SUCCESSIVE PROCESSES OF STATISTICAL OPTIMIZING PROCEDURES

TOSIO KITAGAWA
KYUSHU UNIVERSITY

1. Introduction

The purpose of this paper is to discuss some stochastic aspects of successive processes of controls in connection with optimizing procedures. Various optimizing procedures have been treated by several authors in connection with several different areas belonging to production processes. Thus the response surface analysis aiming to attain an optimal combination of levels of controlled factors was developed by Box, Hunter, and their colleagues [1], [2], and they advocate a certain method for proceeding from some starting point to the optimal point or its neighborhood. The evolutionary operations program introduced by Box [3], [4] has the feature of moving from a routine point of production conditions to a better point in the light of data to be accumulated during the production. The mathematical aspects of response surface analysis and evolutionary operations programs can be formulated more definitely than these authors [1], [2], [3] have done, and indeed we discussed certain mathematical formulations in our two previous papers [21], [22]. It is our viewpoint that, although not all our procedures in these areas can be given in mathematical formulation, there are certainly many situations for which, at least approximately, we can give a rigorous mathematical formulation of our procedure which will lead us to an objective criterion to judge how far our procedures are adequate and efficient.

On the other hand, the recent developments in automatic controls in the production process of plants of various industries are now raising various problems about optimizing procedures which can be defined objectively at least in their main aspects and hence can sometimes be described in mathematical formulations, so far as controlling procedures can be carried out by automatically controlled apparatus. Thus certain rules have been advocated in these areas for changing from a combination of the levels of factors to be controlled to a new combination of them, always aiming to become nearer to an optimal point and to attain it or reach a certain neighborhood of it as fast as possible, that is, with the smallest number of steps needed before realizing its aim. In this connection references can be given to numerous recent works such as Brown [6], Brandon [5], Cosgriff and Emerling [8], Gorn [10], Hooke and van Nice [12], and Lefkowitz and Eckman [25].

Our viewpoint in this paper is to describe such optimizing procedures as successive processes of controls with two emphases. The first emphasis is based upon our recognition of the fact that in current situations our choice of control depends upon information obtained through our observations which, in most real situations of production processes, are under the strong influence of disturbing conditions and circumstances which cannot be fully controlled. Our first emphasis is consequently placed on how to describe the nondeterministic aspects of our processes. This paper is concerned with the situations where stochastic approaches can be expected to have some value in overcoming the nondeterministic aspects of production processes under certain conditions of mass production.

Our second emphasis is to pick out Markov properties from various procedures of our controls. It is obvious that even some of the simplest optimizing procedures cannot be formulated as Markov chains. Nevertheless there are certainly other procedures which are of the Markov type, and it is worthwhile to start with those of the Markov type at least as one possible attack on automatic optimizing procedures.

It shall also be pointed out here that this paper has two particular references. The first reference is to extend the uses of Markov processes to statistical controls to be met in practice, which we started in our previous paper [20]. The second reference is to give a mathematical formulation to the optimizing controlling procedures by logical circuits discussed by Hirai, Asai, and Kitajima [11], who invented electrical apparatus realizing their ideas.

In connection with the first reference it should be added that the uses of Markov processes are advocated by several psychologists [7] in component and pattern models of the mathematical theory of learning. From the cybernetical point of view it is not surprising that our problems derived from engineering circles and others derived from psychological ones may have common aims and common techniques for solving their respective problems, possibly with different emphases placed upon different forms of transition matrices.

In concluding the Introduction the author would point out that this paper is a continuation of his works since 1952 on successive processes of statistical controls. The logical aspects of these works are mentioned in [23]. Although no particular references are made to our previous papers [17] and [19] and to those by our colleagues, Kano [13], [14] and Seguchi [27], these previous papers as well as the present one belong to the same frame of thought emphasizing successive processes of statistical controls in its general sense. Connections with stochastic approximation [26], [9] are also pointed out in our previous paper [19].

2. Controlling procedure of the Markov type

Let a set of all the possible controlled states of our production process be $\{P_1, P_2, \cdots, P_n\}$. Let us assume that to each state P_i there corresponds a certain decision rule D_i by which to determine the transition to a new point P_j

with $1 \leqq j \leqq n$ for the next instance. Our decision rule D_i is defined in principle in the following way.

(i) There corresponds to the point P_i, a set of m_i states $s_i = \{P_{i_1}, P_{i_2}, \cdots, P_{i_{m_i}}\}$, where $1 \leqq i_1 < i_2 < \cdots < i_{m_i} \leqq n$.

(ii) Let us form a statistic x_{i_j} with $j = 1, 2, \cdots, m_i$ at each state P_{i_j} belonging to s_i in virtue of one or more observations on our production process when the state of its controlled factors is assigned to be in the state P_{i_j}, and let us define $x_i = (x_{i_1}, x_{i_2}, \cdots . x_{i_{m_i}})$ as an m_i-dimensional vector.

(iii) A decision function $d[x;\ P_i] = [d_1(x; P_i),\ d_2(x;\ P_i), \cdots, d_n(x; P_i)]$ attached to the state P_i as a function of statistic x is defined for which there is one and only one $j = j_i(x)$ such that $d_j(x; P_i) = 1$ and $d_l(x; p_i) = 0$ if $l \neq j$ in the sense that the transition from the state P_i to the state P_j is implied in view of the statistic x.

Our decision function is a statistical one, because its value depends upon statistics having their respective statistical distribution. Our successive process of statistical controls is now defined as a stochastic process starting from any assigned state to move to another state, which may be or may not be the same state, for the next instance, according to each decision function attached to each of the possible states.

Our successive process of statistical controls is said to be of the Markov type, if the following assumptions are valid.

(1) A sequence of the states $\{P(t)\}$ with $t = 0, 1, 2, \cdots$ is defined according to each decision rule attached to each of the n possible states, that is, the probability for the event that $P(t + 1) = P_j$ under the condition $P(t) = P_i$ is given by the value $P\{d_j[x(t); P_i] = 1\}$ by means of the statistic $x(t)$ obtained at the state $P(t)$ for the time point t.

(2) Each of the n functions $d[\cdot\,; P_i]$ with $i = 1, 2, \cdots, n$ is independent of the time t.

(3) The set of statistics $\{x(t)\}$ with $t = 1, 2, \cdots$ is a set of mutually independent statistics.

Our decision rule will be defined in view of our aim to be attained by means of successive processes of controls. Our aim must be naturally concerned with the populations Π_i attached to each of n controlled states P_i, on which observations are made giving statistics $\{x_i\}$. Let α_i be the population mean of the population Π_i. A maximizing (or minimizing) procedure is one of the procedures formulated in this paragraph in which our aim is to find out the state which will yield us the maximal (or minimal) value of α_i with $i = 1, 2, \cdots, n$. In the following paragraphs several examples are given which illustrate maximizing procedures in various cases.

3. One-dimensional controlling procedure of the Markov type

Let us begin with two examples of general controlling procedure of the Markov type mentioned in section 2.

EXAMPLE 1. A neighborhood s_i attached to each state P_i is defined by

$$\begin{aligned} s_1 &= \{P_1, P_2\}, \\ s_i &= \{P_{i-1}, P_i, P_{i+1}\}, \qquad 2 \leqq i \leqq n-1, \\ s_n &= \{P_{n-1}, P_n\}, \end{aligned} \tag{3.1}$$

that is, s_i consists of a set of three consecutive states except for the two extreme states $i = 1$ and $i = n$, where each two consecutive states are implied respectively.

Let our state be P_i at a time point t. Then let us make independent observations, each one at each state belonging to s_i, which will give us three statistics (x_{i-1}, x_i, x_{i+1}) for $2 \leqq i \leqq n - 1$, and two independent statistics (x_1, x_2) and (x_{n-1}, x_n) for $i = 1$ and $i = n$ respectively.

Various decision functions for maximizing procedures can be defined with reference to the set of these statistics. In what follows, for the sake of simplicity, let us consider the situations where all statistics $\{x_h\}$ for $h \in s_i$, can be assumed without essential loss of generality, to have different values.

