General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

AFCER 69-1022TR

13

COMMENTS ON LINEAR FEATURE EXTRACTION

By

T. L. Henderson
D. G. Lainiotis

Technical Memorandum No. 8
April 15, 1969

INFORMATION SYSTEMS LABORATORY

ELECTRONICS RESEARCH CENTER
THE UNIVERSITY OF TEXAS AT AUSTIN
Austin, Texas 78712

AF-AFOSR-67-0766E

12

COMMENTS ON LINEAR FEATURE EXTRACTION

By

T. L. Henderson
D. G. Lainiotis

Technical Memorandum No. 8
April 15, 1969

INFORMATION SYSTEMS LABORATORY

ELECTRONICS RESEARCH CENTER
THE UNIVERSITY OF TEXAS AT AUSTIN
Austin, Texas 78712

BLANK PAGE

COMMENTS ON LINEAR FEATURE EXTRACTION*

T. L. Henderson and D. G. Lainiotis
Department of Electrical Engineering
and
Electronics Research Center
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

The problem considered is that of finding the best linear transformation to reduce a random data vector z to vector of smaller dimension. It is assumed that the original data are Gaussian under either of two hypotheses, and that one wishes to use the transformed data to distinguish the hypotheses. The Bhattacharya distance is used to measure the information carried by the transformed data. A compromise solution is obtained for the case in which the data has both different means and different covariances under the alternative hypotheses.

---

*Supported by Joint Services Electronics Program Grant AF-AFOSR 766-67 and by NASA Research Grant NGR 44-012-066, and by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under Grant AFOSR 69-1764.

COMMENTS ON LINEAR FEATURE EXTRACTION

We consider the problem of choosing a KxN transformation matrix to extract as much information as possible from an Nx1 data vectorz. The integer K is assumed to be fixed by practical considerations, and is taken to be less than N (since otherwise an invertible transformation could be found which preserves all of the information). It is further assumed that z is a Gaussian random vector under each of two hypotheses H_0 and H_1, and that we wish to extract those features which will enable us to best distinguish between the hypotheses. Specifically,

$$z \sim \eta\,(O, R_0) \text{ under hypothesis } H_O$$

$$z \sim \eta\,(m, R_1) \text{ under hypothesis } H_1\ ,$$

where the assumption of zero mean under the null hypothesis is made with no loss in generality since we could always subtract the mean before applying the transformation. For convenience, we assume that R_0 is positive definite.

The transformed data vector is

$$y = Az$$

where A is the KxN transformation matrix. Thus y has the following statistics:

$$y \sim \eta\,(O, P_O) \text{ under hypothesis } H_O$$

$$y \sim \eta\,(u, P_1) \text{ under hypothesis } H_1,$$

where $P_O \triangleq AR_OA^T$, $P_1 \triangleq AR_1A^T$, and $u \triangleq Am$.

As a measure of the information carried by y we shall use the Bhattacharya distance [*1], which in this case reduces to [*2]

$$B = \frac{1}{8} u^T P^{-1} u + \frac{1}{2} \log \frac{\det P}{(\det P_0 \cdot \det P_1)^{\frac{1}{2}}}, \text{ where } P \triangleq \frac{1}{2}(P_0 + P_1).$$

We shall not here discuss all of the justifications for using this measure, except to say that it provides an upper bound on the probability of error in distinguishing the hypotheses and in many special cases becomes monotonically related to the probability of error [*1,2,3]. We do, however, remark that it obeys most of the rules one would expect from an information measure: (1) It is non-negative (2) It becomes very large if y carries enough information to permit almost-perfect discrimination (3) It goes to zero if y is useless for distinguishing between the hypotheses (4) It will never decrease when the transformation matrix A is augmented by the addition of an extra row (5) It has the same value for choices of A which are equivalent (evquivalent in the sense that would give y's which are related by invertible transformations).

The problem of choosing A to maximize B becomes straight-forward in a couple of special cases, for which the results are fairly-well known:

Case I: If $R_0 = R_1 (\triangleq R)$, then the quantity $s^T z$ is a sufficient statistic for distinguishing between the hypotheses [*4], where s is an Nx1 vector defined by $s \triangleq R^{-1} m$. Thus in this case we could actually take A to be the row vector s^T and retain all of the relevant information with 1xN transformation. It is not difficult to show that, in fact, B attains its maximum whenever s^T is in the space spanned by the rows of A, and the maximum value is $B = \frac{1}{8} m^T R^{-1} m$.

