SEL-64-052

# GEOMETRICAL AND STATISTICAL PROPERTIES OF LINEAR THRESHOLD DEVICES

by

Thomas M. Cover

May 1964

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Technical Report No. 6107-1

Prepared under
ITT Federal Laboratories Grant*

*Partially supported by a National Science Foundation Fellowship

Systems Theory Laboratory
Stanford Electronics Laboratories
Stanford University Stanford, California

## ABSTRACT

The principal purpose of this research is to discover the underlying properties of linear threshold devices and networks of linear threshold devices. Questions of concern include finding the information storage capacity, the ability to implement all functions, the ability to generalize with respect to past data, and the ability to implement large classes of decision surfaces for linear threshold devices operating on a set of pattern vectors.

The most important theoretical developments of this investigation are the theorems counting the number of sides of all dimensions of the solution cones and dual cones for the set of all assignments of inequalities to a system of linear inequalities. The total measure of the cones and boundaries of the cones is also found. Although nothing specific can be said about any given cone, the number and volume of the sides of the totality of cones is an intrinsic property of a system of linear inequalities, depending only on a weak general position requirement on the set of inequalities. These combinatorial geometric results have applications in the theory of games and geometric probability as well as to the analysis linear threshold functions, and tend to make the theory of linear inequalities as concrete, in many respects, as the theory of linear equations.

An important extension of known theory to new domains of application is made by finding the number of dichotomies of a set of pattern vectors that can be separated by large classes of surfaces including hyperspheres, hypercones, and quadrics.

Finally, it is demonstrated that $2d$ is a natural definition of the information storage capacity of a $d$-input linear threshold device. Moreover, it is shown that an infinite, random, linearly separable set of $d$-dimensional pattern vectors can be completely characterized, on the average, by a $2d$ element subset of extreme patterns.

CONTENTS

TABLES

## ILLUSTRATIONS

## ACKNOWLEDGMENT

The author wishes to express his thanks to Dr. N. Nilsson and Professor A. Novikoff for many valuable discussions. He wishes to thank Dr. B. Brown, Dr. J. Koford, and Professor B. Widrow for formulation of several related problems which led to this research.

The author did some of the initial work from a probabilistic point of view with B. Efron, to whom he would like to express his respect and warm thanks for the many stimulating exchanges.

The author wishes especially to thank Professor N. Abramson for his guidance and help during the entire course of this research.

# I. INTRODUCTION

## A. PURPOSE

The basic problem attacked by this investigation is that of analyzing and describing the intrinsic properties of systems of simultaneous linear inequalities—those properties that depend solely on simple nondegeneracy properties of the system, not on the numerical values of the coefficients themselves. Although this work has applications in the fields of linear programming, game theory, and geometric probability, the primary purpose of this research has been to develop the underlying mathematical properties of linear threshold devices and networks of linear threshold devices such as Adaline [Ref. 1] and the Perceptron [Ref. 2]. Since the function of a linear threshold device is described in terms of linear inequalities, the two problems of investigation are formally the same. However, the emphasis will be on the applications of the theory to the behavior of linear threshold devices.

These devices form a class of information processing machines whose primary application at this time is to the problem of pattern recognition—the general problem of assigning objects to categories. The properties developed will be independent, in some sense, of the specific pattern recognition problems to which the linear threshold devices are applied.

Schematically, a linear threshold device, as represented in Fig. 1,

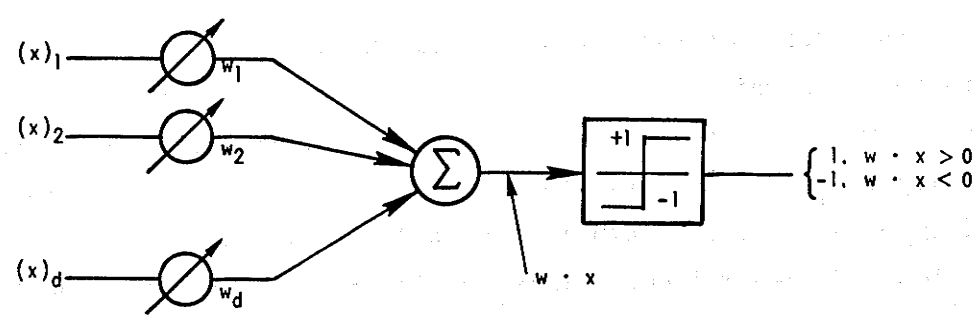



FIG. 1. HOMOGENEOUS LINEAR THRESHOLD UNIT AND IMPLEMENTATION OF SEPARATING HYPERPLANE.

is a d-input device that takes the sign of the sum of the products of the pattern inputs with the corresponding weights. The patterns and the weights can be represented as vectors in a Euclidean d-dimensional space. The linear threshold device then assigns to each pattern vector the sign of the inner product of the pattern vector with the weight vector. Geometrically, such a device separates the space into two regions by a (d-1)-dimensional hyperplane through the origin of the space. This separating hyperplane is orthogonal to the weight vector.

The synthesis problem, concerned primarily with implementing Boolean functions with networks of linear threshold devices, is being pursued by Mattson [Ref. 3], Winder [Ref. 4], and others. The analysis problem, which is the primary concern of this study, will be initially restricted to the analysis of the fundamental building block of linear threshold systems—the linear threshold unit itself.

The following questions are of concern in describing a linear threshold unit.

1. How does a linear threshold device work?
2. What are its capabilities as an information processing element?
3. What is its capacity as an information storage element?
4. On what basis does such an element generalize with respect to past inputs?
5. What families of decision surfaces can it implement?
6. What constraints are put on the internal state of a linear threshold unit by a given problem?

## B. APPROACH

In order to answer the questions posed in the previous section, the pattern recognition problem for linear threshold devices is presented as a problem in solving a system of linear inequalities. See, for example, Refs. 5 and 6. Then the study of a specific pattern recognition problem is imbedded in the study of all systems of linear inequalities. In several cases this imbedding has served to simplify the problem or to expand the domain of application of previously known ideas. For example, when a linear threshold unit operates on a set of pattern vectors that are the vertices of a binary n-cube (as when used in the

synthesis of Boolean functions or in the recognition of quantized patterns), the mathematical analysis of the device becomes simpler when the class of patterns in the domain of operation of the device is allowed to be all finite sets of vectors satisfying a weak nondegeneracy requirement in n-space. Then the problem, when the domain of operation is restricted to the binary n-cube, becomes a special (degenerate) case of the simpler, more general analysis. It was such an approach that allowed Winder [Ref. 4] and Cameron [Ref. 7] to give an upper bound on the number of linearly separable truth functions on n variables. It is this approach that is used herein.

Imbedding also plays an important role when, by the simple device of considering augmented pattern vectors, it is possible to extend the class of separating surfaces treated by present theory from the class of all hyperplanes to the class of all surfaces that are linear in their parameters. The extended class of surfaces includes hypercones, hyperspheres, quadrics, and many general surfaces without names.

Although the analysis is kept on a reasonably abstract and rigorous level in an attempt to identify the fundamental principles in the theory of linear threshold functions, many of the theorems grew out of conjectures arising from empirical research. For example, the conjecture that the capacity of a linear threshold device is twice the number of inputs was suggested experimentally in research on random inputs by Koford [Ref. 8] and was subsequently elaborated on experimentally by Brown [Ref. 9].

The principal contributions developed in this investigation of linear threshold devices are:

1. The demonstration that 2d is a natural definition of the information-storage capacity of a d-input linear threshold device.
2. The counting of the number of dichotomies of a finite set of patterns that can be separated by a surface from a family of surfaces linear in their parameters.
3. The description of the solution cone—the set of weight vectors implementing a given dichotomy of the pattern vectors—in terms of the expected number of sides of all dimensions and the expected boundary areas of all dimensions, with the implications for tolerance requirements on the internal state of a linear threshold device.

4. The description of the dual cone—the set of vectors that are positive linear combinations of the pattern vectors—in terms of the expected number of sides of all dimensions and the expected boundary areas of all dimensions, with implications for the training procedures.

5. The demonstration that a subset of $2d$ patterns in $d$ dimensions is sufficient, on the average, to characterize the classification of an <u>infinite</u>, random, separable set of patterns.

6. The demonstration that, for a large number of patterns, the expected set of solution weight vectors looks like a $d$-cube.

The combinatorial geometric analysis of the solution cones and dual cones for a system of linear inequalities is considered by the author to be the theoretical backbone of the research. Although this work has immediate interpretation in terms of tolerance requirements for linear threshold devices, the most important consequence of a thorough understanding of these results will be a development of an intuitive understanding of linear inequalities and solutions of systems of linear inequalities—a subject which, in many respects, is as concrete as the theory of linear equalities.

## II. DEFINITIONS AND HISTORY OF FUNCTION-COUNTING THEOREMS

### A. DEFINITIONS

Consider a set of patterns represented by a set of vectors in a d-dimensional Euclidean space $E^d$. A homogeneous linear threshold function $f: E^d \to \{-1,0,1\}$ is defined in terms of a parameter or weight vector $w$ for every vector $x$ in this space:

$$f(x;w) = \begin{cases} 1, & w \cdot x > 0 \\ 0, & w \cdot x = 0 \\ -1, & w \cdot x < 0 \end{cases} \tag{2.1}$$

where $w \cdot x$ is understood to mean the inner-product of the vector $w$ and $x$. Any function of this form has the simple implementation by a linear threshold unit indicated schematically in Fig. 1.

Thus every homogeneous linear threshold function naturally dichotomizes the set of pattern vectors into two sets, the set of vectors $x$ such that $f(x;w) = 1$ and the set of vectors $x$ such that $f(x;w) = -1$. These two sets are separated by the hyperplane

$$\left\{x: f(x;w) = 0\right\} = \left\{x: x \cdot w = 0\right\} \tag{2.2}$$

which is the $(d-1)$-dimensional subspace orthogonal to the weight vector $w$. Let $X$ be an arbitrary set of vectors in $E^d$. A dichotomy $\left\{X^+, X^-\right\}$ of $X$ is linearly separable if and only if there exists a weight vector $w$ in $E^d$ and a scalar $t$ such that

$$\begin{aligned} x \cdot w &> t, \quad \text{if } x \in X^+ \\ x \cdot w &< t, \quad \text{if } x \in X^- \end{aligned} \tag{2.3}$$

The dichotomy $\left\{X^+, X^-\right\}$ is said to be homogeneously linearly separable if it is linearly separable with $t = 0$.

By the considerations of Chapter III it will be possible to consider the inhomogeneous case as a special case of the homogeneous case (with one more variable). Therefore, rather than continue to make the distinction between the homogeneous and inhomogeneous cases, the definitions and theorems are here restricted to the homogeneous case and the reader can make the elementary application of Chapter III to generalize the theorems. A vector $w$ satisfying

$$\begin{aligned} w \cdot x &> 0, \qquad x \epsilon X^{+} \\ w \cdot x &< 0, \qquad x \epsilon X^{-} \end{aligned} \tag{2.4}$$

will be called a <u>solution vector</u> and the corresponding orthogonal hyperplane $\{x: x \cdot w = 0\}$ will be called a <u>separating hyperplane</u> for the dichotomy $\{X^{+}, X^{-}\}$. In this, the homogeneous case, the separating hyperplane passes through the origin of the space and is, in fact, the (d-1)-dimensional orthogonal subspace to $w$.

Note that the full generality of the definition of the linear threshold function $f(x;w)$ given in Eq. (2.1) has not been used—the set of patterns $x$ such that $f(x;w) = 0$ form a potential third category. For completeness the case of counting the number of <u>ternary-valued</u> homogeneous linear threshold functions is later treated in Chapter V.

Finally, a set of $N$ vectors is in <u>general position</u> in $d$-space if every $d$ element subset of vectors is linearly independent. That is, if a set of vectors is in general position then any $d$ of the $N$ vectors span the $d$-space and any $k$ vectors generate a $k$-dimensional subspace for $k \leq d$.

## B. THE FUNCTION-COUNTING THEOREM

The foundations have been laid for the presentation of the fundamental function counting theorem which counts the number of homogeneously linearly separable dichotomies of $N$ points in $d$ dimensional.

Theorem 1. (Function-Counting Theorem.) There are $C(N,d)$ homogeneously linearly separable dichotomies of $N$ points in general position in Euclidean $d$-space where

$$C(N,d) = 2 \sum_{k=0}^{d-1} \binom{N-1}{k} . \tag{2.5}$$

The notation $\binom{s}{k}$ denotes the binomial coefficient formally defined for all real $s$ and integer $k$ as the coefficient of $x^k$ in the expansion

$$(1+x)^s = \sum_{k=0}^{\infty} \binom{s}{k} x^k . \tag{2.6}$$

That is,

$$\binom{s}{k} = \frac{s(s-1)\ldots(s-k+1)}{k(k-1)\ldots 2\cdot 1} \tag{2.7}$$

In particular, for integer $s$

$$\begin{aligned} \binom{s}{k} &= 0, \qquad s < k \\ \binom{s}{o} &= 1, \qquad s \geq 0 \end{aligned} \tag{2.8}$$

Thus

$$C(N,d) = 2^N, \qquad N \leq d , \tag{2.9}$$

from which it can be seen that, for $N \leq d$, all dichotomies of $N$ points in $d$ dimensions are homogeneously linearly separable—a result which also follows immediately from the fact that there are fewer inequalities than unknowns. Note that $C(N,d)$ is just the number of sequences of $N-1$ 0's and 1's that contain $d-1$ or fewer 1's.

Theorem 1 has been independently proved in different forms by many authors [Refs. 4, 7, 10, 11, 12], but Winder [Ref. 4], Cameron [Ref. 7],

Perkins, Whitmore, and Willis [Ref. 11], and Joseph [Ref. 10] have emphasized the application of Theorem 1 to counting the number of linearly separable dichotomies of a set. In addition, Winder [Ref. 4] and Cameron [Ref. 7] independently applied Theorem 1 to the vectors that are the vertices of a binary n-cube in order to find an upper bound on the number of linearly separable truth functions of n variables. All the authors listed above have used a variant of a proof, which seems to have first appeared in Schläfli [Ref. 12], of Theorem 2 or its dual statement Theorem 2'.

Theorem 2. N hyperplanes in general position passing through the origin of d-space divide the space into $C(N,d)$ regions.

Theorem 2'. A d-dimensional subspace in general position in N-space intersects $C(N,d)$ orthants.

A set of hyperplanes is in general position in Theorem 2 if every intersection of d hyperplanes is a zero-dimensional subspace (the origin). A d-dimensional subspace is in general position in N-space if every orthogonal projection onto a d-dimensional coordinate axis is d-dimensional.

A proof of Theorem 2 is sketched using geometrical terminology. See Schläfli [Ref. 12], Cameron [Ref. 7], Winder [Ref. 4], and Wendel [Ref. 13] for similar treatments. Proofs in terms of the dual statement, Theorem 2', can be found in Schläfli [Ref. 12] and Joseph [Ref. 10].

Proof of Theorem 2. Let $C(N,d)$ be the number of regions formed by the intersection of N (d-1)-dimensional subspaces in general position in d-space. Consider a $(N+1)^{th}$ (d-1)-dimensional subspace such that the set of N+1 subspaces is in general position. This subspace is intersected by each of the N subspaces in a (d-2)-dimensional subspace. And from the original assumption of general position, it follows that the set of N (d-2)-dimensional subspaces is in general position in this (d-1)-dimensional subspace, and hence divide the new subspace into $C(N,d-1)$ regions. Thus the $(N+1)^{th}$ subspace intersects $C(N,d-1)$ of the $C(N,d)$ regions, forming $C(N,d-1)$ new regions. The total number of regions formed by the N+1 planes is then given by the recurrence relation

$$C(N+1,d) = C(N,d) + C(N,d-1) \quad . \tag{2.10}$$

Using the obvious boundary conditions

$$\begin{aligned} C(N,1) &= 2, \qquad N = 1,2,\ldots \\ C(1,d) &= 2, \qquad d = 1,2,\ldots \end{aligned} \tag{2.11}$$

it is easily verified by induction that $C(N,d)$ is given by Eq. (2.5).

## C. ALGEBRAIC PROOF

An alternative proof of Theorem 1 to those found in the literature is now given along more algebraic lines. First a geometrically obvious lemma is established which will be applied in Chapter VI on generalization. In geometrical terms, lemma 1 says that a new point can be adjoined to both halves of a separable dichotomy to form two new separable dichotomies if and only if there exists a separating hyperplane through the new point which separates the old dichotomy. This is reasonable because if such a hyperplane exists, small displacements of the hyperplane will allow arbitrary classification of the new point without affecting the separation of the old dichotomy.

<u>Lemma 1</u>. Let $\{X^+, X^-\}$ be a dichotomy of $\{x_1, x_2, \ldots, x_N \in E^d\}$ and $x_{N+1}$ a point in $E^d$. Then $\{X^+ \cup \{x_{N+1}\}, X^-\}$ and $\{X^+, X^- \cup \{x_{N+1}\}\}$ are both homogeneously linearly separable if and only if $\{X^+, X^-\}$ is homogeneously linearly separable by a $(d-1)$-dimensional subspace through $x_{N+1}$.

<u>Proof</u>. The dichotomy $\{X^+ \cup \{x_{N+1}\}, X^-\}$ is homogeneously linearly separable if and only if there exists $w$ such that

$$\begin{aligned} w \cdot v &> 0, \qquad x \in X^+ \\ w \cdot x_{N+1} &> 0, \\ w \cdot x &< 0, \qquad x \in X^- \end{aligned} \tag{2.12}$$

and $\{X^+, X^- \cup \{x_{N+1}\}\}$ is homogeneously linearly separable if and only if there exists $w$ such that

$$\begin{aligned} w \cdot x &> 0, \qquad x \epsilon X^+ \\ w \cdot x_{N+1} &< 0, \\ w \cdot x &< 0, \qquad x \epsilon X^- \end{aligned} \tag{2.13}$$

Using the connectedness of the open set $\{w: w \cdot x > 0$ for $x \epsilon X^+$, and $w \cdot x < 0$ for $x \epsilon X^-\}$ of separating vectors for $\{X^+, X^-\}$ and the continuity of the inner product, it is seen that Eqs. (2.12) and (2.13) hold if and only if there exists a vector $w \epsilon E^d$ separating $\{X^+, X^-\}$ such that

$$w \cdot x_{N+1} = 0 \quad . \tag{2.14}$$

Then the hyperplane $\{v: v \cdot w = 0\}$ separates $\{X^+, X^-\}$ and contains the point $x_{N+1}$.