CASE 1. There is one and only one $k(i)$ such that $k(i) \in s_i$ and $x_{k(i)} = \max_{h\in s_i} \{x_h\}$. For any assigned values of statistics $\{x_h\}$, $k(i) \equiv k(i; x)$ is hence uniquely determined as a function of these statistics. Now we define $d_{k(i)}(x; P_i) = 1$ and $d_l(x; P_i) = 0$ for $l \neq k(i)$. For each pair i, with $i, j = 1, 2, \cdots, n$, we can define $p_{i,j} = P\{d_j(x; P_i) = 1\}$ and hence the matrix of transition probabilities $T_1 = (p_{i,j})$, with $i, j = 1, 2, \cdots, n$, is defined.

CASE 2. Our decision functions are here defined in the following way.

(i) For $1 \leqq i \leqq n$ and $|j - i| \geqq 2$, $d_j(x; P_i) = 0$.

(ii) For $2 \leqq i \leqq n - 1$, we have

$$\begin{aligned} d_{i+1}(x; P_i) &= 1 \qquad \text{if} \quad x_{i+1} > x_i, \\ d_{i-1}(x; P_i) &= 1 \qquad \text{if} \quad x_{i-1} > x_i \geqq x_{i+1}, \\ d_i(x; P_i) &= 1 \qquad \text{if} \quad x_i \geqq \max(x_{i+1}, x_{i-1}). \end{aligned} \tag{3.2}$$

(iii) For $i = 1$, we have

$$\begin{aligned} d_2(x; P_1) &= 1 \qquad \text{if} \quad x_2 > x_1, \\ d_1(x; P_1) &= 1 \qquad \text{if} \quad x_1 \geqq x_2. \end{aligned} \tag{3.3}$$

(iv) For $i = n$, we have

$$\begin{aligned} d_{n-1}(x; P) &= 1 \qquad \text{if} \quad x_{n-1} > x_n, \\ d_n(x; P) &= 1 \qquad \text{if} \quad x_n \geqq x_{n-1}. \end{aligned} \tag{3.4}$$

It is to be noted that we can design a sequential procedure of experiments in making observations for obtaining statistics $\{x_i\}$. For instance, for each i in $2 \leqq i \leqq n - 1$ we can proceed in the following way.

(i) Make observations giving us statistics x_i and x_{i+1}.

(ii) If $x_{i+1} > x_i$, then we move the state P_i to the state P_{i+1}.

(iii) If $x_{i+1} \leqq x_i$, then make independent observations giving us the statistic x_{i-1}.

(iv)$_1$ If $x_i \geqq \max(x_{i-1}, x_{i+1})$, then we remain at the state P_i.
(iv)$_2$ If $x_{i-1} > x_i \geqq x_{i+1}$, then we move the state P_i to the state P_{i-1}.
In either of these two cases the matrix of transition probabilities has the form

$$(3.5)\quad (p_{i,j}) = \begin{bmatrix} p_{1,1} & p_{1,2} & 0 & 0 & 0 & \cdots & & 0 & 0 \\ p_{2,1} & p_{2,2} & p_{2,3} & 0 & 0 & \cdots & & 0 & 0 \\ 0 & p_{3,2} & p_{3,3} & p_{3,4} & 0 & \cdots & & 0 & 0 \\ 0 & 0 & p_{4,3} & p_{4,4} & p_{4,5} & \cdots & & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdots & & \cdot & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_{n-1,n-2} & & p_{n-1,n-1} & p_{n-1,n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & & p_{n,n-1} & p_{n,n} \end{bmatrix}$$

In order to discuss our maximizing procedure it is natural to assume the i_0th state for which $p_{i_0,i_0} = 1$.

The state p_{i_0} is an absorbing state in the Markov process. Such a type of matrix has an intimate connection with the matrices which have been discussed in the simple birth and death process. Indeed we can apply some results obtained by Ledermann and Reuter [24] and those by Karlin and McGregor [15].

EXAMPLE 2. A neighborhood s_i attached to each state P_i is defined by

$$(3.6)\quad \begin{aligned} s_1 &= \{P_1, P_2, P_3\}, \\ s_2 &= \{P_1, P_2, P_3, P_4\}, \\ s_i &= \{P_{i-2}, P_{i-1}, P_i, P_{i+1}, P_{i+2}\}, \qquad 3 \leqq i \leqq n-2, \\ s_{n-1} &= \{P_{n-3}, P_{n-2}, P_{n-1}, P_n\}, \\ s_n &= \{P_{n-2}, P_{n-1}, P_n\}. \end{aligned}$$

Associated with each state belonging to s_i is a statistic x. The following two cases 3 and 4 correspond to cases 1 and 2 respectively.

CASE 3. Let $x_{j(i)} = \max_{h \in s_i} \{x_h\}$ and let us define $d_{j(i)}(x; P_i) = 1$ and $d_l(x; P_i) = 0$ for $l \neq j(i)$. Let us put $p_{i,j} = P\{d_j(x; P_i) = 1\}$. Then the matrix of transition probability (p_{ij}) is defined for which $p_{ij} = 0$ for $|i - j| \geqq 3$.

CASE 4. Our decision functions are here defined in the following way.

(i) For $1 \leqq i \leqq n$ and $|j - i| \geqq 3$, $d_j(x; P_i) = 0$.

(ii) For $3 \leqq i \leqq n - 2$, we have

$$(3.7)\quad \begin{aligned} d_{i+1}(x; P_i) &= 1 && \text{if } x_{i+1} > x_i, \\ d_{i-1}(x; P_i) &= 1 && \text{if } x_{i-1} > x_i \geqq x_{i+1}, \\ d_{i+2}(x; P_i) &= 1 && \text{if } x_{i+2} > x_i \geqq \max(x_{i-1}, x_{i+1}), \\ d_{i-2}(x; P_i) &= 1 && \text{if } x_{i-2} > x_i \geqq \max(x_{i+2}, x_{i-1}, x_{i+1}), \\ d_i(x; P_i) &= 1 && \text{if } x_i \geqq \max(x_{i-2}, x_{i+2}, x_{i-1}, x_{i+1}). \end{aligned}$$

(iii) For $i = 2$, we have

$$(3.8)\quad \begin{aligned} d_3(x; P_2) &= 1 && \text{if } x_3 > x_2, \\ d_1(x; P_2) &= 1 && \text{if } x_1 > x_2 \geqq x_3, \\ d_4(x; P_2) &= 1 && \text{if } x_4 > x_2 \geqq \max(x_1, x_3), \\ d_2(x; P_2) &= 1 && \text{if } x_2 \geqq \max(x_4, x_1, x_3). \end{aligned}$$

(iv) For $i = n - 1$, we have

$$(3.9)\quad \begin{aligned} d_n(x; P_{n-1}) &= 1 && \text{if } x_n > x_{n-1}, \\ d_{n-2}(x; P_{n-1}) &= 1 && \text{if } x_{n-2} > x_{n-1} \geqq x_n, \\ d_{n-3}(x; P_{n-1}) &= 1 && \text{if } x_{n-3} > x_{n-1} \geqq \max(x_{n-2}, x_n), \\ d_{n-1}(x; P_{n-1}) &= 1 && \text{if } x_{n-1} \geqq \max(x_{n-3}, x_{n-2}, x_n). \end{aligned}$$

(v) For $i = 1$, we have

$$(3.10)\quad \begin{aligned} d_2(x; P_1) &= 1 && \text{if } x_2 > x_1, \\ d_3(x; P_1) &= 1 && \text{if } x_3 > x_1 \geqq x_2, \\ d_1(x; P_1) &= 1 && \text{if } x_1 \geqq \max(x_3, x_2). \end{aligned}$$

(vi) For $\nu = n$, we have

$$(3.11)\quad \begin{aligned} d_{n-1}(x; P_n) &= 1, && x_{n-1} > x_n, \\ d_{n-2}(x; P_n) &= 1, && x_{n-2} > x_n \geqq x_{n-1}, \\ d_n(x; P_n) &= 1, && x_n \geqq \max(x_{n-1}, x_n). \end{aligned}$$

It is to be noted that we can design a sequential procedure of experiments in making observations for obtaining statistics, as was done in case 2.

