INFORMATION AND CONTROL 19, 41–65 (1971)

Analog Perceptrons: On Additive Representation of Functions

Y. UESAKA

NHK Broadcasting Science Research Laboratories, Tokyo, Japan

The theory of computational geometry in Perceptrons (Rosenblatt, 1962), developed by M. Minsky and S. Papert (1969), is extended to "Analog Perceptrons" with real-valued input and out-put. Mathematically, our problem is to determine the order of a function, i.e., the smallest number of variables necessary to make an additive representation of the function by employing partial functions of the smaller number of variables.

Mathematical tools, called the group-invariance theorem, the classification theorem and the collapsing theorem, are given which are useful for evaluating the order of analog Perceptrons. These are also applied for several analog Perceptrons.

I. INTRODUCTION

M. Minsky and S. Papert (1967, 1969) have developed a fruitful theory of computational geometry of Perceptrons (Rosenblatt, 1962). The central theme of their theory is the classification of certain geometrical properties according to the type of computation necessary to determine whether a given pattern has them. The computational geometry has been mainly motivated by the following considerations:

(a) In a problem of geometrical pattern recognition, to what extent can one use "local" properties—evidences obtained by looking at small portions of a pattern—as a basis for judgments about the "global" character of the pattern?

(b) What are the essential differences between "serial" and "parallel" computation? For example, to what degree can a computation be sped-up by doing several subcomputations at the same time?

Perceptrons in their theory may be said to be rather "digital" in the sense that their inputs and output take only two kinds of values, e.g., "0" and "1". In the present paper, the aim is to extend the theory of computational geometry to "ANALOG" Perceptrons, of which inputs and output may take arbitrary real values. The motivation of this extension stems from the following points of view:

(c) In the actual problem of geometrical pattern recognition, it is desired to deal with figures consisting not only of black picture elements but also gray ones.

(d) Recently, it has been recognized that the information in a nervous system is transmitted in a mode of the continuous type, such as the pule density, rather than of the discrete type. This standpoint stimulates investigation on the analog model of neuron (e.g., Fukushima, 1969). On the other hand, the Perceptron may be regarded as a simplified model of the neuron. Thus the question is: What kinds of differences do there exist between the neuron model of the "digital" type and that of the "analog" type?

As defined below, the output of analog Perceptron is determined by summing up the values of partial functions. There is no weighting coefficient in the summing process and no threshold element in the analog Perceptron. Thus, mathematically, our problem may be reduced to the additive representation of a function if we employ partial functions of the smaller number of variables. This may be regarded as a special case of the 13-th problem of Hilbert (1901).

In Section II, we shall formulate the analog Perceptron, and introduce the central concept of order by following Minsky and Papert. As examples, an elementary method for evaluating the order is demonstrated for simple analog Perceptrons. In Section III, the fundamental property, called the group-invariance theorem, which has an advantage for evaluating the order, is given and is also applied to some analog Perceptrons. In Section IV, a certain class of analog Perceptrons will be classified according to the order. This classification makes the determination of the order easy. In addition, the collapsing theorem (Minsky and Papert, 1969) is verified for the analog Perceptron.

II. Analog Perceptron

We shall conventionally use the following notations: Let E be the set of all real numbers. A finite set, denoted by R, of real-valued variables $x_1, \ldots, x_n$ is called a "retina". An element in E^n—i.e., the direct product of nE's—is interpreted as a geometrical pattern or a figure described on the retina, and variables in R may be regarded as picture-elements or visual cells in the retina.

Variables for patterns on the retina are usually denoted by letters $X, Y, \ldots$.

It seems natural to associate a mapping ψ from E^n to E with the property of a pattern $X = (x_1, \dots, x_n)$. Occasionally, it will be convenient to use the traditional representation of $\psi(X)$ as a function of n real-valued variables such as $\psi(x_1, \dots, x_n)$. In this paper, the word "function" is used to mean only a mapping from $E^{|R|}$ to E, where $|R|$ denotes the number of elements in R. In this context, an expression such as "a function ψ on R..." will be used for the sake of simplicity. In other cases, the word "mapping" is employed.

For a function ψ on the retina R, $\hat{\psi}(x)$ denotes a mapping from E to E which is induced from ψ by fixing the variables in $R - \{x\}$.

DEFINITION 1. A variable x of R is effective to ψ if and only if there is at least one nonconstant mapping $\hat{\psi}(x)$.

For example, x_1 is effective to $\psi(x_1, x_2) = x_1 x_2 / x_2$, but x_2 is not.

DEFINITION 2. A support of ψ, denoted by $s(\psi)$, is the set of all of effective variables to ψ.

The support of ψ means intuitively a set of variables all of which affect the value of ψ. For instance, $s(x_1 + x_2) = \{x_1, x_2\}$, while $s(x_1(x_1 + x_2) - x_1 x_2) = \{x_1\}$.

Subsets of the retina R are usually denoted by letters $A, B, \dots$. For A included in R, $\mathscr{F}(A)$ denotes a set of functions on R of which supports are included in A, i.e.,

$$\mathscr{F}(A) = \{\psi; \psi : E^{|R|} \to E \text{ and } s(\psi) \subset A\}. \tag{1}$$

Now we shall define the analog Perceptron.

DEFINITION 3. Let S be a family of subsets of the retina R. We say that ψ is an analog Perceptron on R with respect to S if for each member A of S there exists φ_A in $\mathscr{F}(A)$ such that

$$\psi(X) = \sum_{A \in S} \varphi_A(X). \tag{2}$$

This is called an additive representation for ψ, and often written more briefly as $\psi = \sum_{A \in S} \varphi_A$. We denote by $\mathscr{A}(S)$ the set of all of analog Perceptrons with respect to S.

Evidently, any function on R is an analog Perceptron with respect to 2^R, where 2^R is the family of all subsets of R. Thus, in spite of "analog Perceptron", the word "function" is occasionally employed for short.

The differences from the Perceptron of Minsky–Papert type are that

(i) Weighting coefficients are embedded into partial functions φ_A's, and that

(ii) A threshold element is removed.

The difference (i) is a natural consequence of the generalization such that inputs and output of the Perceptron can take any real value. In view of (ii) the number of steps in serial computation or number of layers in the Perceptron diminishes. Thus, the analog Perceptron may be suitable for considering the theory of parallel computation, because it becomes more elemental as a parallel computer.

Next, following Minsky and Papert, we shall introduce the central concept of order. For a linearly ordered set A, max A or min A denotes the maximum or the minimum element in A, respectively.

DEFINITION 4. The order of an analog Perceptron ψ, denoted by $o(\psi)$, is the smallest k for which there is a family S satisfying that ψ has an additive representation of $\psi = \sum_{A \in S} \varphi_A$ and for every A in S, $| s(\varphi_A)| \leqslant k$: that is,

$$o(\psi) = \min\{M(S); \psi \in \mathscr{A}(S)\}, \tag{3}$$

where

$$M(S) = \max\{| s(\varphi_A)|; A \in S\}. \tag{4}$$

For example, $x_1 + \cdots + x_n$ is of the order 1. Generally, if there exist $\varphi_1, \ldots, \varphi_n$ such that for $i = 1, \ldots, n$, φ_i is in $\mathscr{F}(\{x_i\})$ and $\psi(X) = \varphi_1(x_1) + \cdots + \varphi_n(x_n)$, then $o(\psi) \leqslant 1$.

In view of the definition, it is seen that an additive representation of ψ with a large order requires at least one partial function that can "look" a large portion of the retina. Thus, the property expressed by such ψ is said to be "global". Conversely, if the order of ψ is small, the property of ψ is said to be "local". This indicates that, in the case of pattern recognition, or processing by a parallel machine, the concept of the order plays an important role for considering the relation between the properties of the pattern and the structure of the machine.

The following form of the definition will be often convenient for evaluating the order.