<u>Proof of Theorem 1</u>. Again the method of proof is induction on $N$ and $d$. Let $C(N,d)$ be the number of homogeneously linearly separable dichotomies of the set $X = \{x_1, x_2, \ldots, x_N\}$. Consider a new point $x_{N+1}$ such that $X \cup \{x_{N+1}\}$ is in general position and consider the $C(N,d)$ homogeneously linearly separable dichotomies $\{X^+, X^-\}$ of $X$. Since $\{X^+, X^-\}$ is separable, either $\{X^+ \cup \{x_{N+1}\}, X^-\}$ or $\{X^+, X^- \cup \{x_{N+1}\}\}$ is separable. However, <u>both</u> dichotomies are separable, by lemma 1, if and only if there exists a separating vector $w$ for $\{X^+, X^-\}$ lying in the $(d-1)$-dimensional subspace orthogonal to $x_{N+1}$. A dichotomy of $X$ is separable by such a $w$ if and only if the projection of the set $X$ onto the $(d-1)$-dimensional orthogonal subspace to $x_{N+1}$ is separable. By the induction hypothesis there are $C(N,d-1)$ such separable dichotomies. Hence

$$C(N+1,d) = C(N,d) + C(N,d-1) \quad . \tag{2.15}$$

Repeated application of Eq. (2.15) to the terms on the right yields

$$C(N,d) = \sum_{k=0}^{N-1} \binom{N-1}{k} C(1,d-k) \quad , \tag{2.16}$$

from which the theorem follows immediately on noting

$$C(1,m) = \begin{cases} 2, & m \geq 1 \\ 0, & m < 1 \quad . \end{cases} \tag{2.17}$$

## III. SEPARABILITY BY ARBITRARY SURFACES

### A. MEASUREMENTS

A change in point of view permits application of the results of Chapter II to classes of separating surfaces that are geometrically different from hyperplanes, but analytically quite similar. Consider a family of surfaces, each of which naturally divides a given space into two regions, and a collection of $N$ points in this space, each of which is assigned to one of two classes $X^+$ or $X^-$. This dichotomy of the points is said to be separable relative to the family of surfaces if there exists at least one surface such that all the $X^+$ points are in one region and all the $X^-$ points are in the other. The crucial property of the family of surfaces, in order that the results of the previous section apply, is that the family can be parameterized in such a way that it is linear in its parameters. Hyperplanes, hyperspheres, and polynomial surfaces are special examples of such families.

Consider the set of $N$ objects $X = \{x_1, \ldots, x_N\}$. The elements of $X$ will be referred to as patterns for intuitive reasons. These patterns need not be considered as vectors in a vector space. On each pattern $x \epsilon X$ a set of real valued <u>measurement functions</u> $\phi_1, \phi_2, \ldots, \phi_d$ comprises the vector of measurements

$$\phi: X \rightarrow E^d \tag{3.1}$$

where $\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_d(x)), \quad x \epsilon X.$

A dichotomy (binary partition) $\{X^+, X^-\}$ of $X$ is <u>$\phi$-separable</u> if there exists a vector $w$ such that

$$\begin{aligned} w \cdot \phi(x) &> 0, \qquad x \epsilon X^+ \\ w \cdot \phi(x) &< 0, \qquad x \epsilon X^- \quad . \end{aligned} \tag{3.2}$$

Observe that the separating surface in the measurement space is the hyperplane $w \cdot \phi = 0$. The inverse image of this hyperplane is the separating

surface $\{x: w \cdot \phi(x) = 0\}$ in the pattern space. The advantage of this general formulation of the problem is that many classes of interesting nonlinear surfaces in the pattern space can be mapped into the class of hyperplanes in another space where the results of Chapter II will apply.

Nonlinear analysis may be suggested by the physical origin of the problem. For example, the pattern separation problem of isolating a cancerous area from good tissue might suggest spheres as the natural class of linear threshold implementable separating surfaces, rather than planes or intersections of planes. If there is obvious correlation in the input measurements, second and third order correlations should be incorporated into new measurements to augment the old. A decision theory approach might suggest a canonical class of separating surfaces for a given class of statistical problems. Even when the suggested class of surfaces is not linear in its parameters, arbitrarily close approximations can be made by $\phi$-surfaces. In all these cases the function counting theorem applied to $\phi$-surfaces will yield the number of $\phi$-separable dichotomies of the pattern set and will provide a basis for comparing linear tests on pattern recognition problems when the number of measurements in the tests differ.

## B. GENERAL POSITION

Definition. Let the vector-valued measurement function $\phi$ be defined on the set of patterns

$$\phi: X = \left\{x_1, \ldots, x_N\right\} \rightarrow R^d \quad . \tag{3.3}$$

Then, a set of patterns $X$ is in $\phi$-general position if the following equivalent conditions hold:

1. Every $d$ element subset of the set of $d$-dimensional measurement vectors $\{\phi(x_1), \ldots, \phi(x_N)\}$ is linearly independent.

1'. Every $d \times d$ submatrix of the $N \times d$ matrix

$$\Phi = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_d(x_1) \\ \phi_1(x_2) & & & \\ \vdots & & & \\ \phi_1(x_N) & \cdots & & \phi_d(x_N) \end{bmatrix} \tag{3.4}$$

has a non-zero determinant.

1". No $d + 1$ patterns lie on the same $\phi$-surface $\{x: \phi(x)\cdot w = 0\}$ in the pattern space.

Clearly Definition 1' is just an explicit algebraic statement of Definition 1. Note that general position is a strengthened rank condition on the matrix $\Phi$ ($\Phi$ has maximal rank $d$ if at least one $d \times d$ submatrix has nonzero determinant). Definition 1" relates general position in the measurement space to general position in the pattern space.

## C. COUNTING THE $\phi$-SEPARABLE DICHOTOMIES

<u>Theorem 3</u>. Let $X = \{x_1, x_2, \ldots, x_N\}$ be in $\phi$-general position where $\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_d(x))$. Then precisely $C(N,d)$ of the $2^N$ dichotomies of $X$ are $\phi$-separable where

$$C(N,d) = 2 \sum_{i=0}^{d-1} \binom{N-1}{i} . \tag{3.5}$$

If the $\phi$-space $\{x: \phi(x)\cdot w = 0\}$ is constrained to contain the set of points $y = \{y_1, y_2, \ldots, y_k\}$, where

1. $\{\phi(y_1), \phi(y_2), \ldots, \phi(y_k)\}$ is linearly independent, and
2. the projection of $\{\phi(x_1), \phi(x_2), \ldots, \phi(x_N)\}$ into the orthogonal subspace to the space spanned by $\{\phi(y_1), \phi(y_2), \ldots, \phi(y_k)\}$ is in general position,

then there are $C(N,d-k)$ $\phi$-separable dichotomies of $X$.

<u>Proof</u>. Every $d$-element subset of the $N$ vectors $\phi(x_1), \ldots, \phi(x_N)$ is linearly independent by hypothesis. Hence, there are precisely $C(N,d)$ homogeneously linearly separable dichotomies of $\{\phi(x_i): i = 1, 2, \ldots, N\}$.

By definition these dichotomies correspond to the $\phi$-separable dichotomies of X.

The second part of the theorem simply states that k independent constraints with respect to a $\phi$-surface reduce the number of degrees of freedom of the surface by k. The condition that the $\phi$-surface contains the set Y is that the weight vector w which characterizes the surface must lie in the (d-k)-dimensional subspace L where

$$L = \left\{w: w \cdot \phi(y_i) = 0, \qquad i = 1,2,\ldots,k\right\} .$$

Let $\hat{\phi}$ be the orthogonal projection of $\phi$ onto L. Then, since

$$w \cdot \phi = w \cdot \hat{\phi} + w \cdot (\phi - \hat{\phi}) = w \cdot \hat{\phi} \tag{3.6}$$

for all w is L, it can be seen that a set of vectors $\{\phi\}$ is separable by a weight vector in L if and only if the set of their projections $\{\hat{\phi}\}$ is separable. Since the vectors $\hat{\phi}(x_1),\ldots,\hat{\phi}(x_N)$ are in $\hat{\phi}$-general position in L by hypothesis (2), there are C(N,d-k) homogeneously linearly separable dichotomies of $\{\phi(x_i): i = 1,2,\ldots,N\}$ by a vector w in L.

## D. POLYNOMIAL SEPARABILITY

A natural generalization of linear separability is polynomial separability. For the ensuing discussion, consider the patterns to be vectors in an m-dimensional space. The measurement function $\phi$ then maps points in m-space into points in d-space.

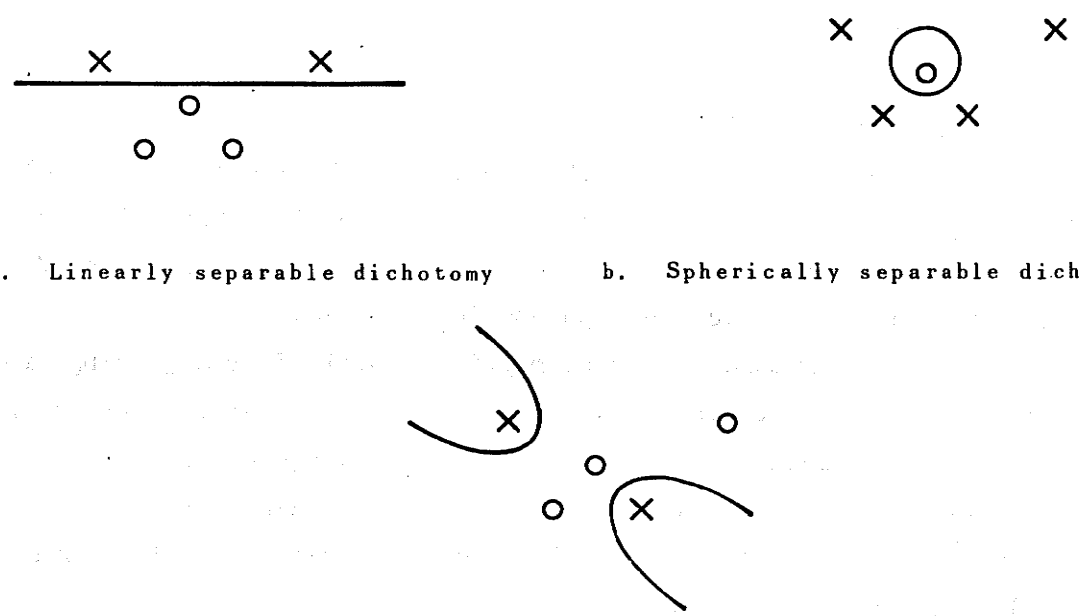
Consider a natural class of mappings obtained by adjoining r-wise products of the pattern vector coordinates. The natural separating surfaces corresponding to such mappings are known as $r^{th}$-order rational varieties. A rational variety of order r obtained in a space of m dimensions is represented by a homogeneous equation in the coordinates $(x)_i$ of the $r^{th}$ degree

$$\sum_{0 \le i_1 \le i_2 \le \cdots \le i_r \le m} a_{i_1 i_2 \cdots i_r} (x)_{i_1} (x)_{i_2} \cdots (x)_{i_r} = 0 , \qquad (3.7)$$

where $(x)_i$ is the $i^{th}$ component of $x$ in $E^m$ and $(x)_o$ is set equal to 1 in order to write the expression in homogeneous form. A simple counting argument gives the number of coefficients $F_m^{(r)}$ in Eq. (3.7) as

$$F_m^{(r)} = \sum_{k=0}^{r} \binom{m+k-1}{k} \qquad (3.8)$$

First-order rational varieties are hyperplanes and second-order rational varieties are quadrics. Hyperspheres are quadrics with certain linear constraints on the coefficients. Figure 2 illustrates three dichotomies of the same set of points. All three dichotomies are quadrically separable (in two dimensions the phrase would be quadratically separable).



a. Linearly separable dichotomy

b. Spherically separable dichotomy

c. Quadrically separable dichotomy

FIG. 2. EXAMPLES OF $\phi$-SEPARABLE DICHOTOMIES OF 5 POINTS IN TWO DIMENSIONS.

In Theorem 3, the mapping $\phi: E^m \to E^{F_m^{(r)}}$ defined by

$$\phi(x) = (1,(x)_1,\ldots,(x)_m,(x_1)^2,\ldots,(x)_i(x)_j,\ldots,(x)_m^r) \tag{3.9}$$

yields the following result: A set of $N$ points in m-space, such that no $F_m^{(r)}$ points lie on the same $r^{th}$-order rational variety, has precisely $C(N,F_m^{(r)})$ dichotomies which are separable by an $r^{th}$-order rational variety. If the variety is constrained to contain $k$ independent points, the number of separable dichotomies is reduced to $C(N,F_m^{(r)}-k)$.

## E. DISCUSSION

Bishop [Ref. 14] has exhaustively found the number $L_m$ of quadrically separable truth functions of $m$ arguments for low $m$. From the foregoing it can be seen that $L_m$ is bounded above by

$$L_m \leq C\left(2^m,\binom{m+1}{2}+m+1\right) \sim 2^{m^3/2+O(m^2 \log m)} . \tag{3.10}$$

Koford [Ref. 15] has observed that augmenting the vector $x \in E^d$ to yield a vector $\phi(x)$ as in Eq. (3.9) is especially easy to implement when the coefficients are binary. In addition, Koford notes that, if the augmented vector $\phi(x)$ is used as an input to a linear threshold device (as in Fig. 3), then the standard training procedure will converge [Ref. 16] (by the Perceptron convergence theorem) in a finite number of steps to a separating $\phi$-surface if one exists.

Table 1 lists several examples of families of separating surfaces. All patterns $x$ should be considered as vectors in an m-dimensional space. The function $\phi(x) = (1,x)$ is a $(m+1)$-dimensional vector. The final column of Table 1 lists the separating capacities of the $\phi$-surfaces—a measure of the expected maximum number of random patterns which can be separated. The separating capacity will be made plausible as a useful idea in the next chapter.

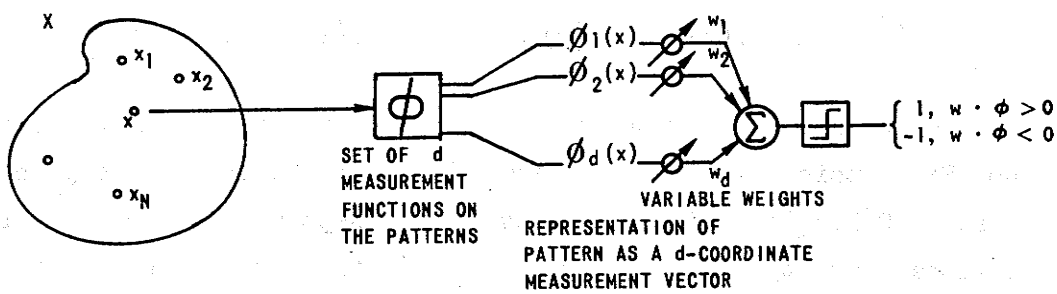



FIG. 3. MEASUREMENT TRANSFORMATION AND IMPLEMENTATION OF SEPARATING $\phi$-SURFACE.

Table I

EXAMPLES OF SEPARATING SURFACES WITH THE CORRESPONDING NUMBER OF SEPARABLE DICHOTOMIES OF N POINTS IN m DIMENSIONS

| MAPPING $\varphi$ DEFINED ON $E^m$ | SEPARATING SURFACE IN PATTERN SPACE | NUMBER OF PARAMETERS OF $\varphi$-SURFACE | GEOMETRICAL MEANING OF $\varphi$-GENERAL POSITION | NUMBER OF $\varphi$-SEPARABLE DICHOTOMIES OF N POINTS | SEPARATING CAPACITY OF $\varphi$-SURFACE |
|---|---|---|---|---|---|
| $\varphi(x) = x$ | Hyperplane through origin | m | Every m points linearly independent | C(N, m) | 2m |
| $\varphi(x) = (1, x)$ | Hyperplane | m + 1 | No m + 1 points on any hyperplane | C(N, m + 1) | 2m + 2 |
| $\varphi(x) = (1, x, \|\|x\|\|^2)$ | Hypersphere | m + 2 | No m + 2 points on any hypersphere | C(N, m + 2) | 2m + 4 |
| $\varphi(x) = (x, \|\|x\|\|)$ | Hypercone with vertex at origin | m + 1 | No m + 1 points on hypercone | C(N, m + 1) | 2m + 2 |
| $\varphi(x)$ as in Eq. (3.9) | Rational rth order variety | $F_m^{(r)} = \sum_{k=0}^{r} \binom{m+k-1}{k}$ | No $F_m^{(r)}$ points on same rth-order rational variety | $C(N, F_m^{(r)})$ | $2F_m^{(r)}$ |

## IV. SEPARABILITY OF RANDOM PATTERNS

### A. THE PROBLEM

In this chapter two kinds of randomness are considered in the pattern dichotomization problem:

1. The patterns are fixed in position but are classified independently with equal probability into one of two categories.
2. The patterns themselves are randomly distributed in space, and the desired dichotomization may either be random of fixed.

Under these conditions the separability of the set of pattern vectors becomes a random event depending on the dichotomy chosen and the position of the patterns. The probability of this random event and the maximum number of random patterns that can be separated by a given family of decision surfaces are to be determined.

It is shown that the expected maximum number of randomly assigned vectors that are linearly separable in $d$ dimensions is equal to $2d$. It is thus possible to conclude that a linear threshold device has an information-storage capacity—relative to learning random dichotomies of a set of patterns—of two patterns per variable weight. This result was originally conjectured and experimentally supported by Koford [Ref. 8], for the case of pattern vectors chosen at random from the set of vertices of a binary $d$-cube. Brown [Ref. 17] found experimentally that the conjecture held for patterns distributed at random in the unit $d$-sphere. Since Brown had stated the problem in a form such that the pattern vectors were in general position with probability 1, the way was open for the direct application of the function-counting theorem to establish the conjecture theoretically. This was done independently by Winder [Ref. 4] and Cover and Efron [Ref. 19].

### B. GENERAL POSITION WITH PROBABILITY 1

There are $C(N,d)$ homogeneously linearly separable dichotomies of a set of $N$ pattern vectors chosen at random according to some probability distribution over $d$-space if and only if the set of pattern vectors is in general position with probability 1. Since the requirement of general position is very weak, it is not surprising to learn that the

class of probability distributions for which general position is satisfied with probability 1 is very large. In fact, this class includes all the "smooth" distributions—the probability distributions that have probability densities and are free from delta functions.

Suppose that the patterns $x_1, x_2, \ldots, x_N$ are chosen independently according to a probability measure $\mu$ on the pattern space. It is easily verified that necessary and sufficient conditions on $\mu$ such that $x_1, x_2, \ldots, x_N$ are in general position in d-space is that the probability be zero that any point fall on any given (d-1)-dimensional subspace. Rewording this statement for $\phi$-surfaces, a set of vectors chosen independently according to a probability measure $\mu$ is in $\phi$-general position with probability 1 if and only if every $\phi$-surface $\left\{x \epsilon E^d : w \cdot \phi(x) = 0\right\}$ has $\mu$ measure zero.

Thus, since, for any $\phi$, every $\phi$-surface has natural Lebesgue measure zero in the pattern space, it is sufficient (but not necessary) for $\mu$ to be absolutely continuous with respect to natural Lebesgue measure in the d-space in order that general position hold with probability 1.

## C. SEPARABILITY OF RANDOM DICHOTOMIES

Suppose that a dichotomy of $X = \left\{x_1, x_2, \ldots, x_N\right\}$ is chosen at random with equal probability from the $2^N$ equiprobable possible dichotomies of X. Let X be in $\phi$-general position with probability 1, and let P(N,d) be the probability that the random dichotomy is $\phi$-separable, where the class of $\phi$-surfaces has d degrees of freedom. Then with probability 1 there are C(N,d) $\phi$-separable dichotomies, and

$$P(N,d) = \left(\tfrac{1}{2}\right)^N C(N,d) = \left(\tfrac{1}{2}\right)^{N-1} \sum_{k=0}^{d-1} \binom{N-1}{k} , \qquad (4.1)$$

which is just the cumulative binomial distribution corresponding to the probability that N-1 flips of a fair coin result in d-1 or fewer successes.

