UNCLASSIFIED

AD NUMBER
AD475844
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; DEC 1965. Other requests shall be referred to Office of Naval Research, Arlington, VA 22203.**
AUTHORITY
ONR ltr 28 Apr 1968

THIS PAGE IS UNCLASSIFIED

December 1965

THE JOHNS HOPKINS UNIVERSITY
DEPARTMENT OF ELECTRICAL ENGINEERING
BALTIMORE, MARYLAND 21218

475844

REPRESENTATION AND ANALYSIS OF SIGNALS
PART XXI. THE INTRINSIC DIMENSIONALITY OF SIGNAL COLLECTIONS

by

Robert S. Bennett

DDC
JAN 10 1966
TISIA E

A report distributed under
Contract Nonr-4010(13) with
the Office of Naval Research.

December 1965

THE JOHNS HOPKINS UNIVERSITY
DEPARTMENT OF ELECRTICL ENGINEERING
BALTIMORE, MARYLAND 21218

REPRESENTATION AND ANALYSIS OF SIGNALS
PART XXI. THE INTRINSIC DIMENSIONALITY OF SIGNAL COLLECTIONS

by

Robert S. Bennett

This research was sponsored by the
Carlyle Barton Laboratory of The
Johns Hopkins University,
Baltimore, Maryland, under USAF
Contract AF33(657)-11029,
Project 4036, Task 403601.

FOREWORD

This work is largely an outgrowth of discussions with Dr. W. H. Huggins, Westinghouse Professor of Electrical Engineering in The Johns Hopkins University, whose inspiration and guidance are gratefully acknowledged. Dr. Stephen S. Wolff, Assistant Professor of Electrical Engineering at the Hopkins, read and made constructive comments on the manuscript.

Gratitude is expressed to many others, in particular to Mrs. Bernice Becher and Miss Phyllis Shipley of the Carlyle Barton Laboratory for their assistance in programming the numerical examples. Mr. C. M. LaPorte and Mr. Russell H. Gordon prepared the illustrations.

The careful typing of the manuscript is due to Miss Barbara Westphal.

ABSTRACT

In view of the trend toward the representation of signals as physical observables, characterized by vectors in an abstract signal space, rather than as time or frequency functions, it is desirable to define dimensionality in a manner which would be independent of the choice of basis for the vectors.

In this work, the dimensionality of a collection of signals is defined as equal to the number of free parameters required in a hypothetical signal generator capable of producing a close approximation to each signal in the collection. Thus defined, dimensionality becomes a relationship between the vectors representing the signals. This relationship need not be a linear one, and does not depend on the basis onto which the vectors are projected in signal measuring processes. It represents a lower bound on the number of coefficients required to describe the signals, no matter how sophisticated the representation scheme, and thus provides an index of the redundancy in a given representation.

A computer program for estimating this dimensionality from the signal coefficients on an arbitrary orthogonal basis is developed. The program, suitable for an IBM 7094 computer, is based on some results from a related multidimensional scaling problem, and utilizes an inverse relationship between the variance in interpoint distance

TR 7.1 63

within a hypersphere, and the dimensionality of the hypersphere. This method is independent of the choice of orthonormal basis, and no prior knowledge of the analytical form of the signals is assumed.

The validity of the program is verified by using it to estimate the dimensionality of signals of known structure, and therefore of known dimensionality.

Page

TABLE OF CONTENTS

I. INTRODUCTION 1

A. Plan of the Report 4

II. FORMULATION OF THE PROBLEM 6

III. GEOMETRICAL CONSIDERATIONS 13

A. Assumptions 13

1. Uniqueness 13

2. Independence 14

3. Holonomic Constraints 14

B. Distances in a Hypersphere 16

IV. DISCUSSION OF APPROACH 20

V. OUTLINE OF COMPUTER ROUTINE 34

VI. EXPERIMENTAL VERIFICATION 40

A. Rectangular Pulses, One Dimension 48

B. Rectangular Pulses, Two Dimensions 50

C. Exponentials, One Dimension 53

D. Exponentials, Two Dimensions 55

E. Dependent Parameters 58

VII. DISCUSSION 61

A. Effects of Computer Parameters 61

1. Alpha 61

2. Gamma 62

3. Beta 62

TABLE OF CONTENTS (CONTD.)

B. Future Work 64

C. Conclusion. 66

VIII. REFERENCES 67

ILLUSTRATIONS

Figure 1	Signal Siftor
Figure 2	Unconstrained Signal Generator
Figure 3	Constrained Signal Generator
Figure 4	Distribution of Distance in a Hypersphere
Figure 5	Two-Dimensional Configuration
Figure 6	Collapsing of a Curve into a Line
Figure 7	Regions Defined by Beta
Figure 8	Simplified Flow Chart
Figure 9	a_1 vs. a_3
Figure 10	a_2 vs. a_4
Figure 11	a_2 vs a_5

I. INTRODUCTION

There has been a tendency in the past for authors of communications papers to speak of a signal synonymously with its representation as a time or frequency function. Thus a particular signal may be designated as a "sine wave", or "square wave", or as a "band-limited" signal. In more recent publications (1, 2)*, it has been demonstrated that such an approach tends to obscure the true nature of signals. In particular, it assumes that in the noise-free case, complete knowledge of the signal is possible, while in fact the only access one has to a physical signal is through a measuring device, or filter, having finite capabilities and therefore able to yield only an approximation to the signal.

To better appreciate signals as physical observables, only partially accessible, it has proven useful to consider them as vectors, $|F\rangle$, on an abstract infinite dimensional signal space, V. (Where practical, Dirac's notation, as adapted by Lai (2), will be followed in this report.) In this representation, the signal energy is characterized by the square of the vector norm. $\langle F|F\rangle$, and the structure by the vector "direction."

---

* Whole numbers in parenthesis refer to references listed beginning on Page 67.

Determining a given relationship between signals, e.g. correlation, summation, etc., thus becomes an operation with the vectors themselves, and is independent of any time or frequency basis. An attribute of signal collections which will be discussed in detail in this document is dimensionality.

The definition of signal dimensionality is at best a difficult task, but several ad hoc definitions are in use, the most common being based on a time-bandwidth product. Here, one speaks of the dimensionality of individual signals, not of classes or collections. If the signal, which is specified as a time function, has negligible energy in the frequency components above B cycles per second, it will have 2B degrees of freedom per second. The dimensionality of a τ second portion of this signal is then defined to be $2B\tau$. The usefulness of such a definition lies in the sampling theorems which permit recovery of the τ second portion of the time function from $2B\tau$ uniformly spaced (in time) amplitude samples, with an error which varies inversely with the product $B\tau$. (3, 4, 5)

Another definition which finds frequent application is based on an orthonormal expansion of the signal, again usually expressed as a time function. The signal is represented as a linear sum of weighted orthonormal components, each

weight being the inner product of the signal with the corresponding component function. The number of such components required to represent the signal to within a specified energy error is then defined to be the dimensionality. The previous definition is a special case of this one, where the orthonormal components are shifted cardinal functions.

Despite their usefulness, such definitions are unsatisfying in view of the trend toward thinking of signals as physical observables rather than as functions. It would seem more satisfactory to define dimensionality in terms of some relationship involving the vectors representing a given collection of signals. Thus defined, dimensionality would become an <u>intrinsic</u> property of the collection.

In order to obtain meaning from a signal, some measurement must be performed upon it. Such a measurement usually has the form of a projection of the signal vector onto prescribed basis vectors, or "patterns", $\langle \phi_1 |$, $\langle \phi_2 |$, $\langle \phi_3 |, \cdots \langle \phi_n |$, which characterize the measuring apparatus. Since the only access an observer has to the signal is through these measurements, a technique must be found for evaluating the dimensionality of the signal collection in terms of these measurements.

A. Plan of the Report

The purpose of this work is twofold:

1. To suggest a generalized definition of signal-collection dimensionality.
2. To develop a technique for estimating this intrinsic dimensionality for collections of signals.

In the present chapter, the problem has been roughly described. In Chapter II, this description will be extended to a precise mathematical formulation, and the generalized definition of dimensionality will be given.

The technique for estimation of signal class dimensionality is based on a relationship between the dimensionality and the geometry of the vectors representing the signals in signal space. Chapter III is in two parts. The first part sets forth certain assumptions about the signal collection which must be valid if the estimation of the dimensionality is to be feasible. The second part discusses an inverse relationship between the variance in interpoint distances in a hypersphere and the dimensionality of the hypersphere.

In Chapter IV, this relationship in hyperspheres is utilized in developing a technique for estimating the dimensionality of signal collections. Nothing is assumed known about the signals with the exception of their spectral coefficients on some arbitrary orthonormal basis. A requirement of the technique is that the final result remain invariant under change of this basis.

In Chapter V, a computer program outline and simplified flow chart are given which realize the technique developed in Chapter IV. To obtain the greatest generality, the program must be an iterative one, and its usefulness is predicated on the availability of a large-scale automatic digital computer.

In Chapter VI, examples of the application of the technique to several collections of signals are given. The computer used in these tests was an IBM 7094.

Chapter VII summarizes and discusses the program and the experimental results.

II. FORMULATION OF THE PROBLEM

A common approach in the representation and analysis of a collection of experimentally obtained physical signals is to find the spectral coefficients of the signal when expanded in a set of basis functions $\langle \tilde{\phi}_i |, i = 1, 2, \ldots$ In order to minimize the effects of slight numerical errors, the basis functions are required to be uncorrelated, for example,

$$\langle \tilde{\phi}_i | \phi_j \rangle = \int dt\, \phi_i^*(t)\, \phi_j(t) = k_i\, \delta_{ij} \tag{1}$$

for all i and j in the case where the basis functions $\phi_i(t)$ are functions of time. For convenience, the k_i are set equal to unity through suitable scale factors associated with the $|\phi_i\rangle$.

If the orthonormal set of basis functions is complete, the time-function representation of a signal may be written as

$$f(t) = \sum_{i=1}^{\infty} |\phi_i\rangle \langle \tilde{\phi}_i | F \rangle \tag{2}$$

or

$$f(t) = \sum_{i=1}^{\infty} a_i |\phi_i\rangle \tag{3}$$

where

$$a_i = \int dt\, \phi_i^*(t)\, f(t) = \langle \tilde{\phi}_i | f \rangle$$

Using a familiar signal representation concept, the signal may be represented as a vector defined on a hyperspace having the $\phi_i(t)$ as a basis. (1) The $\langle \tilde{\phi}_i |$ are considered as unit vectors, and the signal vector $|\underline{F}\rangle$ has coordinates a_i on this basis.

The choice of a suitable basis is somewhat arbitrary. Provided that both bases are complete, that is, all of the signals be wholly within the subspace spanned by the bases, each signal could just as properly be represented by its coordinates in a second basis ψ_i.

$$|F\rangle = \sum_{i=1}^{\infty} a_i |\phi_i\rangle = \sum_{i=1}^{\infty} b_i |\psi_i\rangle \tag{4}$$

The vector representing the signal does not depend on the basis, rather, the coordinates of the vector are dependent on the choice of the basis.