In either of cases 3 and 4 the matrix of transition probabilities has the form

$$(3.12)\quad (p_{ij}) = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & 0 & 0 & 0 & \cdots & 0 & 0 & & \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & 0 & 0 & \cdots & 0 & 0 & & \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,5} & 0 & \cdots & 0 & 0 & & \\ 0 & p_{4,2} & p_{4,3} & p_{4,4} & p_{4,5} & p_{4,6} & \cdots & 0 & 0 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & p_{n-2,n-4} & & p_{n-2,n-3} & & p_{n-2,n-2} & p_{n-2,n-1} & p_{n-2,n} \\ 0 & 0 & 0 & \cdots & 0 & & p_{n-1,n-3} & & p_{n-1,n-2} & p_{n-1,n-1} & p_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & & 0 & & p_{n,n-2} & p_{n,n-1} & p_{n,n} \end{bmatrix}$$

4. Two-dimensional controlling procedures of the Markov type

Let us consider a set of states being defined by two controlled factors, which we label by a pair of integers $P(i, j)$ where $1 \leqq i \leqq m$ and $1 \leqq j \leqq n$. These possible states will be arranged in a series of mn numbers so that $P(i, j)$ corresponds to the number $i + (j - 1)m$. One of the simplest two-dimensional controlling procedures of the Markov type will be discussed in the following example.

EXAMPLE 3. A two-dimensional neighborhood $s(i, j)$ attached to each state $P(i, j)$ is defined by

(i) $s(1, 1) = \{(1, 1), (2, 1), (1, 2)\}$,

(ii) $s(i, 1) = \{(i, 1), (i + 1, 1), (i - 1, 1), (i, 2)\}$, $\quad 2 \leqq i \leqq m - 1$,

(iii) $s(m, 1) = \{(m, 1), (m - 1, 1), (m, 2)\}$,

(iv) $s(1, j) = \{(1, j), (2, j), (1, j - 1), (1, j + 1)\}$, $\quad 2 \leqq j \leqq n - 1$,

(v) $s(1, n) = \{(1, n), (2, n), (1, n - 1)\}$,
(vi) $s(i, n) = \{(i, n), (i + 1, n), (i, n - 1), (i - 1, n)\}, \quad 2 \leqq i \leqq m - 1$,
(vii) $s(m, j) = \{(m, j), (m, j - 1), (m - 1, j), (m, j + 1)\}, 2 \leqq j \leqq n - 1$,
(viii) $s(m, n) = \{(m, n), (m, n - 1), (m - 1, n)\}$,
(ix) $s(i, j) = \{(i, j), (i + 1, j), (i, j - 1), (i - 1, j), (i, j + 1)\}$,
$2 \leqq i \leqq m - 1, 2 \leqq j \leqq n - 1$.

In short we may say that the state (h, l) belongs to the set $s(i, j)$ if and only if $1 \leqq h \leqq m$, $1 \leqq l \leqq n$ and $|h - i| + |l - j| \leqq 1$, for each (i, j) in $1 \leqq i \leqq m$ and $1 \leqq j \leqq n$.

Neighboring states are defined with reference to two-dimensional arrangement, while the numbers of states are referred to the serial number $i + (j - 1)m$. Thus for case (ix) the number of states belonging to $s(i, j)$ are $i + (j - 1)m$, $i + 1 + (j - 1)m$, $(i - 1) + (j - 1)m$, $i + jm$, $i + (j - 2)m$, where the former three are adjacent while the latter two have great distance from each other and from the former three points.

Let our state be (i, j) at the time point t. Then let us make independent observations, at each state (h, l) belonging to $s(i, j)$, giving us a statistic $x_{(h,l)}$ at each state (h, l). As in example 1, let us consider the following two cases.

Case 1. There is one and only one $k(i, j) = k_1(i, j), k_2(i, j)$ such that $k(i, j) \in s(i, j)$ and $x_{k(i,j)} = \max_{h \in s(i,j)} \{x_h\}$. We define

$$d_{k(i,j)}[x; P(i, j)] = 1, \tag{4.1}$$

$$d_{(h,l)}[x; P(i, j)] = 1, \qquad (h, l) \neq k(i, j), \tag{4.2}$$

and

$$P_{(i,j),(h,l)} = P\{d_{(h,l)}[x; P(i, j)] = 1\}. \tag{4.3}$$

It is evident that $P_{(i,j),(h,l)}$ is the transition probability for the controlled state to move from the $[i + (j - 1)m]$th state (i, j) to the $[h + (l - 1)m]$th state (h, l).

Case 2. Our decision functions are here defined similarly as in example 1, case 2. Thus we have $d_{(h,l)}[x; P(i, j)] = 0$ except for (h, l) belonging to the set $s(i, j)$. Now we have classified all the possible states into nine sets (i) to (ix) according to the contents of $s(i, j)$, which may consist of three [(i), (iii), (v) and (viii)]; four [(ii), (iv), (vi), and (vii)]; and five [(ix)] states. Let us now illustrate a set which consists of five states.

For $2 \leqq i \leqq m - 1$ and $2 \leqq j \leqq n - 1$, we have

(a) $d_{(i+1,j)}[x: P(i, j)] = 1$, if $x_{(i+1,j)} > x_{(i,j)}$.
(b) $d_{(i,j+1)}[x; P(i, j)] = 1$, if $x_{(i,j+1)} > x_{(i,j)} \geqq x_{(i+1,j)}$.
(c) $d_{(i-1,j)}[x; P(i, j)] = 1$, if $x_{(i-1,j)} > x_{(i,j)} > \max\{x_{(i,j+1)}, x_{(i,j+1)}\}$.
(d) $d_{(i,j-1)}[x; P(i, j)] = 1$,
if $x_{(i,j-1)} > x_{(i,j)} \geqq \max\{x_{(i-1,j)}, x_{(i,j+1)}, x_{(i,j+1)}\}$.
(e) $d_{(i,j)}[x; P(i, j)] = 1$, if $x_{(i,j)} \geqq \max\{x_{(i,j-1)}, x_{(i-1,j)}, x_{(i,j+1)}, x_{(i,j+1)}\}$.

In either of cases 1 and 2 the matrix of transition probabilities among mn states can be written in the form

$$(4.4)\qquad T = \begin{bmatrix} A_1 & F_1 & 0 & 0 & 0 & \cdots & & 0 & 0 \\ G_2 & A_2 & F_2 & 0 & 0 & \cdots & & 0 & 0 \\ 0 & G_3 & A_3 & F_3 & 0 & \cdots & & 0 & 0 \\ 0 & 0 & G_4 & A_4 & F_4 & \cdots & & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & & \cdots & G_{n-1} & A_{n-1} & F_{n-1} \\ 0 & 0 & 0 & 0 & & \cdots & 0 & G_n & A_n \end{bmatrix}$$

where

$$(4.5)\qquad A_i = \begin{bmatrix} a_1^{(i)} & b_1^{(i)} & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ c_2^{(i)} & a_2^{(i)} & b_2^{(i)} & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & c_3^{(i)} & a_3^{(i)} & b_3^{(i)} & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & c_4^{(i)} & a_4^{(i)} & b_4^{(i)} & \cdot & \cdot & \cdot & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & c_{m-1}^{(i)} & a_{m-1}^{(i)} & b_{m-1}^{(i)} \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & c_m^{(i)} & a_m^{(i)} \end{bmatrix}$$

$$(4.6)\qquad F_i = \begin{bmatrix} f_1^{(i)} & 0 & 0 & \cdots & 0 \\ 0 & f_2^{(i)} & 0 & \cdots & 0 \\ 0 & 0 & f_3^{(i)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & f_n^{(i)} \end{bmatrix}$$

$$(4.7)\qquad G_i = \begin{bmatrix} g_1^{(i)} & 0 & 0 & \cdots & 0 \\ 0 & g_2^{(i)} & 0 & \cdots & 0 \\ 0 & 0 & g_3^{(i)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & g_m^{(i)} \end{bmatrix}$$

and 0 in (4.4) means the $n \times n$ matrices whose elements are all equal to zero.