## D. RANDOM POINTS ON A HYPERSPHERE

One of the first applications of the function-counting theorems to random pattern vectors was by Wendel [Ref. 13], who found the probability that $N$ random points lie in some hemisphere. Let the vectors $x_1, x_2, \ldots, x_N$ on the surface of a d-sphere be in general position with probability 1. In addition let the joint distribution of $x_1, x_2, \ldots, x_N$ be unchanged by the reflection of any subset of the set of vectors through the origin. Under these restrictions Wendel proves that the probability that a set of $N$ vectors randomly distributed on the surface of a d-sphere is contained in some hemisphere is

$$P(N,d) = \left(\tfrac{1}{2}\right)^{N-1} \sum_{k=0}^{d-1} \binom{N-1}{k} . \tag{4.2}$$

The proof of this result follows immediately from the reflection invariance of the joint probability distribution of $X$. This invariance implies that the probability (conditioned on $X$) that a random dichotomy of $X$ be separable is equal to the unconditional probability that a particular dichotomy of $X$ (all $N$ points in one hemisphere) be separable.

## E. SEPARATING CAPACITY OF A SURFACE

Let $\{x_1, x_2, \ldots\}$ be a sequence of random patterns as above and define the random variable $N$ to be the largest integer such that $\{x_1, x_2, \ldots, x_N\}$ is $\phi$-separable, where the $\phi$-surface has $d$ degrees of freedom. Then from Eq. (4.1)

$$\begin{aligned} P_r\{N=n\} &= P(N,d) - P(N-1,d) \\ &= \left(\tfrac{1}{2}\right)^N \binom{N-1}{d-1} , \end{aligned} \tag{4.3}$$

which is just the negative binomial distribution (shifted $d$ units right with parameters $d$ and $\frac{1}{2}$). Thus $N$ corresponds to the waiting time for the $d^{th}$ failure in a series of tosses of a fair coin, and

$$E(N) = 2d$$

$$\text{Median}(N) = 2d \quad . \tag{4.4}$$

The asymptotic probability that $N$ patterns are separable in $d \approx \frac{N}{2} + \frac{\alpha}{2}\sqrt{N}$ dimensions is

$$P(N, \frac{N}{2} + \frac{\alpha}{2}\sqrt{N}) \sim \Phi(\alpha) \tag{4.5}$$

where $\Phi(\alpha)$ is the cumulative normal distribution

$$\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2}\, dx \quad . \tag{4.6}$$

In addition, for $\epsilon > 0$,

$$\lim_{d\to\infty} P(2d(1+\epsilon),d) = 0$$

$$P(2d,d) = \tfrac{1}{2} \tag{4.7}$$

$$\lim_{d\to\infty} P(2d(1-\epsilon),d) = 1$$

as was shown by Winder [Ref. 4]. Thus the probability of separability shows a pronounced threshold effect when the number of patterns is equal to twice the number of dimensions. These results confirm Koford's conjecture [Ref. 8] and suggest that $2d$ is a natural definition of the <u>separating capacity</u> of a family of decision surfaces having $d$ degrees of freedom.

In the following chapters, the fact that $2d$ is indeed a critical number for a system of linear inequalities in $d$ unknowns is established. It has already been shown that fewer than $2d$ random inequalities in $d$ unknowns can usually be solved, while more than $2d$ inequalities usually can not be solved. It is shown that the expected number of extreme inequalities, which are necessary and sufficient to imply the entire set,

tends to $2d$ as the number of consistent inequalities tends to infinity. Hence the amount of sufficient information characterizing an infinite system of inequalities has a finite expectation.

## V. GEOMETRICAL PROPERTIES OF SOLUTION CONE AND DUAL CONE

### A. INTRODUCTORY REMARKS

Consider the set of ternary-valued, homogeneous, linear threshold functions, parameterized by a weight vector $w$, defined on $N$ points $x_1, x_2, \ldots, x_N$ in Euclidean d-space taking on values in the set $\{-1, 0, 1\}$ defined by

$$f(x;w) = \begin{cases} 1, & x \cdot w > 0 \\ 0, & x \cdot w = 0 \\ -1, & x \cdot w < 0, \end{cases} \qquad x \epsilon \{x_1, x_2, \ldots, x_N\}, w \epsilon E^{(d)} \tag{5.1}$$

The number of such functions in this section will be counted, and, in the process, the following two sets will be introduced.

1. The solution set $W$, which consists of all those weight vectors $w$ that correspond to a given function $f: \{x_1, \ldots, x_N\} \rightarrow \{-1, 0, 1\}$.
2. The set $W^*$ of all positive linear combinations of the pattern vectors $\{x_1, x_2, \ldots, x_N\}$.

The sets $W$ and $W^*$ will be shown to be polyhedral convex cones in d-space. The number of boundary faces of $0, 1, 2, \ldots, d-1$ dimensions of $W$ and $W^*$ depends on the function to which the sets correspond.

The expected number and volume of the boundary faces of the solution cone and its dual cone are obtained under the assumption that the linear threshold functions are chosen at random with equal probability. For the special case, when the pattern set itself is randomly distributed according to a uniform distribution on the surface of a d-sphere, the variance of the volume of the solution cone will be calculated.

These results may have applications to the design and construction of linear threshold units such as Adaline [Ref. 1] and Perceptron [Ref. 2]. For example, the shape and size of the solution set $W$ corresponding to a given function to be implemented will dictate certain tolerance requirements on the internal weights in order that:

1. The weight vector will not drift out of the solution cone and begin to implement a different function, and

2. The weight vector will have sufficient resolution to implement any desired linear threshold function.

The count of the actual number of linear threshold functions places a lower bound on the number of internal states of a linear threshold unit in order that it be capable of implementing all those functions for which it is theoretically designed.

The suggested algorithms for changing the internal state of a linear threshold device such as relaxation adaption [Ref. 21] and fixed-increment adaption [Ref. 22] all have in common the feature that the correction vectors are positive linear multiples of the pattern vectors. Hence, if the initial weight vector is zero, $W^*$ is precisely the set of admissible state vectors under the class of all reasonable correction procedures. (It has been proved [Ref. 22] that the two correction procedures mentioned above converge to a solution vector in $W$ in a finite number of steps.) Thus the study of the properties of $W^*$ is related to the study of convergence algorithms. However, from a mathematical standpoint, $W^*$ is even more intimately related to $W$. It is shown that $W^*$ is the dual cone to $W$ and that every k-dimensional boundary plane to $W$ corresponds to a $(d-k)$-dimensional boundary plane for $W^*$. In particular, the extreme rays of $W^*$, corresponding to the extreme patterns, are orthogonal to the dominating boundary hyperplanes of $W$.

One of the most significant applications of this chapter follows from counting the extreme rays of $W^*$ in Proposition 2. For a random† linearly separable dichotomy of the pattern set, the expected number of extreme patterns tends to $2d$ as the number of patterns in d-space tends to infinity. In other words, the expected number of dominating inequalities for a consistent infinite set of inhomogeneous linear inequalities is just twice the number of variables in each inequality. Therefore, since the set of extreme patterns completely characterizes the set of patterns, this result suggests that a finite set of Adaline weights (on the order of $2d^2$) can store <u>all</u> the information with respect to the classification of past, present and future d-dimensional patterns,

†The usual liberty will be taken of using the word "random" to imply an equiprobable distribution over the set in question—in this case, the set of all linearly separable dichotomies of the pattern set.

even as the number of patterns grows without bound. A computer with finite storage capacity, if it has means of updating its memory by replacing the old extreme patterns by the new, will be able to store the essential information in a random pattern set with a probability which is inversely proportional to the capacity of the computer.

## B. NUMBER OF TERNARY-VALUED HOMOGENEOUS LINEAR THRESHOLD FUNCTIONS

Consider a homogeneous linear threshold function, defined on a single vector $x$, taking on values in the set $\{-1,0,1\}$ as defined in Eq. (5.1). This threshold function on $x$ partitions the weight vectors into three equivalence classes—those sets of weight vectors for which $x \cdot w$ is greater than, equal to, and less than zero. Figure 4 illustrates the equivalence classes intersected with the unit d-sphere. Two vectors in the same equivalence class correspond to the same function or assignment of the pattern vector. The (d-1)-dimensional subspace $\{w: x \cdot w = 0\}$ corresponds to the set of weight vectors for which $f(x;w) = 0$. Let

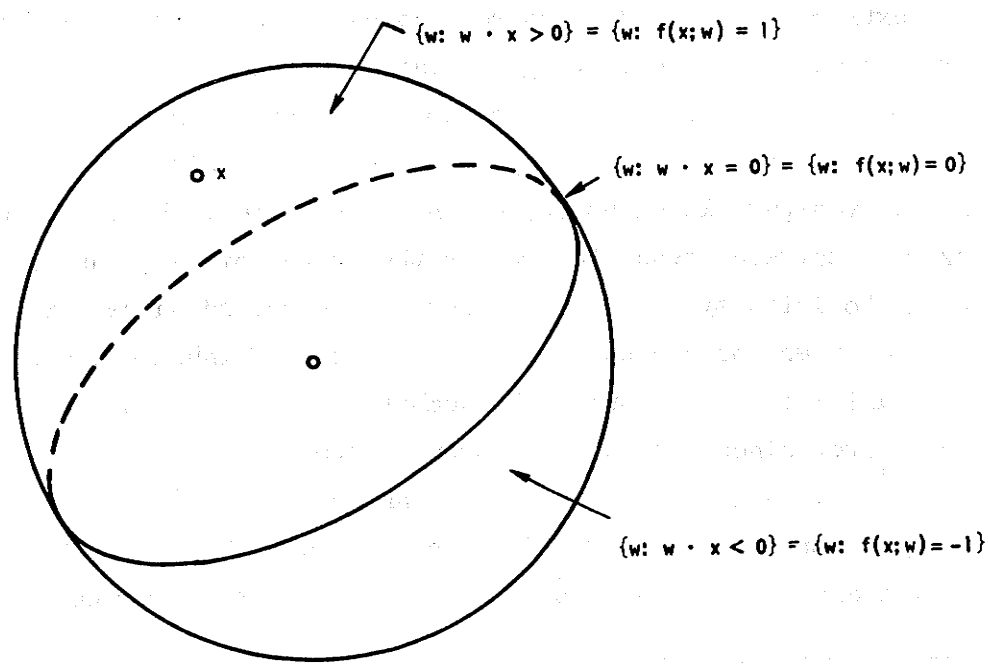



FIG. 4. EQUIVALENCE CLASSES OF WEIGHT VECTORS INDUCED BY THE HOMOGENEOUS LINEAR THRESHOLD FUNCTIONS ON x.

$f = (f_1, f_2, \ldots, f_N)$, $f_i \epsilon \{-1,0,1\}$, denote a function on the $N$ vectors $x_1, x_2, \ldots, x_N$ in d-space where $f(x_i) = f_i$ for each $i$. Let the <u>solution set</u> $W(f)$ corresponding to the function $f$ be defined by

$$W(f) = \left\{w\colon f(x;w) = f, \quad x = x_1, x_2, \ldots, x_N\right\} \tag{5.2}$$

where $f(x;w)$ is defined in Eq. (5.1). Clearly $f$ is a homogeneous linear threshold function on $x_1, x_2, \ldots, x_N$ if and only if $W(f)$ is nonempty, and $W(f)$ is precisely the equivalence class of weight vectors which implements the function $f$.

For example, the shaded region ABCD in Fig. 5 is the intersection of the solution set $W(1,1,1,1)$ with the surface of a three-sphere centered at the origin. The arc AB is the solution set for the function $(1,1,1,0)$, and the point E is the solution set for the function $(1,0,-1,0)$. The function $(1,-1,1,-1)$ has an empty solution set and thus is not a homogeneous linear threshold function on $x_1, x_2, x_3, x_4$.

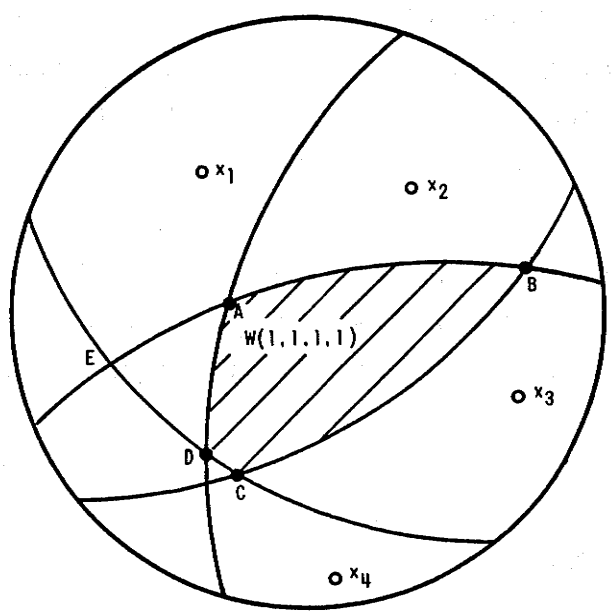



FIG. 5. EQUIVALENCE CLASSES OF WEIGHT VECTORS INDUCED BY THE HOMOGENEOUS LINEAR THRESHOLD FUNCTIONS DEFINED ON 4 POINTS ON THE SURFACE OF A 3-SPHERE.

The following theorem from Chapter II is used here.

<u>Theorem 4</u>. Let $x_1, x_2, \ldots, x_N$ be a set of $N$ vectors in d-space such that any subset of $d$ vectors is linearly independent. Then the $N$ normal subspaces $\left\{v \in E^{(d)} : v \cdot x_i = 0\right\}$ partition d-space into $C(N,d)$ regions where

$$C(N,d) = 2 \sum_{k=0}^{d-1} \binom{N-1}{k} . \tag{5.3}$$

Theorem 4 implies that there are precisely $C(N,d)$ functions $f: \left\{x_1, x_2, \ldots, x_N\right\} \to \left\{-1,1\right\}$ defined on a set of points $\left\{x_1, x_2, \ldots, x_N\right\}$ in general position in d-space where $f$ is of the form

$$f(x) = \begin{cases} 1, & x \cdot w \geq 0 \\ -1, & x \cdot w < 0 \end{cases} . \tag{5.4}$$

Because it is not desired to define $f$ unsymmetrically with respect to the inequality, Theorem 4 will be used to count the number of functions having the symmetric form of Eq. (5.1). At the same time a foundation for the counting of the number of sides of the solution sets will be laid.

<u>Theorem 5</u>. Let $x_1, x_2, \ldots, x_N$ be a set of $N$ vectors in Euclidean d-space such that every subset of $d$ vectors is linearly independent. Let $F$ be the class of functions $f: \left\{x_1, x_2, \ldots, x_N\right\} \to \left\{-1,0,1\right\}$ defined by

$$f(x_i) = \begin{cases} 1, & x_i \cdot w > 0 \\ 0, & x_i \cdot w = 0 \\ -1, & x_i \cdot w < 0, \end{cases} \qquad i = 1,2,\ldots,N \quad , \tag{5.5}$$

where $w$ is any vector in the space. Then there are $Q(N,d)$ functions in $F$ where

$$Q(N,d) = 2 \sum_{k=0}^{N} \sum_{m=0}^{d-k-1} \binom{N}{k}\binom{N-k-1}{m} . \tag{5.6}$$

<u>Proof</u>. Let $F_k$ be the class of functions in $F$ that take on the value zero for precisely $k$ points. There are $\binom{N}{k}$ ways to select the $k$ zero points from the set of $N$ points. Consider, without loss of generality, that $f(x_i) = 0$, $i = 1,2,\ldots,k$ and $f(x_i) = \pm 1$, $i = k+1,k+2,\ldots,N$. The solution cone $W(f)$ has the form

$$W(f) = \left\{w\colon w\cdot x_i = 0,\ i = 1,2,\ldots,k\right\} \bigcap_{i=k+1}^{N} \left\{w\colon w\cdot x_i \gtrless 0\right\} \tag{5.7}$$

Thus $W$ is the intersection of the $(d-k)$-dimensional space orthogonal to $x_1,x_2,\ldots,x_k$ with the cone formed by the intersection of the open half spaces corresponding to the $N-k$ remaining vectors. Rewriting Eq. (5.7) as the intersection of a set of $N-k$ half spaces in a $(d-k)$-dimensional subspace

$$W(f) = \bigcap_{i=k+1}^{N} \left\{w\colon w\cdot x_i \gtrless 0,\quad w\cdot x_j = 0,\quad j = 1,2,\ldots,k\right\}, \tag{5.8}$$

it follows from Theorem 1 and the general position of the vectors $\left\{x_1,\ldots,x_N\right\}$ that there are precisely $C(N-k,d-k)$ functions of the form $f(x_i) = 0$, $i = 1,2,\ldots,k$, $f(x_i) = \pm 1$, $i = k+1,\ldots,N$. Hence, since there are $\binom{N}{k}$ ways to select the $k$ zeroes of $f$, there are $\binom{N}{k}C(N-k,d-k)$ functions in $F_k$, and the theorem follows on summing over $k$.

There are $3^N$ functions defined on $N$ points taking values in the set $\{-1,0,1\}$. $Q(N,d)$ of these functions are the threshold functions as defined in Eq. (5.1). If the number of points $N$ is not greater than the number of dimensions $d$, then it will be shown that $Q(N,d) = 3^N$. Recall that the binomial coefficient $\binom{s}{k}$ for real $s$ and integer $k$ is defined as the coefficient of $x^k$ in the expansion

$$(1+x)^s = \sum_{k=0}^{\infty} \binom{s}{k} x^k . \tag{5.9}$$

In particular, for $N$ an integer,

$$\binom{N}{m} = 0, \quad \text{for} \quad m > N \quad . \tag{5.10}$$

Thus we find, for $N \leq d$,

$$\begin{aligned} 2 \sum_{j=0}^{d-k-1} \binom{N-k-1}{j} &= 2 \sum_{j=0}^{N-k-1} \binom{N-k-1}{j} \\ &= 2^{N-k} \quad . \end{aligned} \tag{5.11}$$

Therefore

$$\begin{aligned} Q(N,d) &= \sum_{k=0}^{N} \binom{N}{k} 2^{N-k} \\ &= 2^N \sum_{k=0}^{N} \binom{N}{k} (\tfrac{1}{2})^k \\ &= 2^N (1 + \tfrac{1}{2})^N = 3^N, \ (N \leq d) \ . \end{aligned} \tag{5.12}$$

It has been verified that all functions with range $\{-1,0,1\}$ defined on $N$ points in general position in $d$ $(N \leq d)$ dimensions may be expressed in the form shown in Eq. (5.1). This is certainly no surprise, for there are fewer equations (or inequalities) than unknowns.

It will now be shown that the limiting ratio of the number of ternary-valued, homogeneous, linear threshold functions $Q(N,d)$ to the number of homogeneous linear threshold functions $C(N,d)$ is $2^{d-1}$ in the limit for a large number of pattern vectors $N$ in a space of fixed dimension $d$. If the asymptotically negligible initial terms are dropped in the sums of the form

$$\sum_{m=0}^{d-k-1} \binom{N-k-1}{m} ,$$

from Eq. (5.6),

$$\lim_{N\to\infty} \frac{Q(N,d)}{C(N,d)} = \lim_{N\to\infty} \sum_{k=0}^{d-1} \frac{\binom{N}{k}\binom{N-k-1}{d-k-1}}{\binom{N-1}{d-1}} . \tag{5.13}$$

Canceling terms and taking the limit gives

$$\lim_{N\to\infty} \frac{Q(N,d)}{C(N,d)} = \sum_{k=0}^{d-1} \binom{d-1}{k} = 2^{d-1} . \tag{5.14}$$

C. THE SOLUTION CONE: COUNTING THE SIDES

Consider the set of $N$ homogeneous linear inequalities in $d$ variables

$$w \cdot x_i > 0, \quad i = 1,2,\ldots,N , \tag{5.15}$$

where $x_1, x_2, \ldots, x_N$ and $w$ are vectors in Euclidean d-space. A vector $w$ satisfying Eq. (5.15) is a <u>feasible vector</u> or <u>solution vector</u> for the set of inequalities. If such a vector exists, the set of inequalities is <u>consistent</u>. In accordance with the definitions in Chapter VA, let $W(1,1,\ldots 1)$ be the <u>solution cone</u> comprised of the solution vectors for the given set of inequalities, i.e.,

$$W(1,1,\ldots,1) = \left\{w\colon\ w \cdot x_i > 0, \quad i = 1,2,\ldots,N\right\} . \tag{5.16}$$

$W$ may be rewritten as a finite intersection of open half spaces

$$W(1,1,\ldots,1) = \bigcap_{i=1}^{N} \left\{w\colon\ w \cdot x_i > 0\right\} . \tag{5.17}$$

The solution set $W$ is a <u>convex cone</u> in the sense that

1. $\alpha v$ belongs to $W$ if $\alpha > 0$ and $v$ belongs to $W$.
2. The sum $v_1 + v_2$ of two vectors $v_1$, $v_2$ belonging to $W$ belongs to $W$.