From a practical standpoint, there is much to be gained from selecting a set of basis functions such that a minimum of vector components suffice to represent the signal to within some error criterion. In practice, signals to be investigated are often empirical rather than being given as analytical functions of time, frequency, or other variables. The spectral coefficients of the signal are obtained by passing the signal through a series of filters, each "matched" to one of the basis patterns, and sampling at the appropriate instant as shown in Figure 1. These sample values are then the projections of the signal in question on each of the basis patterns. To avoid the expense of a large number of filters, the basis functions are chosen such that a

TR 7163

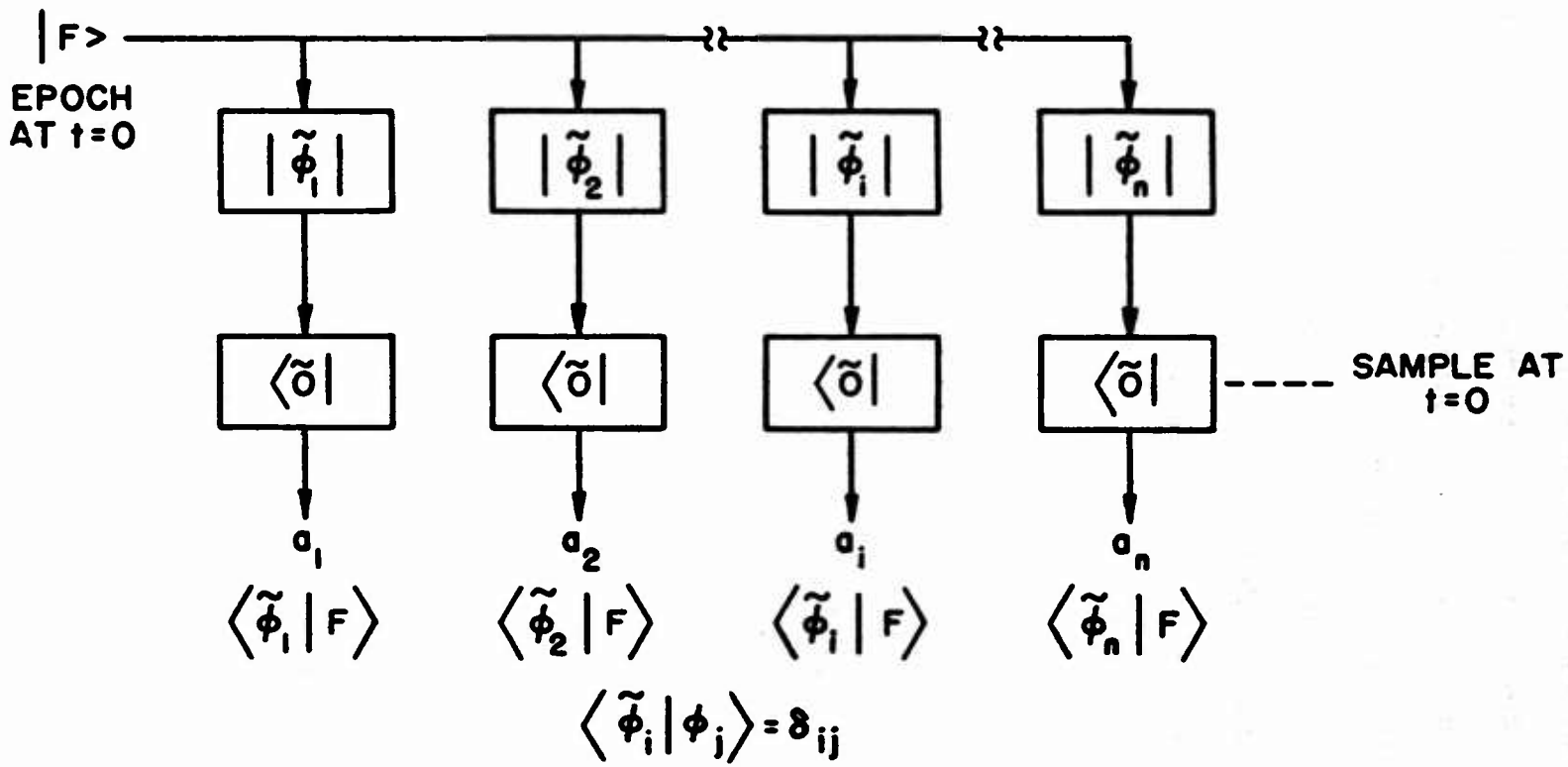


Figure 1 Signal Siftor

small number of basis functions spans the same subspace of V as the signal to within a prescribed energy error.

The desirability of selecting the fewest numbers for representing the signals is more basic than simple considerations of parsimony, however. Consider the model for a generator of one-sided, single-epoch signals shown in Figure 2.

The output of the filter, $|F|$, is the signal which will be applied to the input of the n-dimensional orthonormal filter to determine the a_i. If $|F|0\rangle$ is completely unspecified, the probability that the point representing this signal in the n dimensional space will be within a hypersphere of some specified radius cannot be determined. A collection of observed signals may be thought of as arising from such a filter which is free to vary randomly between consecutive signal outputs. The points in the n-dimensional representation space will then be randomly distributed through the space.

Such a model is clearly inconsistent with real world signal sources, which are not free to vary arbitrarily but are subject to definite constraints. If the signal is noisy, these constraints are "soft", the constraints becoming more well defined as the signal-to-noise ratio increases. Further discussion of the effects of noise on this formulation will be deferred to Chapter III.

These constraints will be reflected in the distribution of signal coordinates in the representation space. For example, a maximum

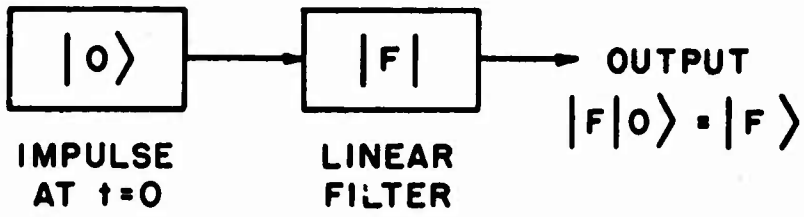


Figure 2 Unconstrained Signal Generator

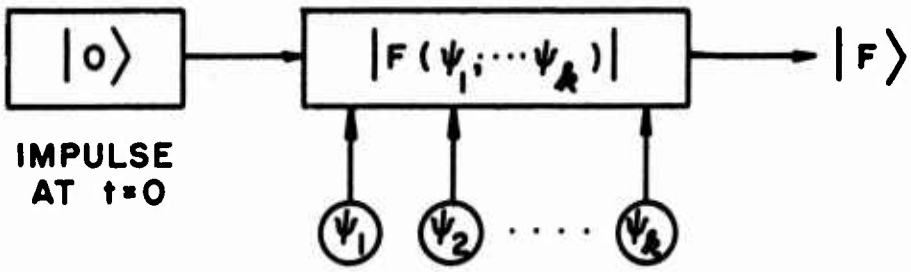


Figure 3 Constrained Signal Generator

TR 7163

energy restriction on the output signals,

$$\langle \tilde{F} \mid F \rangle \leq R \tag{5}$$

would require that the points in the representation space lie within a hypersphere of radius $R^{1/2}$.

A more realistic signal generator model which includes constraints on the variations between successive output signals is shown in Figure 3. The constraints are introduced as a finite number of filter parameters which may vary at random between signals.

Now, if the class of signals so generated is representable to within an acceptable error on a space spanned by n patterns, the representation of the jth signal will be of the form:

$$|F_j|0\rangle = \sum_{i=1}^{n} a_i(\Psi_j)|\phi_i\rangle - |\epsilon_j\rangle \tag{6}$$

where

$$\Psi_j = \left[\psi_1(j), \psi_2(j), \ldots, \psi_k(j)\right]$$

and $|\epsilon_j\rangle$ is the error in representing the jth signal.

The number n is usually referred to as the "dimensionality" of the signal class. For the purpose of this discussion, n will be termed the <u>linear</u> dimensionality of the class. The number k will be defined as the <u>intrinsic</u> dimensionality of the class. The linear and intrinsic dimensionalities are related only by an inequality, $k \leq n$. The value of n can be considerably greater than that of k, as in the case where

$$F_j(\psi_1, \psi_2, t) = \psi_1(j)\, e^{-\psi_2(j)t} \qquad t \geq 0$$
$$= 0 \qquad t < 0 \tag{7}$$

For arbitrary values of ψ_2, expansion of this class of signals on an orthogonal basis will require a great number of coefficients, while $k = 2$.

The difference between n and k represents a redundancy in representation which cannot in general be removed by a linear transformation of basis.

The problem to be considered may now be stated as follows:

Given a collection of signals whose spectral coefficients a_i on a basis $|\phi_i\rangle$ are known, assume that $a_i = a_i(\Psi_j)$ where $\Psi_j = \left[\psi_1(j), \psi_2(j) \ldots \psi_k(j)\right]$ and from these a_i, estimate the value of k.

RR/1163

III GEOMETRICAL CONSIDERATIONS

A. Assumptions

Although the approach to be developed here is intended to achieve the greatest possible generality, it was found necessary to place three restrictions on the hypothetical signal generator:

1. $$|F_p|0\rangle \neq |F_q|0\rangle \quad \text{for all } p \neq q \tag{8}$$

As previously stated, it is advantageous from an instrumentation standpoint to use an orthonormal basis for representing the signals $|F_p\rangle$ and $|F_q\rangle$. Thus the above restriction may be equivalently stated

$$a_i(\Psi_p) \neq a_i(\Psi_q) \quad \text{for all } p \neq q \tag{9}$$

where

$$\Psi_j = \left[\, \psi_1(j),\ \psi_2(j) \ldots \psi_k(j) \,\right]$$

Were this restriction to be invalid, even for just two values of p and q, say p' and q' so that

$$a_i(\Psi_{p'}) = a_i(\Psi_{q'}) \tag{10}$$

then the apparent dimensionality of the collection of signals would be spuriously high. The technique to be developed in Chapter IV operates on the locus of the signal vectors, and, for example, loops in the locus of vectors in the single parameter case will appear to have been produced by a two parameter signal generator.

PREVIOUS PAGE WAS BLANK, THEREFORE NOT FILMED

An equation of the form

$$g(\vec{r}_1, \vec{r}_2, \ldots j) = 0 \tag{14}$$

is not possible here because the terms $\langle \phi_n | n \rangle$ in the $\vec{r}$'s are not deterministic. If the noise is white and has zero mean, the ψ's will determine the most likely coordinates of each $\vec{r}$, and the actual coordinates will have a symmetric probability distribution surrounding this point.

In Chapter IV, it will be postulated that a single-parameter signal generator will produce signal vectors whose locus in the signal space V will be a curve. The intrinsic dimensionality estimation technique will be based on this hypothesis. For large S/N energy ratios, the noise hopefully will do no violence to this hypothesis. However, as the noise energy increases, one would expect a point to be reached where a "threshold" effect occurred and the reliability of the dimensionality measurement would fall off rapidly. In the case of white Gaussian noise, this threshold should occur when the variance of the noise is of the order of magnitude of the least radius of curvature of the locus of the signal vectors $\vec{r}_j$. Since, when finite bases are used, this locus may depend on the basis as well as on the signals, the effect of noise on the measurement cannot be predicted here. However, as a rough indication, the S/N ratio at which white Gaussian noise will have variance exceeding the least radius of curvature of the locus of vectors representing square pulses expanded on an orthonormalized exponential basis will be evaluated in Chapter VII.

B. Distances in a Hypersphere

At this point it is necessary to digress briefly and to consider the distribution of interpoint distances in an N dimensional homogeneous hypersphere. This distribution was first investigated by Deltheil (7) and more recently by Hammersley (8) and Lord (9).

Consider a sphere of radius a and dimension N. Let the distance between two points within the sphere be designated by r, and for convenience let

$$\lambda = r/2a \quad \text{so that } 0 \leq \lambda \leq 1. \tag{15}$$

Deltheil developed an expression for the probability density function for r, and evaluated this for some odd-integer values of N. He did not evaluate the cases where N was even.

Using an involved approach, Hammersley obtained the following compact form of Deltheil's expression:

$$f_N(\lambda) = 2^N N \lambda^{N-1} I_{1-\lambda^2}\left(\tfrac{1}{2}N + \tfrac{1}{2}, \tfrac{1}{2}\right) \tag{16}$$

where

$$I_x(p, q) = \int_0^x Z^{p-1}(1-Z)^{q-1}\, dZ/B(p, q)$$

and B is the Beta function.

This form is good for all values of N. For the particular cases N = 1, N = 2, and N = 3, i.e., a line, a circle, and a solid sphere, the above

expression reduces to easily handled expressions:

$$f_1(\lambda) = 2(1 - \lambda)$$

$$f_2(\lambda) = \left(\frac{16}{\pi}\right) \lambda \left[(\cos^{-1}\lambda) - \lambda(1-\lambda^2)^{1/2} \right] \tag{17}$$

$$f_3(\lambda) = 12\,(2\lambda^2 - 3\lambda^3 + \lambda^5) \qquad \text{for } 0 \le \lambda \le 1$$

Lord obtained the same results through a more general approach, and showed further that the distribution of distances is asymptotically normal as N increases, and furthermore that the second moment of the PDF is given by

$$\overline{\lambda^2} = 2a^2\, N(N + 2)^{-1} \tag{18}$$

From Equation 17, the mean may be calculated, and this is shown in Table 1, along with $\overline{\lambda^2}$, (with a = 1/2 for normalization), together with the variance in the interpoint distances about the mean.

TABLE 1 Distance Distributions in Hyperspheres

N	$\overline{\lambda}$	$\overline{\lambda^2}$	Variance
1	.333333	.166666	.055556
2	.452712	.250000	.045056
3	.514286	.300000	.035510
∞	.707107	.500000	.000000

TR-7163

The probability density functions for $N = 1$, $N = 2$, and $N = 3$ are shown in Figure 4. The function is asymptotically normal and the variance decreases monotonically toward zero as N increases, the limiting case being a δ function at $\lambda = 0.707107$. This result will be of fundamental importance in the sequel.

