \quad (7)$$

This is precisely the function $Q(n)$ discussed in some detail in [3]. The asymptotic behavior of this function was investigated by Ramanujan in 1911 [5], who showed that

$$Q(n) = \frac{n!e^n}{2n^n} - \frac{1}{3} - \frac{4}{135} n^{-1} + \frac{8}{2835} n^{-2} + \frac{1}{8505} n^{-3} + O(n^{-4}). \quad (8)$$

The leading term of (8) can be approximated by using Stirling's formula.

The preceding memory-access model is equivalent to another model that has received considerable study, namely, the behavior of a *random mapping* from the set $\{1, 2, \dots, n\}$ into itself, considering all n^n functions to be equally likely. If we start at any random point a_1 and then compute $a_2 = f(a_1)$, $a_3 = f(a_2)$, etc., until $a_m = f(a_{m-1})$ is first equal to a_j for some $j < m$, the sequences $a_1, a_2, \dots, a_m$ have the same property as in the memory-accessing model, since $f(a_1)$, $f(a_2), \dots, f(a_{m-1})$ are independent of each other and of the choice of a_1 . This problem was apparently first studied by Rapoport [6] and later by many other people (see, for example, [1]). The final value of $m - j$ is called the *period* of the random mapping starting at a_1 and the final value of $j - 1$ is called the length of the *tail*. It is not difficult to prove (cf. [4]) that the average tail length is $(Q(n) - 1)/2$ and the average period length is $(Q(n) + 1)/2$; hence we obtain another proof that $Q(n)$ is the average value of $m - 1$.

REFERENCES

- [1] B. Harris, "Probability distributions related to random mappings," *Ann. Math. Statist.*, vol. 31, pp. 1045-1062, 1960.
- [2] H. Hellerman, *Digital Computer System Principles* (Computer Science Ser.), 2nd ed. New York: McGraw-Hill, 1973, p. 245.
- [3] D. E. Knuth, *The Art of Computer Programming*, vol. 1, *Fundamental Algorithms*, 2nd ed. Reading, Mass.: Addison-Wesley, 1973, pp. 112-119.
- [4] —, *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*. Reading, Mass.: Addison-Wesley, 1969, p. 454, ex. 3.1-12.
- [5] S. Ramanujan, "Questions for solution, number 294," *J. Indian Math. Soc.*, vol. 3, p. 128, 1911; also, *J. Indian Math. Soc.*, vol. 4, pp. 151-152, 1912.
- [6] A. Rapoport, "Cycle distributions in random nets," *Bull. Math. Biophys.*, vol. 10, pp. 145-157, 1948.

Eigenvalue Properties of Projection Operators and Their Application to the Subspace Method of Feature Extraction

C. W. THERRIEN

Abstract—The subspace method proposed by Watanabe *et al.* is an important method of feature extraction in pattern recognition. Some new results described here provide a formulation of the subspace method that is particularly easy to implement and overcomes

Manuscript received July 24, 1974; revised January 22, 1975. This work was sponsored by the Army Ballistic Missile Defense Agency under Contract F19628-73-C0002.

The author is with the Massachusetts Institute of Technology Lincoln Laboratory, Lexington, Mass. 02173.

the various numerical difficulties that were inherent in earlier implementations of the method. The present formulation is also important in that it provides the key to the relation of this method to other methods of feature extraction.

Index Terms—Eigenvalue properties, feature extraction, intersection and union of subspaces, orthogonal subspaces, pattern recognition, projection operators, projections.