A set formed by the intersection of a finite number of half spaces is called a polyhedral convex cone. So $W$ is an open polyhedral convex cone in Euclidean d-space. The closure of $W$, denoted by $\overline{W}$, is the intersection of the closed half spaces which define $W$,

$$\overline{W} = \bigcap_{i=1}^{N} \left\{w: w \cdot x_i \geq 0\right\} \tag{5.18}$$

Equivalently, $\overline{W}$ is the smallest closed set containing $W$. Clearly $\overline{W}$ is a closed polyhedral convex cone.

The nonempty sets $W(f_1, f_2, \ldots, f_N)$, where the $f_i$ take on the value 1 or 0, form a partition of $\overline{W}(1,1,\ldots,1)$. That is, if $f \neq g$, then

$$W(f) \cap W(g) = \emptyset \tag{5.19}$$

and

$$\cup W(f) = \overline{W}(1,1,\ldots,1) \quad , \tag{5.20}$$

where the union is taken over all N-tuples $f$ of 0's and 1's. The $W(f)$'s partition $\overline{W}(1,1,\ldots,1)$ because for each index $i$ the sets $\left\{w: w \cdot x_i > 0\right\}$ and $\left\{w: w \cdot x_i = 0\right\}$ partition $\left\{w: w \cdot x_i \geq 0\right\}$, and the $W(f)$'s are defined as all possible intersections (over $i$) of these elementary sets.

Let $F_k(1,1,\ldots,1)$ be the set of all N-tuples $f = (f_1, f_2, \ldots, f_N)$ such that $f_i \epsilon \{0,1\}$ for each $i = 1,2,\ldots,N$, and $f_i = 0$ for precisely $k$ indices $i$. For example, $F_o(1,1,\ldots,1)$ is just the N-tuples $(1,1,\ldots,1)$. It will be shown that the set of nonempty $W(f)$, $f \epsilon F_{d-k}$, is the set of open k-dimensional boundary faces of $W(1,1,\ldots,1)$; where a k-boundary $B$ of a cone $W$ is defined by the following two properties:

1. (Boundary property) Any $y$ belonging to $\overline{B}$ where $y$ is written as a sum of two vectors $y_1$, $y_2$ in $\overline{W}(1,1,\ldots,1)$ implies that either $y_1$ or $y_2$ is in $\overline{W}(f)$.
2. (Nondegeneracy property) The linear space generated by $B$ is k-dimensional.

$W(f)$ has property 1., as can be seen directly from the definition of $W(1,1,\ldots,1)$ and $W(f)$. Since $y \epsilon \overline{W}(f)$,

$$\begin{aligned} y \cdot x_i &= 0, \quad \text{for all } i \text{ such that } f_i = 0 \\ y \cdot x_i &\geq 0, \quad \text{for all } i \text{ such that } f_i = 1 \end{aligned} \tag{5.21}$$

and since $y_1, y_2 \epsilon \overline{W}(1,1,\ldots,1)$,

$$\begin{aligned} y_1 \cdot x_i &\geq 0, \quad \text{for all } i \\ y_2 \cdot x_i &\geq 0, \quad \text{for all } i \ . \end{aligned} \tag{5.22}$$

So $y = y_1 + y_2$ implies

$$\begin{aligned} y \cdot x_i = y_1 \cdot x_i + y_2 \cdot x_i &= 0, \quad \text{if } f_i = 0 \\ y_1 \cdot x_i + y_2 \cdot x_i &\geq 0, \quad \text{if } f_i = 1 \ . \end{aligned} \tag{5.23}$$

Since $y_1 \cdot x_i$ and $y_2 \cdot x_i$ are nonnegative quantities, Eq. (5.23) implies that

$$\begin{aligned} y_1 \cdot x_i &= 0, \quad \text{if } f_i = 0 \\ y_2 \cdot x_i &= 0, \quad \text{if } f_i = 0 \ . \end{aligned} \tag{5.24}$$

Thus both $y_1$ and $y_2$ are in $\overline{W}(f)$. Property 2. follows from the general position of $\{x_1, x_2, \ldots, x_N\}$, which insures the nondegeneracy of the intersection of the cone $\{w\colon w \cdot x_i > 0 \text{ for all } i \text{ such that } f_i = 1\}$ with the k-dimensional subspace orthogonal to the set of $x_i$'s for which $f_i = 0$.

In Fig. 5, for example, the open arc AB, which is the intersection of the cone $W(1,1,1,0)$ with the surface of a three-sphere centered at the origin, is a two-dimensional boundary face of $W(1,1,1,1)$. The cone

$W(0,1,1,0)$, corresponding to the point B, is a one-dimensional boundary of $W(1,1,1,1)$; and the origin of the space is the unique 0-boundary or vertex of the cone $W(1,1,1,1)$.

With these preliminaries out of the way, it is now possible to count the number of faces of the solution cones. There are $C(N,d)$ nonempty d-dimensional cones formed by the intersection of $N$ $(d-1)$-dimensional subspaces in general position in d-space. Index these cones $W_1, W_2, \ldots, W_{C(N,d)}$. Let $R_k(W_i)$ be the number of k-boundaries of $\overline{W}_i$.

<u>Proposition 1</u>. (Counting the Sides of the Solution Cones.) Let $x_1, x_2, x_3, \ldots x_N$ be vectors in general position in d-space. Let $W_1, W_2, \ldots, W_{C(N,d)}$ be the solution cones for the homogeneous linear threshold functions $f: \{x_1, x_2, \ldots, x_N\} \to \{-1, 1\}$. Then

$$\sum_{i=1}^{C(N,d)} R_k(W_i) = 2^{d-k}\binom{N}{d-k}C(N-d+k,k) \tag{5.25}$$

for $k = 1, 2, \ldots, d$. Hence the expected number of k-boundaries of the solution set of a function chosen according to an equiprobable distribution over the class of homogeneous linear threshold functions is

$$E[R_k(W)] = \frac{2^{d-k}\binom{N}{d-k}C(N-d+k,k)}{C(N,d)} . \tag{5.26}$$

The behavior of Eqs. (5.25) and (5.26) will be studied in Chapter VF, after the dual cone has been discussed. The primary reason for emphasizing the probabilistic interpretation of $R_k$ arises when the patterns themselves are randomly distributed and it is desired, for example, to describe the solution cone of a <u>given</u> dichotomy of the patterns. Proposition 1 holds with probability 1 in the case of random pattern vectors when the set $\{x_1, x_2, \ldots, x_N\}$ is chosen according to some probability distribution such that the set is in general position with probability 1.

Thus, for example, let the $x_i$ be independent, identically distributed, random vectors chosen according to a uniform distribution on the surface of the unit sphere in d-space, and let $W$ be the set of all north poles

on the d-sphere such that the hemisphere corresponding to the north pole contains all the $x_i$, $i = 1,2,\ldots,N$. Then the expected number of sides of $W$ is given in Eq. (5.26).

Proof of Proposition. Let $F_k$ be the set of all N-tuples $f = (f_1, f_2, \ldots, f_N)$ such that $f_i \epsilon \{-1,0,1\}$ for each $i = 1,2,\ldots,N$, and $f_i = 0$ for precisely $k$ indices $i$. Thus there are $\binom{N}{k}2^{N-k}$ elements in $F_k$.

It is desired to count the k-boundaries of $W(f)$ over all $f \epsilon F_o$ for which $W(f)$ is nonempty. The k-boundaries of $W(f)$, $f \epsilon F_o$, are the totality of nonempty k-cones $W(g)$, where $g \epsilon F_{d-k}$, and $g$ agrees with $f$ in the nonzero coordinates. For each $f \epsilon F_o$ there are $\binom{N}{d-k}$ g's in $F_{d-k}$ which agree with $f$ in the nonzero coordinates.

In other words, the set of k-boundaries of the $W(f)$'s, $f \epsilon F_o$, is contained in the set of $W(g)$'s, $g \epsilon F_{d-k}$. Now it is observed that each $W(g)$, $g \epsilon F_{d-k}$, is a k-boundary for exactly $2^{d-k}$ of the d-cones $W(f)$, $f \epsilon F_o$. In fact, $W(g)$ is a k-boundary for those $W(f)$, $f \epsilon F_o$ that agree with $g$ in the nonzero coordinates.

The results developed in the proof of Theorem 5 of Chapter VB yield the number of nonempty $W(g)$, $g \epsilon F_{d-k}$, to be $\binom{N}{d-k}C(N-d+k,k)$. Then, since each $W(g)$ is counted $2^{d-k}$ times

$$\sum_{f \epsilon F_o} R_k(W(f)) = 2^{d-k}\binom{N}{d-k}C(N-d+k,k) \tag{5.27}$$

as desired.

## D. THE DUAL CONE TO THE SOLUTION CONE

Let $B$ be a convex cone defined in a Euclidean d-dimensional space. The dual cone or polar cone of $B$, denoted $B^*$, is defined to be the set of all vectors in the spaces that form an acute angle with every vector in $B$. More precisely,

$$B^* = \{w: w \cdot b \geq 0, \text{ all } b \epsilon B\} \quad . \tag{5.28}$$

If $\overline{W}(f)$ is the polyhedral convex cone of solution vectors to the

set of inequalities

$$\begin{aligned} w \cdot x_i &\geq 0, \quad \text{for all } i \text{ such that } f_i = 1 \\ w \cdot x_i &\leq 0, \quad \text{for all } i \text{ such that } f_i = -1 \end{aligned} \tag{5.29}$$

Then $\overline{W}^*(f)$ may be expressed alternatively as the convex cone <u>spanned</u> by the set of vectors $\{f_1 x_1, f_2 x_2, \ldots, f_N x_N\}$ defined by

$$\overline{W}^*(f) = \left\{w: w = \sum_{i=1}^{N} \alpha_i f_i x_i, \quad \alpha_i \geq 0\right\} . \tag{5.30}$$

The equivalence of Eqs. (5.28) and (5.30) as definitions of $\overline{W}^*(f)$ is very useful in the study of convex cones. The proof, due to Farkas [Ref. 23], is presented in Theorem 3 of Goldman and Tucker [Ref. 6]. Notice that $\overline{W}^*$, the set of all nonnegative linear combinations of the **pattern vectors as expressed in Eq. (5.30), is the set of all admissible training states of an Adaline under those training procedures listed in Chapter VA.**

An example of the convex cone spanned by four vectors in three dimensions is illustrated in Fig. 6. The reader should convince himself that the cone spanned by $x_1, x_2, x_3, x_4$ is indeed the set of all vectors forming a nonnegative inner product with every vector in the cone ABCD in Fig. 5. The one-boundaries of $W^*(1,1,1,1)$ are the rays $\lambda x_i$ for $\lambda \geq 0$ and $i = 1,2,3,4$. The two-boundaries are the four two-dimensional cones spanned by $\{x_1, x_2\}$, $\{x_2, x_3\}$, $\{x_3, x_4\}$, and $\{x_1, x_4\}$. The three-boundary of $W(1,1,1,1)$ is $W(1,1,1,1)$ itself. If there exists no half space containing all the $x_i$, then the cone spanned by the $x_i$ is the whole space. A cone that is not the whole space is a <u>proper</u> cone.

Previous results on the solution cone will carry over to the dual cone because of the following lemma.

<u>Lemma 2</u>. (Correspondence Between Boundaries of Solution Cone and Dual Cone.) Let $x_1, x_2, \ldots, x_N$ be vectors in d-space. Let

$$W = \left\{w: w \cdot x_i \geq 0, \quad i = 1,2,\ldots,N\right\} \tag{5.31}$$

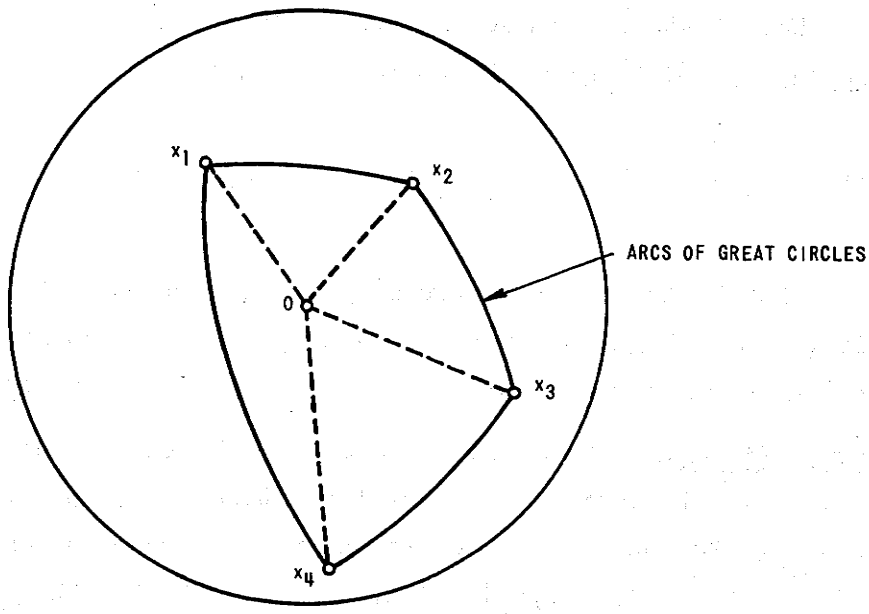



FIG. 6. INTERSECTION OF SPHERE WITH CONVEX CONE SPANNED BY $x_1$, $x_2$, $x_3$, $x_4$. DUAL CONE TO CONE ABCD IN FIG. 5.

and let the polar cone of $W$ be

$$W^* = \left\{v\colon v \cdot w \geq 0, \quad \text{for all } w \epsilon W\right\} , \tag{5.32}$$

which, by previous remarks is also given by

$$W^* = \left\{v\colon v = \sum_{i=1}^{N} \alpha_i x_i, \quad \alpha_i \geq 0\right\} . \tag{5.33}$$

Let $H$ be a subset of $\left\{x_1, x_2, \ldots, x_N\right\}$. Then

$$B = \left\{w\colon w \cdot x = 0, \quad x \epsilon H; \quad w \cdot x \geq 0, \quad x \not\epsilon H\right\} \tag{5.34}$$

is a $(d-k)$-face of $W$ if and only if

$$\tilde{B} = \left\{v\colon v = \sum \alpha_i x_i, \quad \alpha_i \geq 0, \quad \text{and} \quad \alpha_i = 0 \text{ if } x_i \not\epsilon H\right\} \tag{5.35}$$

is a $k$-face of $W^*$. Thus, $R_k(W) = R_{d-k}(W^*)$.

Note that $\tilde{B}$ is not the dual cone to B. In fact, $B^*$ as defined in Eq. (5.28) is given from Eq. (5.34) by

$$B^* = \left\{v: v = \sum \alpha_i x_i,\ \alpha_i \geq 0 \text{ for } x_i \notin H, \text{ and } \alpha_i \text{ unrestricted for } x_i \in H\right\}. \tag{5.36}$$

Thus by the lemma, interchanging k and d-k in the proof of Proposition 1 gives a corresponding proposition counting the number $R_k$ of k-boundaries of the dual cones to the solution cones.

<u>Proposition 2</u>. (Counting the Sides of the Dual Cones.) Let $x_1, x_2, x_3, \ldots, x_N$ be vectors in general position in d-space. Let $W_1, W_2, \ldots, W_{C(N,d)}$ be the solution cones for the homogeneous linear threshold functions $f: \{x_1, x_2, \ldots, x_N\} \to \{-1, 1\}$. Let $W_1^*, W_2^*, \ldots W_{C(N,d)}$ be the corresponding dual cones, i.e., all possible proper cones spanned by $\{\pm x_1, \pm x_2, \ldots, \pm x_N\}$. Then

$$\sum_{i=1}^{C(N,d)} R_k(W_i^*) = 2^k \binom{N}{k} C(N-k, d-k) \tag{5.37}$$

for $k = 0, 1, 2, \ldots, d-1$. Hence the expected number of k-boundaries of the convex cone spanned by the $\pm x_i$'s chosen according to an equiprobable distribution over the set of all proper cones generated by the $\pm x_i$'s is

$$E[R_k(W^*)] = \frac{2^k \binom{N}{k} C(N-k, d-k)}{C(N,d)} . \tag{5.38}$$

The implications of Proposition 2 will not become apparent until the limiting behavior of Eq. (5.38) is found in Sec. F of this chapter. Again it is surprising that the total number of k-boundaries of the $W_i^*$ is essentially independent of the configuration of the pattern vectors which generate them. Proposition 2 has immediate application to many problems in geometrical probability—enabling, for certain distributions, the calculation of the expected number of extreme points of convex polyhedrons generated by random points in d-space.

## E. VOLUME OF THE SOLUTION CONE AND THE DUAL CONE

Let $\mu$ be some finite measure defined on the d-space in which the pattern vectors are defined. It is assumed here that $\mu$ is absolutely continuous with respect to natural Lebesgue measure in this space in order to avoid the slight complications arising when $\mu$ is such that the boundaries of the solution cone or dual cone have nonzero measure. Let us further assume that the $\mu$ measure of the whole space is 1.

In this section the expected measure of the set of k-boundaries of the solution cone and the dual cone corresponding to a random set of consistent linear inequalities is found for all k.† In a special case, where the pattern vectors themselves are randomly distributed according to a uniform probability distribution over the surface of the unit d-sphere, an explicit expression is found for the variance of the measure of the solution cone. In any case a simple probability bound on the measure of the random cones will be found.

The $C(N,d)$ nonempty solution cones of a set of linear threshold functions on $N$ vectors in general position in d-space partition the space. Thus,

$$1 = \mu\left\{\bigcup_{i=1}^{C(N,d)} W_i\right\} = \sum_{i=1}^{C(N,d)} \mu\left\{W_i\right\} , \tag{5.39}$$

and the expected measure of a random nonempty solution cone is

$$E\left\{\mu(W)\right\} = \frac{1}{C(N,d)} . \tag{5.40}$$

Similarly, if $\mu^{(k)}$ is a measure on d-space assigning measure 1 to every k-dimensional subspace, then the measure of the union of all the k-boundaries of the solution cones $W_1, W_2, \ldots, W_{C(N,d)}$ is just $\binom{N}{d-k}$

†The work on finding the expected measure of $W$ and $W^*$ as well as finding the expected number of (d-1)-boundaries of $W$ was begun with B. Efron of Stanford University, with whom a joint paper is being prepared for publication. R. Brown of Stanford University has shown the expected measure of $W$ to be $(\frac{1}{2})^N$ when all $2^N$ dichotomies are equiprobable.

the number of different k-dimensional subspaces corresponding to intersections of subspaces orthogonal to the N vectors. Since each k-boundary is counted $2^{d-k}$ times in this union, and since there are $C(N,d)$ solution cones, the following proposition has been established.