TR 71163

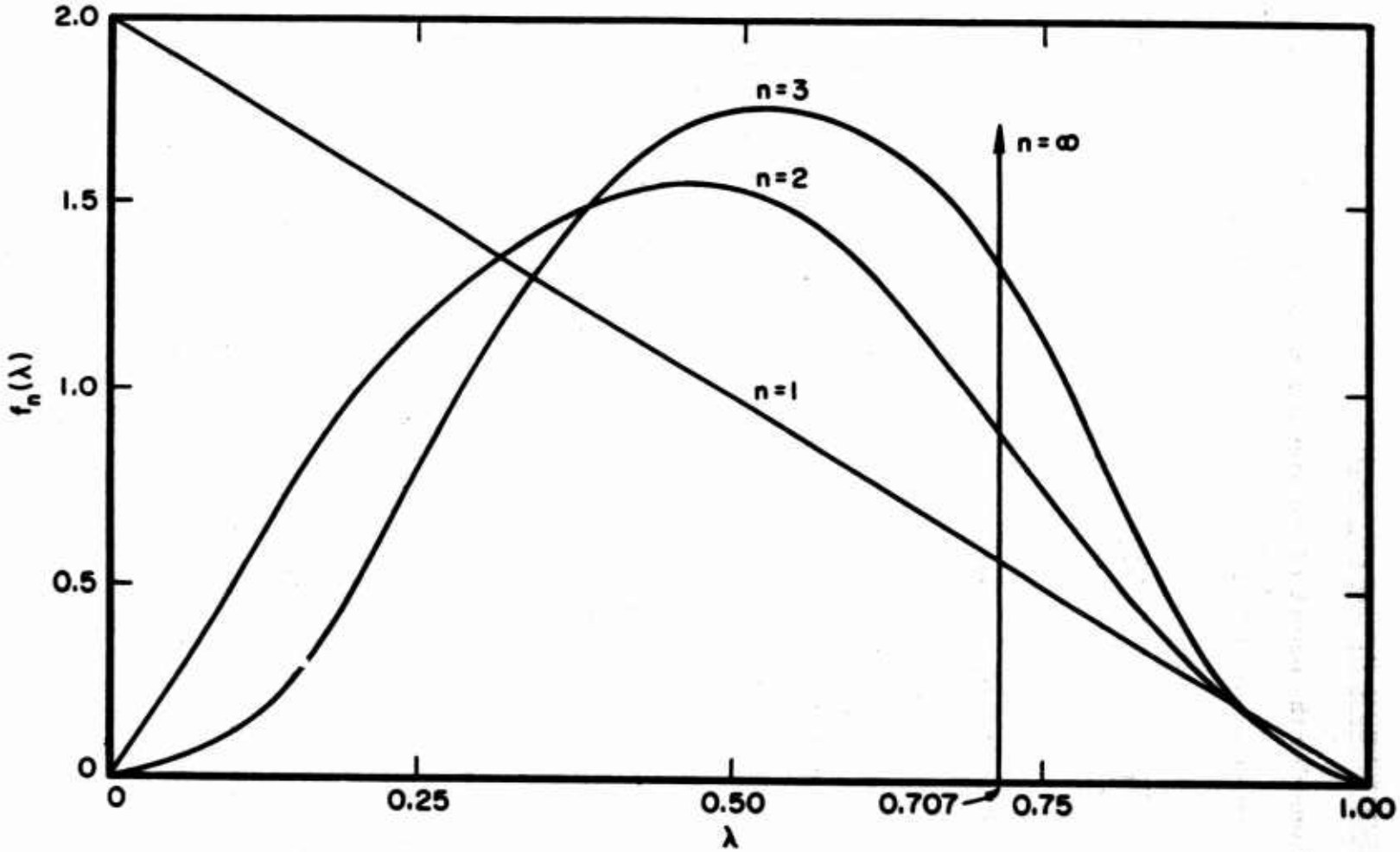


Figure 4 Distribution of Distance in a Hypersphere

IV. DISCUSSION OF APPROACH

Consider the output of the signal generator of Figure 3 as projected on some orthonormal basis complete for this collection

$$|F_j\rangle = \sum_{i=1}^{\infty} a_i (\underline{\Psi}_j) \tag{19}$$

where

$$\Psi_j = \left[\psi_1(j) \ldots \psi_k(j) \right]$$

For each value of J, a point in a signal subspace having the ϕ_i as a basis will be generated, having coordinates $a_1, a_2, \ldots, a_n$. The locus of these points will reflect the value of k, since this locus may be described in terms of k generalized coordinates, provided that the three assumptions in Chapter III are valid for the signal generator $|F\rangle$. The problem of determining the value of k is thus equivalent to determining whether this locus is a curve, (k = 1), a surface, (k = 2), or an m dimensional solid (k = m). Peano's continuous mapping of an interval onto the whole of a square shows that the dimension of a space cannot be defined as the number of parameters required to describe the space. (10) That a single parameter signal generator will produce a curve in signal space cannot therefore be rigorously shown. For continuously varying ψ's, this will be taken as an assumption under the restrictions imposed in the preceding chapter, subject to experimental verification. From this point on, the problem of estimating the value of k from the spatial distribution of the points representing the signals

will be taken to be equivalent to finding the number of generalized coordinates needed to describe the locus of these points.

A related problem occurs in the field of Experimental Psychology, and is known as the problem of Multidimensional Scaling. Briefly, this problem is stated as follows:

> "Given the experimental dissimilarities of n objects, find a set of n points whose interpoint distances are a monotone function of these dissimilarities" (11, 12).

A computer routine for multidimensional scaling which is based on the inverse relationship between interpoint distance variance and the dimension of a hypersphere has been written by Roger Shepard. (12) This program begins with a collection of n points, $P_1, P_2, \ldots, P_n$, each representing some psychological quantity: color perception, interpretation of facial expression, etc. The only knowledge of the relationship between the P_i is in the form of a similarity ranking. A configuration is sought such that this ranking is inversely duplicated by the distances between the P_i. That is, if the quantities represented by P_i and P_k are known to be quite similar, then the distance between points P_i and P_k should be small. If the similarity between P_i and P_k is denoted by S_{ik}, and the distance between points P_i and P_k is denoted by d_{ik}, then for all $n(n-1)/2$ pairs of points, the ranking of distances and dissimilarities is to be preserved.

$$\underline{S_{ik}} < \underline{S_{lm}} \rightarrow \underline{d_{ik}} > \underline{d_{lm}} \tag{20}$$

It would be expected that given only the set of inequalities ranking the S_{ik}, the corresponding configuration of points having the required ranking of d_{ik} would be far from unique. Surprisingly, it can be demonstrated that for moderate values of n, ($n \approx 50$), and for a final configuration dimensionality of 3, the resulting configuration is very well defined. (13) At the end of this analysis not only is it possible to obtain the proper distance ranking, but actual measurements of these distances as well. That measurements can be obtained from non-metric ranking data is not too startling when it is considered that the full set of 1225 inequalities is highly redundant if the configuration is of small dimensionality.

The method used is as follows: The n points are first located at the n vertices of a regular n-1 dimensional simplex. The points are then perturbed in directions which make the ranking of the $(n^2-n)/2$ interpoint distances conform to the desired ranking. Next the points are shifted so that d_{ik} larger than the mean are increased, and d_{ik} smaller than the mean are decreased. These two steps are iterated until the configuration becomes stationary. As the iterations progress, the ranking is maintained, and the dimensionality of the configuration decreased from n to its final value.

An anomaly in some of Shepard's results can be exploited in solving the intrinsic dimensionality problem. If the configuration of

TR 71-163

points at some iteration in the multidimensional scaling program is an arc of less than 180 degrees, this will be stretched out into a line. Similarly, a hemispherical shell will be deformed into a plane surface. In the case of multidimensional scaling, this yields a spuriously low dimensionality for the final configuration. In the problem considered here, however, this is just what is needed. Provided that the ranking is not preserved over too much of the configuration at a time, curves in n-dimensional space collapse into lines, surfaces into planes, and so forth. The intrinsic dimensionality of the signal source thus is reflected as the spatial dimensionality of the final configuration. The procedure consists of iterating two processes which together collapse the configuration of points representing the signals $|F_j\rangle$ on the ϕ basis.

The first process increases the variance in interpoint distances. Consider a configuration of points in a plane as in Figure 5.

The projections of the vector from 0 to point P_j onto the X basis are given by a_{j1} and a_{j2} where the first subscript denotes the point and the second subscript denotes the basis vector.

For an orthonormal basis, the distance between points i and j is given by

$$d_{ij} = \left[(a_{i1} - a_{j1})^2 + (a_{i2} - a_{j2})^2 \right]^{1/2} \tag{21}$$

or, in n dimensions

$$d_{ij}^2 = \sum_{k=1}^{n} (a_{ik} - a_{jk})^2 \tag{22}$$

TR 71 163

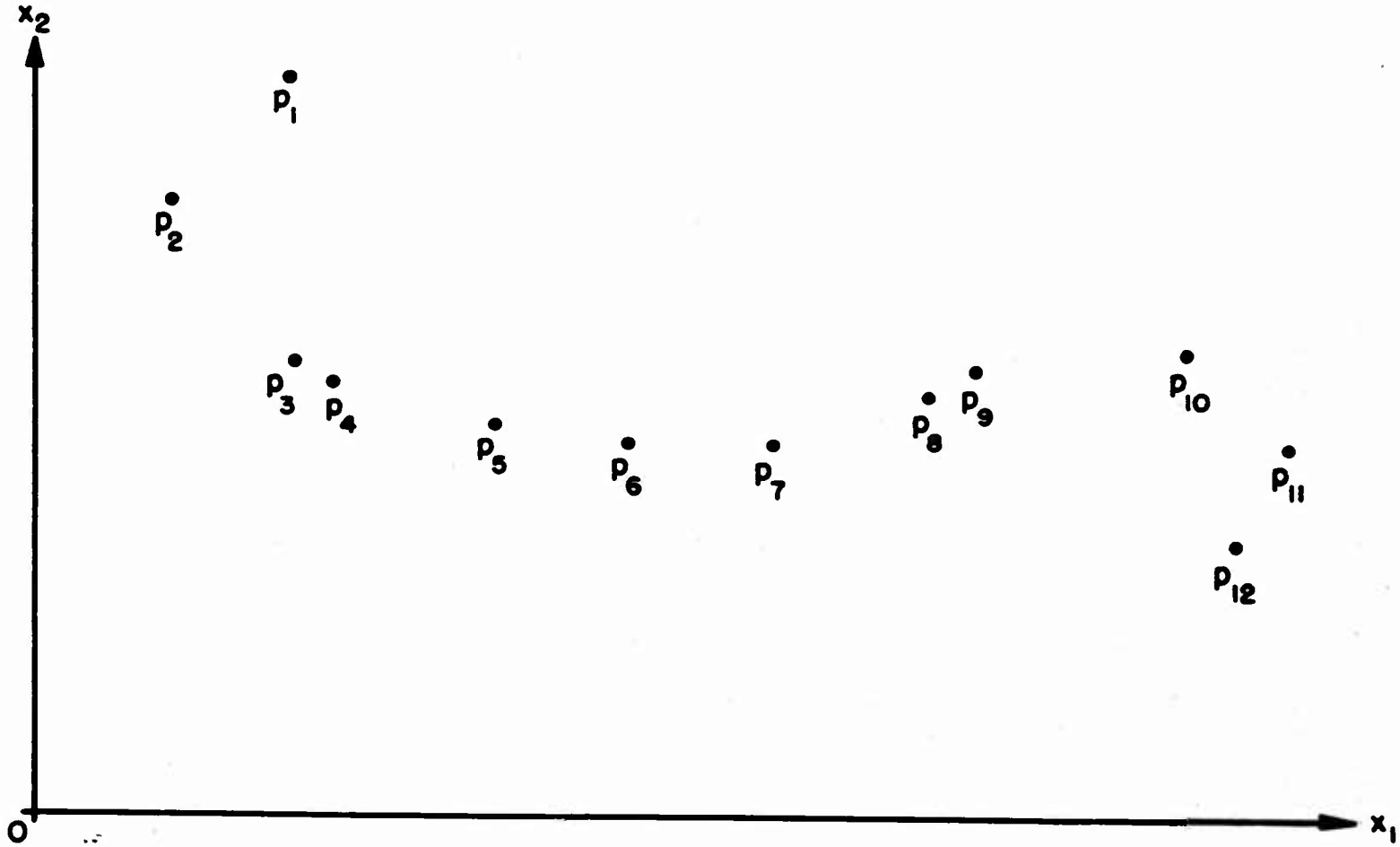


Figure 5 Two-Dimensional Configuration

Let the unit vector in the X_j direction be u_j. Then the vector from P_i to P_j is

$$\vec{A}_{ij} = (a_{j1} - a_{i1})\, u_1 + (a_{j2} - a_{i2})\, u_2 \tag{23}$$

or, in n dimensions

$$\vec{A}_{ij} = \sum_{k=1}^{n} (a_{jk} - a_{ik})\, u_k \ . \tag{24}$$

Let the arithmetic mean of the d_{ij} be $\bar{d}$.