The constants $\{a_j^{(i)}\}$, $\{b_j^{(i)}\}$, $\{c_j^{(i)}\}$, $\{f_j^{(i)}\}$, and $\{g_j^{(i)}\}$ mean the following probabilities respectively.

$$(4.8)\qquad P\{d_{(i,j)}[x; P(i,j)]\} = a_j^{(i)},$$

$$(4.9)\qquad P\{d_{(i+1,j)}[x; P(i,j)]\} = b_j^{(i)},$$

$$(4.10)\qquad P\{d_{(i-1,j)}[x; P(i,j)]\} = c_j^{(i)},$$

$$(4.11)\qquad P\{d_{(i,j+1)}[x; P(i,j)]\} = f_j^{(i)},$$

$$(4.12)\qquad P\{d_{(i,j-1)}[x; P(i,j)]\} = g_j^{(i)},$$

for $1 \leqq i \leqq m$ and $1 \leqq j \leqq n$.

EXAMPLE 4. A two-dimensional neighborhood $s(i,j)$ attached to each state $P(i,j)$ is now defined in the following manner: a state (h, l) belongs to the set $s(i,j)$ if and only if $1 \leqq h \leqq m$, $1 \leqq l \leqq n$ and $|h - i| + |l - j| \leqq 2$. More definitely we may enumerate all the possible cases according to the combination of i and j, as we have done in example 3, which however we omit here. Two cases will be mentioned here, each of which corresponds to each of the cases 1 and 2 in example 3 respectively.

CASE 1. This can be defined quite similarly as in case 1 of example 3, except for the difference of the set $s(i, j)$.

CASE 2. Let $x_{(h,l)}$ be the statistic obtained from observations performed under the controlled state (h, l) belonging to $s(i, j)$. Let us give an ordering of all the sets belonging to $s(i, j)$ which begin with the state (i, j).

For instance, let our ordering beginning with the number 0 be given in the following manner: 0: (i, j);

1: $(i + 1, j)$; 2: $(i, j + 1)$; 3: $(i - 1, j)$; 4: $(i, j - 1)$;
5: $(i + 2, j)$; 6: $(i, j + 2)$; 7: $(i - 2, j)$; 8: $(i, j - 2)$.

Let us write for the moment $x(k) = x_{(h,l)}$, provided that k is the number of the order corresponding to (h, l). Further, let us define $x^*(k) = \max \{x(1), x(2), \cdots, x(k)\}$ with $1 \leqq k \leqq 8$. Now our decision functions are defined by means of the statistic $x^*(k)$ in the following way:

(a) $d_{(1)}[x; P(i, j)] = 1$, if $x(1) > x(0)$.
(b) $d_{(2)}[x; P(i, j)] = 1$, if $x(2) > x(0) \geqq x^*(1) = x(1)$.
(c) $d_{(k)}[x; P(i, j)] = 1$, if $x(k) > x(0) \geqq x^*(k - 1)$ for $3 \leqq k \leqq 8$.
(d) $d_{(0)}[x; P(i, j)] = 1$, if $x(0) \geqq x^*(8)$.

In either of the cases 1 and 2 the matrix of transition probabilities among mn states can be written in the form

$$
(4.13)\quad T = \begin{bmatrix}
A_1 & F_1 & F_1' & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\
G_2 & A_2 & F_2 & F_2' & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\
G_3' & G_3 & A_3 & F_3 & F_3' & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\
0 & G_4' & G_4 & A_4 & F_4 & F_4' & 0 & \cdot & \cdot & \cdot & 0 & 0 \\
0 & 0 & G_5' & G_5 & A_5 & F_5 & F_5' & \cdot & \cdot & \cdot & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & G_{n-1}' & G_{n-1} & A_{n-1} & F_{n-1} \\
0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & G_n' & G_n & A_n
\end{bmatrix}
$$

where

$$
(4.14)\quad A_i = \begin{bmatrix}
a_1^{(i)} & b_1^{(i)} & b_1'^{(i)} & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
c_2^{(i)} & a_2^{(i)} & b_2^{(i)} & b_2'^{(i)} & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
c_3'^{(i)} & c_3^{(i)} & a_3^{(i)} & b_3^{(i)} & b_3'^{(i)} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
0 & c_4'^{(i)} & c_4^{(i)} & a_4^{(i)} & b_4^{(i)} & b_4'^{(i)} & 0 & \cdot & \cdot & \cdot & 0 \\
0 & 0 & c_5'^{(i)} & c_5^{(i)} & a_5^{(i)} & b_5^{(i)} & b_5'^{(i)} & \cdot & \cdot & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & c_{n-1}'^{(i)} & c_{n-1}^{(i)} & a_{n-1}^{(i)} & b_{n-1}^{(i)} \\
0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & c_n'^{(i)} & c_n^{(i)} & a_n^{(i)}
\end{bmatrix}
$$

and F_i, F_i', G_i, and G_i' are $n \times n$ diagonal matrices defined similarly as those in (4.6) and (4.7).

5. Monads governing translations and rotations in two-dimensional areas

In our two-dimensional controlling procedures of the Markov type in the previous paragraph we are concerned with the decision functions and allied transition probabilities among the states belonging to the set $s(i, j)$ associated with a state $P(i, j)$. The decision functions are so formulated as to decide the next state to which the present state will be moved as a translation under each one of the specified, mutually exclusive conditions. In some controlling procedures our decision procedure has to be considered from another point of view, that is, with reference to a local system of state coordinates in which notions of rotations as well as translations are explicitly introduced. In order to avoid the complexity of our controlling procedure in its totality and to give insights into its local aspects, it is adequate and convenient to define a localization of the given controlling procedure satisfying the following two conditions.

(a) To describe the same local procedure with that of the original two-dimensional controlling procedure of the Markov type.

(b) To give a separation of the controlling procedure within the set of states $s(i, j)$ from other states which do not belong to the set of states $s(i, j)$.

We shall call such a local mechanism of controlling procedure a neuronic monad associated with each state of a controlling procedure. Since there are mn states in a two-dimensional controlling procedure of the Markov type, there exist mn neuronic monads in total.

The reason why we use a tentative terminology "monad" (in an incompletely specified way) is that any two neuronic monads cannot be directly connected with each other, and that they can be modified so as to form a net system of neurons which will represent the whole aspect of two-dimensional controlling procedure of the Markov type with regard to their global behaviors as well as to their local behaviors.

In the present paragraph we shall illustrate a formulation of neuronic monads.

EXAMPLE 5. With reference to the terminologies used in example 3, let us consider a state $P(i, j)$ such that $2 \leqq i \leqq m - 1$ and $2 \leqq j \leqq n - 1$, that is, a state belonging to $P(i, j)$. The other cases (i) to (viii) can be discussed quite similarly. Five states belonging to the set $s(i, j)$ will be numbered as follows:

TABLE I

	1	2	3	4	0	1′	2′	3′	4′
1	1	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
1′	q_1	0	0	0	q_0	0	$q_{2'}$	0	0
2′	0	r_2	0	0	r_0	0	0	$r_{3'}$	0
3′	0	0	s_3	0	s_0	0	0	0	$s_{4'}$
4′	0	0	0	t_4	t_0	$t_{1'}$	0	0	0

0: (i, j); 1: $(i + 1, j)$, 2: $(i, j + 1)$; 3: $(i - 1, j)$; 4: $(i, j - 1)$. The transient states implying the four rotations around the four points 1, 2, 3, and 4 are numbered in the following way: $1'$: $(1 \to 2)$; $2'$: $(2 \to 3)$; $3'$: $(3 \to 4)$; $4'$: $(4 \to 1)$. The set of nine alternatives 0, 1, 2, 3, 4, $1'$, $2'$, $3'$, and $4'$ is called a set of local states associated with the state 0: (i, j). Let us consider the table of the transition probabilities, shown in table I. Here we have

$$q_1 + q_0 + q_{2'} = 1, \qquad q_j \geqq 0,\ j = 1, 0, 2', \tag{5.1}$$

$$r_2 + r_0 + r_{3'} = 1, \qquad r_j \geqq 0,\ j = 2, 0, 3', \tag{5.2}$$

$$s_3 + s_0 + s_{4'} = 1, \qquad s_j \geqq 0,\ j = 3, 0, 4', \tag{5.3}$$

$$t_4 + t_0 + t_{1'} = 1, \qquad t_j \geqq 0,\ j = 4, 0, 1', \tag{5.4}$$

Here each of the transformations $0 \to i$ for $i = 1, 2, 3, 4$ is a translation, and each of the transformations $i \to (i + 1)$ for $i = 1, 2, 3$ and $4 \to 1$ is a rotation. The matrix given in table I gives us the monad associated with the state 0.