Proposition 3. (The Volume of the Solution Cones.) The sum of the measures of the sets $B_i^{(k)}$ of k-boundaries of the nonempty solution cones $W_i$ for a set of homogeneous linear threshold functions defined on N vectors in general position in d-space is

$$\sum_{i=1}^{C(N,d)} \mu^{(k)}\left\{B_i^{(k)}\right\} = 2^{d-k}\binom{N}{d-k}, \qquad k = 1,2,\ldots,d \tag{5.41}$$

and the expected measure of the k-boundaries of a random nonempty solution cone is

$$E\mu^{(k)}\left\{B^{(k)}\right\} = \frac{2^{d-k}\binom{N}{d-k}}{C(N,d)} \tag{5.42}$$

Although it is in general a difficult problem to find the measure of the dual cone $W^*$ in terms of the definition of the solution cone $W$, the fact that the solution cones partition the space together with the crucial fact that every (d-1)-dimensional subspace intersects the same number of solution cones will allow the computation of the total measure of all the proper dual cones.

Proposition 4. (The Volume of the Dual Cones.) Let $x_1, x_2, \ldots, x_N$ be vectors in general position in d-space. Let $W_1, W_2, \ldots, W_{C(N,d)}$ be the solution cones for the homogeneous linear threshold function $f\colon \left\{x_1, x_2, \ldots, x_N\right\} \to \left\{-1,1\right\}$. Let $W_1^*, W_2^*, \ldots, W_{C(N,d)}^*$ be the corresponding dual cones. Then for a measure $\mu$, assigning measure one to d-space,

$$\sum_{i=1}^{C(N,d)} \mu\left\{W_i^*\right\} = \binom{N-1}{d-1} \tag{5.43}$$

and hence the expected measure of a random (proper) dual cone is

$$E\mu\{W^*\} = \frac{\binom{N-1}{d-1}}{C(N,d)} \tag{5.44}$$

<u>Proof</u>. Define the indicator function $f_i$ of the cone $W_i^*$ on the d-dimensional space as follows:

$$f_i(v) = \begin{cases} 1, & v \in W_i^* \\ 0, & v \notin W_i^* \end{cases} . \tag{5.45}$$

Then

$$\mu(W_i^*) = \int f_i(v)\, d\mu(v) , \tag{5.46}$$

and

$$\begin{aligned} \sum_{i=1}^{C(N,d)} \mu(W_i^*) &= \sum_{i=1}^{C(N,d)} \int f_i(v)\, d\mu(v) \\ &= \int \sum_{i=1}^{C(N,d)} f_i(v)\, d\mu(v) . \end{aligned} \tag{5.47}$$

But for almost every vector $v$ it can be shown that

$$\sum_{i=1}^{N} f_i(v) = \binom{N-1}{d-1} \tag{5.48}$$

independent of $v$. Thus Eq. (5.47) becomes

$$\sum_{i=1}^{C(N,d)} \mu(W_i^*) = \int \binom{N-1}{d-1} d\mu(v) = \binom{N-1}{d-1} \tag{5.49}$$

because the $\mu$-measure of the whole space is assumed to be 1.

To establish Eq. (5.48), let the vector $v$ partition the set $S$ of solution cones into three sets defined by

$$\underline{S}^+ = \left\{W \epsilon \underline{S}: v \cdot w > 0, \text{ all } w \epsilon W\right\}$$

$$\underline{S}_o = \left\{W \epsilon \underline{S}: v \cdot w = 0, \text{ some } w \epsilon W\right\} \tag{5.50}$$

$$\underline{S}^- = \left\{W \epsilon \underline{S}: v \cdot w < 0, \text{ all } w \epsilon W\right\} .$$

There are $C(N,d)$ cones in $\underline{S}$; and there are $C(N,d-1)$ cones in $\underline{S}_o$ by the general position of $v$ with respect to $\left\{x_1, x_2, \ldots, x_N \epsilon E^d\right\}$ for almost every $v$. The number of cones in $\underline{S}^+$ is equal to the number of cones in $\underline{S}^-$ because the reflection of a solution cone through the origin is also a solution cone (of the complementary function). Thus the number of elements in $S^+$ is

$$\tfrac{1}{2}\Big(C(N,d) - C(N,d-1)\Big) = \binom{N-1}{d-1} . \tag{5.51}$$

Finally, neglecting a set of points of $\mu$-measure zero, $v$ is in $W_i$ (and $f_i(v) = 1$), if and only if $W_i$ is in $S^+$. Thus $f_i(v)$ is equal to 1 for precisely $\binom{N-1}{d-1}$ indices $i$ and Eq. (5.48) and the proposition are established.

A separate argument will be required to find the measure of the boundaries of the dual cones.

<u>Proposition 5</u>. (The Volume of the Boundaries of the Dual Cones.) Let $x_1, x_2, \ldots, x_N$ be in general position in d-space. Let $W_1^*, W_2^*, \ldots, W_{C(N,d)}^*$ be the set of proper cones spanned by sets of $N$ vectors of the form $\left\{\pm x_1, \pm x_2, \ldots, \pm x_N\right\}$. Then for a measure $\mu^{(k)}$ assigning measure 1 to every k-dimensional subspace, the sum of the measures of the set $B_i^{(k)}$ of k-boundaries of the proper cones $W_i^*$ is

$$\sum_{i=1}^{C(N,d)} \mu^{(k)}\left\{B_i^{(k)}\right\} = \binom{N}{k} C(N-k, d-k), \qquad k = 0,1,2,\ldots,d-1 . \tag{5.52}$$

Thus the expected measure of the set of k-boundaries of a random proper cone is

$$E_\mu^{(k)}\left\{B^{(k)}\right\} = \frac{\binom{N}{k}C(N-k,d-k)}{C(N,d)}, \qquad k = 0,1,2,\ldots,d-1 . \qquad (5.53)$$

Proof. There are $\binom{N}{k}$ ways to select a k-element subset H of $\{x_1,x_2,\ldots,x_N\}$. Each of the $2^k$ cones generated by $\{\pm x_i : x_i \in H\}$ is a proper k-dimensional cone. Moreover, these $2^k$ cones partition the k-dimensional subspace spanned by H. Hence the sum of the measures of the cones generated by H is 1, by hypothesis on $\mu^{(k)}$.

For each of the $2^k$ cones generated by H there are, by the function-counting theorem, $C(N-k,d-k)$ homogeneously linearly separable dichotomies of $H^c$ (the remaining N-k vectors) in the (d-k)-dimensional space orthogonal to that spanned by H. In these cases, and only these cases, is the generated cone a boundary face. Thus there are $\binom{N}{k}C(N-k,d-k)$ sets of $2^k$ boundary k-faces, and each such set of $2^k$ boundary k-faces has total measure 1. Therefore, Eq. (5.52) is verified.

The variance of the measure of the solution cone can be obtained in the special, but interesting, case when $\mu$ is the natural measure of the surface area of a d-dimensional sphere of unit area centered at the origin of the d-space. Thus the measure of a cone is the solid angle subtended by the cone.

Define the indicator function

$$f(v) = \begin{cases} 1, & v \in W(1,1,\ldots,1) \\ 0, & v \notin W(1,1,\ldots,1) \end{cases}, \qquad (5.54)$$

where, as usual, W is the solution cone of vectors w such that

$$w \cdot x_i > 0, \qquad i = 1,2,\ldots,N .$$

Then

$$\mu\{W\} = \int f(v)\, d\mu(v) \tag{5.55}$$

and, squaring this and taking expected values,

$$E\mu^2\{W\} = E \int f(v)\, f(v')\, d\mu(v)\, d\mu(v') \tag{5.56}$$

But $f(v)f(v') = 1$ only if

$$\begin{aligned} v \cdot x_i &> 0, \qquad i = 1,2,\ldots,N \\ v' \cdot x_i &> 0, \qquad i = 1,2,\ldots,N \ . \end{aligned} \tag{5.57}$$

Then if the $x_i$ are independent identically distributed according to $\mu$, the probability that all $x_i$, $i = 1,2,\ldots,N$ lie in the wedge $\{w\colon w \cdot v > 0,\ w \cdot v' > 0\}$ (and thus $f(v)f(v') \neq 0$) is equal to the $N^{th}$ power of the area of the wedge. Thus Eq. (5.56) becomes

$$E\mu^2\{W\} = \frac{\int_0^{\frac{1}{2}} y^N \sin^{d-2} 2\pi y dy}{\int_0^{\frac{1}{2}} \sin^{d-2} 2\pi y dy} , \tag{5.58}$$

which can also be expressed in terms of gamma functions. However, a probability bound on $\mu(w)$ which is almost as informative as Eq. (5.58) will be derived in Proposition 6.

Finally, since the number of k-boundaries $R_k(W)$ and the measure of the set of k-boundaries $\mu^{(k)}\{B^{(k)}\}$ of a random, nonempty solution cone are <u>positive</u> random variables, the following proposition bounding the probability that $R$ and $\mu$ exceed a given constant is an elementary consequence of Eqs. (5.26) and (5.42) expressing the expected values of $R$ and $\mu$.

<u>Proposition 6</u>. For $t > 0$, for a random nomempty solution cone $W$ of a set of homogeneous linear threshold functions on $N$ vectors in general position in $d$ dimensions, the distribution of the number $R_k$ and measure $\mu^{(k)}$ of the set of k-boundaries of $W$ obey

$$\Pr\left\{R_k(W) \geq t\bar{R}\right\} \leq \frac{1}{t} \tag{5.59}$$

$$\Pr\left\{\mu^{(k)}\left\{B^{(k)}\right\} \geq t\bar{\mu}\right\} \leq \frac{1}{t}, \qquad k = 1, 2, \ldots, d \tag{5.60}$$

where

$$\bar{R} = \frac{2^{d-k}\binom{N}{d-k}C(N-d+k,k)}{C(N,d)} \tag{5.61}$$

and

$$\bar{\mu} = \frac{2^{d-k}\binom{N}{d-k}}{C(N,d)} \quad . \tag{5.62}$$

<u>Proof</u>. If $R$ is a positive random variable with distribution $F$ and mean $\bar{R}$, then

$$\bar{R} = \int_0^\infty r dF(r) \geq t\int_t^\infty dF(r) = t\Pr\{R \geq t\} \quad . \tag{5.63}$$

<u>Example</u>. Consider a set of 200 pattern vectors in general position in a 100-dimensional space. There are $2^{200}$ dichotomies of the pattern set of which precisely $C(200, 100) = 2^{199}$ are homogeneously linearly separable. Therefore a random dichotomy is separable with probability $\frac{1}{2}$. Then the average measure of a solution cone, conditioned on the event that the pattern set is homogeneously linearly separable, is

$$E\mu^{(100)}\{W\} = \left(\tfrac{1}{2}\right)^{199} \tag{5.64}$$

and, by Eq. (5.60), the conditional probability is less than $\left(\frac{1}{2}\right)^{44}$ that the measure of the solution cone is greater than $\left(\frac{1}{2}\right)^{150}$. Thus with overwhelming probability a random solution cone is quite small. (For comparison, an orthant in 100-space has natural measure $\left(\frac{1}{2}\right)^{100}$.)

## F. LIMITING BEHAVIOR OF SIZE AND SHAPE OF SOLUTION CONE AND DUAL CONE

In this section the previous sections are tied together by finding limiting values of the expected number of k-boundaries and the expected values of volumes of solution cones and dual cones for large numbers of patterns in a large dimensional space. The results obtained will simplify greatly the understanding and visualization of the nature of the solution cone and dual cone. It will be possible to show that, in a sense, the average set of solution vectors looks like a (d-1)-cube, and the average convex set generated by the pattern vectors looks like the standard geometrical dual of a (d-1)-cube. Moreover, the pronounced change in character of the description of the solution and dual cones as the ratio of patterns to dimensions exceeds 2 will be apparent, emphasizing again that the capacity $N/d = 2$ of a linear threshold device plays a critical role in the description of the performance of the device.

In the following the limit of the ratios of two cumulative binomial distributions will be important. The author is indebted to B. Elspas[†] for the following useful result.

<u>Proposition 7</u>. Let

$$G(N,d) = \sum_{i=0}^{d} \binom{N-1}{i} \Big/ \binom{N-1}{d} \tag{5.65}$$

and let

$$G^*(\beta) = \lim_{d \to \infty} G(\beta d, d) \quad . \tag{5.66}$$

Then

$$G^*(\beta) = \begin{cases} \infty \, , & 0 \leq \beta \leq 2 \\ \dfrac{\beta-1}{\beta-2} \, , & \beta \geq 2 \end{cases} \tag{5.67}$$

[†]B. Elspas of Stanford Research Institute; personal communication.

Proof. Observe that $G(N,d)$ satisfies the recursion

$$\begin{aligned} G(N,d) &= 1 + \frac{d}{N-d} G(N,d-1) \\ &= 1 + \frac{1}{\beta-1} G(N,d-1) \end{aligned} \tag{5.68}$$

for $\beta = N/d$. For $\beta-1 > 1$, the finite limiting value $G^*(\beta)$ of $G(N,d)$ is the solution of

$$G^*(\beta) = 1 + \frac{1}{\beta-1} G^*(\beta) \quad . \tag{5.69}$$

Thus

$$G^*(\beta) = \frac{\beta-1}{\beta-2} \qquad \text{for} \quad \beta > 2 \quad . \tag{5.70}$$

Now

$$G^*(\beta) = \infty \qquad \text{for} \quad 0 \leq \beta \leq 2 \tag{5.71}$$

because

$$\lim_{d \to \infty} G(2d,d) = \infty \tag{5.72}$$

and $G(\beta d,d)$ is monotonic decreasing with $\beta$ in the range $0 < \beta \leq 2$. In fact, for $0 < \beta < 1$ and finite $d$,

$$G(\beta d,d) = \infty \tag{5.73}$$

follows from the fact that the denominator $\binom{N}{d-1}$ of $G(N,d)$ is equal to zero by definition for $N < d-1$.

Let $W$ be the solution cone for a set of $N$ homogeneous inequalities in $d$ unknowns,

$$w \cdot x_i \gtrless 0, \qquad i = 1,2,\ldots,N \quad , \tag{5.74}$$

where it is assumed that the set $\{x_1, \ldots, x_N\}$ is in general position and that the set of inequalities is chosen from an equiprobable distribution over the class of $C(N,d)$ consistent sets of inequalities. Recall that the random variables $R_k(W)$ and $R_k(W^*)$ are defined to be the number of k-boundaries of $W$ and $W^*$ respectively. In particular, $R_1(W^*)$ is the number of extreme rays of $W^*$. Each extreme ray corresponds to a positive multiple of one of the $x_i$. Thus, for $N \geq d$,

$$d \leq R_1(W^*) \leq N \quad , \tag{5.75}$$

for all proper cones $W^*$. The subset $Z$ of extreme vectors in $\{x_1, x_2, \ldots, x_N\}$ completely characterizes the set, with respect to a given homogeneous linear threshold function $f$, in the sense that, for all $w$ satisfying

$$w \cdot x_i = f_i, \qquad \text{for all } i \text{ such that } x_i \epsilon Z \quad , \tag{5.76}$$

the following complete system of inequalities must hold:

$$w \cdot x_i = f_i, \qquad \text{for all } x_i \epsilon \{x_1, x_2, \ldots, x_N\} \quad . \tag{5.77}$$

The vectors in $Z$ form the boundary matrix investigated by Mays [Ref. 22]. The expected number of extreme vectors of $\{x_1, x_2, \ldots, x_N\}$ will be shown to be bounded as $N$ approaches infinity.

<u>Proposition 8</u>. The asymptotic number of k-boundaries of $W$ and $W^*$ are

$$\lim_{N \to \infty} E\{R_k(W)\} = 2^{d-k}\binom{d-1}{d-k} \tag{5.78a}$$

$$\lim_{N \to \infty} E\{R_k(W^*)\} = 2^{k}\binom{d-1}{k} \tag{5.78b}$$

Proof. From Eq. (5.38),

$$\lim_{N\to\infty} E\left\{R_k(W^*)\right\} = \frac{2^k\binom{N}{k}C(N-k,d-k)}{C(N,d)}$$

$$= \lim_{N\to\infty} \frac{2^k\binom{N}{k}\sum_{i=0}^{d-k-1}\binom{N-k-1}{i}}{\sum_{i=0}^{d-1}\binom{N-1}{i}} \tag{5.79}$$

which, dropping the asymptotically negligible initial terms of each of the two summations,

$$= \lim_{N\to\infty} \frac{2^k\binom{N}{k}\binom{N-k-1}{d-k-1}}{\binom{N-1}{d-1}}$$

$$= \lim_{N\to\infty} 2^k\binom{d-1}{k}\frac{N}{N-k} \tag{5.80}$$

$$= 2^k\binom{d-1}{k} .$$

Equation (5.78a) follows from the identity

$$R_k(W) = R_{d-k}(W^*) , \tag{5.81}$$

which is a consequence of the duality of $W$ and $W^*$. Q.E.D.

Thus the expected number of boundaries for $W$ and $W^*$ tends to a limit that is characteristic of the dimension of the space in which the pattern vectors lie. Moreover, this limit suggests that the average solution cone is like a $(d-1)$-cube. To see this, determine the number of sides of each dimension of a $(d-1)$-dimensional cube. If an $n$-cube has $m_j(n)$ $j$-boundaries, $j = 0,1,2,\ldots,n$, then by translating the $n$-cube parallel to itself and connecting the vertices, the number of $j$-boundaries of a $(n+1)$-cube is given by

$$m_j(n+1) = 2m_j(n) + m_{j-1}(n), \qquad j = 0,1,2,\ldots,n+1 \tag{5.82}$$

resulting from counting the $j$-boundaries in initial and final position and the $j$-boundaries swept out by the $(j-1)$-boundaries.

Multiplying by an indeterminate $x^j$ and summing over $j$ gives

$$M_{n+1}(x) = 2M_n(x) + xM_n(x) \quad , \tag{5.83}$$

where

$$M_n(x) = \sum_{j=0}^{\infty} m_j(x)\, x^j \quad . \tag{5.84}$$

Hence

$$M_n(x) = (2+x)M_{n-1}(x) = (2+x)^n \tag{5.85}$$

and the number of $j$-boundaries of an $n$-cube is just the coefficient of $x^j$ in $(2+x)^n$:

$$m_j(n) = \binom{n}{n-j} 2^{n-j} \quad . \tag{5.86}$$

Thus, by comparison with Eq. (5.78a), it can be seen that the intersection of the polyhedral solution cone $W$ with the unit sphere in $d$-space is a polyhedral body (now in a $(d-1)$-dimensional variety), which has

precisely the same expected number of j-boundaries as does a (d-1)-cube. Thus, the prototype solution cone (for a large number of inequalities) intersects the d-sphere in a (d-1)-cube.

Similarly, the intersection of $W^*$ with the unit d-sphere looks like the dual of a (d-1)-cube (the dual of a 3-cube is an octahedron). It is perhaps surprising that so few of the N inequalities are essential in the determination of W and $W^*$.

When these results are applied to the inhomogeneous case of a large number of random lines distributed at random in the plane, the expected region formed is found to be a quadrilateral, which agrees with Kendall and Moran [Ref. 24, p. 57].

Note that the expected number of extreme inequalities (the number of patterns in the boundary matrix of Mays [Ref. 22]) tends to 2(d-1) in the limit for an infinite consistent set of homogeneous linear inequalities in d unknowns.

In the following proposition the number of extreme patterns in a random dichotomy is examined when both the number of patterns N and the number of dimensions d is large, but the ratio of N to d is finite.