To increase the variance in the d_{ij}, those distances $d_{ij} < \bar{d}$ should be reduced, and $d_{ij} > \bar{d}$ should be increased. This may be done incrementally as follows.

Consider

$$\frac{\alpha(d_{ij} - \bar{d})}{\bar{d}} = \Delta_{ij} = \text{the expansion factor} \tag{25}$$

by which points P_i and P_j should be moved apart, ignoring for the moment all other pairs of points which include either P_i or P_j. It is desired to increase the magnitude of $\vec{A}_{ij}$ by the factor Δ_{ij}. Thus the point P_i should be moved a distance $\Delta_{ij}\, d_{ij}/2$ in the direction $-\vec{A}_{ij}$, and P_j should be moved a distance $\Delta_{ij}\, d_{ij}/2$ in the direction $+\vec{A}_{ij}$.

The final coordinates for the point P_j are thus given by

$$a_{jk} + \frac{\Delta_{ij}}{2} (a_{jk} - a_{ik}). \tag{26}$$

When the entire collection of points is considered, the shifting of the point P_j is governed by the vector sum of the $\vec{A}_{ij}$ weighted by the corresponding $\Delta_{ij}/2$, as shown in Figure 6. In this case the final position of P_j is

$$a_{jk} + 1/2 \sum_{i=1}^{m} (a_{jk} - a_{ik}) \Delta_{ij} \tag{27}$$

where m is the total number of points.

The second process restores the ranking of interpoint distances within a small spherical region local to each point, of radius $\beta\bar{d}$. To initially determine this set of inequalities consider the point P_i. The distances to each other point, d_{ij} for $i \neq j$, are known from the calculations for the first process. Those points for which $d_{ij} > \beta\bar{d}$ are ignored, and for the remaining points, the interpoint distances are ranked. This is repeated for each point in the collection. In all m chains of inequalities are obtained, with lengths depending on the choice of the coefficient β.

$2\beta\bar{d}$ is the diameter of a hypersphere of dimension n centered on the point P_i. Consider the simple case where this hypersphere includes 3 points, which will be designated P_i, P_k, and P_j. There are three interpoint distances, d_{ij}, d_{ik}, and d_{kj}. Suppose further that

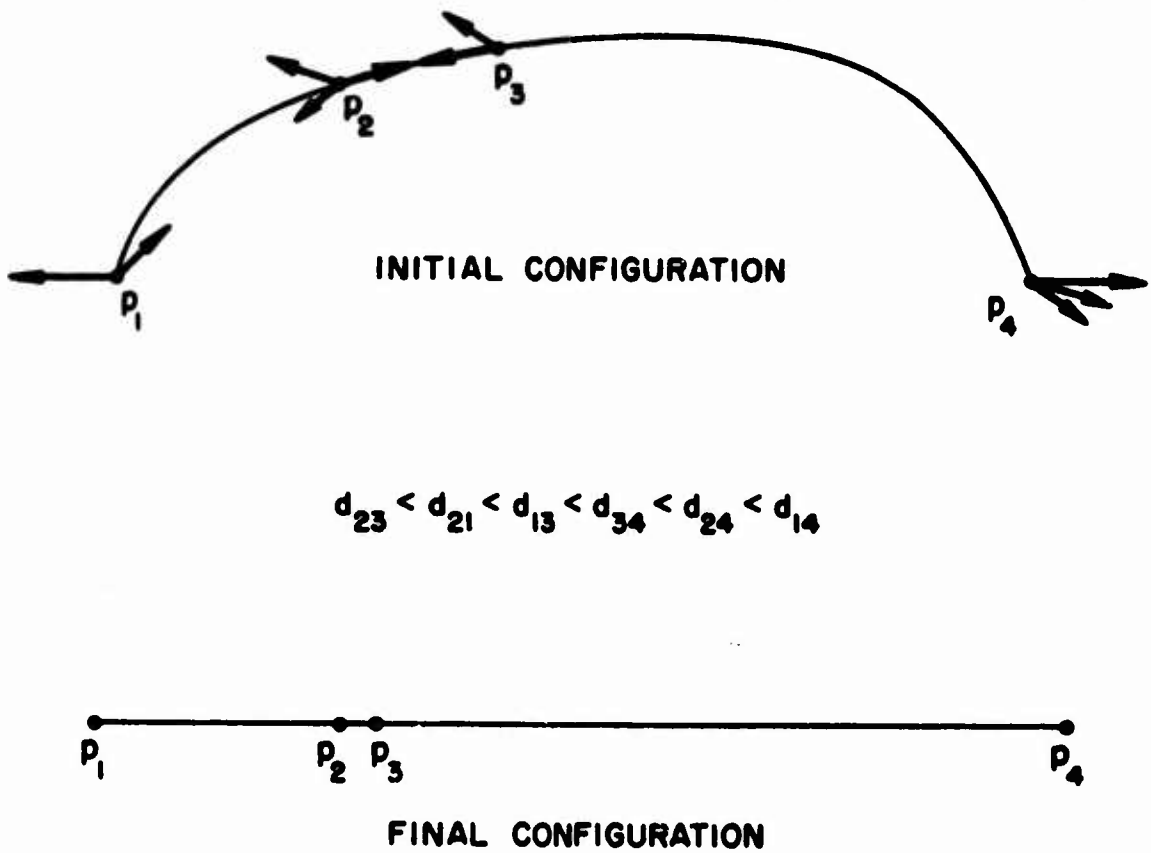


Figure 6 Collapsing of a Curve into a Line

$d_{ij} < d_{ik} < d_{kj}$. (28) Since the goal of this process is to preserve this inequality, but not the actual values of the d_{ij}'s, etc., the d's are now replaced by the numerical values of their rank in the inequality, R_{abc}, where a designates the point at the hypersphere center, and b, c designate the two points whose interpoint distance is being ranked. For example, in the case above, a = i and

$$\begin{aligned} R_{ikj} &= 1 \\ R_{iik} &= 2 \\ R_{iij} &= 3 \end{aligned} \tag{29}$$

For a hypersphere containing N points, the values of R_{abc} will run from 1 to $(N^2 - N)/2$ since $d_{ij} = d_{ji}$ and the d_{ii} are zero and not ranked. The value of a runs from 1 to m, the number of points in the collection.

Once the initial values for the R's have been determined, the hyperspheres will no longer be required and all further calculations will use the same values of b and c for each a.

After each iteration of the variance-increasing process, the ranking procedure is repeated, i.e., for the example where $R_{ikj} = 1$, $R_{iik} = 2$, and $R_{iij} = 3$, the distances between points are computed and ranked after the variance-increasing shifts have taken place, and these ranks are designated R'_{ikj}, R'_{iik}, and R'_{iij}. If within the ith chain of inequalities no violations of the original ranking have

occurred,

$$R_{ikj} = R'_{ikj} = 1$$

$$R_{iik} = R'_{iik} = 2 \qquad (30)$$

$$R_{iij} = R'_{iij} = 3$$

If, on the other hand, the inequalities no longer hold, then for at least two pairs of points $R \neq R'$. The more violent the scrambling of the inequalities, the greater will be the difference $R - R'$. This difference will be symbolized by

$$D_{ijk} \triangleq (R_{ijk} - R'_{ijk})\gamma \qquad (31)$$

where γ is a constant.

When $D_{ijk} = 0$, the ranking of the interpoint distance d_{jk} is correct. For positive integer values of D_{ijk}, the distance between points P_j and P_k is too great, and for negative values it is too small. The procedure for restoring proper ranking is very similar to that used to increase variance, except that $-D_{ijk}$ will be used in place of Δ_{jk}. The minus sign arises as a simple consequence of the definition of D_{ijk}. Also, an additional summation will be needed to accommodate the extra subscript.

Consider a pair of points, P_j and P_k within a hypersphere of radius $\beta\bar{d}$ centered on point P_i before the first iteration. To tend to restore the distance between P_j and P_k to its proper rank in the ith

inequality chain, the coordinates for the point P_j are shifted from a_{jl} to

$$a_{jl} - \frac{D_{ijk}}{2} (a_{jl} - a_{kl}) \tag{32}$$

In the event that the interpoint distance d_{ij} is out of place in several of the inequality chains, the i subscript will be summed over the collection of points. This yields as the final set of coordinates for point P_j,

$$a_{jl} \rightarrow a_{jl} - \sum_{i=1}^{m} \frac{D_{ikj}}{2} (a_{jl} - a_{kl}) \tag{33}$$

This yields the final position of P_j provided that only its distance from P_k is considered. In the desired application, the distances from other points must be considered as well, and the shifting of P_j will depend on the vector sum of all possible D's. This vector sum is taken in the same way as in the variance increasing process, and the required final position for P_j is given by:

$$a_{jl} \rightarrow a_{jl} - \sum_{i=1}^{m} \sum_{k=1}^{m} \frac{D_{ikj}}{2} (a_{jl} - a_{kl}) \tag{34}$$

For each value of i, there will be many values of k and j for which D_{ikj} is not defined because either point P_k or P_j, or both, lie outside the initial hypersphere of radius $\beta \bar{d}$. To facilitate computation, these D_{ikj} will be defined to be zero.

Iterating these two processes should eventually lead to a configuration which no longer changes, the two processes cancelling out. Further iterations may rotate the configuration, but the "shape" should remain fixed, as should the variance in interpoint distances.

Once the final configuration has been obtained, it remains to identify the linear dimensionality of this configuration. A method for doing this is based on two theorems from matrix theory. (14)

Theorem 1 The rank of a normal matrix is equal to the number of non-zero eigenvalues possessed by the matrix.

Theorem 2 The rank of any Gramian matrix of vectors is equal to the linear dimensionality of the space spanned by the vectors. Combining these theorems yields:

Corollary The linear dimensionality of the space spanned by a set of vectors is equal to the number of non-zero eigenvalues possessed by the Gramian matrix.

In the case of a matrix whose elements are inner products, the requirement that the matrix be normal and positive semidefinite is automatically filled.

The remaining part of the procedure to be followed is as follows:

Starting from any point j, calculate the matrix of inner products of vectors from that point to every point in the configuration

$$\langle r_{jk} | r_{jl} \rangle = b'_{kl}(j)$$

IR-7 163

The number of non-zero eigenvalues of this matrix will then be equal to the linear dimensionality of the configuration, and therefore will be equal to the intrinsic dimensionality of the original collection of signals. Taking into account that $|r_{jj}\rangle \equiv 0$, and that

$$\langle \tilde{r}_{jk} | r_{jl} \rangle = \langle \tilde{r}_{jl} | r_{jk} \rangle$$

the matrix B' is

$$\begin{bmatrix} \langle r_{j1}|r_{j1}\rangle & \langle \tilde{r}_{j2}|r_{j1}\rangle & \ldots 0 \ldots & \langle \tilde{r}_{jm}|r_{j1}\rangle \\ \langle \tilde{r}_{j2}|r_{j1}\rangle & \langle r_{j2}|\tilde{r}_{j2}\rangle & \ldots 0 \ldots & \langle \tilde{r}_{jm}|r_{j2}\rangle \\ 0 & 0 & \ldots 0 \ldots & 0 \\ \langle \tilde{r}_{jm}|r_{j1}\rangle & \ldots\ldots\ldots\ldots & 0 \ldots & \langle \tilde{r}_{jm}|r_{jm}\rangle \end{bmatrix} = B' \tag{35}$$

The equation to be solved is

$$B' - I\lambda = 0 \tag{36}$$

For m signals projected on an n dimensional basis, at least (m-n) of the λ's so obtained must be zero. This will be verified in Chapter VI. Of the remaining n eigenvalues, k will be non-zero. The matrix B' will of course be different for different choices of the jth point, but the final number of non-zero eigenvalues, k, will be constant. Several standard techniques for finding eigenvalues of

TR #F63

symmetric matrices are available. The technique used in this work is a variant of Jacobi's method. During the variance maximizing process, the centroid of the configuration, C, is not constrained. This is of no consequence as far as the final dimensionality is concerned, but it is convenient that the centroid of the configuration be used as the origin of the vectors rather than arbitrary point j. A formula which can be used to calculate the inner product matrix, using the centroid as the vector origin is given in Equation 37. (15)

$$b_{jk} = \frac{1}{2m} \left[\sum_{j=1}^{m} d_{jk}^2 + \sum_{k=1}^{m} d_{jk}^2 - \frac{2}{n} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} d_{jk}^2 - n\, d_{jk}^2 \right] \tag{37}$$

where: b_{jk} = element of matrix B of scalar products referred to centroid

$$b_{jk} = \langle \tilde{r}_{cj} | r_{ck} \rangle \tag{38}$$

d_{jk} = distance between j^{th} and k^{th} point

m = total number of points in the configuration.