A set of decision functions on the basis of which these transition probabilities are given may be defined in various different ways. However, we are here content with giving one possible way.

(1A) Let our controlled state be 0: (i, j). Then we move our state to the state $1'$.

(2A) Let us make observations at two controlled states 0: (i, j) and 1: $(i + 1, j)$ which will give us statistics $x^{(1)}_{(i,j)}$ and $x^{(1)}_{(i+1,j)}$ respectively.

(3A) If $x^{(1)}_{(i+1,j)} > x^{(1)}_{(i,j)}$, then we move our state $1'$ to the new state 1: $(i + 1, j)$.

(3B) If $x^{(1)}_{(i+1,j)} = x^{(1)}_{(i,j)}$, then we move our state $1'$ to the state 0.

(3C) If $x^{(1)}_{(i+1,j)} < x^{(1)}_{(i,j)}$, we move our state $1'$ to the state $2'$.

(4A) For the case (3C) let us make new observations at two controlled states 0: (i, j) and 2: $(i, j + 1)$, which will give new statistics $x^{(2)}_{(i,j)}$ and $x^{(2)}_{(i,j+1)}$.

(5A) If $x^{(2)}_{(i,j+1)} > x^{(2)}_{(i,j)}$, then we move our state $2'$ to the state 2: $(i, j + 1)$.

(5B) If $x^{(2)}_{(i,j+1)} = x^{(2)}_{i,j}$, then we move our state $2'$ to the state 0: (i, j).

(5C) If $x^{(2)}_{(i,j+1)} < x^{(2)}_{(i,j)}$, then we move our state $2'$ to the state $3'$.

(6A) For the case (5C), let us make new observations at two controlled states 0: (i, j) and 3: $(i - 1, j)$, which will give new statistics $x^{(3)}_{(i,j)}$ and $x^{(3)}_{(i-1,j)}$.

(7A) If $x^{(3)}_{(i-1,j)} > x^{(3)}_{(i,j)}$, then we move our state $3'$ to the state 3: $(i - 1, j)$.

(7B) If $x^{(3)}_{(i-1,j)} = x^{(3)}_{(i,j)}$, then we move our state $3'$ to the state 0: (i, j).

(7C) If $x^{(3)}_{(i-1,j)} < x^{(3)}_{(i,j)}$, then we move our state $3'$ to the state $4'$.

(8A) For the case (7C) let us make new observations at two controlled states 0: (i, j) and 4: $(i, j - 1)$, which will give us new statistics $x^{(4)}_{(i,j)}$ and $x^{(4)}_{(i,j-1)}$ respectively.

(9A) If $x^{(4)}_{(i,j-1)} > x^{(4)}_{(i,j)}$, then we move our state $4'$ to the state 4: $(i, j - 1)$.

(9B) If $x^{(4)}_{(i,j-1)} = x^{(4)}_{(i,j)}$, then we move our state $4'$ to the state 0: $(0, j)$.

(9C) If $x^{(4)}_{(i,j-1)} < x^{(4)}_{(i,j)}$, then we move our state $4'$ to the state $1'$.

Consequently we have

$$P\{x^{(1)}_{(i+1,j)} > x^{(1)}_{(i,j)}\} = q_1(i, j), \tag{5.5}$$

$$P\{x^{(1)}_{(i+1,j)} = x^{(1)}_{(i,j)}\} = q_0(i, j), \tag{5.6}$$

(5.7) $$P\{x^{(1)}_{(i+1,j)} < x^{(1)}_{(i,j)}\} = q_{2'}(i, j),$$

(5.8) $$P\{x^{(2)}_{(i,j+1)} > x^{(2)}_{(i,j)}\} = r_2(i, j),$$

(5.9) $$P\{x^{(2)}_{(i,j+1)} = x^{(2)}_{(i,j)}\} = r_0(i, j),$$

(5.10) $$P\{x^{(2)}_{(i,j+1)} < x^{(2)}_{(i,j)}\} = r_{3'}(i, j),$$

(5.11) $$P\{x^{(3)}_{(i-1,j)} > x^{(3)}_{(i,j)}\} = s_3(i, j),$$

(5.12) $$P\{x^{(3)}_{(i-1,j)} = x^{(3)}_{(i,j)}\} = s_0(i, j),$$

(5.13) $$P\{x^{(3)}_{(i-1,j)} < x^{(3)}_{(i,j)}\} = s_{4'}(i, j),$$

(5.14) $$P\{x^{(4)}_{(i,j-1)} > x^{(4)}_{(i,j)}\} = t_4(i, j),$$

(5.15) $$P\{x^{(4)}_{(i,j-1)} = x^{(4)}_{(i,j)}\} = t_0(i, j),$$

(5.16) $$P\{x^{(4)}_{(i,j-1)} < x^{(4)}_{(i,j)}\} = t_{1'}(i, j),$$

where we have written $q_k(i, j)$, $r_k(i, j)$, $s_k(i, j)$, and $t_k(i, j)$ instead of q_k, r_k, s_k, and t_k respectively in order to show that these probabilities may in general depend upon the state (i, j).

Let us denote the matrix defined by table I by $L(i, j)$ or simply by L. Then we have

(5.17) $$A_L(\lambda) \equiv |\lambda E - L(i, j)| = (\lambda - 1)^4 \begin{vmatrix} \lambda & -1 & 0 & 0 & 0 \\ -q_0 & \lambda & -q_{2'} & 0 & 0 \\ -r_0 & 0 & \lambda & -r_{3'} & 0 \\ -s_0 & 0 & 0 & \lambda & -s_{4'} \\ -t_0 & -t_{1'} & 0 & 0 & \lambda \end{vmatrix}$$

In particular, when $q_0 = r_0 = s_0 = t_0 = 0$, we have

(5.18) $$A_L(\lambda) = (\lambda - 1)^4 \lambda(\lambda^4 - q_{2'}r_{3'}s_{4'}t_{1'}).$$

The matrix $L(i, j)$ defines an absorbing Markov chain having absorbing boundaries consisting of $1: (i + 1, j)$, $2: (i, j + 1)$, $3: (i - 1, j)$, and $4: (i, j - 1)$. The various well-known theorems regarding an absorbing Markov chain can be directly applied to a discussion on the mean number of steps for our state $0: (i, j)$ and to each one of the four absorbing states 1, 2, 3, and 4 respectively (see [16], chapter III). Some of these means may be infinite according to the values of q, r, s, and t. For instance, when $q_{2'} = r_{3'} = s_{4'} = 1$, we have an infinite sequence of rotations among the four states $1'$, $2'$, $3'$, and $4'$ if we start with any of the five states 0, $1'$, $2'$, $3'$, and $4'$. On the other hand, for the situation when all $q_{2'}$, $r_{3'}$, $s_{4'}$, $t_{1'} < 1$, one or more of the four translations, that is, movements from the states $0: (i, j)$ to the states $1: (i + 1, j)$, $2: (i, j + 1)$, $3: (i - 1, j)$, and $4: (i, j - 1)$ will become dominant, that is, a probability for the state to move from 0 to either one of the states 1, 2, 3, and 4 will become large.