THEOREM 1. *For a family S of subsets of R, $\| S \|$ denotes* $\max\{| A |; A \in S\}$. *Then*

$$o(\psi) = \min\{\| S \|; \psi \in \mathscr{A}(S)\}. \tag{5}$$

Proof. Suppose that ψ is in $\mathscr{A}(S)$. Then there is an additive representation for ψ such that $\psi = \sum_{A\in S} \varphi_A$ and for A of S, $s(\varphi_A) \subset A$. Hence,

$$| s(\varphi_A)| \leqslant | A | \leqslant \| S \|. \tag{6}$$

Taking the maximum of the left side of (6) with respect to A of S, we have that, by the definition of the order,

$$o(\psi) \leqslant \max\{| s(\varphi_A)|;\, A \in S\} \leqslant \| S \|. \tag{7}$$

Taking the minimum of the right side of (7) with respect to S under the constraint such that ψ is in $\mathscr{A}(S)$, we have that

$$o(\psi) \leqslant \min\{\| S \|;\, \psi \in \mathscr{A}(S)\}. \tag{8}$$

Next, suppose that ψ is in $\mathscr{A}(T)$. Then there is a representation such that $\psi = \sum_{A\in T} \varphi_A$. Let $T_0 = \{s(\varphi_A);\, A \in T\}$. We shall define new partial functions χ_B's such that, for each B in T_0 ,

$$\chi_B = \sum_{B=s(\varphi_A)} \varphi_A{}^*,$$

where

$$\varphi_A{}^* = \begin{cases} \varphi_A \text{ , if } A = s(\varphi_A); \\ \text{the partial function obtained from } \varphi_A \text{ by} \\ \text{inserting 0 into variables in } A - s(\varphi_A) \text{, if } A \subsetneq s(\varphi_A). \end{cases}$$

Note that χ_B is in $\mathscr{F}(B)$ and $B = s(\chi_B)$. Since in $A - s(\varphi_A)$ there is no effective variable to ψ, $\psi = \sum_{B\in T_0} \chi_B$, and

$$| B | = | s(\chi_B)| \leqslant o(\psi). \tag{9}$$

Taking the maximum of the left side of (9) with respect to B in T_0 , we have that $\| T_0 \| \leqslant o(\psi)$. Noting that $\| T_0 \| \geqslant \min\{\| S \|;\, \psi \in \mathscr{A}(S)\}$, we finally have

$$o(\psi) \geqslant \min\{\| S \|;\, \psi \in \mathscr{A}(S)\}. \tag{10}$$

Combining (8) with (10), we have (5). Q.E.D.

Let S and T be families of subsets of R. We define a relation such that $S < T$ if and only if for every A in S there is B in T such that $A \subset B$. In addition, we define as $S \sim T$ if and only if $S < T$ and $T < S$. For each S, we define

$$ml(S) = \{A;\, A \in S \text{ and } \forall B \in S,\, B \neq A \Rightarrow B \not\supset A\}. \tag{11}$$

For example, let $S = \{\{x_1, x_2, x_3\}, \{x_1, x_2\}, \{x_2, x_4\}\}$ and let $T = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2, x_4\}\}$. Then $T < S$, $ml(S) = \{\{x_1, x_2, x_3\}, \{x_2, x_4\}\} \sim S$ and $ml(T) = T$.

LEMMA 1. *Let S and T be families of subsets of R. Then,*

(i) *if* $S < T$, *then* $\mathscr{A}(S) \subset \mathscr{A}(T)$;
(ii) *if* $S \sim T$, *then* $\mathscr{A}(S) = \mathscr{A}(T)$; *and*
(iii) $\mathscr{A}(ml(S)) = \mathscr{A}(S)$.

Proof. (i) Suppose that ψ is in $\mathscr{A}(S)$. Then there is a representation for ψ such that $\psi = \sum_{A \in S} \varphi_A$ and for every A in S, φ_A is in $\mathscr{F}(A)$. Since, by assumption, for every A in S, there is B in T such that $B \supset A$, so for every A in S there is B in T such that at least one of the members in $\mathscr{F}(B)$ is identical to φ_A. Let $T_B = \{A; A \subset B \text{ and } A \in S\}$. Thus, defining each B in T as

$$\chi_B = \begin{cases} \sum\limits_{A \in T_B} \varphi_A, & \text{if } B \in ml(T) \text{ and } T_B = \varnothing; \\ 0, & \text{otherwise,} \end{cases}$$

we can write ψ as $\psi = \sum_{B \in T} \chi_B$. Hence ψ is in $\mathscr{A}(T)$.

(ii) Since by assumption $S < T$, $\mathscr{A}(S) \subset \mathscr{A}(T)$ in view of (i). Similarly, $\mathscr{A}(S) \supset \mathscr{A}(T)$. Hence, $\mathscr{A}(S) = \mathscr{A}(T)$.

(iii) Obvioulsy, $ml(S) \sim S$. Thus, by (ii) we have the assertion. Q.E.D.

THEOREM 2. *In the definition of the order, we may restrict the range of S's to the family of ml(S)'s. Namely, let* $M = \{ml(S); S \subset 2^R\}$. *Then, for every function* ψ *on R,*

$$o(\psi) = \min\{\| T \|; T \in M \text{ and } \psi \in \mathscr{A}(T)\}. \tag{12}$$

Proof. It is easily seen that for every S in 2^R, $\| S \| = \| ml(S) \|$. Thus, the theorem follows (iii) of the Lemma 1. Q.E.D.

When we want to estimate the order of an analog Perceptron, this theorem becomes useful for simplifying the job of evaluation. For example, consider a problem to show that $o(\psi) = | R |$. It is enough, for this purpose, to show that there is no set of partial functions φ_1 in $\mathscr{F}(A_1), \ldots, \varphi_n$ in $\mathscr{F}(A_n)$ such that $\psi = \varphi_1 + \cdots + \varphi_n$, where $A_1 = R - \{x_1\}, \ldots, A_n = R - \{x_n\}$. In fact, if the order of ψ were smaller than $| R |$, then there would be an additive representation for ψ such that $\psi = \sum_{A \in S} \varphi_A$ and $\| S \| < | R |$. In view of Theorem 2, we may choose $\{A_1, \ldots, A_n\}$ as S satisfying that $\| S \| < | R |$. Thus, there exists a set of φ_1 in $\mathscr{F}(A_1), \ldots, \varphi_n$ in $\mathscr{F}(A_n)$ such that $\psi = \sum_i \varphi_i$, and this is a contradiction.

The following examples demonstrate an elementary method for evaluating the order of simple analog Perceptrons.

PROPOSITION 1. *An analog Perceptron defined as*

$$\text{mult}(X) = x_1 \cdots x_n \tag{13}$$

is of order n.

Proof. By induction on n. Initial step is obvious. Inductive step: Suppose that the order of mult on $R = \{x_1, \ldots, x_n\}$ is smaller than n. Then, in view of Theorem 2, mult must be in $\mathscr{A}(S)$, where $S = \{R - \{x_1\}, \ldots, R - \{x_n\}\}$. In other words, there exist φ_1 in $\mathscr{F}(R - \{x_1\}), \ldots, \varphi_n$ in $\mathscr{F}(R - \{x_n\})$ such that

$$\begin{aligned} \text{mult}(X) &= \varphi_1(x_2, \ldots, x_n) + \varphi_2(x_1, x_3, \ldots, x_n) \\ &\quad + \cdots + \varphi_n(x_1, \ldots, x_{n-1}). \end{aligned} \tag{14}$$

Inserting 1 or 0 into x_n of (14), we have, respectively, that

$$\begin{aligned} \text{mult}(x_1, \ldots, x_{n-1}) &= \varphi_1(x_2, \ldots, x_{n-1}, 1) \\ &\quad + \cdots + \varphi_{n-1}(x_1, \ldots, x_{n-2}, 1) + \varphi_n(x_1, \ldots, x_{n-1}), \end{aligned} \tag{15}$$

or

$$0 = \varphi_1(x_2, \ldots, x_{n-1}, 0) + \cdots + \varphi_{n-1}(x_1, \ldots, x_{n-2}, 0) + \varphi_n(x_1, \ldots, x_{n-1}). \tag{16}$$

Subtracting sidewise (16) from (15), we have a representation for mult on $\{x_1, \ldots, x_{n-1}\}$:

$$\text{mult}(x_1, \ldots, x_{n-1}) = \chi_1(x_2, \ldots, x_{n-1}) + \cdots + \chi_{n-1}(x_1, \ldots, x_{n-2}), \tag{17}$$

where for $i = 1, \ldots, n - 1$

$$\chi_i(y_1, \ldots, y_{n-2}) = \varphi_i(y_1, \ldots, y_{n-2}, 1) - \varphi_i(y_1, \ldots, y_{n-2}, 0).$$

Equation (17) shows that the order of mult on $\{x_1, \ldots, x_{n-1}\}$ is smaller than $n - 1$, and this contradicts to the inductive hypothesis. Thus, the proof completes by induction. Q.E.D.