It is now possible to set forth an outline for a computer program which will accept the coefficients of a collection of signals expanded on any basis, and from these coefficients estimate the intrinsic dimensionality of the collection.

V. OUTLINE OF COMPUTER ROUTINE

Step Number	Step Description
1	Set values for α, β, γ, and ϵ. Note: α determines the rate at which the interpoint distance variance increases; β defines the region about each point over which interpoint distance ranking is to be preserved; γ determines the vigor with which the program resists inequality violations; and ϵ sets the stopping criterion.
2	Read in coefficients of signals expanded on some orthonormal basis. Coordinates of the jth signal on the kth basis vector $\equiv a_{jk}$ where $j = 1, 2, \ldots m$; $k = 1, 2, \ldots n$.
3	Calculate interpoint distances $$d_{ij} = \left[\sum_{k=1}^{n} (a_{ik} - a_{jk})^2 \right]^{1/2} \qquad (39)$$
4	Calculate the arithmetic mean of $d_{ij} \triangleq \bar{d}$ $$\bar{d} = \sum_{i=1}^{m} \sum_{j \neq 1} d_{ij}/(m^2 - m) \qquad (40)$$

TR 71-163

Step Number	Step Description
5	Normalize data to make $\bar{d} = 1$.
6	Compute variance in normalized $d_{ij} \stackrel{\Delta}{=}$ VAR
7	VARO = variance from previous iteration.
	$\lvert$VARO - VAR $\rvert < \varepsilon$?
	Yes, go to step 20. No, go to step 8.
8	First iteration ?
	Yes, go to step 9, No, go to step 11.
9	For each point i, store numbers of point j, k such that both $d_{ij} < \beta\bar{d}$ and $d_{ik} < \beta\bar{d}$. $d_{ii} \equiv 0$, $\bar{d} = 1$ by 5. (See Figure 7)
10	Rank all possible pairs of points obtained in step 9 in decreasing order of d_{ij}, and calculate R_{ijk}.
11	Calculate Δ_{ij}

$$\Delta_{ij} = \alpha(d_{ij} - \bar{d})/\bar{d} = \alpha(d_{ij} - 1). \tag{25}$$

12	Compute new coordinates.

$$a_{jk} \rightarrow a_{jk} + \frac{1}{2} \sum_{i=1}^{m} (a_{jk} - a_{ik}) \Delta_{ij} \tag{27}$$

13	Compute new values of d_{ij}.

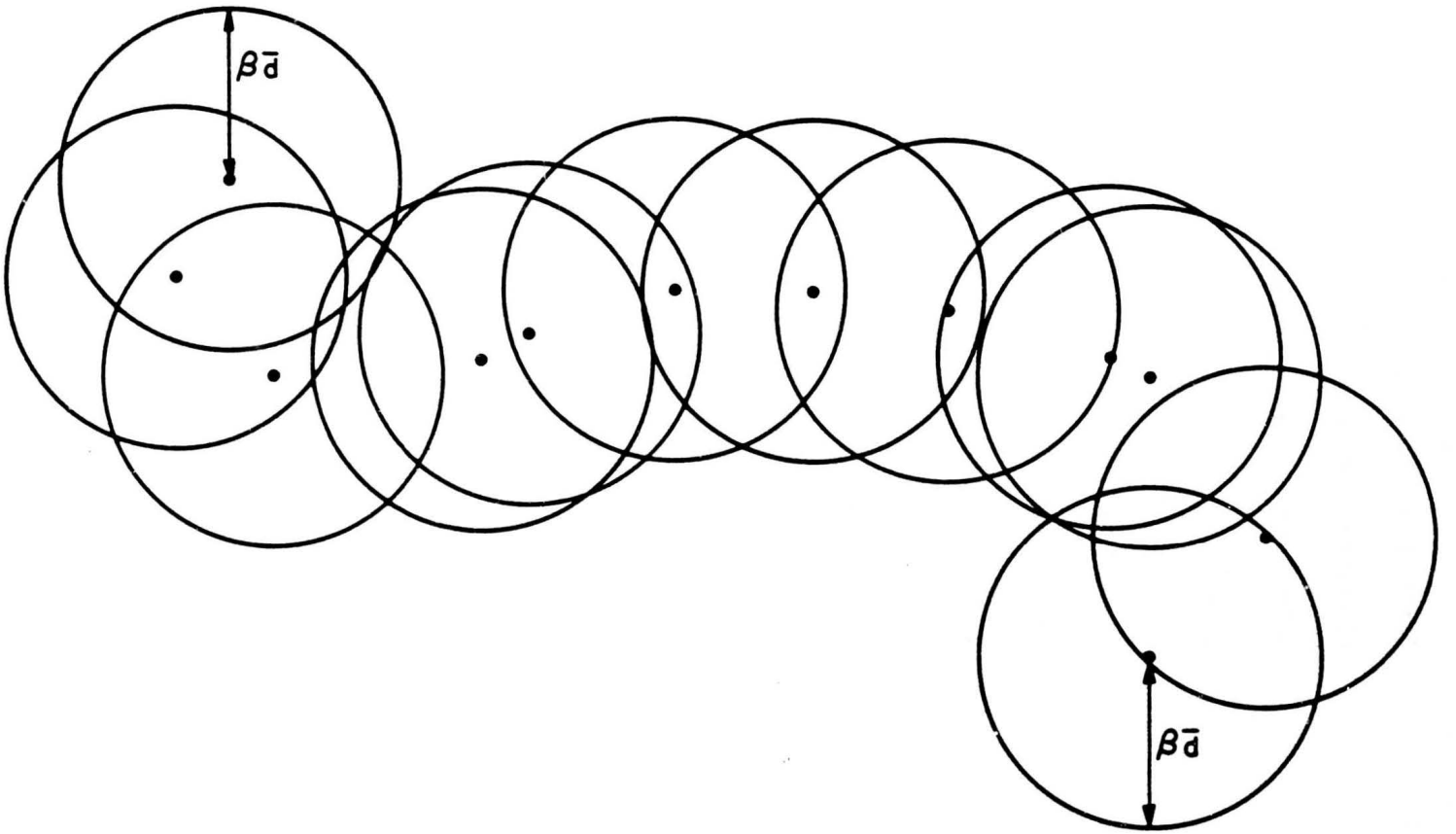


Figure 7 Regions Defined by Beta

Step Number	Step Description
14	Compute new value of $\bar{d}$.
15	Normalize a_{jk}'s to set new $\bar{d} = 1$.
16	Compute R'_{ijk}.
17	Calculate $D_{ijk} = \gamma(R_{ijk} - R'_{ijk})$ (31)
18	Compute new coordinates.

$$a_{jl} \rightarrow a_{jl} - \sum_{i=1}^{m} \sum_{k=1}^{m} \frac{D_{ijk}}{2} (a_{jl} - a_{kl}) \tag{34}$$

19	Return to step 3.
20	Calculate B matrix

$$b_{jk} = \frac{1}{2m} \left[\sum_{j=1}^{m} d^2_{jk} + \sum_{k=1}^{m} d^2_{jk} - \frac{2}{n} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} d^2_{jk} - nd^2_{jk} \right] \tag{37}$$

21	Find eigenvalues of B, using the method of Jacobi. This is a standard subroutine.
22	Write out final coordinates, D matrix, B matrix.

Step Number	Step Description
23	Write out eigenvalues of B.
24	End.

A simplified flow chart for this program appears in Figure 8.

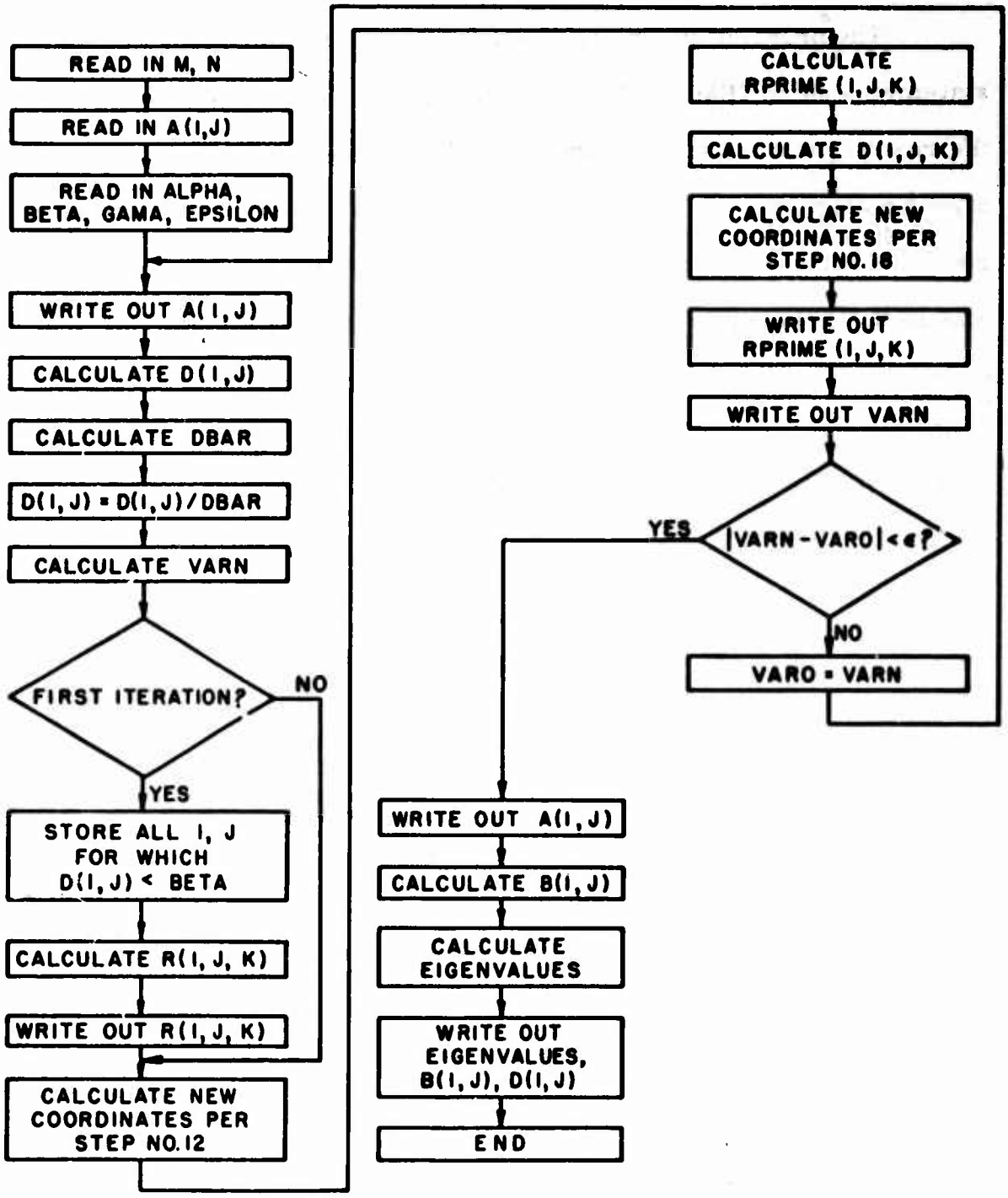


Figure 8 Simplified Flow Chart

VI. EXPERIMENTAL VERIFICATION

The program previously outlined was compiled from Fortran statements on an IBM 7094 digital computer. This chapter describes the results obtained for several examples of one and two-dimensional signal collections. In all cases, the basis used consists of one-sided, real, decaying exponentials, orthogonalized on the interval 0 - ∞. The exponents in the basis are -t, -2t, -3t, -4t, and -5t. These exponentials may be orthogonalized by either Schmidt's method or the method of Kautz. (16) In the time domain, the resulting orthonormal basis is described by Equation 41.

$$\phi_1 = (2)^{1/2} (e^{-t})\, u(t) \tag{41}$$

$$\phi_2 = (4)^{1/2} (-2e^{-t} + 3e^{-2t})\, u(t)$$

$$\phi_3 = (6)^{1/2} (3e^{-t} - 12e^{-2t} + 10\, e^{-3t})\, u(t)$$

$$\phi_4 = (8)^{1/2} (-4e^{-t} + 30e^{-2t} - 60e^{-3t} + 35e^{-4t})\, u(t)$$

$$\phi_5 = (10)^{1/2} (5e^{-t} - 60e^{-2t} + 210e^{-3t} - 280e^{-4t} + 126e^{-5t})\, u(t)$$

or equivalently,

$$\Phi_1(s) = \frac{1}{s+1} \cdot (2)^{1/2}$$

$$\Phi_n(s) = \left[\Phi_{n-1}(s) \left(\frac{s+1-n}{s+n} \right) \right] \left[\frac{n}{n-1} \right]^{1/2} \tag{42}$$

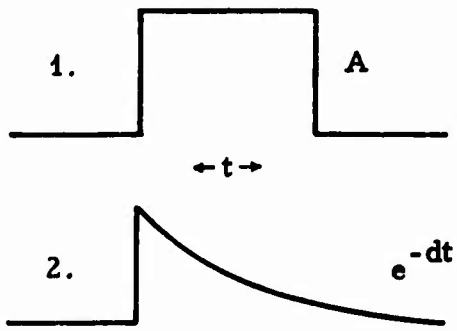
for $n \neq 1$

TR 1 63

This basis will be referred to as the Kautz basis in the remainder of the paper. The Kautz basis was chosen for the tests because of the ease with which the coefficients of signal expansions may be computed.