EXAMPLE 6. With reference to example 4, let us consider a state (i, j) such that $3 \leqq i \leqq m - 2$ and $3 \leqq j \leqq n - 2$. The other cases having $s(i, j)$, in which the number of elements is less than nine, can be treated quite similarly. Now the eight transient states implying the eight rotations around the eight states 1, 2, 3, $\cdots$, 8 are numbered in the following way: i': $(i \to i + 1)$ for $i = 1, 2, \cdots , 7$, and $8'$: $(8 \to 1)$. Now we have the set of 17 alternatives in the order 1, 2, 3, 4, $\cdots$, 8, 0, $1'$, $2'$, $3'$, $\cdots$, $8'$, which is called a set of local states associated with the state 0: (i, j). Now let us define the 17×17 matrix of the transition probabilities such that

$$L(i, j) = \begin{bmatrix} L_1(i, j) & 0 \\ L_2(i, j) & L_3(i, j) \end{bmatrix} \tag{5.19}$$

where $L_1(i, j)$ is the 8×8 unit diagonal matrix having units as the diagonal elements and zero elsewhere, 0 is the 8×9 zero matrix whose elements are all zero, while $L_2(i, j)$ and $L_3(i, j)$ are defined by

$$L_2(i, j) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_{1'1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & q_{2'2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{3'3} & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{8'8} \end{bmatrix} \tag{5.20}$$

$$L_3(i, j) = \begin{bmatrix} p_0 & p_{1'} & p_{2'} & p_{3'} & p_{4'} & p_{5'} & p_{6'} & p_{7'} & p_{8'} \\ q_{1'0} & 0 & q_{1'2'} & 0 & 0 & 0 & 0 & 0 & 0 \\ q_{2'0} & 0 & 0 & q_{2'3'} & 0 & 0 & 0 & 0 & 0 \\ q_{3'0} & 0 & 0 & 0 & q_{3'4'} & 0 & 0 & 0 & 0 \\ q_{4'0} & 0 & 0 & 0 & 0 & q_{4'5'} & 0 & 0 & 0 \\ q_{5'0} & 0 & 0 & 0 & 0 & 0 & q_{5'6'} & 0 & 0 \\ q_{6'0} & 0 & 0 & 0 & 0 & 0 & 0 & q_{6'7'} & 0 \\ q_{7'0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{7'8'} \\ q_{8'0} & q_{8'1'} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.21}$$

We have

$$\begin{aligned} A_L(\lambda) &\equiv |\lambda E - L(i, j)| \\ &= (\lambda - 1)^8(\lambda - p_0)|\lambda E - L_3(i, j)|. \end{aligned} \tag{5.22}$$

In particular, when $q_{i'0} = 0$ for $i = 1, \cdots , 8$, we have

$$A_{L_3}(\lambda) \equiv |\lambda E - L_3(i, j)| = \lambda^9 - q_{1'2'}q_{2'3'}q_{3'4'} \cdots q_{7'8'}q_{8'1}. \tag{5.23}$$

Similar observations on the matrix $L(i, j)$ can be readily obtained as in example 5.

6. Two-dimensional controlling nets of neurons

The neuronic monads introduced in section 5 are in some sense isolated from each other. We now give some examples of two-dimensional controlling procedures based upon nets of neurons. By a net of neurons we mean a set of local mechanisms of controlling procedures each of which is associated with one state (point) of the two-dimensional lattice, and has neuronic functions governing both translations and rotations at the point which bear a certain relation to each other. Let us illustrate our method by the following example.

EXAMPLE 7. Let us consider a matrix of transition probabilities which is of the form defined in (4.4) to (4.7) in its global connection, with the important difference that $a_j^{(i)}, b_j^{(i)}, c_j^{(i)}, f_j^{(i)}, g_j^{(i)}$ for $i = 1, 2, \cdots, n; j = 1, 2, \cdots, m$ are now not constants, but 5×5 matrices. Indeed, let us now define

$$(6.1) \qquad a_j^{(i)} = \begin{bmatrix} p_{j0}^{(i)} & p_{j1'}^{(i)} & p_{j2'}^{(i)} & p_{j3'}^{(i)} & p_{j4'}^{(i)} \\ q_{j0}^{(i)} & 0 & q_{j2'}^{(i)} & 0 & 0 \\ r_{j0}^{(i)} & 0 & 0 & r_{j3'}^{(i)} & 0 \\ s_{j0}^{(i)} & 0 & 0 & 0 & s_{j4'}^{(i)} \\ t_{j0}^{(i)} & t_{j1'}^{(i)} & 0 & 0 & 0 \end{bmatrix}$$

$$(6.2) \qquad b_j^{(i)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ q_{j1}^{(i)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(6.3) \qquad c_j^{(i)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ s_{j3}^{(i)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(6.4) \qquad f_j^{(i)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ r_{j2}^{(i)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(6.5) \qquad g_j^{(i)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ t_{j4}^{(i)} & 0 & 0 & 0 & 0 \end{bmatrix}$$

In fact we can now denote each of $4mn$ possible alternatives by a set of three integers (i, j, k) such that $1 \leqq i \leqq m$, $1 \leqq j \leqq n$, $0 \leqq k \leqq 4$, and the alternative (i, j, k) has the order $5m(j - 1) + 5(i - 1) + k + 1$ in the ordering previously introduced. Now $a_j^{(i)}, b_j^{(i)}, c_j^{(i)}, f_j^{(i)}$, and $g_j^{(i)}$ are concerned with the transformations of the set of the alternatives $\{(i, j, k)\} (0 \leqq k \leqq 4)$ into the sets $\{(i, j, k)\} (0 \leqq k \leqq 4)$, $\{(9i + 1, j, k)\} (0 \leqq k \leqq 4)$, $\{(i - 1, j, k)\} (0 \leqq k \leqq 4)$, $\{(i, j + 1, k)\} (0 \leqq k \leqq 4)$, and $\{(i, j - 1, k)\} (0 \leqq k \leqq 4)$ respectively.

7. Two-dimensional net of stimulated input-output neurons without path memory

Let us associate with each point (x, y) five vectors $0(x, y)$, $\xi_1^1(x, y)$, $\xi_2^1(x, y)$, $\xi_3^1(x, y)$, and $\xi_4^1(x, y)$, each of which has its endpoint at the point (x, y). $\xi_1^1(x, y)$ is the vector of unit length whose origin is located at the point $(x + 1, y)$, which we shall denote by $\xi_1^1(x, y) = (x + 1, y) \rightarrow (x, y)$. Similar notations will be used to define $\xi_2^1(x, y) = (x, y + 1) \rightarrow (x, y)$, $\xi_3^1(x, y) = (x - 1, y) \rightarrow (x, y)$, $\xi_4^1(x, y) = (x, y - 1) \rightarrow (x, y)$, while $\xi_0^1(x, y) = (x, y) \rightarrow (x, y)$. We call these the five input vectors associated with the point (x, y). Now the output vectors associated with the point (x, y) are the five vectors each of whose origin is located at the point (x, y) such that $\eta_1^1(x, y) = (x, y) \rightarrow (x + 1, y)$, $\eta_2^1(x, y) = (x, y) \rightarrow (x, y + 1)$, $\eta_3^1(x, y) = (x, y) \rightarrow (x - 1, y)$, $\eta_4^1(x, y) = (x, y) \rightarrow (x, y - 1)$ while $\eta_0^0(x, y) = (x, y) \rightarrow (x, y)$, which is identical with $\xi_0^0(x, y)$.

The uses of these notions are concerned with controlling procedures regarding two-dimensional states. Decision functions governing input and output relations are now defined as follows.