I. INTRODUCTION

A number of methods for feature extraction in pattern recognition [1]–[6] have been proposed that are based in one way or another on the Karhunen–Loève expansion [6], [7] and thus implicitly on the second moment properties of the class distributions. One method of particular interest was proposed by Watanabe first under the acronym CLAFIC [1] and later (with some refinements) under the acronym MOSS (for Method of Orthogonal SubSpaces) [2], [3]. This method attempts to find a set of subspaces of the observation space that relate to features of the classes and employs so-called orthogonal projection operators to represent the subspaces. A critical issue in the application of the method is the computation of projection operators corresponding to the intersection and union of subspaces. The subspace method, as originally developed, provided only iterative algorithms for the computation of the bilateral intersection of (two) projection operators. The algorithms had to be carried out with extreme caution to avoid numerical instability and this rendered the entire feature extraction method somewhat difficult and computationally expensive to apply.

In this paper some results based on eigenvalue properties are derived that can be applied to compute the intersection or union of any number of projection operators in a single step and completely eliminate the numerical difficulties of the iterative algorithms. Moreover, the results lead to a very streamlined and computationally efficient formulation of the subspace method and provide the key to its relation to other methods of feature extraction [8].

II. EIGENVALUE PROPERTIES OF PROJECTION OPERATORS

Let X be an n -dimensional linear vector space defined over the field of real numbers. A linear transformation P on X is called an orthogonal projection operator if it is both idempotent and symmetric, that is

$$P^2 = P \quad (1)$$

and

$$P^T = P. \quad (2)$$

An orthogonal projection operator corresponds one-to-one with a subspace of X as follows. Let $\mathcal{S}(P)$ denote the range and $\mathcal{N}(P)$ denote the null space of P . Then from (1) and (2) $\mathcal{S}(P)$ and $\mathcal{N}(P)$ are complementary orthogonal subspaces [9], [10]; that is

$$X = \mathcal{S}(P) \oplus \mathcal{N}(P) \quad (3)$$

with

$$\mathcal{S}(P) \perp \mathcal{N}(P). \quad (4)$$

Thus any vector $x \in X$ has a unique representation as the sum of two orthogonal components

$$x = x_p + x_1 \quad (5)$$

where $x_p \in \mathcal{S}(P)$ and $x_1 \in \mathcal{N}(P)$. Observe¹ that

¹ Since $x_1 \in \mathcal{N}(P)$ $Px = Px_p + 0$. Since $x_p \in \mathcal{S}(P)$ $x_p = Py \exists y \in X$. Thus $Px = PPy = Py = x_p$.

$$x_p = Px \quad (6)$$

and that x_p may be viewed geometrically as the orthogonal projection of x into $\mathcal{S}(P)$.

It also follows from (1) through (6) that if $x \in \mathcal{S}(P)$ then $x_p = x$ and so x is an eigenvector of P corresponding to an eigenvalue of 1. Further, if the dimension of $\mathcal{S}(P)$ is m , then any set of orthonormal vectors $\{e_1, e_2, \dots, e_m\}$ that span $\mathcal{S}(P)$ are eigenvectors of P corresponding to eigenvalues of 1, and any set of orthonormal vectors $\{e_{m+1}, e_{m+2}, \dots, e_n\}$ that span $\mathcal{N}(P)$ are eigenvectors of P corresponding to eigenvalues of 0. Thus the eigenvalues of P are all 1 and 0; the eigenvectors $\{e_1, e_2, \dots, e_n\}$ form an orthonormal basis of X ; and P can be expressed [9] as

$$P = \sum_{j=1}^m e_j e_j^T. \quad (7)$$

This shows that $\mathcal{S}(P)$ uniquely defines P as well as *vice versa*; thus one can refer to $\mathcal{S}(P)$ as the subspace “corresponding to” P .