Proposition 9. (The Asymptotic Number of Extreme Inequalities.) If, as before, $R_1(W^*)$ is the number of extreme vectors of the convex cone spanned by N d-dimensional vectors of random sign, then

$$\lim_{\substack{\beta=N/d \\ d\to\infty}} E\left\{\frac{R_1(W^*)}{2d}\right\} = \begin{cases} \frac{\beta}{2}, & 0 \le \beta \le 2 \\ 1, & \beta \ge 2 \end{cases} . \tag{5.87}$$

Proof. Note that the limiting behavior of $E\left\{\frac{R_{d-1}(W)}{2d}\right\}$ is also given by (5.87). From Eq. (5.38),

$$E\left\{\frac{R_1(W^*)}{2d}\right\} = \frac{2N\ C(N-1,d-1)}{2d\ C(N,d)} . \tag{5.88}$$

Substituting the recursion relation

$$C(N,d) = C(N-1,d) + C(N-1,d-1) \tag{5.89}$$

gives

$$E\left\{\frac{R_1(W^*)}{2d}\right\} = \frac{N}{d} \frac{C(N-1,d-1)}{C(N-1,d) + C(N-1,d-1)} .$$

It can be verified by substitution that

$$E\left\{\frac{R_1(W^*)}{2d}\right\} = \frac{N(G(N-1,d-1)-1)}{d(2G(N-1,d-1)-1)}$$

where $G(N,d)$ is defined in Eq. (5.65). Thus, by Proposition 7,

$$\lim_{\substack{\beta=N/d \\ d\to\infty}} E\left\{\frac{R_1(W^*)}{2d}\right\} = \frac{\beta(G^*(\beta)-1)}{2G^*(\beta)-1}$$

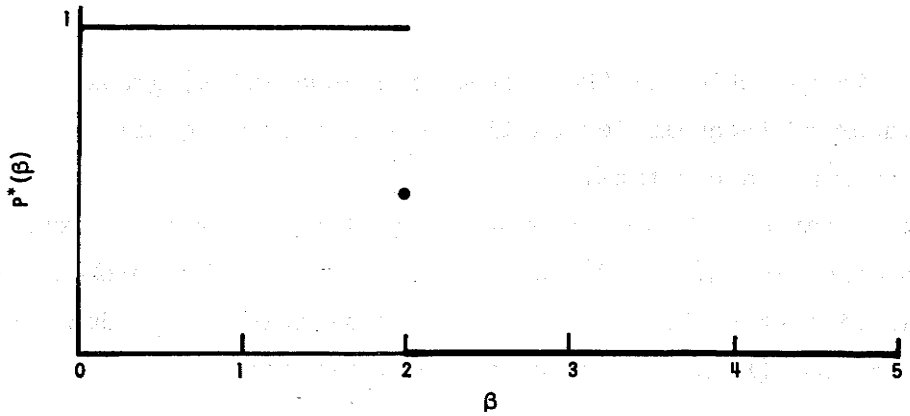
$$= \begin{cases} \frac{\beta}{2}, & 0 \le \beta \le 2 \\ 1, & \beta \ge 2 . \end{cases} \tag{5.92}$$

First, note that the limiting average number of extreme vectors is independent of $\beta$ for $\beta \geq 2$. The capacity has played a role again. In fact, it is now possible to draw a parallel between the theory of systems of linear equalities and the theory of systems of linear inequalities.

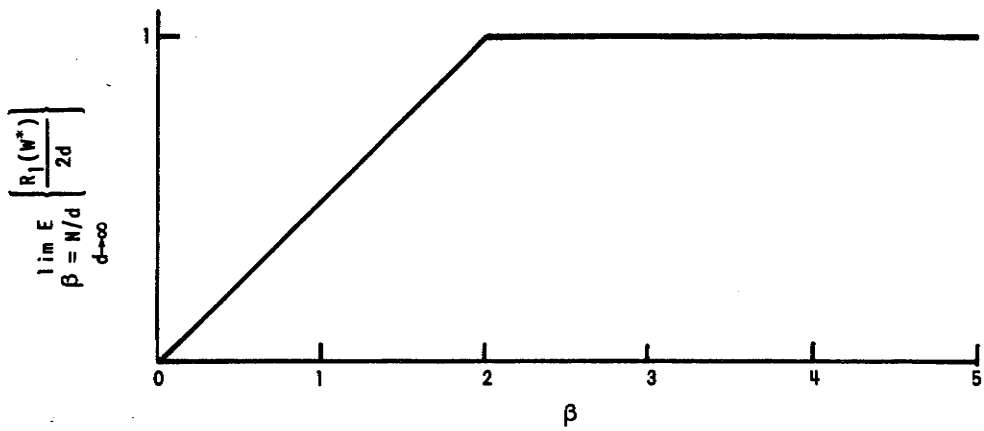
In a system of linear equalities in $d$ variables, the "capacity" is $d$—the maximum number of consistent equations. In a system of (random) inequalities the (statistical) capacity has been shown (in Chapter IV) to be $2d$ (in the limit for large $d$). Moreover, while the number of independent equations grows linearly with the number of equations until the number of equations equals $d$ and then remains constant, the asymptotic expected number of "independent" inequalities (i.e., the

extreme inequalities of the system of inequalities) grows linearly with the number of inequalities until the number of inequalities equals 2d, and then remains constant.

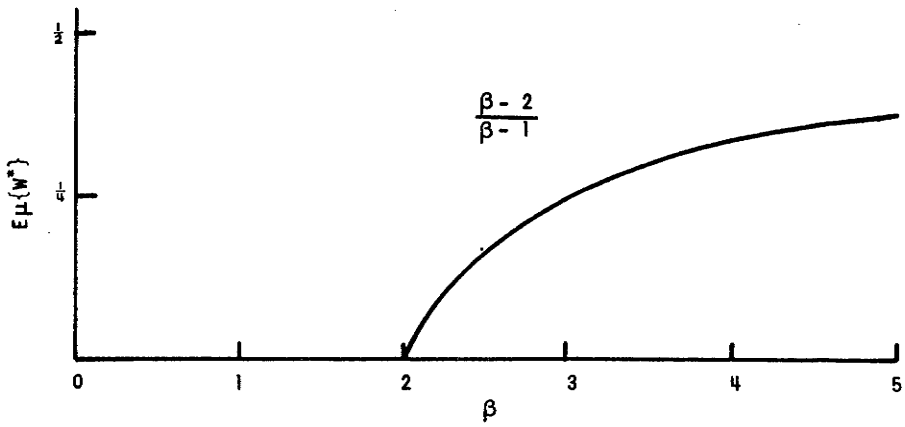
The behavior of the asymptotic expected number of extreme inequalities, as expressed in Eq. (5.87) is shown in Fig. 7b. The probability of separability (Fig. 7a) and the expected measure of the dual cone to the solution cone (Fig. 7c) are shown for comparison.




a. Limiting probability $P^*(\beta)$ that $\beta d$ points are separable in d dimensions




b. Limiting expected proportion of extreme inequalities in $\beta d$ inequalities in d unknowns




c. Limit of expected volume of random proper convex cone spanned by $\beta d$ vectors in d dimensions

FIG. 7. LIMITING GEOMETRIC PROBABILITIES.

## VI. GENERALIZATION AND LEARNING

### A. DEFINITIONS OF GENERALIZATION

Consider a set $X$ of $N$ pattern vectors in general position in d-space. This set of pattern vectors, together with a dichotomy of the set into two categories $X^+$ and $X^-$ will constitute a training set. On what basis can a new point $x_{n+1}$ be categorized into one of the two training categories? This is the problem of generalization.

If each pattern vector is drawn according to some probability distribution, then methods of decision theory provide schemes of classification based on criteria such as minimizing the probability of error. For example, if patterns in $X^+$ are chosen from a normal distribution with mean $\mu^+$ and covariance matrix $K^+$, while patterns in $X^-$ are chosen from a normal distribution with mean $\mu^-$ and covariance matrix $K^-$, then the optimal decision surface with respect to minimizing the probability of error is a quadric surface, which degenerates to a hyperplane if $K^+ = K^-$. Recall from Chapter III that augmented linear threshold devices can implement quadrics and will converge, using standard training algorithms, to a separating quadric when one exists.

Consider the problem of generalizing from the training set with respect to a given admissible family of decision surfaces—that family of surfaces that can be implemented by linear threshold devices. By some process, a decision surface from the admissible class will be selected which correctly separates the training set into the desired categories. Then the new pattern will be assigned to the category lying on the same side of the decision surface.

An important related problem to that of choosing a good decision surface from the admissible class is the problem of selecting the natural admissible class of decision surfaces. Presumably, in physical problems, some basis for selection will exist. For example, for two categories normally distributed with different means and identical covariances, the class of all hyperplanes would be the natural admissible class of separating surfaces. It would not be wise to let the class of all quadrics by the admissible class because, although a hyperplane is a special quadric,

the additional degrees of freedom of quadrics would cause longer convergence times and greater probability of ambiguous response. A very weak definition of generalization has been given here, in that no statistical knowledge of the pattern distribution is assumed, and no metrical concepts of closeness are introduced. There is no "best" separating decision surface among the class of separating surfaces.

Clearly, for some dichotomies of the set of training patterns, the assignment of category will not be unique. However, it is generally believed that, after a "large number" of training patterns, the state of a linear threshold device is sufficiently constrained to yield a unique response to a new pattern. If all the training sets are equally likely, it will be shown in the next section that the number of training patterns must exceed the statistical capacity of the linear threshold device before unique generalization becomes probable.

## B. AMBIGUITY

The classification of a pattern $y$ with respect to the training set $\{X^+, X^-\}$ is said to be <u>ambiguous relative to a given class of $\phi$-surfaces</u>, if there exists one $\phi$-surface in the class that induces the dichotomy $\{X^+ \cup \{y\}, X^-\}$ and another $\phi$-surface in the class that induces the dichotomy $\{X^+, X^- \cup \{y\}\}$. That is, there exist two $\phi$-surfaces, both correctly separating the training set, but yielding different classifications of the new pattern $y$. Thus, if $w_1$ and $w_2$ are the parameter weight vectors for the two $\phi$-surfaces, then

$$w_1 \cdot x > 0 \quad \text{and} \quad w_2 \cdot x > 0, \qquad \text{for} \quad x \epsilon X^+$$

$$w_1 \cdot x < 0 \quad \text{and} \quad w_2 \cdot x < 0, \qquad \text{for} \quad x \epsilon X^-$$

and either

$$w_1 \cdot y > 0 \quad \text{and} \quad w_2 \cdot y < 0 \tag{6.1}$$

or

$$w_1 \cdot y < 0 \quad \text{and} \quad w_2 \cdot y > 0 \quad .$$

In addition, $y$ is said to be ambiguous with respect to the training set if the training set is not separable.

In Fig. 8, for example, points $y_1$ and $y_2$ are unambiguous and point $y_3$ is ambiguous with respect to the training set $\{X^+, X^-\}$

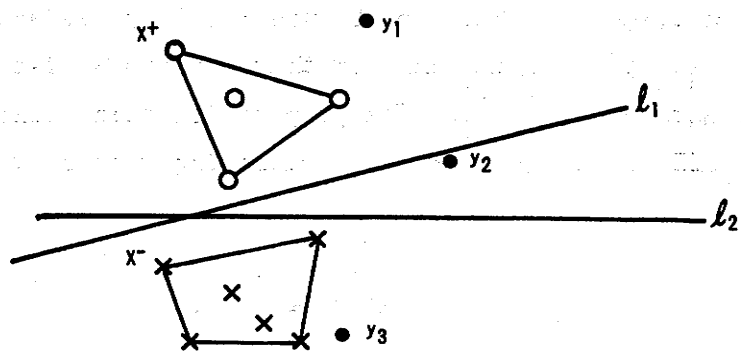



FIG. 8. AMBIGUOUS GENERALIZATION.

relative to the class of all lines in the plane (not necessarily through the origin). Points $y_1$ and $y_2$ are uniquely classified into sets $X^+$ and $X^-$ respectively by any line separating $X^+$ and $X^-$, while $y_2$ is classified either way by lines $\ell_1$ and $\ell_2$.

The following proposition establishes that, with respect to random dichotomies, the probability that a new pattern is ambiguous with respect to a random dichotomy of the training set is independent of the configuration of the pattern vectors.

<u>Proposition 10</u>. Let $X \cup \{y\} = \{x_1, x_2, \ldots, x_N, y\}$ be in $\phi$-general position in d-space, where $\phi = (\phi_1, \phi_2, \ldots, \phi_d)$. Then $y$ is ambiguous with respect to $C(N, d-1)$ dichotomies of $X$ relative to the class of all $\phi$-surfaces. Hence, if each of the $\phi$-separable dichotomies of $X$ has equal probability, then the probability $F(N,d)$ that $y$ is ambiguous with respect to a random $\phi$-separable dichotomy of $X$ is

$$F(N,d) = \frac{C(N,d-1)}{C(N,d)} = \frac{\sum_{k=0}^{d-2} \binom{N-1}{k}}{\sum_{k=0}^{d-1} \binom{N-1}{k}} \tag{6.2}$$

Proof. From lemma 1 of Chapter II, the point $y$ is ambiguous with respect to $\{X^+, X^-\}$ if and only if there exists a $\phi$-surface containing $y$ which separates $\{X^+, X^-\}$. The proposition then follows from Theorem 3 of Chapter III, on noting that the separating vector $w$ is constrained by

$$w \cdot \phi(y) = 0 \quad . \tag{6.3}$$

Applying the proposition to the example in Fig. 8, where $X$ has 10 points, it can be seen that each of the $y_i$ is ambiguous with respect to $C(10,2) = 20$ dichotomies of $X$ relative to the class of all lines in the plane. Now $C(10,3) = 92$ dichotomies of $X$ are separable by the class of all lines in the plane. Thus a new pattern is ambiguous with respect to a random, linearly separable, dichotomy of $X$ with probability

$$F(10,3) = \frac{C(10,2)}{C(10,3)} = \frac{5}{23} \quad . \tag{6.4}$$

C. LIMITING FORM OF PROBABILITY OF AMBIGUITY

Proposition 7 of Chapter V will be applied to find the limiting form of $F(N,d)$ for large $N$ and $d$. Write, using the definition in Eq. (5.60),

$$F(N,d) = \frac{\sum_{k=0}^{d-2} \binom{N-1}{k}}{\sum_{k=0}^{d-1} \binom{N-1}{k}} = 1 - \frac{\binom{N-1}{d-1}}{\sum_{k=0}^{d-1} \binom{N-1}{k}} = 1 - \frac{1}{G(N,d-1)} \quad . \tag{6.5}$$

Hence we have the asymptotic probability of ambiguity

$$F^*(\beta) = \lim_{d\to\infty} F(\beta d, d) = 1 - \frac{1}{G^*(\beta)} \tag{6.6}$$

or

$$F^*(\beta) = \begin{cases} 1 \ , & 0 \leq \beta \leq 2 \\ \frac{1}{\beta-1} \ , & \beta \geq 2 \end{cases} \tag{6.7}$$

by the conclusion of Proposition 7. The graph of $F^*(\beta)$ is shown in Fig. 9.

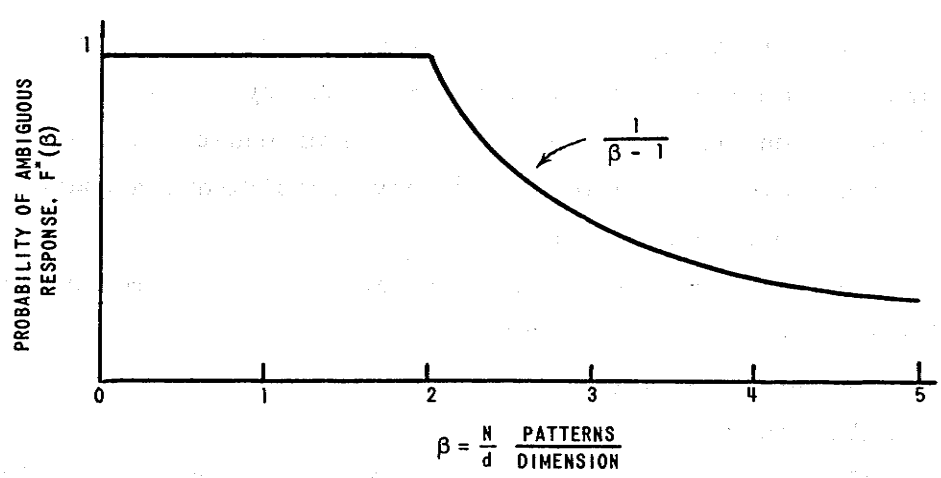



FIG. 9. ASYMPTOTIC PROBABILITY OF AMBIGUOUS GENERALIZATION.

Note the relatively large number of training patterns required for unambiguous generalization. If it is recalled that the capacity of a

linear threshold device is $\beta = 2$ patterns per variable weight, it can again be seen that the capacity is a critical number in the description of the behavior of a linear threshold device. Indeed, in this case, the probability of ambiguity in generalization remains high even after the probability of a consistent training set tends toward zero.

## D. GENERALIZING WITH RESPECT TO RANDOM PATTERNS

In the event that the patterns themselves are randomly distributed, the comments of Chapter IV concerning randomly distributed patterns and random dichotomies of the pattern set apply in full to this chapter. The crucial condition is that the pattern set by in general position with probability 1.

Thus, if a linear threshold device is trained on a set of $N$ points chosen at random according to a uniform distribution on the surface of a unit sphere in $d$-space, and these points are classified independently with equal probability into one of two categories, then it is readily seen that the probability of error on a new pattern similarly chosen, conditioned on the separability of the entire set, is just $\frac{1}{2}F(N,d)$.

## E. NECESSARY CONDITIONS FOR MEANINGFUL EXPERIMENTS IN GENERALIZATION

Consider an experiment consisting of training a 100-input linear threshold device on 70 patterns, followed by a test on 20 new patterns. It is evident that, since there are fewer equations than unknowns (i.e., fewer patterns than dimensions), any classification whatsoever can be made of the 20 new patterns.

The following three quantities must be specified in order to complete the description of the experiment:

1. The training algorithm;
2. The training sequence;
3. The initial state vector of the linear threshold device.

The effects of these factors on the response of the device to the new patterns can be quite remarkable. Concerning factor 3, for example, it can be shown that there exist initial state vectors, correctly separating the training set, such that any of the $2^{20}$ dichotomizations of

the new patterns can be made. If, on the other hand, the initial weight vector is zero, then the class of training algorithms that form only positive linear combinations of the pattern vectors will generate a solution vector that lies in the intersection of the solution cone with the dual cone—a restriction that may reduce the number of classifications that can be made on the new set of patterns.

Suppose that the training algorithm and the initial weight vector have been specified, but the training sequence is chosen at random. Then the response of the device to the new pattern set will be random. In this case, it is suggested that the problem be presented many times with sequences chosen at random in order that a proper statistical analysis can be made of the device performance.

Finally, when there is no natural distinction as to which $K$ of $N$ patterns in $d$ dimensions shall be used to train the device in order to recognize the remaining $N-K$ patterns, a condition 4 specifying which subset is used as a training set should be made, because it has been shown in Chapter V that, on the average, the specification of the classification of $2d$ patterns (the extreme patterns) is sufficient to establish the unambiguous classification of the remainder. Therefore, if the extreme patterns formed the training set, the linear threshold device would respond perfectly to the rest of the patterns. At the very least, condition 4 could be replaced by the condition that a series of experiments be performed on suitably random choices of training subsets.