Case II: If m=0, then the maximum value for B will be attained if we take the rows of A to be the first K eigenvectors $\phi_1, \phi_2, \ldots, \phi_K$ (transposed) of

the matrix equation [*3] $R_1\phi = \lambda R_0\phi$, where the eigen-solutions $(\phi_i, \lambda_i)_{i=1}^{N}$ are ordered so that $\lambda_1 + 1/\lambda_1 \geq \lambda_2 + 1/\lambda_2 \geq \ldots\ldots \geq \lambda_N + 1/\lambda_N$. It is easily shown that there exist N eigen-solutions which satisfy the "orthonormality" condition $\phi_i^T R_0 \phi_j = \delta_{ij}$ [*5]. Moreover, with this choice of A, the components of y will be statistically independent under either hypothesis.

In the general case the problem is much more difficult. Gradient techniques could be used to find the optimum A, but such techniques would prove extremely impractical when N is large. We shall develop here a slightly sub-optimal solution to the problem, which is relatively easy to find in a given practical problem.

If we write $B = B_I + B_{II}$, where $B_I = \frac{1}{8} u^T P^{-1} u$; $B_{II} = \frac{1}{2} \ln \frac{\det P}{(\det P_0 \cdot \det P_1)^{\frac{1}{2}}}$, then we can note the following facts: First we note that B_{II} is completely independent of m, and is simply the value of B that would have been obtained under the assumption m = 0. Thus we must have $B_{II} \leq B_{II}^*$, where B_{II}^* is the value of B_{II} attained when A is chosen as in Case II. Furthermore, B_I is just the value of B that would have been obtained if we had assumed equal covariance matrices, taking the average value $R = \frac{1}{2}(R_0 + R_1)$. Therefore, $B_I \leq B_I^* \triangleq \frac{1}{8} m^T R^{-1} m$, and equality will be acheived whenever the row space of A contains the vector $s^T = m^T R^{-1}$.

At this point we should mention one special case: if it should happen that s^T is contained in the row space of the matrix A chosen as in Case II , then both equalities will obtain and the maximum B will have been achieved. (This may actually occur, at least approximately, in some practical cases.)

If we write $B^{opt} = B_I^{opt} + B_{II}^{opt}$ to denote the values attained when A is chosen to maximize B, we must have $B_I^{opt} \leq B_I^*$ and $B_{II}^{opt} \leq B_{II}^*$. But suppose we choose A as in Case II. Then we would have $B = B_I + B_{II}^*$, where, excepting the special case, $B_I < B_I^*$. Now suppose we use the (K+1)xN matrix A^{aug} formed by adjoining s^T onto A as its K+1 row; i.e., $A^{aug} \triangleq \begin{bmatrix} A \\ s^T \end{bmatrix}$. Then $B^{aug} = B_I^* + B_{II}^{aug}$, since the row space now contains s^T. Furthermore, it must be true that $B_{II}^* \leq B_{II}^{aug}$ since we know that (in Case II) adding a new row to the matrix transormation never decreases the information. Hence $B^{aug} \geq B^{opt}$.

The result can be summarized as follows: Although we cannot easily find the best KxN matrix A^{opt}, we can always use the (K+1)xN matrix A^{aug} to extract just as much (and possibly more) information at the cost of requiring that we be permitted to use one extra component in the transformed data vector y.

It should be noted that another interpretation of this result is provided by simply identifying J=K+1; J-1=K, and stating that although A^{aug} may not be quite as good as the best JxN matrix, it certainly yields as much information as any (J-1)xN matrix, and may be therefore regarded as a slightly suboptimal solution to the problem of finding the best JxN matrix.

In many real problems this compromise solution should be good enough, provided that increasing or decreasing K slightly does not drastically affect the amount of information we can extract. This condition would be met if N were quite large and K had been predetermined by a law of diminishing returns, (i.e. if K had been selected to be large enough so that further increase would

yield relatively little additional information).