Let $|x|$ represent the Euclidean norm or “magnitude” of x defined by

$$|x| = (x^T x)^{1/2} = \left(\sum_{j=1}^n (x_j)^2 \right)^{1/2} \quad (8)$$

where $(x)_j$ is the j th component of x . Then it follows from (5) that

$$|x_p| \leq |x| \quad (9)$$

with equality only when $x_p = x$. It also follows from (5) and (6) that the squared magnitude of the vector projection in $\mathcal{S}(P)$ is given by

$$|x_p|^2 = x^T P x. \quad (10)$$

The operations of complementation, intersection, and union are defined for projection operators by similar operations on the corresponding subspaces,² that is,

$$\mathcal{S}(\bar{P}) = \overline{\mathcal{S}(P)} = \mathcal{N}(P) \quad (11a)$$

$$\mathcal{S}(P_1 \cap P_2) = \mathcal{S}(P_1) \cap \mathcal{S}(P_2) \quad (11b)$$

$$\mathcal{S}(P_1 \cup P_2) = \mathcal{S}(P_1) \cup \mathcal{S}(P_2). \quad (11c)$$

The following notation will also be used:

$$\mathcal{S}\left(\bigcap_{k=1}^K P_k\right) = \bigcap_{k=1}^K \mathcal{S}(P_k) \equiv \mathcal{S}(P_1) \cap \mathcal{S}(P_2) \cap \dots \cap \mathcal{S}(P_K) \quad (12a)$$

$$\mathcal{S}\left(\bigcup_{k=1}^K P_k\right) = \bigcup_{k=1}^K \mathcal{S}(P_k) \equiv \mathcal{S}(P_1) \cup \mathcal{S}(P_2) \cup \dots \cup \mathcal{S}(P_K). \quad (12b)$$

It can be shown [3] that the subspaces, and thus the projection operators satisfy the law of dualization:

$$\bigcup_{k=1}^K P_k = \overline{\bigcap_{k=1}^K \bar{P}_k}. \quad (13)$$

Two projection operators P_1 and P_2 are said to be mutually orthogonal (written $P_1 \perp P_2$) if their corresponding subspaces are orthogonal

$$\mathcal{S}(P_1) \perp \mathcal{S}(P_2). \quad (14)$$

Condition (14) holds if and only if P_1 and P_2 satisfy

$$P_1 P_2 = P_2 P_1 = 0. \quad (15)$$

It is easily verified that

² The intersection of two subspaces is the set of vectors contained in both subspaces. The union of two subspaces is the set of all linear combinations of vectors in the two subspaces.

$$\bar{P} = I - P \quad (16)$$

where I is the identity matrix and that $\bar{P} \perp P$. The other operations (11b), (11c), and (12) have traditionally been more difficult to implement computationally. The following procedure is presented as a solution.

Define the generating matrix

$$\Gamma = \sum_{k=1}^K a_k P_k \quad (17)$$

where the P_k are arbitrary projection operators and the a_k are any real numbers satisfying

$$0 < a_k < 1 \quad k = 1, 2, \dots, K \quad (18a)$$

$$\sum_{k=1}^K a_k = 1. \quad (18b)$$

Since Γ is symmetric, its eigenvectors are linearly independent and span X [9]. Accordingly, one can always determine a set of n eigenvectors that are orthonormal (even if Γ has repeated eigenvalues). Further, if λ is an eigenvalue and e is the corresponding (normalized) eigenvector of Γ , then λ can be expressed [see (10)] as

$$\lambda = e^T \Gamma e = \sum_{k=1}^K a_k e^T P_k e = \sum_{k=1}^K a_k |e_{(k)}|^2 \quad (19)$$

where $e_{(k)}$ is the projection of e into $S(P_k)$. Since e has unit magnitude, the terms $|e_{(k)}|^2$ of (19) lie between 0 and 1. As a result the eigenvalue λ lies in the same interval, that is $0 \leq \lambda \leq 1$.

Let the eigenvalues of Γ be ordered such that

$$\begin{aligned} \lambda_j &= 1 & 1 \leq j \leq m \\ 0 < \lambda_j < 1 & m+1 \leq j \leq m' \\ \lambda_j &= 0 & m'+1 \leq j \leq n. \end{aligned} \quad (20)$$

The following theorems show that the eigenvalues of Γ equal to 1 and 0 (if any exist) relate to the intersection and union subspaces, respectively.