PROPOSITION 2. *Analog Perceptrons defined as*

$$\begin{aligned}\max(X) &= \max\{x_1, \ldots, x_n\},\\ \min(X) &= \min\{x_1, \ldots, x_n\}\end{aligned} \tag{18}$$

are both of order n.

Proof. First we shall show the above in the case that the domain of max is restricted to $[0, 1]^n$. In induction on n, the initial step is obvious. Inductive step: Suppose that the order of the restricted max on $R = \{x_1, \ldots, x_n\}$ is smaller than n. Then, similarly as in the proof of the Proposition 1, there is an additive representation for max such that

$$\max(X) = \varphi_1(x_2, \ldots, x_n) + \varphi_2(x_1, x_3, \ldots, x_n) + \cdots + \varphi_n(x_1, \ldots, x_{n-1}). \tag{19}$$

Inserting 0 or 1 into x_n of (19), we have, respectively, that

$$\begin{aligned}\max(x_1, \ldots, x_{n-1}) = {} & \varphi_1(x_2, \ldots, x_{n-1}, 0)\\ & + \cdots + \varphi_{n-1}(x_1, \ldots, x_{n-2}, 0) + \varphi_n(x_1, \ldots, x_{n-1}),\end{aligned} \tag{20}$$

or

$$1 = \varphi_1(x_2, \ldots, x_{n-1}, 1) + \cdots + \varphi_{n-1}(x_1, \ldots, x_{n-2}, 1) + \varphi_n(x_1, \ldots, x_{n-1}). \tag{21}$$

Subtracting sidewise (21) from (20), we have a representation for max on $\{x_1, \ldots, x_{n-1}\}$:

$$\max(x_1, \ldots, x_{n-1}) = \chi_1(x_2, \ldots, x_{n-1}) + \cdots + \chi_{n-1}(x_1, \ldots, x_{n-2}) + 1, \tag{22}$$

where for $i = 1, \ldots, n-1$

$$\chi_i(y_1, \ldots, y_{n-2}) = \varphi_i(y_1, \ldots, y_{n-2}, 0) - \varphi_i(y_1, \ldots, y_{n-2}, 1).$$

Equation (22) shows that the order of the restricted max on $\{x_1, \ldots, x_{n-1}\}$ is smaller than $n - 1$, and this contradicts to the inductive hypothesis. Hence the restricted max is of the order $|R|$.

Now we shall remove the restriction on the domain of max. Suppose that the order of max on $R = \{x_1, \ldots, x_n\}$ is smaller than n. Then, there is a representation such as (19); and, furthermore, for $0 \leqslant x_1, \ldots, x_{n-1} \leqslant 1$, (20) and (21) should hold. Thus, in view of (22), the order of the restricted max should be smaller than $|R|$. This contradicts the result above. Thus the proof is complete, similarly as in case of the function min. Q.E.D.

III. Group-Invariance Theorem

In this section, the group-invariance theorem corresponding to the Theorem 2.3 in Minsky–Papert (1969), which is powerful for estimating the order of several analog Perceptrons, is given.

For a subset A of the retina R, we denote by $\pi(A)$ the set of all permutations formed up from variables in A. For example, when $A = \{x_1, x_2, x_3\}$,

$$\pi(A) = \{(x_1, x_2, x_3), (x_1, x_3, x_2), (x_2, x_1, x_3), (x_2, x_3, x_1), (x_3, x_1, x_2), (x_3, x_2, x_1)\}.$$

Members of $\pi(A)$ are usually denoted by letters $\sigma, \tau, \ldots$. In addition, if ψ is in $\mathscr{F}(A)$ and for σ in $\pi(A)$, $\sigma = (x_{i_1}, \ldots, x_{i_m})$, then $\psi(\sigma)$ means $\psi(x_{i_1}, \ldots, x_{i_m})$.

Consider a group G of permutations on the retina $R = \{x_1, \ldots, x_n\}$. When g in G is

$$g = \begin{pmatrix} x_1 & \cdots & x_n \\ x_{i_1} & \cdots & x_{i_n} \end{pmatrix},$$

then we write for $k = 1, \ldots, n$ as $gx_k = x_{i_k}$. For a subset A of R and g in G, the set gA is defined as

$$gA = \{gx;\, x \in A\}. \tag{23}$$

For subsets A, B of R, we introduce a relation $\underset{G}{\sim}$ such that $A \underset{G}{\sim} B$ if and only if there is a member g in G for which $gA = B$. It is easily seen that $\underset{G}{\sim}$ is an equivalent relation. When σ in $\pi(A)$ is $(x_{i_1}, \ldots, x_{i_m})$, we denote $(gx_{i_1}, \ldots, gx_{i_m})$ by $g\sigma$.

Let A and B be subsets of R. Let σ be in $\pi(A)$. Then, by $\pi(A \to B, \sigma)$ we mean a set of $g\sigma$'s such that $gA = B$ and g is in G:

$$\pi(A \to B, \sigma) = \{g\sigma;\, gA = B \text{ and } g \in G\}. \tag{24}$$

For g in G, we define as

$$g\pi(A \to B, \sigma) = \{g\tau;\, \tau \in \pi(A \to B, \sigma)\}. \tag{25}$$

We introduce a relation $\underset{G}{\sim}$ as follows: $\pi(A \to B, \sigma) \underset{G}{\sim} \pi(A' \to B', \sigma')$ if and only if there is a member g in G for which $g\pi(A \to B, \sigma) = \pi(A' \to B', \sigma')$.

Lemma 2. *Let A and B be subsets of R. If $A \underset{G}{\sim} B$, then for every subset C of R and σ in $\pi(C)$, $\pi(C \to A, \sigma) \underset{G}{\sim} \pi(C \to B, \sigma)$.*

Proof. If either $\pi(C \to A, \sigma)$ or $\pi(C \to B, \sigma)$ is empty, then so is the other. Thus, in this case, the lemma trivially holds.

Let τ be in $\pi(C \to A, \sigma)$. Then, by notation, there is h in G such that $hC = A$ and $\tau = h\sigma$. Since by assumption there is g in G such that $gA = B$, $g\tau = (gh)\,\sigma$ is in $\pi(C \to B, \sigma)$. In fact, $(gh)\,C = gA = B$. Hence $g\pi(C \to A, \sigma) \subset \pi(C \to B, \sigma)$.

Let τ be in $\pi(C \to B, \sigma)$. Then by notation there is h in G such that $hC = B$ and $\tau = h\sigma$. Noting that G is a group, $\tau = g(g^{-1}h)\sigma$ is in $g\pi(C \to A, \sigma)$ for the above g. In fact, $(g^{-1}h)C = g^{-1}B = A$. Hence $g\pi(C \to A, \sigma) \supset \pi(C \to B, \sigma)$. Q.E.D.

As an immediate consequence from the above, we have the following.