The signals used are of two types:

1. Rectangular pulses
2. Decaying real one-sided exponentials



All signals are of the single epoch type. Inner products of the Kautz basis with the two basic signals used in testing this program are tabulated below.

1. $s(t) = u(t)$, the unit step beginning at $t = 0$

$$a_1 = \langle \tilde{\phi}_1 | s \rangle = 1.414 \tag{43}$$

$$a_2 = \langle \tilde{\phi}_2 | s \rangle = -1.000$$

$$a_3 = \langle \tilde{\phi}_3 | s \rangle = .815$$

$$a_4 = \langle \tilde{\phi}_4 | s \rangle = -.707$$

$$a_5 = \langle \tilde{\phi}_5 | s \rangle = .632$$

2. $s(t) = u(t-\tau)$, the unit step beginning at $t = \tau$

$$a_1 = (2)^{1/2}\, e^{-\tau}$$

$$a_2 = -(4)^{1/2}\,(2e^{-\tau} - 3/2\, e^{-2\tau})$$

$$a_3 = (6)^{1/2}\,(3e^{-\tau} - 6e^{-2\tau} + 10/3\, e^{-3\tau})$$

$$a_4 = -(8)^{1/2}\,(4e^{-\tau} - 15e^{-2\tau} + 20\, e^{-3\tau} - 35/4\, e^{-4\tau})$$

$$a_5 = (10)^{1/2}\,(5e^{-\tau} - 30e^{-2\tau} + 70e^{-3\tau} - 70e^{-4\tau} + \frac{126}{5}\, e^{-5\tau})$$

3. $s(t) = e^{-dt}\, u(t)$, the one sided decaying real exponential having epoch at $t = 0$.

$$a_1 = (2)^{1/2}\left(\frac{1}{1+d}\right)$$

$$a_2 = (4)^{1/2}\left(\frac{-2}{1+d}\right) + \frac{3}{2+d})$$

$$a_3 = (6)^{1/2}\left(\frac{3}{1+d}\right) - \frac{12}{2+d} + \frac{10}{3+d}\Big)$$

$$a_4 = (8)^{1/2}\left(\frac{-4}{1+d}\right) + \frac{30}{2+d} - \frac{60}{3+d} + \frac{35}{4+d}\Big)$$

$$a_5 = (10)^{1/2}\left(\frac{5}{1+d}\right) - \frac{60}{2+d} + \frac{210}{3+d} - \frac{280}{4+d} + \frac{126}{5+d}\Big)$$

In the initial runs, a cutoff limit of 20 iterations was written into the program to insure against excessive wasted time in the event that the configuration did not converge.

TR 71-163

As was stated in Chapter IV, the matrix equation

$$B' - I\lambda = 0 \tag{36}$$

has m eigenvalues, of which at least (m-n) should be zero. As verification that this part of the program was correct, a collection of 20 random decaying exponentials was expanded on the Kautz basis, and the B matrix was calculated directly from this input data, without the collapsing operation being performed. The total number of eigenvalues so obtained should have been 20, of which 15 should have been zero. The eigenvalues calculated are shown in Table 2.

TABLE 2
Eigenvalues

1.	7.33867	11.	0.00000
2.	1.93994	12.	0.00000
3.	1.49168	13.	0.00000
4.	0.32322	14.	0.00000
5.	0.14997	15.	0.00000
6.	0.00000	16.	0.00000
7.	0.00000	17.	0.00000
8.	0.00000	18.	0.00000
9.	0.00000	19.	0.00000
10.	0.00000	20.	0.00000

TR-/ I 63

This proved to be the case in every run, that is, eigenvalues 6 to 20 were always zero. Therefore, in the examples to follow, only the first five eigenvalues are shown.

In Chapter IV, the postulate was put forth that a single parameter class of signals should generate a locus of points in signal space which describes a curve. To gain some insight into the shape such a curve might have, a set of 20 rectangular pulses was expanded on the Kautz basis. The pulse width was varied from 0.1 to 2.0 seconds, and the height was held constant at unity. It is, of course, not possible to depict the resulting curve on a single plot, but three projections of the curve, on the ϕ_1-ϕ_3, ϕ_2-ϕ_4, and the ϕ_2- ϕ_5 planes are shown in Figures 9, 10, and 11. These projections show the great nonlinearity of the curve, and also indicate that about a 10 to 1 signal to noise energy ratio might be expected to cause the program to collapse the configuration to a line.

In the examples to follow, signals of one or two parameters were expanded on the Kautz basis, and their intrinsic dimensionalities were determined by the program. The same set of control parameters, α, β, γ, and ϵ were used in all cases. Examples 1 and 3 were single parameter classes of signals, while Examples 2 and 4 had two independently varying parameters. Example 5 shows the effect of dependence in the variation of two parameters. It is essentially the same class of signals as in Example 2, but the program identifies it as a single-parameter class due to the dependent parameter variation.