Theorem 1: A vector e is contained in $S(\cap_{k=1}^K P_k)$ if and only if it is an eigenvector of Γ corresponding to an eigenvalue of 1.

Proof: $\rightarrow$: If

$$e \in S(\cap_{k=1}^K P_k)$$

then

$$e \in S(P_k), \quad k = 1, 2, \dots, K$$

thus

$$\Gamma e = \sum_{k=1}^K a_k P_k e = \sum_{k=1}^K a_k e = e.$$

$\leftarrow$: If e is an eigenvector of Γ corresponding to an eigenvalue of 1 then

$$e = \Gamma e = \sum_{k=1}^K a_k P_k e = \sum_{k=1}^K a_k e_{(k)}.$$

If there exists any $S(P_k)$ for which $e \notin S(P_k)$, then from (9)

$$|e| \leq \sum_{k=1}^K a_k |e_{(k)}| < \sum_{k=1}^K a_k |e| = |e|.$$

Since the last inequality is strict, this relation is a contradiction. Thus e must be contained in each $S(P_k)$ and so is in the intersection.

Corollary 1: If $\{e_1, e_2, \dots, e_n\}$ is the set of orthonormal eigenvectors of Γ corresponding to eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ satisfying (20), then the eigenvectors e_j , $j = 1, 2, \dots, m$ span $S(\cap_{k=1}^K P_k)$ and $\cap_{k=1}^K P_k$ is given by

$$\cap_{k=1}^K P_k = \sum_{j=1}^m e_j e_j^T. \quad (21)$$

The latter provides a computational procedure for determining the intersection of projection operators. A second theorem which follows directly from the first relates to the union.

Theorem 2: A vector e is in the complement of $S(\cap_{k=1}^K P_k)$ [i.e. $e \in \mathcal{N}(\cap_{k=1}^K P_k)$] if and only if it is an eigenvector of Γ corresponding to an eigenvalue of 0 [i.e., $e \in \mathcal{N}(\Gamma)$].

Proof: $\leftarrow$: From (13) and (16), the complement of the union can be written as

$$\overline{\cup_{k=1}^K P_k} = \cap_{k=1}^K (I - P_k).$$

$$\text{Form } \hat{\Gamma} = \sum_{k=1}^K a_k (I - P_k) = I - \Gamma.$$

By Theorem 1 a vector e is in the complement of the union iff it is an eigenvector of $\hat{\Gamma}$ corresponding to an eigenvalue of 1. But according to the above relation this will be true iff e is an eigenvector of Γ corresponding to an eigenvalue of 0.

Corollary 2: If $\{e_1, e_2, \dots, e_n\}$ is the set of orthonormal eigenvectors of Γ corresponding to eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ satisfying (20), then the eigenvectors e_j , $j = m'+1, m'+2, \dots, n$ span the complement of $S(\cap_{k=1}^K P_k)$ and $\cup_{k=1}^K P_k$ is given by

$$\cup_{k=1}^K P_k = I - \sum_{j=m'+1}^n e_j e_j^T = \sum_{j=1}^{m'} e_j e_j^T. \quad (22)$$

Note that the results of these two theorems and their corollaries do not depend on the particular choice of the $\{a_k\}$. The same intersection and union subspaces are generated for any choice of the $\{a_k\}$ satisfying (18).

The results (21) and (22), can be applied to the subspace method of feature extraction. The application leads to a very simple implementation of the subspace method and eliminates the numerical difficulties inherent in the iterative methods that were previously used to compute the intersection of projection operators.

III. APPLICATION OF RESULTS TO THE SUBSPACE METHOD OF FEATURE EXTRACTION

Assume that the patterns, which are to be classified into one of K distinct classes, are described by vectors in an n -dimensional linear vector space X (the "observation space"). The goal of the subspace method of feature extraction is to develop a set of subspaces of X that relate to important distinguishing characteristics of each class. If P_i' is the projection operator corresponding to such a subspace, then a feature z_i may be defined by

$$z_i = |x_{(i)}|^2 = x^T P_i' x \quad (23)$$

where x is a vector to be classified and $x_{(i)}$ denotes the projection of x into the subspace $S(P_i')$. The subspaces $S(P_i')$ are generated as follows.