In order to find A^{aug} we must find s, and determine the first K eigenvectors ϕ_i described under Case II. Finding s is not too difficult, and is actually analogous to finding the matched filter in a problem with continuous (rather than discrete, i.e. vector) data. For finding the ϕ_i's we first note that

$$R_1\phi = \lambda R_0\phi \quad <=> \quad R_0^{-1}R_1\phi = \lambda\phi \quad <=> \quad R_1^{-1}R_0\phi = \frac{1}{\lambda}\phi \text{ (if } R^{-1} \text{ exists)}.$$

Matrix iteration techniques can be used to find the eigenvectors[*6]. However, when N is extremely large such techniques are unreasonable (and it becomes impossible to invert the matrices). For this reason a practical method has been developed for finding the eigenvectors under the fairly general assumption that the components of z can be regarded as the output sequence of discrete linear dynamic system excited by a "white" Gaussian noise sequence (with a different model for each hypothesis). The details are too involved to be presented in this correspondence[*5].

In closing we remark that these results apply equally well to the case of continuous data $z = z(t)$, $t \in (a,b)$, provided a complete set of eigenfunctions $\phi_i(t)$ exist for the integral equation[*3,7]

$$\int_a^b R_1(t,\tau)\,\phi(\tau)\,d\tau = \lambda \int_a^b R_0(t,\tau)\,\phi(\tau)\,d\tau,$$

where R_0 and R_1 are the hypothesis-conditional covariance functions of z(t). (State variable techniques have been developed for solving this integral equation)[*8].

REFERENCES

[1] D. G. Lainiotis, "Optimal Feature Extraction In Pattern Recognition," IEEE International Information Theory Symposium Abstracts, San Remo, Italy, September 1967.

[2] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Transactions on Communications Technology, Vol. 15, No. 1, pp. 52-60, February 1967.

[3] T. T. Kadota and L. A. Shepp, "On The Best Set of Linear Observables for Discriminating Two Gaussian Signals," IEEE Transactions on Information Theory, Vol. IT-13, No. 2, April 1967. [Note: Kadota and Shepp use the Hellinger integral H = exp (-B)]

[4] H. L. Van Trees, Detection, Estimation, and Modulation Theory: New York, John Wiley and Sons, 1968.

[5] T. L. Henderson and D. G. Lainiotis, "Digital Matched Filters for Detecting Gaussian Signals in Gaussian Noise," IEEE International Information Theory Symposium Abstracts, Ellenville, New York, January 1969.

[6] A. Ralston, A First Course in Numerical Analysis, (New York: McGraw Hill, 1965) p. 480 ff.

[7] C. R. Baker, "Simultaneous Reduction of Covariance Operators and The Noise-in-Noise Problems of Communication Theory," Ph.D. Dissertation, Department of Electrical Engineering, University of California at Los Angeles, 1967.

[8] D. G. Lainiotis and T. L. Henderson, " Application of State Variable Techniques to Optimal Feature Extraction," Proceedings of The IEEE, Vol. 56, No. 12, pp. 2175-2176, December 1968.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

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

1 ORIGINATING ACTIVITY *(Corporate author)*
The University of Texas at Austin
Electronics Research Center
Austin, Texas 78712

2a REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b GROUP

3 REPORT TITLE

COMMENTS ON LINEAR FEATURE EXTRACTION

4 DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Scientific Interim

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

T. L. Henderson
D. G. Lainiotis

6 REPORT DATE	7a TOTAL NO. OF PAGES	7b NO. OF REFS
15 April 1969	8	8

8a CONTRACT OR GRANT NO.
AF-AFOSR-67-0766E

b. PROJECT NO.
4751

c.
6144501F

d
681305

9a ORIGINATOR'S REPORT NUMBER(S)
JSEP Technical Memorandum 8

9b OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
AFOSR-69- 2TR

10 AVAILABILITY/LIMITATION NOTICES
This document has been approved for public release ale; its distribution is unlimited

11 SUPPLEMENTARY NOTES
TECH, OTHER

12 SPONSORING MILITARY ACTIVITY JSEP through
AF Office of Scientific Research (SREE)
1400 Wilson Boulevard
Arlington, Virginia 22209

13 ABSTRACT

The problem considered is that of finding the best linear transformation to reduce a random data vector z to vector of smaller dimension. It is assumed that the original data are Gaussian under either of two hypotheses. The Bhattacharya distance is used to measure the information carried by the transformed data. A compromise solution is obtained for the case in which the data has both different means and different covariances under the alternative hypotheses.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
FEATURE EXTRACTION BHATTACHARYA DISTANCE MEASURE DATA COMPRESSION GAUSSIAN RANDOM PROCESSES						

INSTRUCTIONS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

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

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

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

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

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