TR 7 F63

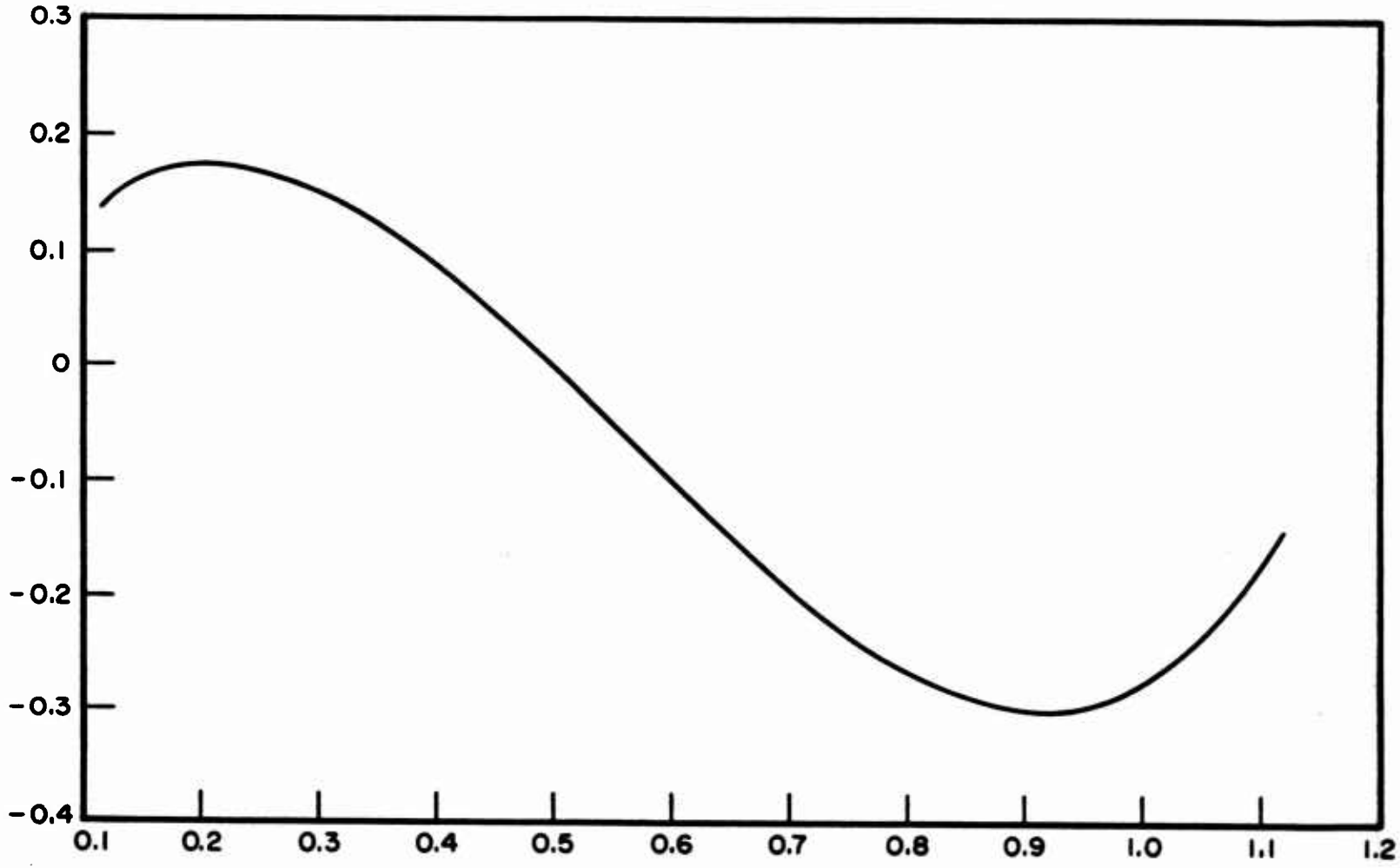


Figure 9 a_1 vs. a_3

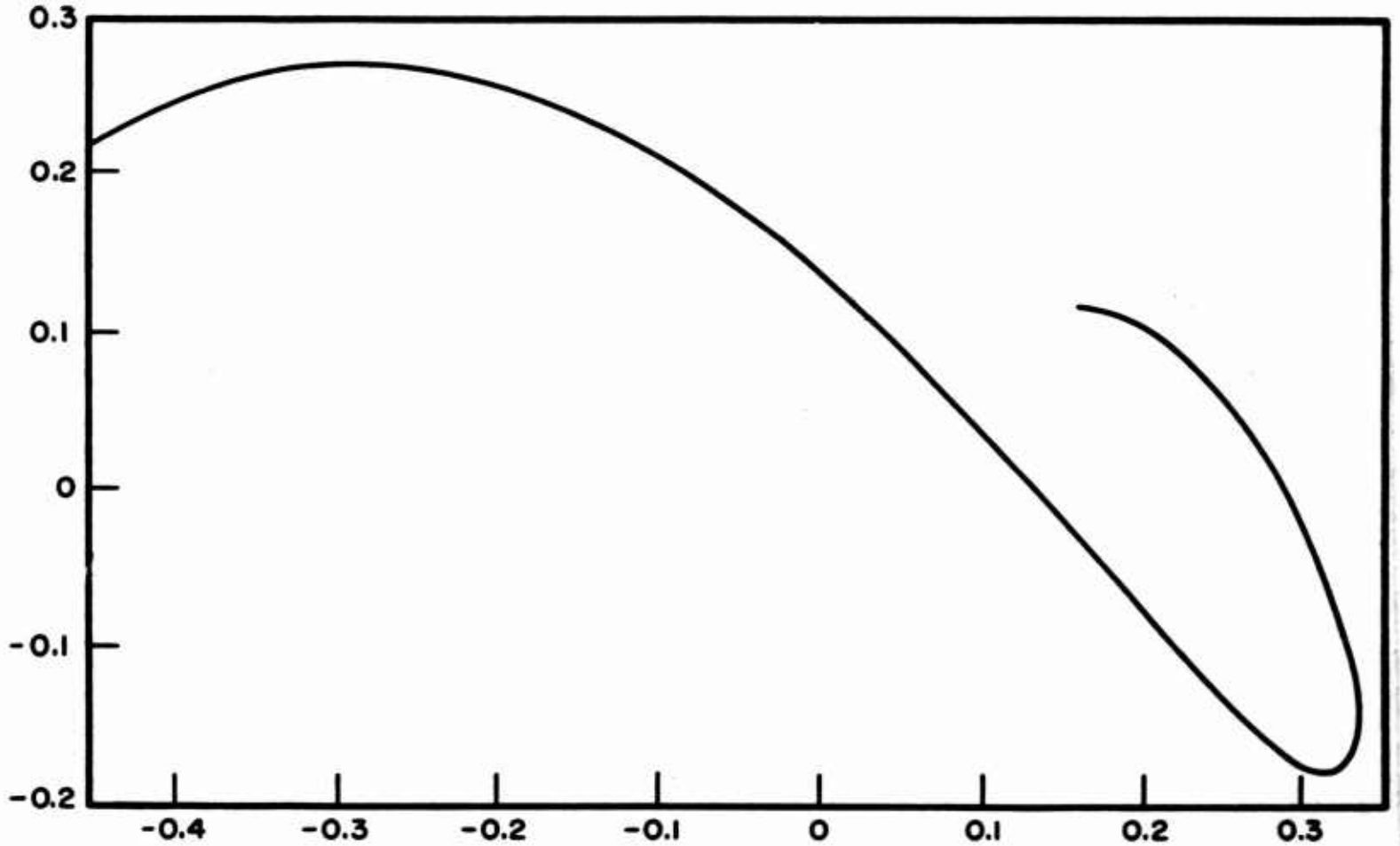


Figure 10 a_2 vs. a_4

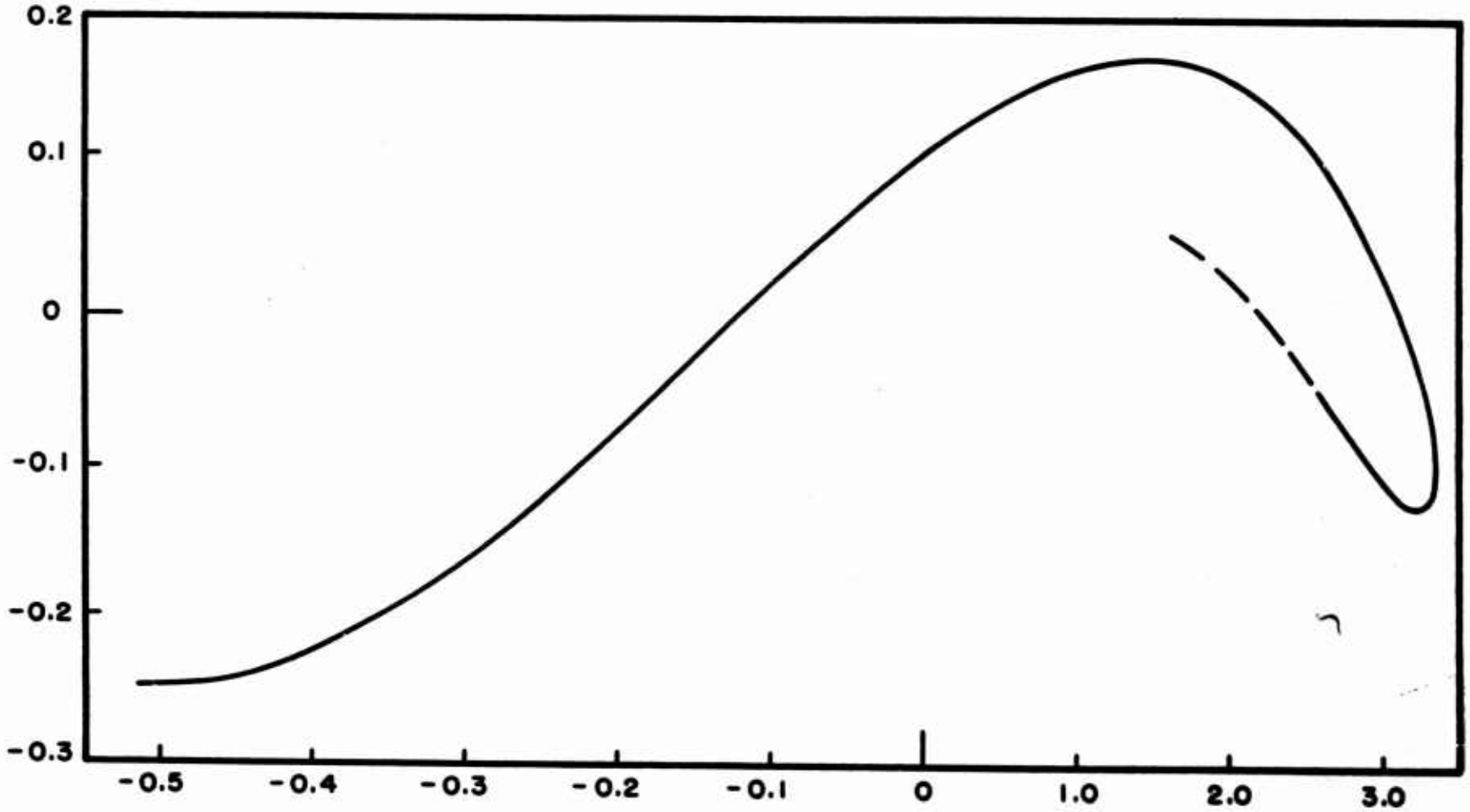


Figure 11 a_2 vs. a_5

A. Example 1

Signal used - Rectangular pulses of unit amplitude with width τ varied from 0.1 to 2.0 seconds in equal steps. (This is the signal collection depicted in Figures 9, 10, and 11) 20 signals.

Parameter values: $\alpha = 0.05$ $\gamma = 0.04$

$\beta = 0.75$ $\epsilon = 0.005$

Number of iterations required = 8

Running time = 3 hundredths hour

Eigenvalues obtained: 35.914352
.000008
.000000
.000000
.000000

TR 71 163

EXAMPLE 1

Coefficients on Kautz Basis

Signal Number	a_1	a_2	a_3	a_4	a_5
1	.1346	.1632	.1514	.1116	.0578
2	.2564	.2640	.1707	.04165	-.0582
3	.3665	.3168	.1188	-.0649	-.1230
4	.4662	.3333	.0352	-.1453	-.1044
5	.5564	.3225	-.0557	-.1800	-.3306
6	.6381	.2917	-.1395	-.1710	+.0521
7	.7119	.2466	-.2083	-.1289	.1226
8	.7788	.1916	-.2588	-.0663	.1638
9	.8392	.1304	-.2905	.0052	.1722
10	.8940	.0655	-.3043	.0763	.1514
11	.9435	-.0009	-.3023	.1405	.1086
12	.9883	-.0674	-.2866	.1937	.0523
13	1.029	-.1327	-.2599	.2336	-.0098
14	1.065	-.1960	-.2243	.2597	-.0712
15	1.099	-.2568	-.1822	.2721	-.1268
16	1.129	-.3147	-.1352	.2721	-.1734
17	1.156	-.3694	-.0852	.2611	-.2092
18	1.180	-.4208	-.0335	.2408	-.2334
19	1.203	-.4688	+.0189	.2128	-.2460
20	1.223	-.5136	+.0709	.1789	-.2479

B. Example 2

Signals used - Rectangular pulses with both amplitude, A, and width, τ, varied as shown below:

Signal Number	τ	A	Signal Number	τ	A
1	0.1	0.6	11	0.1	1.0
2	0.4	0.6	12	0.4	1.0
3	0.7	0.6	13	0.7	1.0
4	1.0	0.6	14	1.0	1.0
5	1.3	0.6	15	1.3	1.0
6	0.1	0.8	16	0.1	1.2
7	0.4	0.8	17	0.4	1.2
8	0.7	0.8	18	0.7	1.2
9	1.0	0.8	19	1.0	1.2
10	1.3	0.8	20	1.3	1.2

Parameter Values: $\alpha = 0.05$ $\gamma = 0.04$

$\beta = 0.75$ $\epsilon = 0.005$

Number of iterations required = 5

Running time = 3 hundredths hour

Eigenvalues obtained: 27.56159

1.19406

.00001

.00000

.00000

TR 7 163

EXAMPLE 2

Coefficients on Kautz Basis

Signal Number	a_1	a_2	a_3	a_4	a_5
1	.0808	.0981	.0909	.0670	.0347
2	.2780	.1995	.0212	-.0872	-.0627
3	.4275	.1484	-.1250	-.0774	.0726
4	.5360	.0393	-.1827	.0458	.0909
5	.6170	-.0795	-.1559	.1401	-.0059
6	.1078	.1306	.1211	.0893	.0462
7	.3730	.2666	.0282	-.1162	-.0835
8	.5695	.1975	-.1666	-.1031	.0981
9	.7152	.0524	-.2434	.0610	.1211
10	.8232	-.1062	-.2079	.1869	-.0078
11	.1346	.1632	.1514	.1116	.0578
12	.4662	.3333	.0352	-.1453	-.1044
13	.7119	.2466	-.2083	-.1289	.1226
14	.8940	.0655	-.3043	.0763	.1514
15	1.029	-.1327	-.2599	.2336	-.0098
16	.1616	.1961	.1818	.1340	.0694
17	.5560	.3990	.0423	-.1744	-.1253
18	.8550	.2968	-.2500	-.1547	.1471
19	1.072	.0786	-.3655	.0916	.1818
20	1.234	-.1591	-.3119	.2802	-.0119

C. Example 3

Signals used - Real decaying one-sided exponentials of unit amplitude e^{-dt}

$\frac{1}{d}$ varied from 0.1 to 2.0 seconds in equal steps. 20 signals.

Parameter Values: $\alpha = 0.05$ $\gamma = 0.04$
$\beta = 0.75$ $\epsilon = 0.005$

Number of iterations required = 8

Running time = 4 hundredths hour

Eigenvalues obtained: 38.517118
.000010
.000000
.000000
.000000

TR 7163

EXAMPLE 3

Coefficients on Kautz Basis

Signal Number	a_1	a_2	a_3	a_4	a_5
1	.12856	.1364	.1030	.0594	.032
2	.2357	.1905	.0874	.022	.003
3	.32636	.2019	.0522	.001	0.
4	.40405	.1905	.0210	-.001	0.
5	.4714	.1667	0.	.0003	0.
6	.53033	.1363	-0.012	.0020	0.
7	.58232	.1029	-0.016	.0045	0.
8	.62853	.06840	-0.015	.0065	0.
9	.66990	.03380	-0.0091	.0040	0.
10	.70711	0.	0.	0.	0.
11	.74078	-.0327	0.011	-0.0051	0.
12	.77139	-.0642	0.024	-0.0120	0.
13	.79934	-.0942	0.0377	-0.0210	0.
14	.82495	-.1228	0.0519	-.0291	0.
15	.84853	-.1500	0.0669	-.0385	.03
16	.87028	-.1758	0.0816	-.0478	.03
17	.89043	-.2004	0.0968	-.0591	.06
18	.90914	-.2236	0.1110	-.0696	.06
19	.92655	-.2457	0.1260	-.0795	.06
20	.94281	-.2667	0.1400	-.0897	.06

TR 71163

D. Example 4

Signals used - Real decaying one-sided exponentials with decrement and amplitude varied as shown below.

Signal Number	d	A	Signal Number	d	A
1	10	0.6	11	10	1.0
2	2.5	0.6	12	2.5	1.0
3	1.5	0.6	13	1.5	1.0
4	1.0	0.6	14	1.0	1.0
5	.75	0.6	15	.75	1.0
6	10	0.8	16	10	1.2
7	2.5	0.8	17	2.5	1.2
8	1.5	0.8	18	1.5	1.2
9	1.0	0.8	19	1.0	1.2
10	.75	0.8	20	.75	1.2

TR-7163

Parameter Values: $\alpha = 0.05$ $\gamma = 0.04$

$\beta = 0.75$ $\epsilon = 0.0005$

Number of iterations required = 5

Running time = 3 hundredths hour

Eigenvalues obtained: 44.8376

.1519

.0010

.0000

.0000

EXAMPLE 4

Coefficients on Kautz Basis

Signal Number	a_1	a_2	a_3	a_4	a_5
1	.0772	.0818	.0618	.0356	.0192
2	.2425	.1143	.0126	-.0006	0.
3	.3494	.0617	-.0096	.0027	0.
4	.4243	0.	0.	0.	0.
5	.4796	-.0565	.0226	-.0126	0.
6	.1029	.1091	.0824	.0475	.0256
7	.3233	.1524	.0168	-.0008	0.
8	.4658	.0823	-.0128	.0036	0.
9	.5657	0.	0.	0.	0.
10	.6394	-.0753	.0302	-.0168	0.
11	.1286	.1364	.1030	.0594	.0320
12	.4041	.1905	.0210	-.0010	0.
13	.5823	.1029	-.0160	.0045	0.
14	.7071	0.	0.	0.	0.
15	.7993	-.0942	.0377	-.0210	0.
16	.1543	.1636	.1236	.0712	.0384
17	.4849	.2286	.0252	-.0012	0.
18	.6988	.1235	-.0192	.0054	0.
19	.8485	0.	0.	0.	0.
20	.9592	-.1130	.0452	-.0252	0.