LEMMA 3. *Let A and B be subsets of R. Let ψ be a mapping from $E^{|A|}$ to E. If $A \underset{G}{\sim} B$, then for every subset C of R and σ in $\pi(C)$, there is g in G such that*

$$\sum_{\tau \in \pi(C \to A, \sigma)} \psi(g\tau) = \sum_{\tau \in \pi(C \to B, \sigma)} \psi(\tau). \tag{26}$$

Using this lemma, we shall show the group-invariance theorem for the analog Perceptron.

DEFINITION 5. Let S be a family of subsets of the retina R. Let G be a group of permutations on R. We say that S is closed under G if for every A in S and g in G, the set gA is also in S.

DEFINITION 6. Let ψ be a function on R. Let G be a group of permutations on R. We say that ψ is invariant under G if for every g in G and X on R, $\psi(gX) = \psi(X)$, where $gX = (gx_1, \ldots, gx_n)$.

THEOREM 3. *Let* (i) *G be a group of permutations on R,*

(ii) *S be a family of subsets of R and closed under G, and*

(iii) *ψ be in $\mathscr{A}(S)$ and invariant under G.*

Then, there exists an additive representation for ψ such that $\psi = \sum_{A \in S} \varphi_A$ and the partial function φ_A depends only on the G-equivalence classes of S. Namely, we can choose the partial functions such that, for φ_A and φ_B, if $A \underset{G}{\sim} B$, then there is g in G for which $\varphi_A(gX) = \varphi_B(X)$.

Proof. Since ψ is in $\mathscr{A}(S)$, there is an additive representation for ψ such that

$$\psi(X) = \sum_{C \in S} \chi_C(X) = \sum_{C \in S} \chi_C(\sigma_C),$$

where σ_C is in $\pi(C)$. Furthermore, since ψ is invariant under G, for every g in G,

$$\psi(X) = \psi(gX) = \sum_{C \in S} \chi_C(g\sigma_C).$$

Summing sidewise up this equation for all g in G, we have that

$$|G|\,\psi(X) = \sum_{g \in G} \left\{ \sum_{C \in S} \chi_C(g\sigma_C) \right\} = \sum_{C \in S} \left\{ \sum_{g \in G} \chi_C(g\sigma_C) \right\}. \tag{27}$$

Noting that for every A and C in S, and for every σ_C in $\pi(C)$

$$\pi(C \to A, \sigma_C) = \{g\sigma_C\, ;\, g \in G\} \cap \pi(A)$$

and

$$\bigcup_{A \in S} \pi(C \to A, \sigma_C) = \{g\sigma_C\, ;\, g \in G\},$$

we see that

$$\sum_{g \in G} \chi_C(g\sigma_C) = \sum_{A \in S} \left\{ \sum_{\tau \in \pi(C \to A, \sigma_C)} \chi_C(\tau) \right\}.$$

Applying this to (27), we have a new representation for ψ such that

$$\psi = \sum_{A \in S} \varphi_A\,,$$

where

$$\varphi_A(X) = \frac{1}{|G|} \sum_{C \in S} \left\{ \sum_{\tau \in \pi(C \to A, \sigma_C)} \chi_C(\tau) \right\}. \tag{28}$$

It remains only to show that for every A and B in S if $A \underset{G}{\sim} B$, then there is g in G such that $\varphi_A(gX) = \varphi_B(X)$. In view of Lemma 3, there is g in G such that, for every C in S and ρ in $\pi(C)$,

$$\sum_{\tau \in \pi(C \to A, \rho)} \chi_C(g\tau) = \sum_{\tau \in \pi(C \to B, \rho)} \chi_C(\tau).$$

Hence, from (28), we have

$$\begin{aligned} \varphi_A(gX) &= \frac{1}{|G|} \sum_{C \in S} \left\{ \sum_{\tau \in \pi(C \to A, \sigma_C)} \chi_C(g\tau) \right\} \\ &= \frac{1}{|G|} \sum_{C \in S} \left\{ \sum_{\tau \in \pi(C \to B, \sigma_C)} \chi_C(\tau) \right\} = \varphi_B(X). \end{aligned}$$

Q.E.D.

Now we shall demonstrate with examples how to apply the group-invariance theorem to the evaluation of the order.

THEOREM 4. *Let G be a transitive group—i.e., for every pair of variables x, y in R there is at least one g in G such that $gx = y$—of permutations on R. Let ψ be a function on R and invariant under G. Then, $o(\psi) \leqslant 1$ if and only if, for every $x_1, \ldots, x_n$,*

$$n\psi(x_1, \ldots, x_n) = \psi(x_1, \ldots, x_1) + \cdots + \psi(x_n, \ldots, x_n). \tag{29}$$

Proof. In view of the definition of the order, the "if" part is obvious. Suppose that $o(\psi) \leqslant 1$. Then there is a set of φ_i's such that $s(\varphi_i) \subset \{x_i\}$ for $i = 1, \ldots, n$, and ψ can be written as

$$\psi(X) = \varphi_1(x_1) + \cdots + \varphi_n(x_n).$$

Let $S = \{\{x_1\}, \ldots, \{x_n\}\}$. Since G is transitive, S is closed under G and $S/\underset{G}{\sim} \simeq \{S\}$. Hence, by the group-invariance theorem, we can choose the identical function φ as $\varphi_1, \ldots, \varphi_n$:

$$\psi(x_1, \ldots, x_n) = \varphi(x_1) + \cdots + \varphi(x_n). \tag{30}$$

Inserting x into x_i of (30), we have that $\varphi(x) = \psi(x, \ldots, x)/n$. Thus, using (30) again, we obtain (29). Q.E.D.

By this result it is easily seen that the order of functions $(x_1 + \cdots + x_n)^m$ $(m \geqslant 2)$, $|x_1 + \cdots + x_n|$, $\operatorname{sign}(x_1 + \cdots + x_n)$,... etc. are larger than 1.

Next we discuss a Perceptron that determines the uniformness of a pattern: Let "uniform" be a function on R defined as

$$\operatorname{uniform}(X) = \begin{cases} 1, & \text{if } x_1 = \cdots = x_n; \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

LEMMA 4. *Let R_1 be a set of variables $y_1, \ldots, y_{n-1}$. By* prop(k) *we mean a proposition such that some k variables in R_1 take the same value a, and the values of the residual variables and a are mutually distinct. If there is a mapping φ from E^{n-1} to E such that*

$$\operatorname{uniform}(X) = \varphi(x_2, \ldots, x_n) + \varphi(x_1, x_3, \ldots, x_n) + \cdots + \varphi(x_1, \ldots, x_{n-1}), \tag{32}$$

then, for $k = 1, \ldots, n-1$, prop(k) *implies*

$$\varphi(y_1, \ldots, y_{n-1}) = (-1)^{n-k-1} \frac{(n-k-1)!\, k!}{n!}. \tag{33}$$

Proof. By induction on $k = n - 1, n - 2, \ldots, 1$. Initial step: Let $x_1 = \cdots = x_n = x$ in (32). Since uniform$(X) = 1$, we see that $\varphi(x, \ldots, x) = 1/n$. Hence (33) holds for $k = n - 1$. Inductive step: Let $x_1 = \cdots = x_k = x$ in (32) for some k. Then (32) becomes

$$\begin{aligned}
&\text{uniform}(\underbrace{x, \ldots, x}_{k}, x_{k+1}, \ldots, x_n) \\
&= k\varphi(\underbrace{x, \ldots, x}_{k-1}, x_{k+1}, \ldots, x_n) + \varphi(\underbrace{x, \ldots, x}_{k}, x_{k+2}, \ldots, x_n) \\
&\quad + \varphi(\underbrace{x, \ldots, x}_{k}, x_{k+1}, x_{k+3}, \ldots, x_n) + \cdots \cdots \cdots \cdots \cdots \\
&\quad + \varphi(\underbrace{x, \ldots, x}_{k}, x_{k+1}, \ldots, x_{n-2}, x_n) + \varphi(\underbrace{x, \ldots, x}_{k}, x_{k+1}, \ldots, x_{n-1}).
\end{aligned}$$