An optimal subspace for representation of each class is first generated using the Karhunen-Loève expansion. In particular, if it is desired to represent vectors of Class i with some maximum relative mean square error ϵ ;

$$\epsilon \geq \epsilon_i = \frac{E_i[|x - x_{(i)}|^2]}{E_i[|x|^2]} \quad (24)$$

where E_i denotes expectation using the density of Class i , then the subspace of smallest dimension m_i that achieves condition (24) is spanned by the eigenvectors of the class correlation matrix corresponding to its m_i largest eigenvalues [6]. If ϵ is specified, then m_i can be determined from (24) and the relation

$$\epsilon_i = \left(\sum_{j=i+1}^n \lambda_j^{(i)} \right) / \left(\sum_{j=1}^n \lambda_j^{(i)} \right) \quad (25)$$

where $\lambda_j^{(i)}, j = 1, 2, \dots, n$ are the eigenvalues of the class correlation matrix

$$R_i = E_i[xx^T] \quad (26)$$

assumed to be ordered such that $\lambda_1^{(i)} \geq \lambda_2^{(i)} \geq \dots \geq \lambda_n^{(i)}$. Projection operators $P_i, i = 1, 2, \dots, K$ corresponding to the representation subspaces are then generated using (7).

Feature subspaces are formed from the representation subspaces by removing the intersection of these subspaces. If P_i' represents the projection operator of the i th feature subspace then³

$$P_i' = P_i \cap \left(\bigcap_{k=1; k \neq i}^K \bar{P}_k \right) = P_i \cap \left(\bigcap_{k=1; k \neq i}^K (I - P_k) \right). \quad (27)$$

The projection operator corresponding to the portion of the observation space not included in any of the feature subspaces is given by

$$P_{K+1}' = \bigcup_{i=1}^K P_i' = I - \bigcup_{i=1}^K P_i'. \quad (28)$$

In the final step, vectors in the observation space are transformed via an orthonormal transformation B to

$$x^* = Bx \quad (29)$$

and the projection operators are transformed to

$$P_i^* = BP_i'B^T \quad i = 1, 2, \dots, K+1 \quad (30)$$

with $\mathcal{S}(P_i^*) = \mathcal{S}(P_i)$. The transformation B is chosen such that the P_i^* takes on a particularly simple form with 1's on selected diagonal elements and 0's elsewhere. Thus the computation of the features z_i involves only squares of the components of x^* .⁴

It will now be shown that the results of the previous section lead directly to the generation of the P_i' and the orthonormal transformation B , and to the computation of the features z_i^* . If $P_i, i = 1, 2, \dots, K$ are the projection operators corresponding to the representation subspaces (derived as discussed earlier in this section) then the projection operators for the feature subspaces can be derived using (27) and Corollary 1. Form the generating matrix

$$\Gamma_i = a_i P_i + \sum_{k=1; k \neq i}^K a_k (I - P_k) \quad (31)$$

where the a_k satisfy (18). In order to simplify notation for the ensuing development denote the eigenvectors of Γ_i corresponding to eigenvalues of 1 by $u_j, j = l_{i-1} + 1, l_{i-1} + 2, \dots, l_i$ with $l_0 = 0$. Then, by Corollary 1, the feature space projection operators P_i' for $i = 1, 2, \dots, K$ are given by

$$P_i' = \sum_{j=l_{i-1}+1}^{l_i} u_j u_j^T. \quad (32)$$

The projection operator P_{K+1}' corresponding to the complementary subspace [i.e., that not included in any of the $\mathcal{S}(P_i'), i = 1, 2, \dots, K$] is computed by forming

$$\Gamma_{K+1} = \sum_{i=1}^K a_i P_i'. \quad (33)$$

³ An alternative definition for the feature subspaces is

$$P_i' = P_i \cap \left(\bigcup_{k=1; k \neq i}^K P_i \cap P_k \right).$$

This definition is not the same as (27) since the operation of intersection of subspaces does not distribute over the operation of union [3]. Although both methods have been cited in the earlier literature [2] we shall adhere to definition (27) because of the greater computational simplicity that it affords.