Let an input be given by $\xi_h^1(x, y)$. For instance, we understand by the input vector $\xi_1^1(x, y)$ the fact that we are now in the state (x, y) and that we were in the state $(x + 1, y)$ at the previous step. Let the four neighboring states of the state (x, y) be 1: $(x + 1, y)$; 2: $(x, y + 1)$; 3: $(x - 1, y)$; 4: $(x, y - 1)$ respectively.

By a state h we mean a state k such that $h \sim k \pmod 4$, where $1 \leqq k \leqq 4$. We shall begin with the definition of decision functions associated with the state (x, y). Let our input state be $\xi_h^1(x, y)$. We proceed in the following sequential way.

(1) Let us make observations at the state 0 and at the state $h + 2$ which will give us statistics x_0 and x_{h+2} respectively.

(2A) We have an output vector $\eta_{h+2}^1(x, y)$ if $x_{h+2} > x_0$.

(2B) If $x_{h+2} \leqq x_0$, then let us make observations at the state $h + 3$ which will give us statistic x_{h+3}.

(3A) We have an output vector $\eta_{h+3}^1(x, y)$ if $x_{h+2} \leqq x_0 < x_{h+3}$.

(3B) If $\max (x_{h+2}, x_{h+3}) \leqq x_0$, then let us make observations at the state $h + 1$ which will give us statistic x_{h+1}.

(4A) We have an output vector $\eta_{h+1}^1(x, y)$ if $\max (x_{h+2}, x_{h+3}) \leqq x_0 < x_{h+1}$.

(4B) If $\max (x_{h+2}, x_{h+3}, x_{h+1}) \leqq x_0$, then we have an output vector $\eta_0^0(x, y)$.

We adopt an abbreviated notation to state each of the conditions given in our sequential steps as follows.

(2A) $x_0 < x_{h+2}$ by $0(h + 2)$.

(3A) $x_{h+2} \leqq x_0 < x_{h+3}$ by $(h + 2)0(h + 3)$.

(4A) $\max (x_{h+2}, x_{h+3}) \leqq x_0 < x_{h+1}$ by $\overline{(h + 2)(h + 3)}0(h + 1)$.

(4B) $\max (x_{h+2}, x_{h+3}, x_{h+1}) \leqq x_0$ by $\overline{(h + 2)(h + 3)(h + 1)}0$.

Let us denote by $\emptyset$ the impossibility of a transition and by 1 its complement, a transition certainty. Then the conditions for the transitions are given in table II.

TABLE II

	η_0^0	η_1^1	η_2^1	η_3^1	η_4^1
ξ_0^0	1	ϕ	ϕ	ϕ	ϕ
ξ_1^1	(243)0	ϕ	02	$(\overline{24})-3$	204
ξ_2^1	(134)0	01	ϕ	103	$(\overline{13})04$
ξ_3^1	(412)0	401	$(\overline{41})02$	ϕ	04
ξ_4^1	$(\overline{321})0$	$(\overline{32})01$	302	03	ϕ

We shall now describe the whole aspect of combining these neurons governing input and output vectors associated with each state of the whole domain. In order to state the relation among these neurons, we conventionally introduce a set of identifications such that

$$\eta_1^1(x, y) \equiv \xi_3^1(x + 1, y), \tag{7.1}$$

$$\eta_2^1(x, y) \equiv \xi_4^1(x, y + 1), \tag{7.2}$$

$$\eta_3^1(x, y) \equiv \xi_1^1(x - 1, y), \tag{7.3}$$

$$\eta_4^1(x, y) \equiv \xi_2^1(x, y - 1), \tag{7.4}$$

$$\eta_0^0(x, y) \equiv \xi_0^0(x, y). \tag{7.5}$$

Under these conventions to each state (x, y) corresponds the 5×5 matrix whose columns as well as rows consist of $\xi_0^0(x, y)$, $\xi_1^1(x, y)$, $\xi_2^1(x, y)$, $\xi_3^1(x, y)$, and $\xi_4^1(x, y)$ in ascending order from 0 to 4.

Let us consider the domain consisting of lattice points such that $1 \leqq x \leqq m$ and $1 \leqq y \leqq n$. The domain must be divided into nine sets just as in example 3, and certain modifications of definitions of input and output vectors must be introduced for the other domains (i) to (viii) similar to those in example 3. Since five vectors are associated with each one of the mn states, we have now a $5mn \times 5mn$ matrix similar to (4.4) with the essential differences that each of $a_i^{(j)}$, $b_i^{(j)}$, $f_i^{(j)}$, and $g_i^{(j)}$ are not constants, but now 5×5 matrices such as

$$a_i^{(j)} = \begin{array}{c} \\ \xi_0^0(i,j) \\ \xi_1^1(i,j) \\ \xi_2^1(i,j) \\ \xi_3^1(i,j) \\ \xi_4^1(i,j) \end{array} \begin{array}{c} \begin{array}{ccccc} \xi_0^0(i,j) & \xi_1^1(i,j) & \xi_2^1(i,j) & \xi_3^1(i,j) & \xi_4^1(i,j) \end{array} \\ \begin{bmatrix} * & * & * & * & * \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{7.6}$$

$$b_i^{(j)} = \begin{array}{c} \\ \xi_0^0(i,j) \\ \xi_1^1(i,j) \\ \xi_2^1(i,j) \\ \xi_3^1(i,j) \\ \xi_4^1(i,j) \end{array} \begin{array}{c} \begin{array}{ccccc} \xi_0^0(i+1,j) & \xi_1^1(i+1,j) & \xi_2^1(i+1,j) & \xi_3^1(i+1,j) & \xi_4^1(i+1,j) \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & * & 0 \end{bmatrix} \end{array} \tag{7.7}$$

$$
(7.8)\quad c_i^{(j)} = \begin{array}{c} \\ \xi_0^0(i,j) \\ \xi_1^1(i,j) \\ \xi_2^1(i,j) \\ \xi_3^1(i,j) \\ \xi_4^1(i,j) \end{array}
\begin{array}{c} \begin{array}{ccccc} \xi_0^0(i-1,j) & \xi_1^1(i-1,j) & \xi_2^1(i-1,j) & \xi_3^1(i-1,j) & \xi_4^1(i-1,j) \end{array} \\
\left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \end{array}\right] \end{array}
$$

$$
(7.9)\quad f_i^{(j)} = \begin{array}{c} \\ \xi_0^0(i,j) \\ \xi_1^1(i,j) \\ \xi_2^1(i,j) \\ \xi_3^1(i,j) \\ \xi_4^1(i,j) \end{array}
\begin{array}{c} \begin{array}{ccccc} \xi_0^0(i,j+1) & \xi_1^1(i,j+1) & \xi_2^1(i,j+1) & \xi_3^1(i,j+1) & \xi_4^1(i,j+1) \end{array} \\
\left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & * \end{array}\right] \end{array}
$$

$$
(7.10)\quad g_i^{(j)} = \begin{array}{c} \\ \xi_0^0(i,j) \\ \xi_1^1(i,j) \\ \xi_2^1(i,j) \\ \xi_3^1(i,j) \\ \xi_4^1(i,j) \end{array}
\begin{array}{c} \begin{array}{ccccc} \xi_0^0(i,j-1) & \xi_1^1(i,j-1) & \xi_2^1(i,j-1) & \xi_3^1(i,j-1) & \xi_4^1(i,j-1) \end{array} \\
\left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}
$$

where * means certain constants which may not be zero.

8. Two-dimensional net of stimulated input-output neurons with path memory

Let us associate with each point (x, y) nine vectors $\xi_0^0(x, y)$, $\xi_h^i(x, y)$ for $i = 1, 2$; $h = 1, 2, 3, 4$, each of which has its endpoints at the point (x, y), where the upper suffix i denotes the length of the vector and the lower suffix h denotes one of the 4 directions given in section 7. For instance $\xi_3^1(x, y)$ is identical with $\xi_3^1(x, y)$ introduced in section 7, and $\xi_3^2(x, y) = (x - 2, y) \rightarrow (x, y)$. These nine vectors are called input vectors associated with the point (x, y). The nine output vectors associated with the point (x, y), which we denote by $\eta_0^0(x, y)$ and $\eta_h^i(x, y)$ for $i = 1, 2$; $h = 1, 2, 3, 4$, are defined similarly. For instance the vector $\eta_4^1(x, y) = (x, y) \rightarrow (x, y - 1)$ and $\eta_4^2(x, y) = (x, y) \rightarrow (x, y - 2)$.