Suppose that prop(k) holds. Then, by the inductive hypothesis,

$$\begin{aligned}
\varphi(\underbrace{x, \ldots, x}_{k}, x_{k+2}, \ldots, x_n) = \cdots &= \varphi(\underbrace{x, \ldots, x}_{k}, x_{k+1}, \ldots, x_{n-1}) \\
&= (-1)^{n-k-1} \frac{(n-k-1)!\, k!}{n!},
\end{aligned}$$

and, by definition, uniform$(\underbrace{x, \ldots, x}_{k}, x_{k+1}, \ldots, x_n) = 0$. Hence we have that

$$0 = k\varphi(\underbrace{x, \ldots, x}_{k-1}, x_{k+1}, \ldots, x_n) + (-1)^{n-k-1} \frac{(n-k-1)!\, k!}{n!},$$

or equivalently

$$\varphi(\underbrace{x, \ldots, x}_{k-1}, x_{k+1}, \ldots, x_n) = (-1)^{n-(k-1)-1} \frac{(n-(k-1)-1)!\,(k-1)!}{n!}. \tag{34}$$

Noting that φ is invariant under the permutation group on R_1, in view of Theorem 5 which will be shown later, (34) confirms the validity of (33) for $k - 1$. This completes the inductive proof. Q.E.D.

PROPOSITION 3. $o(\text{uniform}) = |R|$.

Proof. Suppose that $o(\text{uniform}) < |R|$. Then, by the group invariance theorem, there is an additive representation for uniform such as (32). Thus, letting $x_1, \ldots, x_n$ be mutually distinct, in view of Lemma 4,

$$\varphi(x_{i_1}, \ldots, x_{i_{n-1}}) = (-1)^{n-2}(n-2)!/n!,$$

where $1 \leqslant i_1 < \cdots < i_{n-1} \leqslant n$. But, in this case, obviously uniform$(X) = 0$.

Hence

[the left side of (32)] $= 0 \neq n \times (-1)^{n-2}(n-2)!/n! =$ [the right side of (32)],

and this is a contradiction! Q.E.D.

IV. Classification Theorem

Through in this section, by G we mean the group of all permutations on the retina R. A problem of classifying Perceptrons which are invariant under G will be discussed from the order point of view. The classification theorem follows several lemmata which gives the necessary and sufficient condition that the order of G-invariant Perceptron is smaller than or equal to a given number. An application of this theorem will demonstrate that the orders of the Perceptrons discussed already are more easily determined. In addition, after giving the collapsing theorem, we shall discuss a few analog Perceptrons.

Let φ be a mapping from E^m to E, where E is the set of all of real numbers. For the sake of simplicity, we use the following notation: For integers r, s such that $r = 1,\ldots, m$ and $r \leqslant s$,

$$\sum^{s} \varphi_r[a] = \sum_{1 \leqslant i_1 < \cdots < i_r \leqslant s} \varphi(x_{i_1}, \ldots, x_{i_r}, \underbrace{a, \ldots, a}_{m-r}), \tag{35}$$

where a is a constant number. When $r = 0$, for $s = 1, 2, \ldots$,

$$\sum^{s} \varphi_0[a] = \varphi(a, \ldots, a).$$

For example, in case of $m = 3$,

$$\sum^{2} \varphi_1[a] = \varphi(x_1, a, a) + \varphi(x_2, a, a),$$

$$\sum^{2} \varphi_2[a] = \varphi(x_1, x_2, a),$$

$$\sum^{3} \varphi_2[5] = \varphi(x_1, x_2, 5) + \varphi(x_1, x_3, 5) + \varphi(x_2, x_3, 5),$$

$$\sum^{4} \varphi_3[a] = \varphi(x_1, x_2, x_3) + \varphi(x_1, x_2, x_4) + \varphi(x_1, x_3, x_4) + \varphi(x_2, x_3, x_4),$$

... ..., etc.

LEMMA 5. *Let ψ be a function on R. If there is an additive representation for ψ such that $\psi = \sum^n \varphi_m[a]$, then for $k = 0, 1,\ldots, m$*

$$\psi(x_1, \ldots, x_k, \underbrace{a, \ldots, a}_{n-k}) = \sum_{j=0}^{k} \left\{ \begin{bmatrix} n-k \\ m-j \end{bmatrix} \sum^{k} \varphi_j[a] \right\}, \tag{36}$$

where

$$\begin{bmatrix} n \\ m \end{bmatrix} = \begin{cases} \binom{n}{m}, & \text{if } n \geqslant m; \\ 0, & \text{if } n < m. \end{cases} \tag{37}$$

Proof. By notation

$$\psi(x_1, \ldots, x_n) = \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n} \varphi(x_{i_1}, \ldots, x_{i_m}). \tag{38}$$

Insert a into $x_{k+1}, x_{k+2}, \ldots, x_n$ of (38). Then, we can see that in the right side of (38), there exist $\left[\begin{smallmatrix} n-k \\ m-k \end{smallmatrix}\right] \varphi(x_1, \ldots, x_k, \underbrace{a, \ldots, a}_{m-k})$'s; that for $1 \leqslant i_1 < \cdots < i_{k-1} \leqslant k$, there exist $\left[\begin{smallmatrix} n-k \\ m-k+1 \end{smallmatrix}\right] \varphi(x_{i_1}, \ldots, x_{i_{k-1}}, \underbrace{a, \ldots, a}_{m-k+1})$'s, ... ; and that there exist $\left[\begin{smallmatrix} n-k \\ m \end{smallmatrix}\right] \varphi(a, \ldots, a)$'s. Thus, by notation, we have the representation (36). Q.E.D.

LEMMA 6. *Let φ be a mapping from E^m to E. Let $\chi(x_1, \ldots, x_s) = \sum^s \varphi_r[a]$. Then, for every integers s, t and r such that $0 \leqslant r \leqslant s \leqslant t$ and $r \leqslant m$,*

$$\sum^{t} \chi_s[a] = \binom{t-r}{s-r} \sum^{t} \varphi_r[a]. \tag{39}$$

Proof. For the sake of simplicity, we write $\varphi(x_1, \ldots, x_n, \underbrace{a, \ldots, a}_{m-n})$ as $\varphi(1, \ldots, n)$. By $[n]$ we mean a set of integers $1, 2, \ldots, n$. By I we mean a set of mappings from $[r]$ to $[t]$ such that for i in I, $i(1) < \cdots < i(r)$, i.e.,

$$I = \{i; i : [r] \to [t] \text{ and } i(1) < \cdots < i(r)\}.$$

Similarly, we define as follows:

$$J = \{j; j : [r] \to [s] \text{ and } j(1) < \cdots < j(r)\},$$
$$K = \{k; k : [s] \to [t] \text{ and } k(1) < \cdots < k(s)\}.$$

$$\begin{tikzcd} {[r]} \arrow[r, "j"] \arrow[rr, bend right] & {[s]} \arrow[r, "k"] & {[t]} \end{tikzcd}$$

Note that $|I| = \binom{t}{r}$, $|J| = \binom{s}{r}$ and $|K| = \binom{t}{s}$. We denote by k^* a mapping induced from k in K by restricting the domain of k to the image of j, denoted by $\mathrm{Im}(j)$. Then, for every j in J and k in K, k^*j is a mapping from $[r]$ to $[t]$. Using these notations, we may write (39) as follows:

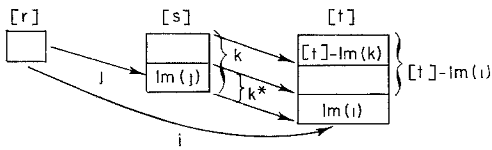
$$\sum_{k\in K}\sum_{j\in J}\varphi(k^*j(1),\dots,k^*j(r)) = \binom{t-r}{s-r}\sum_{i\in I}\varphi(i(1),\dots,i(r)). \tag{40}$$

Thus we shall show this equation.