⁴ When the alternative definition for the feature subspaces is used, a similar transformation B that is not orthonormal can be derived, see [3].

Let $u_j, j = l_K + 1, l_K + 2, \dots, n$ denote the eigenvectors of Γ_{K+1} corresponding to eigenvalues of 0. They by (28) and Corollary 2 the projection operator P_{K+1}' is also given by (32) with $l_{K+1} = n$.

Consider now the linear transformation defined by

$$B = \begin{bmatrix} \leftarrow u_1^T \rightarrow \\ \leftarrow u_2^T \rightarrow \\ \vdots \\ \leftarrow u_n^T \rightarrow \end{bmatrix}. \quad (34)$$

Since the subspaces $\mathcal{S}(P_i')$ are orthogonal and complete, the vectors $u_j, j = 1, 2, \dots, n$ satisfy

$$u_j^T u_k = \delta_{jk} \quad (35)$$

where δ_{jk} is the Kronecker delta. It is easy to show that the matrix B of (34) has the properties previously claimed. First note that if (32) and (34) are substituted in (30) then the projection operators P_i' assume the form

$$P_i^* = BP_i'B^T = \sum_{j=l_{i-1}+1}^{l_i} B u_j u_j^T B^T$$

$$= \begin{bmatrix} 0 & & & & & & \\ & 0 & & & & & \\ & & \ddots & & & & \\ & & & 0 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \\ & & & & & & & & 0 \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 0 \end{bmatrix} \begin{matrix} \leftarrow l_{i-1} + 1. \\ \\ \\ \leftarrow l_i \end{matrix} \quad (36)$$

Further, if $x \in \mathcal{S}(P_i')$ then one can write

$$x = \sum_{j=l_{i-1}+1}^{l_i} c_j u_j \quad (37)$$

where

$$c_j = u_j^T x \quad j = l_{i-1} + 1, \dots, l_i. \quad (38)$$

From (29), (34), (35), and (37) the transformed vector x^* is given by

$$x^* = Bx = \sum_{j=l_{i-1}+1}^{l_i} c_j B u_j = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ c_{l_{i-1}+1} \\ \vdots \\ \vdots \\ c_{l_i} \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}. \quad (39)$$

TABLE I
DETECTION PROBABILITY USING SUBSPACE FEATURES

False Alarm Probability	Target 1	Target 2	Target 4	Target 6	Target 7
0.05	0.88	0.92	1.00	0.77	0.60
0.10	0.99	1.00	1.00	0.89	0.77

TABLE II
DETECTION PROBABILITY IN THE OBSERVATION SPACE

False Alarm Probability	Target 1	Target 2	Target 4	Target 6	Target 7
0.05	1.00	0.99	1.00	1.00	0.94
0.10	1.00	1.00	1.00	1.00	0.98

It follows from (36) and (39) that

$$x^* = P_i^* x^*$$

that is, $x^* \in \mathcal{S}(P_i^*)$.

Because of the form (36) of the P_i^* , the computation of the features is almost trivial. In particular (23) reduces to

$$z_i = \sum_{j=i-1+1}^{l_i} |(x^*)_j|^2 \quad i = 1, 2, \dots, K+1. \quad (40)$$

Thus the feature transformation is completely determined by the matrix B and the indices $l_i, i = 1, 2, \dots, K$.