A major point to be emphasized is that a unique response on new patterns cannot be made, <u>even with fixed data</u>, unless factors 1, 2, 3, and 4 are defined. If these factors are not precisely given, but rather are unknown or random, then the experiment should be repeated several times on the same data in order to provide a basis for a good statistical description of the performance of the device.

## VII. MINIMUM COMPLEXITY OF A NETWORK

In this chapter the function counting theorem of Chapter II will be applied to a large class of networks of linear threshold units in order to place a lower bound on the number of variable weights in a universal network. A network will be called universal with respect to a set of $N$ pattern vectors if the network can implement each of the $2^N$ functions from the pattern set to $\{-1,1\}$. Cameron [Ref. 7], Winder [Ref. 18] and Joseph [Ref. 10] have studied several specific network organizations of linear threshold units and have determined lower bounds on the number of linear threshold units (gates) in a universal network.

It has been shown that a single linear threshold unit has a capacity of two patterns per variable weight. Hence it is natural to ask for the capacity of a network of linear threshold units in terms of the total number of variable weights.

Consider a class $\underline{A}$ of networks of linear threshold units imbedded in fixed but arbitrary circuitry, as depicted in Fig. 10.

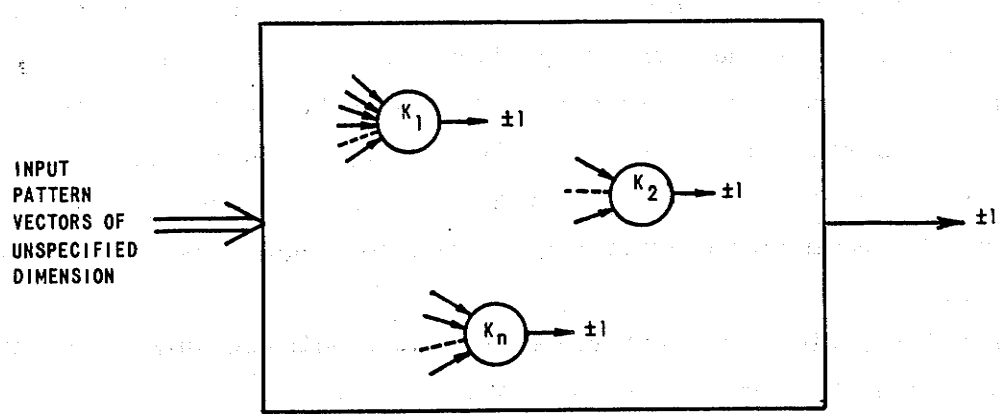



FIG. 10. NETWORK OF LINEAR THRESHOLD UNITS IMBEDDED IN ARBITRARY BUT FIXED CIRCUITRY.

However, in order to avoid problems of timing and stability it is required that there be no feedback from the output of any linear threshold unit to its input. Order the linear threshold units in any way that is not inconsistent with the flow of signal from input to output of the network, and let $K_i$ denote the number of variable weights in the $i^{th}$ linear

threshold unit. Note that it is not required that all the inputs to the $i^{th}$ linear threshold unit be utilized, nor is it required that all the dimensions of the input patterns be accomodated by the network. Let $T$ denote the total number of weights in a given network in the class $\underline{A}$. That is,

$$T = \sum_i K_i \quad . \tag{7.1}$$

Proposition 11. If a network in $\underline{A}$ containing a total of $T$ variable weights is universal with respect to an input set of $N$ patterns, then

$$T > \frac{N}{1 + \log_2 N} \quad . \tag{7.2}$$

Proof. An upper bound is to be placed on the number of states of any T-weight network in $\underline{A}$. Consider the $r^{th}$ linear threshold unit, with the weight vectors of the first $r-1$ linear threshold units fixed. The $r^{th}$ unit receives input $K_r$-tuples $\{x_1', x_2', \ldots, x_N'\}$ corresponding to the set of $N$ input states $\{x_1, x_2, \ldots, x_N\}$ to the network. Then, from the function-counting theorem, there are at most $C(N,K_r)$ different states of the $r^{th}$ unit with respect to $\{x_1', x_2', \ldots, x_N'\}$. (There are precisely $C(N,K_r)$ states if every $K_r$ element subset of $\{x_1', x_2', \ldots, x_N'\}$ is linearly independent.) Hence, an upper bound $r(N,T)$ on the number of states of the network is

$$r(N,T) = \max_{\sum_i K_i = T} \prod C(N,K_i) \quad . \tag{7.3}$$

Consider the crude bound on $C(N,K)$ holding for all positive integers $N$ and $K$:

$$C(N,K) = 2 \sum_{m=0}^{K-1} \binom{N-1}{m} \leq 2 \sum_{m=0}^{K-1} N^m \leq 2N^K \quad . \tag{7.4}$$

Thus

$$r(N,T) \leq \max_{\Sigma K_i = T} \Pi \, (2N^{K_i}) \tag{7.5}$$

or

$$r(N,T) \leq (2N)^T \quad . \tag{7.6}$$

There are $2^N$ functions mapping $\{x_1, x_2, \ldots, x_N\}$ to $\{-1, 1\}$. Thus the number of states of a universal network must exceed $2^N$. That is,

$$(2N)^T > 2^N \tag{7.7}$$

or

$$T > \frac{N}{1 + \log_2 N} \tag{7.8}$$

## VIII. SUMMARY AND CONCLUSIONS

Most of the foregoing analysis has been devoted to determining the intrinsic properties of linear threshold functions. The properties obtained, when suitably interpreted, provide an intuitive and reasonably concrete understanding of the nature of linear threshold functions and systems of simultaneous linear inequalities.

The essential contribution of this research to the theory of linear inequalities is the description of the set of all solution cones and dual cones corresponding to a system of $N$ homogeneous inequalities in $d$ unknowns. The sum of the number of $k$-boundaries of the solution cones and dual cones and the sum of the volumes of the solution cones and dual cones were found to be independent of the pattern configuration. For large $N$ and inhomogeneous inequalities the solution set can be said to "look like" a $d$-cube.

Applications of these geometrical results can be found in the theory of games, linear programming, and geometrical probability, in addition to the theory of linear threshold devices. Also, there are immediate applications of the geometrical analysis of the solution cones to expected tolerance requirements on the weight vectors of linear threshold devices.

Two distinct ideas of the capacity of a linear threshold device have been developed. The first idea—developed in Chapter IV—concerns the number of random patterns that a linear threshold device can be expected to separate. It is seen that the probability is quite small that more than $2d$ random patterns in $d$ dimensions can be separated. In fact, the number of additional patterns which are separable in excess of the capacity is a good index of the degree of correlation between pattern category and pattern position. Following this reasoning, different classes of separating surfaces can be tested on separating random patterns, and the class with the highest ratio of separated patterns to degrees of freedom of the separating surface will then be the most natural class of surfaces for that particular problem.

The second idea concerning the storage capacity of networks of

linear threshold devices resulted from counting the expected number of extreme patterns in a random, separable, dichotomy. It was found that the expected number of extreme patterns was equal to $2d$, in the limit, for an infinite number of linearly separable pattern vectors in $d$ dimensions, and that the probability that the number of extreme patterns exceeds $2dt$ is less than $1/t$. The implication for pattern recognition devices is that the essential information in an infinite training set can be expected to be stored in a computer of finite storage capacity.

Finally, the extension of the function counting theorem to so-called nonlinear threshold functions opens a wide area of application. Although the option of augmenting pattern vectors in order to generate nonlinear decision surfaces has always been open, it is now possible to evaluate the probability of separability by hyperspheres and other surfaces given the null hypothesis that pattern position and category are independent. Thus pattern recognition schemes having different dimensions and having different families of decision surfaces can be compared.

## IX. SUGGESTIONS FOR FUTURE WORK

### A. COUNTING THE NUMBER OF CONFIGURATIONS

In the previous chapters characteristics of systems of linear inequalities that are essentially configuration-free have been investigated. That is, subject only to the weak requirement of general position, the characteristics of the set of linear threshold functions are independent of the precise configuration of the set of pattern vectors on which the functions are defined. Thus, for example, the number of linear threshold functions and the total volume and the number of k-boundaries of the solution cones and dual cones depends only on the number of patterns $N$ and the dimension $d$ of the space.

However, many counting problems defy easy analysis because they are not configuration-free. Such problems include:

1. Counting the number assignments of inequalities that are within $r$ or fewer of being consistent (when $r = 0$, there are $C(N,d)$ such assignments).
2. Counting the number of dichotomies that can be achieved by forming unions of regions formed by several hyperplanes (as in a Madaline or Perceptron).

Two indexed families of points in a vector space are said to have differenent configurations (with respect to the class of all homogeneous linear threshold functions) if there exists a homogeneous linear threshold function on one family that does not exist on the other. The author has a geometric characterization of the properties of a set of points that determine its configuration, and has counted the number of configurations of $N$ points in 3 dimensions, but work remains to be done on the general problem of counting the number of configurations.

If the pattern vectors are randomly distributed, it is possible in some cases to find the relative probabilities of the various configurations. In a simple case this problem is known as Sylvester's problem—finding the probability that one of four points taken at random in a convex domain lies within the triangle formed by the other three—and Delthiel [Ref. 25] has carried out the calculation of this probability in many regular cases.

## B. COUNTING THE FUNCTIONS REALIZABLE BY NETWORKS

Although the function counting theorems allow development of upper bounds on the capacities of networks of linear threshold devices [Refs. 7, 10, 18], there is hope that exact answers can be obtained for special networks such as Madalines or Perceptrons.

If a set of pattern vectors is not separable, there are two possible procedures for increasing the complexity of the separating surface in order to separate the set using primarily implementation by linear threshold devices. The first scheme—described in Chapter III—is to augment the pattern vector with appropriate nonlinear functions of itself and separate the new set of augmented patterns by hyperplanes in the higher dimensional space. All of the old theory continues to apply. The second scheme is to pass several planes through the original pattern space and to form a dichotomy of the pattern space by taking appropriate unions of the regions thus created. The latter scheme is essentially the Madaline or Perceptron approach. An important series of questions concerns the comparison of the two schemes in terms of

1. Flexibility or universality—how many functions can be implemented?
2. Naturalness of the separating surfaces—will the generalizations be good?
3. Matching of machine capacity to the dimension of the problem—will a "full" machine tend to give a unique response to new patterns?

## REFERENCES


1. B. Widrow and M.E. Hoff, "Adaptive Switching Circuits," TR No. 1553-1, Contract Nonr 225(24), Stanford Electronics Laboratories, Stanford, Calif., Jun 1960.

2. F. Rosenblatt, "Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanism," Spartan Books, Washington, D.C., 1962.

3. R.L. Mattson, "The Analysis and Synthesis of Adaptive Systems Which Use Networks of Threshold Elements," Rept SEL-62-132, (TR No. 1553-1), Stanford Electronics Laboratories, Stanford, Calif., Dec 1962.

4. R. O. Winder, "Threshold Logic," Ph.D. Dissertation, Princeton University, Princeton, New Jersey, 1962.

5. H. Weyl, "The Elementary Theory of Convex Polyhedra," Contributions to the Theory of Games, Vol. I, Annals of Mathematics Studies, No. 24, Princeton, 1950, pp. 3-18.

6. A. J. Goldman and A. W. Tucker, "Polyhedral Convex Cones," Linear Inequalities and Related Systems, Annals of Mathematics Studies, No. 38, Princeton, 1956, pp. 19-40.

7. S. H. Cameron, "Proceedings of the Bionics Symposium," Wright Air Development Division, Tech. Report 60-600, 1960, pp. 197-212.

8. B. Widrow, "Generalization and Information Storage in Networks of Adaline 'Neurons'," Self Organizing Systems, Spartan Books, New York City, 1962, p. 442.

9. R. Brown, "Logical Properties of Adaptive Networks," Stanford Electronics Laboratories Quarterly Research Review No. 4, Rept SEL-63-043, 1963, pp. III-6 to III-9.

10. R. D. Joseph, "The Number of Orthants in n-Space Intersected by a s-Dimensional Subspace," Tech. Memorandum 8, Project PARA, Cornell Aeronautical Laboratory, Buffalo, New York, 1960.

11. E. A. Whitmore and D. G. Willis, "Division of Space by Concurrent Hyperplanes," Unpublished internal report, Lockheed Missiles and Space Division, Sunnyvale, 1960.

12. L. Schläfli, Gesammelte Mathematische Abhandlungen I, Verlag Birkhauser, Basel, Switzerland, 1950, pp. 209-212.

13. J. G. Wendel, "A Problem in Geometric Probability, "Mathematica Scandinavica, 11, 1962, pp. 109-111.

14. A. B. Bishop, "Adaptive Pattern Recognition," WESCON, Report of Session 1.5, 1963.

15. J. Koford, "Adaptive Network Organization," Stanford Electronics Laboratories Quarterly Research Review No. 3, SEL-63-009, 1962, p. III-6.

16. A. Novikoff, "On Convergence Proofs for Perceptrons," Symposium on Mathematical Theory of Automata, Polytechnic Institute of Brooklyn, April 1963, Proceedings in press.

17. R. Brown, "Logical Properties of Adaptive Networks," Stanford Electronics Laboratories Quarterly Research Review No. 1, SEL-62-109, 1962, pp. 87-88.

18. R. O. Winder, "Bounds on Threshold Gate Realizeability," IEEE Transactions on Electronic Computers, EC-12, Oct 1963, pp. 561-564.

19. B. Efron and T. Cover, "Linear Separability of Random Vectors," Unpublished Internal Report, Stanford Research Institute, Menlo Park, Calif., Mar 1963.

20. T. Cover, "Classification and Generalization Capabilities of Linear Threshold Units," Rome Air Development Center Tech., Documentary Report No. RADC-TDR-64-32, Feb 1964.

21. T. S. Motzkin and I. J. Schoenberg, "The Relaxation Method for Linear Inequalities," Canadian J. of Math., 6, 1954, pp. 393-404.

22. C. H. Mays, "Adaptive Threshold Logic," Rept SEL-63-027 (TR No. 1557-1), Stanford Electronics Laboratories, Stanford, Calif., Apr 1963.

23. J. Farkas, "Über Die Theorie Der Einfachen Ungleichungen," J. Reine Angew. Math., 124, 1902, pp. 1-24.

24. M. G. Kendall and P. A. P. Moran, Geometrical Probability, Griffin's Statistical Monographs and Courses, Ed. by M. G. Kendall, Hafner, N.Y., 1963.

25. R. Deltheil, Probabilités Géometriques. Traite du Calcul des Probabilités et de ses Applications, Gauthier-Villars, Paris, 1926.

SYSTEMS THEORY LABORATORY
DISTRIBUTION LIST
February 1964

GOVERNMENT

USAELRDL
Ft. Monmouth, N.J.
1 Attn: Dr. H. Jacobs, SIGRA/SL-PF

Procurement Data Division
USAS Equipment Support Agency
Ft. Monmouth, N.J.
1 Attn: Mr. M. Rosenfeld

Commanding General, USAELRDL
Ft. Monmouth, N.J.
5 SIGRA/SL-SC, Bldg. 42
1 TDC, Evans Signal Lab Area

Commanding Officer, ERDL
Ft. Belvoir, Va.
1 Attn: Tech. Doc. Ctr.

Commanding Officer
Frankford Arsenal
Bridge and Tacony St.
Philadelphia 37, Pa.
1 Attn: Library Br., 0270, Bldg. 40

Ballistics Research Lab
Aberdeen Proving Ground, Md.
2 Attn: V. W. Richard, BML

Chief of Naval Research
Navy Dept.
Washington 25, D.C.
2 Attn: Code 427
1 Code 420
1 Code 463

Commanding Officer, USAERU
P.O. Box 205
1 Mt. View, Calif.

Commanding Officer
ONR Branch Office
1000 Geary St.
1 San Francisco 9, Calif.

Chief Scientist
ONR Branch Office
1030 E. Green St.
1 Pasadena, Calif.

Office of Naval Research
Branch Office Chicago
230 N. Michigan Ave.
1 Chicago 1, Ill.

Commanding Officer
ONR Branch Office
207 W. 24th St.
New York 11, N.Y.
1 Attn: Dr. I. Rowe

U.S. Naval Applied Science Lab.
Tech. Library
Bldg. 291, Code 9832
Naval Base
1 Brooklyn, N.Y. 11251

Chief Bureau of Ships
Navy Dept.
Washington 25, D.C.
1 Attn: Code 691A1
1 Code 686
1 Code 607 NTDS
1 Code 687D
1 Code 732, A. E. Smith
1 Code 681A

Officer in Charge, ONR
Navy 100, Bx. 39, Fleet P.O.
16 New York, N.Y.

U.S. Naval Research Lab
Washington 25, D.C.
6 Attn: Code 2000
1 5240
1 5430
1 5200
1 5300
1 5400
1 5266, G. Abraham
1 2027
1 5260
1 6430

Chief, Bureau of Naval Weapons
Navy Dept.
Washington 25, D.C.
1 Attn: RAAV-6
1 RUDC-1
2 RREN-3
1 RAAV-44

Chief of Naval Operations
Navy Dept.
Washington 25, D.C.
1 Attn: Code Op 945Y

Director, Naval Electronics Lab
1 San Diego 52, Calif.

USN Post Graduate School
1 Monterey, Calif.
1 Attn: Tech. Reports Librarian
1 Prof. Gray, Electronics Dept.
1 Dr. H. Titus

Weapons Systems Test Div.
Naval Air Test Center
Patuxent River, Md.
1 Attn: Library

U.S. Naval Weapons Lab
Dahlgren, Va.
1 Attn: Tech. Library

Naval Ordnance Lab.
Corona, Calif.
1 Attn: Library
1 H. H. Wieder, 423

Commander, USN Air Dev. Ctr.
Johnsville, Pa.
1 Attn: NADC Library
1 AD-5

Commander
USN Missile Center
Pt. Mugu, Calif.
1 Attn: NO3022

Commanding Officer
U.S. Army Research Office
Box CM, Duke Station
Durham, N.C.
3 Attn: CRD-AA-IP

Commanding General
U.S. Army Materiel Command
Washington 25, D.C.
1 Attn: AMCRD-DE-E
1 AMCRD-RS-PE-E

Department of the Army
Office, Chief of Res. and Dev.
The Pentagon
Washington 25, D.C.
1 Attn: Research Support Div., Rm. 3D442

Office of the Chief of Engineers
Dept. of the Army
Washington 25, D.C.
1 Attn: Chief, Library Br.