For a given i in I we estimate the number of k^*j's, denoted by $N(i)$, such that $k^*j = i$. Since every i, j and k are one-to-one, there exists j such that $k^*j = i$ if and only if $\mathrm{Im}(k)$ includes $\mathrm{Im}(i)$, or, equivalently, $[t] - \mathrm{Im}(i)$ includes $[t] - \mathrm{Im}(k)$. Hence, $N(i)$, i.e., the number of k's satisfying this, is obtained by a simple combinatorial calculation as

$$N(i) = \binom{|[t]-\mathrm{Im}(i)|}{|[t]-\mathrm{Im}(k)|} = \binom{t-r}{t-s} = \binom{t-r}{s-r}$$

for every i in I.



This means that the left side of (40) has $\binom{t-r}{s-r}$ $\varphi(i(1),\dots,i(r))$'s for every i in I, of which the total number, i.e.,

$$\sum_{i\in I} N(i) = \binom{t-r}{s-r}\binom{t}{r}$$

is equal to the total number of φ's in the left side of (40), because

$$\binom{t-r}{s-r}\binom{t}{r} = \binom{s}{r}\binom{t}{s} = |J|\cdot|K|.$$

Hence the left side of (40) is exactly equal to the summation of $\binom{t-r}{s-r}$ $\varphi(i(1),\dots,i(r))$'s. This completes the proof. Q.E.D.

LEMMA 7. *For $n = 1, 2, 3,\dots$ and $m = 1,\dots, n-1$,*

$$N(n, m) = \sum_{i=0}^{m}(-1)^i\binom{m}{i}\begin{bmatrix} n-1-i \\ m-1 \end{bmatrix} = 0. \tag{41}$$

Proof. *Case* (i): $n/2 \geqslant m \geqslant 1$. First, for $m = 1$,

$$N(n, 1) = \binom{1}{0}\binom{n-1}{0} - \binom{1}{1}\binom{n-2}{0} = 0.$$

Next, using the well-known relation

$$\binom{m}{i} = \binom{m-1}{i} + \binom{m-1}{i-1}, \tag{42}$$

$N(n, m)$ is rewritten as

$$N(n, m) = \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i}\binom{n-1-i}{m-1} + \sum_{i=1}^{m} (-1)^i \binom{m-1}{i-1}\binom{n-1-i}{m-1}.$$

Furthermore, applying the relation $\binom{n-1-i}{m-1} = \binom{n-2-i}{m-1} + \binom{n-2-i}{m-2}$ for the 1st term of the above, we have that

$$\begin{aligned} N(n, m) = &\left\{ \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i}\binom{n-2-i}{m-1} \right. \\ &\left. + \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i}\binom{n-2-i}{m-2} \right\} \\ &- \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i}\binom{n-2-i}{m-1} = N(n-1, m-1). \end{aligned}$$

Hence,

$$\begin{aligned} N(n, m) &= N(n-1, m-1) = N(n-2, m-2) = \cdots \\ &= N(n-m+1, 1) = 0. \end{aligned}$$

Case (ii): $n > m > n/2$. In this case $N(n, m)$ is written as

$$N(n, m) = \sum_{i=0}^{n-m} (-1)^i \binom{m}{i}\binom{n-1-i}{m-1}.$$

Let $n - m = r$. Then,

$$N(n, m) = M(r, m) = \sum_{i=0}^{r} (-1)^i \binom{m}{i}\binom{m-1+r-i}{m-1}.$$

Thus, it is enough for proof to show that $M(r, m) = 0$ for $m > r \geqslant 1$. First, for $r = 1$, $M(1, m) = \binom{m}{0}\binom{m}{m-1} - \binom{m}{1}\binom{m-1}{m-1} = 0$. Next, using (42), $M(r, m)$ is rewritten as

$$M(r, m) = \sum_{i=0}^{r} (-1)^i \binom{m-1}{i}\binom{m-1+r-i}{m-1}$$

$$+ \sum_{i=1}^{r} (-1)^i \binom{m-1}{i-1}\binom{m-1+r-i}{m-1}.$$

Furthermore, applying the relation $\binom{m-1+r-i}{m-1} = \binom{m-2+r-i}{m-1} + \binom{m-2+r-i}{m-1}$ for the 1st term in the above, we have

$$M(r, m) = \left\{\sum_{i=0}^{r-1} (-1)^i \binom{m-1}{i}\binom{m-2+r-i}{m-1}\right.$$

$$\left. + \sum_{i=0}^{r-1} (-1)^i \binom{m-1}{i}\binom{m-2+r-i}{m-2}\right\}$$

$$- \sum_{i=0}^{r-1} (-1)^i \binom{m-1}{i}\binom{m-2+r-i}{m-1} = M(r-1, m-1).$$

Hence,

$$\begin{aligned} M(r, m) &= M(r-1, m-1) = M(r-2, m-2) = \cdots \\ &= M(1, m-r+1) = 0. \end{aligned} \qquad \text{Q.E.D.}$$

LEMMA 8. *Let ψ be a function on R. If there is a mapping φ from E^m to E by which ψ is represented as $\psi(X) = \sum^n \varphi_m[a]$, then, conversely, φ is expressed by ψ as*

$$\varphi(x_1, \ldots, x_m) = \sum_{k=0}^{m} \left\{(-1)^{m-k} \frac{n-m}{n-k} \sum^m \psi_k[a]\right\}. \tag{43}$$

Proof. For $1 \leqslant i_1 < \cdots < i_k \leqslant m$, inserting $x_{i_1}, \ldots, x_{i_k}$ into $x_1, \ldots, x_k$ of (36), respectively, and summing sidewise up, we have that

$$\sum^m \psi_k[a] = \sum_{j=0}^{k} \left\{\begin{bmatrix} n-k \\ m-j \end{bmatrix} \sum^m \left(\sum^k \varphi_j[a]\right)_k [a]\right\}.$$

In view of Lemma 6, the above is rewritten as

$$\Psi_k = \begin{bmatrix} n-k \\ m \end{bmatrix}\binom{m}{k}\Phi_0 + \begin{bmatrix} n-k \\ m-1 \end{bmatrix}\binom{m-1}{k-1}\Phi_1 + \cdots + \begin{bmatrix} n-k \\ m-k \end{bmatrix}\binom{m-k}{0}\Phi_k ,$$

where

$$\Psi_k = \sum^m \psi_k[a] \qquad \text{and} \qquad \Phi_j = \sum^m \varphi_j[a]$$

for $k, j = 0, 1,..., m$. Thus we have a linear equation $\Psi = A\Phi$, where

$$\Psi^t = (\Psi_0 , \Psi_1 ,..., \Psi_m),$$
$$\Phi^t = (\Phi_0 , \Phi_1 ,..., \Phi_m),$$

and

$$A = \begin{pmatrix} a_{00} & & & 0 \\ a_{10} & a_{11} & & \\ \vdots & & \ddots & \\ a_{m0} & a_{m1} & \cdots & a_{mm} \end{pmatrix}, \qquad a_{kh} = \begin{bmatrix} n-k \\ m-h \end{bmatrix}\binom{m-h}{k-h}.$$

Let B be the inverse of A. Then the m-th row of B is given by

$$(b_{m0} , b_{m1} ,..., b_{mm}) = (n-m)\left(\frac{(-1)^m}{n}, \frac{(-1)^{m-1}}{n-1}, ..., \frac{1}{n-m}\right). \tag{44}$$

In fact, the (m, j)-element of BA, denoted by c_{mj}, is written as

$$c_{mj} = \sum_{i=j}^{m} b_{mi}a_{ij} = \sum_{i=j}^{m} (-1)^{m-i}\frac{n-m}{n-i}\begin{bmatrix} n-i \\ m-j \end{bmatrix}\binom{m-j}{i-j}$$

Thus, $c_{mm} = 1$. For $j = 0, 1,..., m-1$,

$$\frac{1}{n-i}\begin{bmatrix} n-i \\ m-j \end{bmatrix} = \frac{1}{m-j}\begin{bmatrix} n-i-1 \\ m-j-1 \end{bmatrix}.$$