The computational steps involved in applying the subspace method of feature extraction are as follows. First, the correlation matrix for each class is formed. Second, the eigenvectors of the correlation matrix are determined and the projection operators for the representation subspaces are formed using (7). Third, the generating matrices Γ_i are formed and the appropriate eigenvectors are extracted to form the feature space projection operators. Finally, the eigenvectors from which the projection operators were derived are used to form the linear transformation B in terms of which the features are finally expressed.

The major computational advantage of this formulation of the subspace method show up in the third step. The projection operator for each feature subspace is derived by simply forming the appropriate generating matrix and determining its eigenvectors. Moreover, these eigenvectors are precisely those needed for forming the linear transformation B in the last step.

IV. FEATURE EXTRACTION APPLIED TO RADAR SIGNATURE CLASSIFICATION

Results from the application of the subspace method to the classification of radar signatures are reported here. The data to be classified consisted of 300 simulated radar signatures from each of several distinct objects in ballistic trajectories. Each signature was represented in the observation space by a 32-dimensional vector corresponding to 16 principally polarized and 16 orthogonally polarized returns received by the radar. Feature extraction was applied, and the signatures were classified pairwise against the signatures from a single target representing a reentry vehicle (RV).

For the results shown here, an RV and five other target objects were used to derive features. Four nonempty feature subspaces resulted, providing four features. A quadratic classifier was used and performance was evaluated using the "leave-one-out" (L) method [7]. Tables I and II show the performance of the classifier in the feature space and in the observation space respectively. The tables list the probability of detection (probability of correctly classifying an RV) for specific values of the probability of false alarm (probability of classifying a non-RV target as the RV). Performance in

the feature space is reasonable for Targets 1, 2, and 4 considering the reduction in dimensionality. Performance for Targets 6 and 7 (those more difficult to classify) shows considerable degradation.

REFERENCES

- [1] S. Watanabe, P. F. Lambert, C. A. Kulikowski, and R. Walker, "Evaluation and selection of variables in pattern recognition," in *Computer and Information Sciences*, vol. 2, J. T. Tou, Ed. New York: Academic Press, 1967.
- [2] C. A. Kulikowski and S. Watanabe, "Multiclass subspace methods in pattern recognition," in *Proc. Nat. Electron. Conf.*, Chicago, Ill., 1970.
- [3] S. Watanabe and N. Pakvasa, "Subspace method in pattern recognition," in *Proc. 1st Int. Joint Conf. Pattern Recognition*, Washington, D. C., 1973.
- [4] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loeve expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. C-19, pp. 311-318, Apr. 1970.
- [5] J. Kittler and P. C. Young, "A new approach to feature selection based on the Karhunen-Loeve expansion," *Pattern Recognition J.*, vol. 5, pp. 335-352, 1973.
- [6] Y. T. Chien and K. S. Fu, "On the generalized Karhunen-Loeve expansion," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 518-520, July 1967.
- [7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic Press, 1972.
- [8] C. W. Therrien, W. H. Schoendorf, G. L. Carayannopoulos, and B. J. Burdick, "Application of feature extraction to radar signature classification," in *Proc. 2nd Int. Joint Conf. Pattern Recognition*, Lyngby, Denmark, 1974.
- [9] L. A. Zadeh and C. A. Desoer, *Linear System Theory*. New York: McGraw-Hill, 1963.
- [10] K. Hoffman and R. Kunze, *Linear Algebra*. Englewood Cliffs, N. J.: Prentice-Hall, 1961.

A Unified Approach to Feature Selection and Learning in Unsupervised Environments

A. L. LAKSHMINARASIMHAN AND B. V. DASARATHY

Abstract—Here the twin problems of feature selection and learning are tackled simultaneously to obtain a unified approach to the problem of pattern recognition in an unsupervised environment.

Manuscript received November 19, 1973; revised October 13, 1974.

A. L. Lakshminarasimhan is with the School of Automation, Indian Institute of Science, Bangalore, India.

B. V. Dasarathy is with the Computer Sciences Corporation, Huntsville, Ala. 35802.