Office of the Asst. Secy. of Defense
Washington 25, D.C.
1 (AE) Pentagon Bldg., Rm 3D984

Hqs., USAF (AFRDR-NU,3)
The Pentagon, Washington 25, D.C.
1 Attn: Mr. H. Mulkey, Rm 4D335

Chief of Staff, USAF
Washington 25, D.C.
2 Attn: AFDRT-ER

Hq., USAF
Dir. of Science and Technology
Electronics Div.
Washington 25, D.C.
1 Attn: AFRST-EL/CS, Maj. E.N. Myers

Aeronautical Systems Div.
Wright-Patterson AFB, Ohio
1 Attn: Lt. Col. L. M. Butsch, Jr. ASRNE-2
1 ASRNE-2, D. R. Moore
1 ASRNR-32
1 ASRNE-1, Electronic Res. Br. Elec. Tech. Lab
1 ASRNCF-2, Electromagnetic and Comm. Lab
3 ASNXR
1 ASNXRR (Library)
6 ASRNE-32

Commandant
AF Institute of Technology
Wright-Patterson AFB, Ohio
1 Attn: AFIT (Library)

Executive Director
AF Office of Scientific Res.
Washington 25, D.C.
1 Attn: SREE

AFWL (WLL)
2 Kirtland AFB, New Mexico

Director
Air University Library
Maxwell AFB, Ala.
1 Attn: CR-4582

Commander, AF Cambridge Res. Labs
ARDC, L. G. Hanscom Field
Bedford, Mass.
1 Attn: CRTOTT-2, Electronics

Hqs., AF Systems Command
Andrews AFB
Washington 25, D.C.
1 Attn: SCTAE

Asst. Secy. of Defense (R and D)
R and D Board, Dept. of Defense
Washington 25, D.C.
1 Attn: Tech. Library

Office of Director of Defense
Dept. of Defense
Washington 25, D. C.
1 Attn. Research and Engineering

Institute for Defense Analyses
1666 Connecticut Ave.
Washington 9, D.C.
1 Attn: W. E. Bradley

Defense Communications Agency
Dept. of Defense
Washington 25, D. C.
1 Attn: Code 121A, Tech. Library

Advisory Group on Electron Devices
346 Broadway, 8th Floor East
New York 13, N.Y.
2 Attn: H. Sullivan

Advisory Group on Reliability of
Electronic Equipment
Office Asst. Secy. of Defense
The Pentagon
1 Washington 25, D.C.

Commanding Officer
Diamond Ordnance Fuze Labs
Washington 25, D.C.
2 Attn: ORDTL 930, Dr. R.T. Young

Diamond Ordnance Fuze Lab.
U.S. Ordnance Corps
Washington 25, D.C.
1 Attn: ORDTL-450-638
Mr. R.H. Comyn

U.S. Dept. of Commerce
National Bureau of Standards
Boulder Labs
Central Radio Propagation Lab.
1 Boulder, Colorado
2 Attn: Miss J.V. Lincoln, Chief
RWSS

NSF, Engineering Section
1 Washington, D.C.

Information Retrieval Section
Federal Aviation Agency
Washington, D.C.
1 Attn: MS-112, Library Branch

DDC
Cameron Station
Alexandria 4, Va.
30 Attn: TISIA

U.S. Coast Guard
1300 E. Street, N.W.
Washington 25, D.C.
1 Attn: EEE Station 5-5

Office of Technical Services
Dept. of Commerce
1 Washington 25, D.C.

Director
National Security Agency
Fort George G. Meade, Md.
1 Attn: R42

NASA, Goddard Space Flight Center
Greenbelt, Md.
1 Attn: Code 611, Dr. G.H. Ludwig
1 Chief, Data Systems Divisions

NASA
Office of Adv. Res. & Tech.
Federal Office Bldg. 10-B
600 Independence Ave.
Washington, D.C.
1 Attn: Mr. Paul Johnson

Chief, U.S. Army Security Agency
Arlington Hall Station
2 Arlington 12, Virginia

SCHOOLS

*U of Aberdeen
Dept. of Natural Philosophy
Marischal College
Aberdeen, Scotland
1 Attn: Mr. R.V. Jones

U of Arizona
EE Dept.
Tucson, Ariz.
1 Attn: R.L. Walker
1 D.J. Hamilton

*U of British Columbia
Vancouver 8, Canada
1 Attn: Dr. A.C. Soudack

*No AF or Classified Reports.

California Institute of Technology
Pasadena, Calif.
1 Attn: Prof. R.W. Gould
1 Prof. L.M. Field, EE Dept.
1 D. Braverman, EE Dept.

California Institute of Technology
4800 Oak Grove Drive
Pasadena 3, Calif.
1 Attn: Library, Jet Propulsion Lab.

U. of California
Berkeley 4, Calif.
1 Attn: Prof. R.M. Saunders, EE Dept.
Dr. R.K. Wakerling,
Radiation Lab. Info. Div.,
Bldg. 30, Rm. 101

U of California
Los Angeles 24, Calif.
1 Attn: C.T. Leondes, Prof. of
Engineering, Engineering Dept.
1 R.S. Elliott, Electromagnetics
Div., College of Engineering

U of California, San Diego
School of Science and Engineering
La Jolla, Calif.
1 Attn: Physics Dept.

Carnegie Institute of Technology
Schenley Park
Pittsburg 13, Pa.
1 Attn: Dr. E.M. Williams, EE Dept.

Case Institute of Technology
Engineering Design Center
Cleveland 6, Ohio
1 Attn: Dr. J.B. Neswick, Director

Cornell U
Cognitive Systems Research Program
Ithaca, N.Y.
1 Attn: F. Rosenblatt, Hollister Hall

Thayer School of Engr.
Dartmouth College
Hanover, New Hampshire
1 Attn: John W. Strohbehn
Asst. Professor

Drexel Institute of Technology
Philadelphia 4, Pa.
1 Attn: F.B. Haynes, EE Dept.

U of Florida
Engineering Bldg., Rm. 336
Gainsville, Fla.
1 Attn: M.J. Wiggins, EE Dept.

Georgia Institute of Technology
Atlanta 13, Ga.
1 Attn: Mrs. J.H. Crosland, Librarian
1 F. Dixon, Engr.Experiment Station

Harvard U
Pierce Hall
Cambridge 38, Mass.
1 Attn: Dean H. Brooks, Div of Engr. and
Applied Physics, Rm. 217
2 E. Farkas, Librarian, Rm. 303A,
Tech. Reports Collection

U of Hawaii
Honolulu 14, Hawaii
1 Attn: Asst. Prof. K. Najita, EE Dept.

U of Illinois
Urbana, Ill.
1 Attn: P.D. Coleman, EE Res. Lab.
1 W. Perkins, EE Res. Lab.
1 A. Albert, Tech.Ed.,EE Res. Lab.
1 Library Serials Dept.
1 Prof.D.Alpert,Coordinated Sci.Lab.

*Instituto de Pesquisas da Marinha
Ministerio da Marinha
Rio de Janeiro
Estado da Guanabara, Brazil
1 Attn: Roberto B. da Costa

Johns Hopkins U
Charles and 34th St.
Baltimore 18, Md.
1 Attn: Librarian, Carlyle Barton Lab.

Johns Hopkins U
8621 Georgia Ave.
Silver Spring, Md.
1 Attn: N.H. Choksy
1 Mr. A.W. Nagy, Applied
Physics Lab.

Linfield Research Institute
McMinnville, Ore.
1 Attn: G.N. Hickok, Director

Marquette University
College of Engineering
1515 W. Wisconsin Ave.
Milwaukee 3, Wis.
1 Attn: A.C. Moeller, EE Dept.

M I T
Cambridge 39, Mass.
1 Attn: Res. Lab. of Elec., Doc.Rm.
1 Miss A. Sils, Libn.Rm 4-244,
LIR
1 Mr. J.E. Ward, Elec.Sys.Lab.

M I T
Lincoln Laboratory
P.O. Box 73
1 Attn: Lexington 73, Mass.
1 Navy Representative
1 Dr. W.I. Wells
1 Kenneth L. Jordon, Jr.

U of Michigan
Ann Arbor, Mich.
1 Attn: Dir., Cooley Elec. Labs.,
N. Campus
1 Dr. J.E. Rowe,Elec.Phys.Lab.
1 Comm. Sci.Lab.,180 Frieze Bldg.

U of Michigan
Institute of Science and Technology
P.O. Box 618
Ann Arbor, Mich.
1 Attn: Tech. Documents Service
1 W. Wolfe--IRIA--

U of Minnesota
Institute of Technology
Minneapolis 14, Minn.
1 Attn: Prof. A. Van der Ziel,
EE Dept.

U of Nevada
College of Engineering
Reno, Nev.
1 Attn: Dr. R.A. Manhart, EE Dept.

Northeastern U
The Dodge Library
Boston 15, Mass.
1 Attn: Joyce E. Lunde, Librarian

Northwestern U
2422 Oakton St.
Evanston, Ill.
1 Attn: W.S. Toth Aerial
Measurements Lab.

U of Notre Dame
South Bend, Ind.
1 Attn: E. Henry, EE Dept.

Ohio State U
2024 Niel Ave.
Columbus 10, Ohio
1 Attn: Prof. E.M. Boone, EE Dept.

Oregon State U
Corvallis, Ore.
1 Attn: H.J. Oorthuys, EE Dept.

Polytechnic Institute
333 Jay St.
Brooklyn, N.Y.
1 Attn: L. Shaw, EE Dept.

Polytechnic Institute of Brooklyn
Graduate Center, Route 110
Farmingdale, N.Y.
1 Attn: Librarian

Purdue U
Lafayette, Ind.
1 Attn: Library, EE Dept.

Rensselaer Polytechnic Institute
Troy, N.Y.
1 Attn: Library, Serials Dept.

*U of Saskatchewan
College of Engineering
Saskatoon, Canada
1 Attn: Prof. R.E. Ludwig

Syracuse U
Syracuse 10, N.Y.
1 Attn: EE Dept.

*Uppsala U
Institute of Physics
Uppsala, Sweden
1 Attn: Dr. P.A. Tove

U of Utah
Salt Lake City, Utah
1 Attn: R.W. Grow, EE Dept.

U of Virginia
Charlottesville, Va.
1 Attn: J.C. Wyllie, Alderman Library

U of Washington
Seattle 5, Wash.
1 Attn: A.E. Harrison, EE Dept.

Worchester Polytechnic Inst.
Worchester, Mass.
1 Attn: Dr. H.H. Newell

Yale U
New Haven, Conn.
1 Attn: Sloane Physics Lab.
1 EE Dept.
1 Dunham Lab., Engr. Library

INDUSTRIES

Avco Corp.
Res. Lab.
2385 Revere Beach Parkway
Everett 49, Mass.
1 Attn: Dr. Gurdon Abell

Argonne National Lab.
9700 South Cass
Argonne, Ill.
1 Attn: Dr. O.C. Simpson

Admiral Corp.
3800 Cortland St.
Chicago 47, Ill.
1 Attn: E.N. Roberston, Librarian

Airborne Instruments Lab.
Comac Road
Deer Park, Long Island, N.Y.
1 Attn: J. Dyer, Vice-Pres.&Tech.Dir.

Amperex Corp.
230 Duffy Ave.
Hicksville, Long Island, N.Y.
1 Attn: Proj.Engineer, S. Barbasso

Autonetics
Div. of North American Aviation, Inc.
9150 E. Imperial Highway
Downey, Calif.
1 Attn: Tech. Library 3040-3

Bell Telephone Labs.
Murray Hill Lab.
Murray Hill, N.J.
1 Attn: Dr. J.R. Pierce
1 Dr. S. Darlington
1 Mr. A.J. Grossman

Bell Telephone Labs., Inc.
Technical Information Library
Whippany, N.J.
1 Attn: Tech. Repts. Librn.,
Whippany Lab.

*Central Electronics Engineering
Research Institute
Pilani, Rajasthan, India
1 Attn: Om P. Gandhi - Via: ONR/London

Columbia Radiation Lab.
538 West 120th St.
1 New York, New York

Convair - San Diego
Div. of General Dynamics Corp.
San Diego 12, Calif.
1 Attn: Engineering Library

Cook Research Labs.
6401 W. Oakton St.
1 Attn: Morton Grove, Ill.

Cornell Aeronautical Labs., Inc.
4455 Genessee
Buffalo 21, N.Y.
1 Attn: Library

Eitel-McCullough, Inc.
301 Industrial Way
San Carlos, Calif.
1 Attn: Research Librarian

Ewan Knight Corp.
East Natick, Mass.
1 Attn: Library

Fairchild Semiconductor Corp.
4001 Junipero Serra Blvd.
Palo Alto, Calif.
1 Attn: Dr. V.H. Grinich

General Electric Co.
Defense Electronics Div., LMED
Cornell University, Ithaca, N.Y.
1 Attn: Library - Via: Commander,
ASD W-P AFB, Ohio, ASRNGW
D.E. Lewis

General Electric TWT Products Sec.
601 California Ave.
Palo Alto, Calif.
1 Attn: Tech. Library, C.G. Lob

General Electric Co. Res. Lab
P.O. Box 1088
Schnectady, N.Y.
1 Attn: Dr. P.M. Lewis
1 R.L. Shuey, Mgr. Info.
Studies Sec.

General Electric Co.
Electronics Park
Bldg. 3, Rm. 143-1
Syracuse, N.Y.
1 Attn: Doc. Library, Y. Burke

Gilfillan Brothers
1815 Venice Blvd.
Los Angeles, Calif.
1 Attn: Engr. Library

The Hallicrafters Co.
5th and Kostner Ave.
1 Attn: Chicago 24, Ill.

Hewlett-Packard Co.
1501 Page Mill Road
1 Attn: Palo Alto, Calif.

Hughes Aircraft
Malibu Beach, Calif.
1 Attn: Mr. Iams

Hughes Aircraft Ct.
Florence at Teale St.
Culver City, Calif.
1 Attn: Tech.Doc.Cen., Bldg 6,
Rm. C2048

Hughes Aircraft Co.
P.O. Box 278
Newport Beach, Calif.
1 Attn: Library, Semiconductor Div.

IBM, Box 390, Boardman Road
Poughkeepsie, N.Y.
1 Attn: J.C. Logue, Data Systems Div.

IBM, Poughkeepsie, N.Y.
1 Attn: Product Dev.Lab., E.M. Davis

IBM ASD and Research Library
Monterey and Cottle Roads
San Jose, Calif.
1 Attn: Miss M. Griffin, Bldg.025

ITT Federal Labs.
500 Washington Ave.
Nutley 10, N.J.
1 Attn: Mr. E. Mount, Librarian

Laboratory for Electronics, Inc.
1075 Commonwealth Ave.
Boston 15, Mass.
1 Attn: Library

LEL, Inc.
75 Akron St.
Copiague, Long Island, N.Y.
1 Attn: Mr. R.S. Mautner

Lenkurt Electric Co.
San Carlos, Calif.
1 Attn: M.L. Waller, Librarian

Librascope
Div. of General Precision, Inc.
808 Western Ave.
Glendale 1, Calif.
1 Attn: Engr. Library

Lockheed Missiles and Space Div.
P.O. Box 504, Bldg. 524
Sunnyvale, Calif.
1 Attn: Dr. W.M. Harris, Dept.67-30
G.W. Price, Dept. 67-33

Melpar, Inc.
3000 Arlington Blvd.
Falls Church, Va.
1 Attn: Librarian

Microwave Associates, Inc.
Northwest Industrial Park
Burlington, Mass.
1 Attn: K. Mortenson
1 Librarian

*No AF or Classified Reports.

Microwave Electronics Corp.
4061 Transport St.
Palo Alto, Calif.
1 Attn: S.F. Kaisel
M.C. Long

Minneapolis-Honeywell Regulator Co.
1177 Blue Heron Blvd.
Riviera Beach, Fla.
1 Attn: Semiconductor Products Library

Monsanto Research Corp.
Station B, Box 8
Dayton 7, Ohio
1 Attn: Mrs. D. Crabtree

Monsanto Chemical Co.
800 N. Linbergh Blvd.
St. Louis 66, Mo.
1 Attn: Mr. E. Orban, Mgr. Inorganic Dev.

*Dir., National Physical Lab.
Hilside Road
New Delhi 12, India
1 Attn: S.C. Sharma - Via: ONR/London

*Northern Electric Co., Lmtd.
Research and Development Labs.
P.O. Box 3511, Station "C"
Ottawa, Ontario, Canada
1 Attn: J.F. Tatlock
Via: ASD, Foreign Release Office
W-P AFB, Ohio
Mr. J. Troyal (ASYF)

Northronics
Palo Verdes Research Park
6101 Crest Road
Palos Verdes Estates, Calif.
1 Attn: Tech. Info. Center

Pacific Semiconductors, Inc.
14520 So. Aviation Blvd.
Lawndale, Calif.
1 Attn: H.Q. North

Philco Corp.
Tech. Rep. Division
P.O. Box 4730
Philadelphia 34, Pa.
1 Attn: F.R. Sherman, Mgr. Editor

Philco Corp.
Jolly and Union Meeting Roads
Blue Bell, Pa.
1 Attn: C.T. McCoy
1 Dr. J.R. Feldmeier

Polarad Electronics Corp.
43-20 Thirty-Fourth St.
Long Island City 1, N.Y.
1 Attn: A.H. Sonnenschein

Radio Corp. of America
RCA Labs., David Sarnoff Res. Cen.
Princeton, N.J.
2 Attn: Dr. J. Sklansky

RCA Labs., Princeton, N.J.
1 Attn: H. Johnson

RCA, Missile Elec. and Controls Dept.
Woburn, Mass.
1 Attn: Library

The Rand Corp.
1700 Main St.
Santa Monica, Calif.
1 Attn: Helen J. Waldron, Librarian

Raytheon Manufacturing Co.
Microwave and Power Tube Div.
Burlington, Mass.
1 Attn: Librarian, Spencer Lab.

Raytheon Manufacturing Co.
Res. Div., 28 Seyon St.
Waltham, Mass.
1 Attn: Dr. H. Statz
1 Mrs. M. Bennett, Librarian
1 Research Div. Library

Roger White Electron Devices, Inc.
Tall Oaks Road
1 Laurel Hedges, Stamford, Conn.

Sandia Corp.
Sandia Base, Albuquerque, N.M.
1 Attn: Mrs. B.R. Allen, Librarian

Sperry Rand Corp.
Sperry Electron Tube Div.
1 Gainesville, Fla.

Sperry Gyroscope Co.
Div. of Sperry Rand Corp.
Great Neck, N.Y.
1 Attn: L. Swern (MS3T105)

Sperry Gyroscope Co.
Engineering Library
Mail Station F-7
Great Neck, Long Island, N.Y.
1 Attn: K. Barney, Engr. Dept. Head

Sperry Microwave Electronics
Clearwater, Fla.
1 Attn: J.E. Pippin, Res. Sec. Head

Sylvania Electric Products, Inc.
500 Evelyn Ave.
1 Mt. View, Calif.

Sylvania Electronics Systems
100 First Ave.
Waltham 54, Mass.
1 Attn: Librarian, Waltham Labs.
1 Mr. E.E. Hollis

Technical Research Group
1 Syosett, L.I., N.Y.

Texas Instruments, Inc.
Semiconductor-Components Div.
P.O. Box 205
Dallas 22, Tex.
1 Attn: Library
2 Dr. W. Adcock

Texas Instruments, Inc.
P.O. Box 6015
Dallas 22, Tex.
1 Attn: M.E. Chun, Apparatus Div.

Texas Instruments
6017 E. Calle Tuberia
Phoenix, Arizona
1 Attn: R.L. Pritchard

Texas Instruments, Inc.
Corporate Research and Engineering
Technical Reports Service
P.O. Box 5474
1 Attn: Dallas 22, Tex

Tektronix, Inc.
P.O. Box 500
Beaverton, Ore.
4 Attn: Dr. J.F. DeLord, Dir. of Research

Varian Associates
611 Hansen Way
Palo Alto, Calif.
1 Attn: Tech. Library

Weitermann Electronics
4549 North 38th St.
1 Milwaukee 9, Wisconsin

Westinghouse Electric Corp.
Friendship International Airport
Box 746, Baltimore 3, Md.
1 Attn: G.R. Kilgore, Mgr. Appl. Res. Dept. Baltimore Lab.

Westinghouse Electric Corp.
3 Gateway Center
Pittsburgh 22, Pa.
1 Attn: Dr. G.C. Sziklai

Westinghouse Electric Corp.
P.O. Box 284
Elmira, N.Y.
1 Attn: S.S. King

Zenith Radio Corp.
6001 Dickens Ave.
Chicago 39, Ill.
1 Attn: J. Markin

*No AF or Classified Reports.