Hence, in view of Lemma 7, for $j = 0, 1,..., m-1$,

$$\begin{aligned} c_{mj} &= \frac{n-m}{m-j}\sum_{i=j}^{m}(-1)^{m-i}\binom{m-j}{i-j}\begin{bmatrix} n-i-1 \\ m-j-1 \end{bmatrix} \\ &= (-1)^{m-j}\frac{n-m}{m-j}\sum_{i=0}^{m-j}(-1)^{i}\binom{m-j}{i}\begin{bmatrix} n-j-1-i \\ m-j-1 \end{bmatrix} \\ &= (-1)^{m-j}\frac{n-m}{m-j}N(n-j, m-j) = 0. \end{aligned}$$

Using (44), we have finally

$$\varphi(x_1, \ldots, x_m) = \Phi_m = \sum_{k=0}^{m} b_{mk}\Psi_k$$

$$= \sum_{k=0}^{m} \left\{(-1)^{m-k} \frac{n-m}{n-k} \sum^{m} \psi_k[a]\right\}. \qquad \text{Q.E.D.}$$

THEOREM 5. *Let ψ be a function on R and invariant under G. Let m be an integer such that $0 \leqslant m \leqslant |R| - 1$. If there is a mapping φ from E^m to E by which ψ is representable as $\psi = \sum^n \varphi_m[a]$, then $\varphi(y_1, \ldots, y_m)$ is invariant under the group G_0 of all permutations on $\{y_1, \ldots, y_m\}$.*

Proof. In view of (43) in Lemma 8, it is enough for proof to show that, for $k = 0, 1, \ldots, m$, $\sum^m \psi_k[a]$ is invariant under G_0. We denote by $[m]$ a set of integers $1, 2, \ldots, m$ and by G_m the group of all permutations on $[m]$. For a given k, let S be a family of sets $\{i_1, \ldots, i_k\}$ such that $i_1, \ldots, i_k$ are in $[m]$ and mutually distinct. We regard g in G_m as a mapping from S to S such that $g\{i_1, \ldots, i_k\} = \{gi_1, \ldots, gi_k\}$. Then, since g is a permutation on $[m]$, every g in G_m is a one-to-one and onto mapping from S to S. This concludes that $\sum^m \psi_k[a]$ is invariant under G_0, because of ψ being invariant under G. Q.E.D.

This theorem was used for proof of Lemma 4. Now we shall give the classification theorem.

THEOREM 6. *Let ψ be a function on R and invariant under G. Then, for $m = 0, 1, \ldots, |R| - 1$, $o(\psi) \leqslant m$ if and only if, for every $x_1, \ldots, x_n$ and a,*

$$\psi(x_1, \ldots, x_n) = \sum_{k=0}^{m} \left\{(-1)^{m-k} \binom{n-1-k}{m-k} \sum^{n} \psi_k[a]\right\}. \tag{45}$$

Specially, in case of $m = n - 1$, (45) becomes

$$\psi(x_1, \ldots, x_n) = (-1)^{n-1} \sum_{k=0}^{n-1} \left\{(-1)^k \sum^{n} \psi_k[a]\right\}. \tag{46}$$

Proof. Noting that by notation for $k = 0, 1, \ldots, m$, $o(\sum^n \psi_k[a]) \leqslant k$, the "if" part is obvious. We shall prove the "only if" part. Since ψ is invariant under G, in view of the group-invariance theorem, there is a mapping φ

from E^m to E by which ψ is representable as $\psi = \sum^n \varphi_m[a]$. Thus, by Lemma 8, φ is expressed as follows:

$$\varphi(x_1, \ldots, x_m) = \sum_{k=0}^{m} \left\{(-1)^{m-k} \frac{n-m}{n-k} \sum^{m} \psi_k[a]\right\}.$$

By inserting all of $x_{i_1}, \ldots, x_{i_m}$ such that $1 \leqslant i_1 < \cdots < i_m \leqslant n$ into $x_1, \ldots, x_m$ of the above, respectively, and summing sidewise up, we have that

$$\psi(X) = \sum^{n} \varphi_m[a] = \sum_{k=0}^{m} \left\{(-1)^{m-k} \frac{n-m}{n-k} \sum^{n} \left(\sum^{m} \psi_k[a]\right)_m [a]\right\}.$$

Employing Lemma 6 and the relation

$$\frac{n-m}{n-k}\binom{n-k}{m-k} = \binom{n-1-k}{m-k},$$

We finally obtain (45). Q.E.D.

As an application of this theorem, we shall give shorter proofs for mult, max, min and uniform being all of order $|R|$. Since all of these Perceptrons are invariant under G, the classification theorem is applicable. Suppose that the order of those is smaller than $|R|$. Then, for every $x_1, \ldots, x_n$ and a, (46) must hold.

Case (i): $\psi =$ mult. Let $x_1, \ldots, x_n$ be 1 and let $a = 0$. Then mult$(X) = 1$ and, for $k = 0, 1, \ldots, n-1$, mult$(x_{i_1}, \ldots, x_{i_k}, a, \ldots, a) = 0$, where $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Hence, in (46) [the left side] $= 1 \neq 0 =$ [the right side], and this is a contradiction.

Case (ii): $\psi =$ max. Let $x_1 = \cdots = x_n = b < a$. Then max$(X) = b$ and for $k = 0, 1, \ldots, n-1$, max$(x_{i_1}, \ldots, x_{i_k}, a, \ldots, a) = a$, where $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Hence, in (46) [the left side] $= b$ and

$$\begin{aligned}\text{[the right side]} &= \sum_{k=0}^{n-1} \left\{(-1)^{n-1-k} \binom{n}{k} a\right\} \\ &= (-1)^{n-1} a \left\{\sum_{k=0}^{n} (-1)^k \binom{n}{k} - (-1)^n\right\} = a.\end{aligned}$$

This shows a contradiction.

Case (iii): $\psi =$ min. Let $x_1 = \cdots = x_n = b > a$. Then, the similar argument as in case of max leads to a contradiction.

Case (iv): ψ = uniform. Let $x_1, ..., x_n$ and a be mutually distinct. Then uniform$(X) = 0$, uniform$(a, ..., a) = 1$ and for $k = 1, ..., n-1$, uniform$(x_{i_1}, ..., x_{i_k}, a, ..., a) = 0$, where $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Hence, in (46) [the left side] $= 0 \neq (-1)^{n-1} =$ [the right side], and this a contradiction.

Next we shall show that both of the two Perceptrons, equal$(X; P)$ and tolerance$(X; P, \epsilon)$, are of the order $|R|$, where, letting $P = (p_1, ..., p_n)$ be a pattern on R,

$$\text{equal}(X; P) = \begin{cases} 1, & \text{if for } i = 1, ..., n,\ x_i = p_i\,; \\ 0, & \text{otherwise,} \end{cases}$$

$$\text{tolerance}(X; P, \epsilon) = \begin{cases} 1, & \text{if for } i = 1, ..., n,\ |x_i - p_i| \leqslant \epsilon; \\ 0, & \text{otherwise.} \end{cases}$$

Before evaluating the order, we discuss about the change of order which is induced by the transformation of a retina. By $E^{|R|}$ we mean the set of all patterns on R. Let $R_1 = \{x_1, ..., x_n\}$ and $R_2 = \{y_1, ..., y_m\}$. We write a mapping f from $E^{|R_1|}$ to $E^{|R_2|}$ as follows: For X in $E^{|R_1|}$,

$$Y = (y_1, ..., y_m) = (f_1(X), ..., f_m(X)) = f(X).$$

Let ψ_2 be a function on R_2. We define ψ_1 as

$$\psi_1(X) = \psi_2(Y) = \psi_2(f(X)).$$

Then, we say that ψ_1 is a function on R_1 induced by f from a function ψ_2 on R_2.