TR / 163

E. Example 5

Signals used - Rectangular pulses with both amplitude and width varied, but with $\tau = 2 \sin A$, and τ varied from 0.1 to 2.0 seconds in equal steps. 20 signals.

Parameter Values: $\alpha = 0.05$ $\gamma = 0.04$

$\beta = 0.75$ $\epsilon = 0.005$

Number of iterations required = 8

Running time = 4 hundredths hour

Eigenvalues obtained: 41.03028
.00001
.00000
.00000
.00000

EXAMPLE 5

Coefficients on Kautz Basis

Signal Number	a_1	a_2	a_3	a_4	a_5
1	.0067	.0082	.0076	.0056	.0029
2	.0256	.0264	.0171	.0042	-.0058
3	.0553	.0478	.0179	-.0098	-.0186
4	.0942	.0673	.0072	-.0294	-.0211
5	.1408	.0816	-.0141	-.0455	-.0836
6	.1946	.0889	-.0426	-.0522	-.0159
7	.2549	.0883	-.0746	-.0462	.0439
8	.3201	.0788	-.1064	-.0272	.0673
9	.3911	.0608	-.1354	.0024	.0803
10	.4681	.0343	-.1593	.0400	.0793
11	.5501	-.0005	-.1762	.0819	.0633
12	.6365	-.0434	-.1846	.1247	.0337
13	.7275	-.0938	-.1837	.1652	-.0069
14	.8243	-.1517	-.1736	.2010	-.0551
15	.9309	-.2175	-.1543	.2305	-.1074
16	1.045	-.2914	-.1252	.2520	-.1606
17	1.174	-.3750	-.0865	.2651	-.2124
18	1.322	-.4713	-.0375	.2697	-.2614
19	1.508	-.5877	.0237	.2668	-.3084
20	1.921	-.8068	.1110	.2810	-.3894

TR 1163

These results are in complete agreement with the "correct" results. Cases 1 and 3, the single-parameter cases, produced single non-zero eigenvalues. Cases 2 and 4, with two independently varying parameters, produced two non-zero eigenvalues. Case 5, with two varying parameters, produced a single eigenvalue. This is in agreement with the discussion in Chapter III, Equation 12.

TR-4163

VII. DISCUSSION

A. Effect of Computer Parameters

The discussion and outlining of the computer program in Chapters IV and V avoided comment on the effects of the parameters α, β, and γ on the final results. A few such comments follow. Of course, the values of these "constants" will depend to an extent on the input data, and since this cannot be predicted, some adjustment may be necessary during the computation. From Shepard's experience with the Multidimensional Scaling problem, (13) a few indications as to reasonable starting values for α and γ were obtained. The value of β had to be considered separately, as nothing in multidimensional scaling corresponds to β.

1. Alpha

α determines the rate at which the interpoint distance variance increases, or, equivalently by Equation 16, the rate at which the dimensionality of the configuration decreases. Large values of α collapse the structure more rapidly, but unfortunately also do violence to the ranking of distances within the inequality chains. It is therefore prudent to use small values for α, and a few more iterations. To this end, Shepard used program parameter values corresponding to values of α from 0.01 to 0.05. Since the goals of the program described here and Shepard's program are widely different, a one-to-one correspondence in parameters does not exist, and the values $\alpha = 0.01$ to 0.05 should be used only as a starting point. For the 20-signal examples

TR-163

of Chapter VI, $\alpha = 0.05$ was satisfactory, although this may be too high if more than 20 signals are included in the collection.

2. Gamma

As a striking example of the danger in blindly using Shepard's parameter values, consider γ, the parameter determining the vigor with which the program resists inequality violations. Shepard uses 0.2 for this parameter, while in the intrinsic dimensionality program, instabilities resulted if γ exceeded 0.05. In multidimensional scaling, only one chain of inequalities is used, while here the number varies as an inverse function of β. The result is that an inversion of rank in any one chain of inequalities is likely to occur in several others. This will multiply the effect of γ several times, and since γ is a feedback parameter, instability may result if γ is not held to a low value. $\gamma = 0.04$ proved satisfactory for the experimental work in Chapter VI.

3. Beta

Now the second parameter, β, is more difficult to discuss. Perhaps the best way to approach it is to look at an analogy with a parameter in "clustering" problems. (17, 18) Young, in performing an analysis on a collection of signals, considers their first-order correlations, and forms a cross-correlation matrix, symmetric with 1's on the diagonal. If any element exceeds a threshold, it is replaced by unity, if not, with zero. All further work is then done with this matrix. Thus the matrix, and hence the clusters obtained, depend

heavily on the choice of threshold. All that Young could suggest was that if N is the number of clusters obtained and η was the threshold $0 < \eta < 1$, then if the function $N(\eta)$ showed small values of first derivative for a wide range of η, then the threshold should be put in this range.

Now, a similar phenomenon will be observed as the value of β is varied. For very small β, each chain of inequalities contains only a few distances to be ranked. The configuration is very loosly defined and may collapse into a line no matter what the original configuration. On the other hand, large values of β place nearly every possible distance, d_{ij}, in each chain. From Shepard's results this means that the program would repeatedly reproduce the original data with no collapsing.

This may be thought of as arising from the discrete nature of the input data. Through any finite collection of points in space, many curves, surfaces, etc., may be passed. The problem is one of finding a "best" number of dimensions for a discrete collection in the sense that as the number of signal samples increases and the collection of points begins to approximate a continuum, it is desired that the dimensionality of that continuum be the same as the "best" estimate. One approach is to follow Young, and determine whether the value of k is invariant under a wide variation of β. If so, the midpoint of this range of β should be used, and the resulting value of k taken as the best estimate of the true value. In this manner, the value of β for the examples of Chapter VI was set at 0.75.

B. Future Work

Extensions of this work are needed in several areas. The first of these is the study of the effect of noise added to the input data. This was briefly discussed in Chapter III, but a complete investigation has not been undertaken. It is expected that a threshold effect will occur when additive noise is present in the data, and both theoretical and experimental work should be performed to verify or disprove this supposition. Various probability distributions for the noise should be considered, as well as the relationship between signal-to-noise ratio; number of signal samples; and frequency of errors in estimating dimensionality.

Another area of needed research involves Assumption 2 of Chapter III, Equation 11. This is the assumption which involves a complete independence of the parameter variations in the signal generator. It has been stated that independent variations will yield the correct results, while two parameters which are functionally dependent will appear to the computer program to be a single parameter, thus reducing the apparent dimensionality by one. No mention has been made of a possible statistical relationship between two parameters. This would be the case where the statistics of the variation of one parameter are modified in accordance with the value of a second parameter. Errors here would depend on the correlation between the two variations, and an investigation of such errors would be of value.

Another important area for more work is the question of stability of the computer program. As previously mentioned, β is a feedback parameter, and wherever feedback is involved, the stability problem arises. In early runs with the program described here, convergence was not obtained. A reduction in the value of beta corrected the situation, and convergence was obtained. A means for eliminating this trial-and-error process, or at least automating it, would be of value. Excessive values of β resulted in divergence of the configuration, accompanied by a sharp decrease in the interpoint distance variance. This decrease was monitored, and when it occurred, the program was stopped. It may be possible to continuously monitor the variance changes from one iteration to the next, and modify β as the computer run progresses. This is only an ad hoc approach, however, and a full investigation of stability criteria would result in the saving of considerable computer time which must be wasted now before a satisfactory value for β is established for each set of data. Such an investigation should be carried out in conjunction with the additive noise investigation because the stability of the program will almost certainly be dependent on the noise present in the input data.

Of course, the most important future work with this program will be in applying it to actual signal analysis problems. A few such problems are suggested in the conclusion.

C. Conclusion

A definition of the intrinsic dimensionality of a signal collection has been formulated, and a computer program capable of evaluating this attribute of the collection has been written and evaluated. The program for making these estimations has several advantages over the conventional linear signal analysis techniques.

1. It uses the data points themselves, not an inferred continuous plot.
2. It is insensitive to changes in the basis on which the signals are projected.
3. It assumes no prior knowledge of the signals, except that most of their energy can be represented on the basis used.

REFERENCES

1. Huggins, W. H., "Representation and Analysis of Signals", Part I, "The Use of Orthogonalized Exponentials", Report No. AFCR-TR-57-357, The Johns Hopkins University, September, 1957.

2. Lai, D. C., "Representation and Analysis of Signals", Part VI, "Signal-Space Concepts and Dirac's Notation", Report No. AFCR-TN-60-167, The Johns Hopkins University, January, 1960.

3. Hartley, R. V. L., Bell System Technical Journal, Volume 7, Page 535, 1928.

4. Gabor, D., Journal of the Institute of Electrical Engineers, Part III, Page 429, 1946.

5. Shannon, C. E., Bell System Technical Journal, Volume 27, Pages 379 and 623, 1948.

6. Goldstein, H., Classical Mechanics, Addison-Wesley Publishing Company, Page 11, 1953.

7. Deltheil, R., Probabilites Geometriques, Traite du Calcul des Probabilites et de ses Applications: Tome II, Fasciscule II, Gauthier-Villars, 1926.

8. Hammersley, J. M., "The Distribution of Distances in a Hypersphere", Ann. Math. Stat., Volume 21, Pages 447-452, 1950.

9. Lord, R. D., "The Distribution of Distances in a Hypersphere", Ann. Math. Stat., Volume 25, Pages 794-798, 1954.

10. Hurewicz, W., and Wallman, H., "Dimension Theory", Princeton University Press, Volume 19, 1941.

11. Coombs, C. H., and Kao, R. C., "On a Connection between Factor Analysis and Multidimensional Scaling", Psychometrika, Volume 25, Pages 219-321, 1960.

TR-7 163

REFERENCES (CONTD.)

12. Shepard, R. N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function I", Psychometrika, Volume 27, Pages 125-140, 1962.

13. Shepard, R. N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function II", Psychometrika, Volume 27, Pages 219-246, 1962.

14. Young, G., and Householder, A. S., "Discussion of a Set of Points in Terms of their Mutual Distances", Psychometrika, Volume 3, Pages 19-22, 1938.

15. Torgerson, W. S., Theory and Methods of Scaling, John Wiley and Sons, Chapter XI, "Multidimensional Scaling", 1958.

16. Kautz, W. H., "Transient Synthesis in the Time Domain", IRE Transactions on Circuit Theory, CT-1, No. 3, Pages 29-39, September, 1954.

17. Bonner, R. E., "On Some Clustering Techniques", IBM Journal of Research and Development, Volume 8, Pages 22-32, 1964.

18. Young, T. Y., "Characteristic Vectors of Boolean Matrices", CBL/63/IMR-4, Carlyle Barton Laboratory, The Johns Hopkins University, 1963.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

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

1. ORIGINATING ACTIVITY (Corporate author)
Department of Electrical Engineering
The Johns Hopkins University
Baltimore, Maryland 21218

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
THE INTRINSIC DIMENSIONALITY OF SIGNAL COLLECTIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (Last name, first name, initial)
Bennett, Robert S.

6. REPORT DATE
December, 1965

7a. TOTAL NO. OF PAGES
66

7b. NO. OF REFS
18

8a. CONTRACT OR GRANT NO.
Nonr-4010(13)

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
21

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
TR/163

10. AVAILABILITY/LIMITATION NOTICES
Qualified requesters may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES
Distributed to recipients of reports under Contract Nonr-4010(13) with permission of the author.

12. SPONSORING MILITARY ACTIVITY
Systems Engineering Group
Research and Technology Division
Air Force Systems Command
Wright-Patterson AFB, Ohio

13. ABSTRACT

In view of the trend toward the representation of signals as physical observables, characterized by vectors in an abstract signal space, rather than as time or frequency functions, it is desirable to define dimensionality in a manner which would be independent of the choice of basis for the vectors.

In this work, the dimensionality of a collection of signals is defined as equal to the number of free parameters required in a hypothetical signal generator capable of producing a close approximation to each signal in the collection. Thus defined, dimensionality becomes a relationship between the vectors representing the signals. This relationship need not be a linear one, and does not depend on the basis onto which the vectors are projected in signal measuring processes. It represents a lower bound on the number of coefficients required to describe the signals, no matter how sophisticated the

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Vector spaces						
Signal spaces						
Dimensionality						
Parameter estimation						

INSTRUCTIONS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

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

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

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

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

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

DD FORM 1 JAN 64 1473 (BACK)

Unclassified
Security Classification