THEOREM 7. *Let ψ_1 be a function on R_1 induced by $f: E^{|R_1|} \to E^{|R_2|}$ from a function ψ_2 on R_2. If for $i = 1, ..., m$, $|s(f_i)| \leqslant 1$, then $o(\psi_1) \leqslant o(\psi_2)$.*

Proof. Let there be an additive representation for ψ_2 such that $\psi_2(Y) = \sum_{B \in T} \chi_B(Y)$. Let

$$S = \{A; A = s(\chi_B(f(X))) \text{ and } B \in T\},$$

and, for A in S, φ_A be defined as follows: If $A = s(\chi_B(f(X)))$, then, $\varphi_A(X) = \varphi_B(f(X))$. Then ψ_1 can be written as $\psi_1(X) = \sum_{A \in S} \varphi_A(X)$. From this definition, if $\varphi_A(X) = \chi_B(f(X))$, then $s(\varphi_A) \subset \bigcup_i s(f_i)$, where $\bigcup_i$ means a summation with respect to i such that y_i is in $s(\chi_B)$. Hence $|s(\varphi_A)| \leqslant \sum_i |s(f_i)|$, where $\sum_i$ means the same as $\bigcup_i$. Since by assumption $|s(f_i)| \leqslant 1$, we obtain finally that

$$|s(\varphi_A)| \leqslant |s(\chi_B)| \leqslant o(\psi_2).$$

Taking the maximum of the left side in the above with respect to A in S, we have that $o(\psi_1) \leqslant o(\psi_2)$. Q.E.D.

This theorem corresponds to the collapsing theorem by Minsky–Papert (1969).

COROLLARY 1. *Let* ψ_1 *be a function on* R_1 *induced by* $f : E^{|R_1|} \to E^{|R_2|}$ *from a function* ψ_2 *on* R_2. *If*

(i) f *is one-to-one and onto,*
(ii) *for* $i = 1,..., m$, $| s(f_i)| \leqslant 1$, *and*
(iii) *for* $j = 1,..., n$, $| s(f_i^{-1})| \leqslant 1$,

then $o(\psi_1) = o(\psi_2)$, *where* f^{-1} *is written as follows: For* Y *in* $E^{|R_2|}$

$$X = (x_1 ,..., x_n) = (f_1^{-1}(Y),..., f_n^{-1}(Y)) = f^{-1}(Y).$$

Proof. Applying Theorem 7 for f and f^{-1}, we obtain, respectively, that $o(\psi_1) \leqslant o(\psi_2)$ and that $o(\psi_2) \leqslant o(\psi_1)$. Q.E.D.

PROPOSITION 4. $o(\text{equal}) = | R |$.

Proof. Let $\text{eq}(X) = \text{equal}(X; 0)$, where $0 = (0,..., 0)$. First we shall show that $o(\text{eq}) = | R |$. Note that eq is invariant under G, while equal is not. Thus, supposing that $o(\text{eq}) < | R |$, for every $x_1 ,..., x_n$ and a, Eq. (46) of the classification theorem holds. Let $x_1 = \cdots = x_n \neq 0$ and let $a = 0$ for (46). Then

$$\text{eq}(X) = 0, \quad \text{eq}(a,..., a) = 1 \qquad \text{and for} \quad k = 1,..., n - 1,$$

$\text{eq}(x_{i_1} ,..., x_{i_k}, a,..., a) = 0$, where $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Hence in (46) [the left side] $= 0 \neq (-1)^{n-1} =$ [the right side], and this is a contradiction.

Next, let $\hat{R} = \{y_1 ,..., y_n\}$. We regard eq as a function on $\hat{R}$. Defining $f : E^{|R|} - E^{|\hat{R}|}$ as follows: For $i = 1,..., n$,

$$y_i = f_i(X) = x_i - p_i ,$$

$\text{equal}(X; P) = \text{eq}(Y)$, i.e., equal is a function on R induced from eq by f. Since obviously f is one-to-one and onto, and for $i = 1,..., n$, $| s(f_i)| = | s(f_i^{-1})| = 1$, in view of Corollary 1 $o(\text{equal}) = o(\text{eq}) = | \hat{R} | = | R |$. Q.E.D.

PROPOSITION 5. $o(\text{tolerance}) = | R |$.

Proof. Let $\text{tol}(X; \epsilon) = \text{tolerance}(X; 0, \epsilon)$. First, we shall show that

$o(\mathrm{tol}) = |R|$. Note that tol is invariant under G, while tolerance is not. Thus, supposing that $o(\mathrm{tol}) < |R|$, for every $x_1, \ldots, x_n$ and a, Eq. (46) holds. Let $|x_1| \leqslant \epsilon, \ldots, |x_n| \leqslant \epsilon$ and $|a| > \epsilon$ for (46). Then $\mathrm{tol}(X; \epsilon) = 1$ and for $k = 0, 1, \ldots, n-1$, $\mathrm{tol}(x_{i_1}, \ldots, x_{i_k}, a, \ldots, a; \epsilon) = 0$, where $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Hence in (46) [the left side] $= 1 \neq 0 =$ [the right side], and this is a contradiction.

Next, let $\hat{R} = \{y_1, \ldots, y_n\}$. We regard tol as a function on $\hat{R}$. Defining $f : E^{|R|} \to E^{|\hat{R}|}$ as follows: For $i = 1, \ldots, n$

$$y_i = f_i(X) = x_i - p_i,$$

by similar argument to Proposition 4 we can conclude that $o(\mathrm{tolerance}) = o(\mathrm{tol}) = |\hat{R}| = |R|$. Q.E.D.

V. Concluding Remarks

It is investigated the kinds of mathematical tools that are effective for evaluating the order of analog Perceptrons. As a result, the group-invariance theorem, the classification theorem, and the collapsing theorem are given, which were applied to several Perceptrons.

Mathematically, the evaluation of the order is deeply concerned with the 13-th problem of Hilbert. So, some of the contributions to the problem—e.g., Kolmogorov (1958)—will be usefull for the analysis of multilayered analog Perceptrons.

From the pattern-recognition point of view, it is desired to expand our theory to the analog Perceptrons that express the geometrical property of "two-dimensional" figures. On the other hand, from the computational point of view, it may give an insight into the theory of computation if we connect our theory with the complexity of computation appearing in the theory of serial computation, e.g., Winograd (1965, 1967) or Spira (1969).

Acknowledgment

This paper was prepared while the author was in the Research Institute of Fundamental Information Science of Kyushu University at Fukuoka, Japan. He wishes to express his gratitude to Prof. T. Kitagawa and the entire staff of the Institute for their valuable comments and discussion, and for their kindness he would not forget.

Received: September 15, 1970

References

Fukushima, K. (1969), Visual feature extraction by a multilayered network of analog threshold elements, *IEEE Trans.* **SSC-5**, 322–333.

Hilbert, D. (1901), Mathematische probleme, *Archiv. Math. Phys.* **1**, 213–237.

Kolmogorov, A. N. (1957), On the representation of continous functions with several variables as a composition of continuous functions with one variable, Report of Science Academy of S.S.S.R. **114**, 953–956.

Minsky, M., and Papert, S. (1967), Linearly unrecognizable patterns, *in* "Proceedings of Symposia in Applied Mathematics," XIX, Mathematical Aspects of Computer Science, 176–217.

Minsky, M., and Papert, S. (1969), "Perceptrons—An Introduction to Computational Geometry," Massachusetts Institute of Technology Press, Cambridge, Mass.

Rosenblatt, F. (1962), "Principles of Neurodynamics," Spartan Books, Washington, D. C.

Spira, D. H. (1969), The time required for group multiplication, *J. Assoc. Comput. Mach.* **16**, 235–243.

Winograd, S. (1965), On the time required to perform addition, *J. Assoc. Comput. Mach.* **12**, 277–285.

Winograd, S. (1967), On the time required to perform multiplication, *J. Assoc. Comput. Mach.* **14**, 793–802.