We are now concerned with a walk of a particle from one point to another when the rule is given by a set of decision functions which have peculiar features different from those given in section 7 in the sense that it will take account of the previous two points (states) occupied by the particle—that is, it will have a certain memory of its history of walking route. In consequence our input state is not one input vector, but a combination of two connected input vectors representing the two previous positions of the particle. Now two connected input vectors associated with the point (x, y) will be denoted by $\xi_h^i\xi_h^i(x, y)$, $\xi_h^i\xi_0^0(x, y)$, and $\xi_0^0\xi_h^i(x, y)$ respectively, and are called input states.

Let us denote by $\psi_h^i(x, y)$ the starting point of an input vector $\xi_h^i(x, y)$, and by $\theta_l^j(x, y)$ the endpoint of an output vector $\eta_l^j(x, y)$.

Hence $\xi_l^j\xi_h^i(x, y)$ is the ordered connection of the two input vectors $\xi_l^j[\psi_h^i(x, y)]$ and $\xi_h^i(x, y)$ in this order, $\xi_0^0\xi_h^i(x, y)$ that of the two input vectors $\xi_0^0[\psi_h^i(x, y)]$ and $\xi_h^i(x, y)$, and $\xi_l^j\xi_0^0(x, y)$ that of two input vectors $\xi_l^j[\psi_h^j(x, y)]$ and $\xi_0^0(x, y)$.

Similarly our output states are two connected output vectors, an output vector connected with the last vector of an input state; that is, each output state is a combination of one of the nine input vectors $\xi_0^0(x, y)$ and $\xi_h^i(x, y)$ with one of the nine output vectors $\eta_0^0(x, y)$ and $\eta_l^j(x, y)$, which we denote by $\xi_0^0\eta_l^j(x, y)$, $\xi_h^i\eta_0^0(x, y)$, $\xi_h^i\eta_l^j(x, y)$, and $\xi_0^0\eta_0^0(x, y)$ respectively.

A neuron associated with the point (x, y) gives us a decision rule by which we decide the transformation from one of the input states associated with the point (x, y) to one of the output states associated with the point (x, y).

Let us give the table of all possible output states with reference to a given output state. For this purpose we have to give a set of decision rules governing the transitions among input vectors and output vectors. There are various possible decision functions. The following is one example.

(a) For the input state $\xi_h^1\xi_h^1(x, y)$: the input state consists of two input vectors $\xi_h^1(x, y)$ and $\xi_h^1[\psi_h^1(x, y)]$ and hence of one route connecting three points $\psi_h^1[\psi_h^1(x, y)]$, $\psi_h^1(x, y)$, and (x, y).

(1) Make observations at the state (x, y) and at the state $\theta_{h+2}^2(x, y)$ which will give us statistics $z^0(x, y)$ and $z_{h+2}^2(x, y)$ respectively.

(2A) If $z_{h+2}^2(x, y) > z^0(x, y)$, we have the output vector $\eta_{h+2}^2(x, y)$ and hence the output state $\xi_h^1\eta_{h+2}^2(x, y)$.

(2B) If $z_{h+2}^2(x, y) \leqq z^0(x, y)$, then let us make observations at the state $\theta_{h+3}^2(x, y)$ which will give us the statistic $z_{h+3}^2(x, y)$.

(3A) If $z_{h+2}^2(x, y) \leqq z^0(x, y) < z_{h+3}^2(x, y)$, we have the output vector $\eta_{h+3}^2(x, y)$ and hence the output state $\xi_h^1\eta_{h+3}^2(x, y)$.

(3B) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y)] \leqq z^0(x, y)$, then let us make observations at the state $\theta_{h+1}^2(x, y)$ which will give us the statistic $z_{h+1}^2(x, y)$.

(4A) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y)] \leqq z^0(x, y) < z_{h+1}^2(x, y)$, we have the output vector $\eta_{h+1}^2(x, y)$ and hence the output state $\xi_h^1\xi_{h+1}^2(x, y)$.

(4B) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y), z_{h+1}^2(x, y)] \leqq z^0(x, y)$, then let us make observations at the state $\theta_{h+2}^1(x, y)$ which will give us the statistic $z_{h+2}^1(x, y)$.

(5A) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y), z_{h+1}^2(x, y)] \leqq z^0(x, y) < z_{h+2}^1(x, y)$, we have the output vector $\eta_{h+2}^1(x, y)$ and hence the output state $\xi_h^1\eta_{h+2}^1(x, y)$.

(5B) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y), z_{h+1}^2(x, y), z_{h+2}^1(x, y)] \leqq z^0(x, y)$, then let us make observations at the state $\theta_{h+3}^1(x, y)$ which will give us the statistic $z_{h+3}^1(x, y)$.

(6A) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y), z_{h+1}^2(x, y), z_{h+2}^1(x, y)] \leqq z^0(x, y) < z_{h+3}^1(x, y)$, we have the output vector $\eta_{h+3}^1(x, y)$ and hence the output state $\xi_h^1\eta_{h+3}^1(x, y)$.

(6B) If $\max [z_{h+2}^2(x, y), z_{h+3}^2(x, y), z_{h+1}^2(x, y), z_{h+2}^1(x, y), z_{h+3}^1(x, y)] \leqq z^0(x, y)$, then let us make observations at the state $\theta_{h+1}^1(x, y)$ which will give us the statistic $z_{h+1}^1(x, y)$.

(7A) If max $[z^2_{h+2}(x, y), z^2_{h+3}(x, y), z^2_{h+1}(x, y), z^1_{h+2}(x, y), z^1_{h+3}(x, y)] \leqq z^0(x, y) < z^1_{h+1}(x, y)$, we have the output vector $\eta^1_{h+1}(x, y)$ and hence the output state $\xi^1_h\eta^1_{h+1}(x, y)$.

(7B) If other cases occur, that is, max $[z^2_{h+2}(x, y), z^2_{h+3}(x, y), z^2_{h+1}(x, y), \cdots, z^1_{h+1}(x, y)] \leqq z^0(x, y)$, we have the output vector $0(x, y)$ and hence the output state $\xi^1_h 0(x, y)$.

Our decision rule can be summarized in the abbreviated table III.

TABLE III

OUTPUT STATES FOR THE INPUT STATE $\xi^1_h\xi^1_h$

Cases	Output State	Conditions for Decision
(2A)	$\xi^1_h\eta^2_{h+2}$	$0(h+2)^2$
(3A)	$\xi^1_h\eta^2_{h+3}$	$(h+2)^2 0(h+3)^2$
(4A)	$\xi^1_h\eta^2_{h+1}$	$\overline{(h+2)^2(h+3)^2} 0(h+1)^2$
(5A)	$\xi^1_h\eta^1_{h+2}$	$\overline{(h+2)^2(h+3)^2(h+1)^2} 0(h+2)$
(6A)	$\xi^1_h\eta^1_{h+3}$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)} 0(h+3)$
(7A)	$\xi^1_h\eta^1_{h+1}$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)(h+3)} 0(h+1)$
(7B)	$\xi^1_h 0$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)(h+3)(h+1)} 0$

Now we proceed to other cases having other input states. We may and shall use the abbreviated summaries of our decision rules by means of the notations just used for the case (a) having $\xi^1_h\xi^1_h(x, y)$ as input state.

(b) For the input state $\xi^1_{h+i}\xi^1_h(x, y)$ with $i = 1, 3$, see table IV.

TABLE IV

OUTPUT STATES FOR THE INPUT STATE $\xi^1_{h+i}\xi^1_h(x, y)$

Cases	Output State	Conditions for Decision
(2A)	$\xi^1_h\eta^1_{h+2}$	$0(h+2)$
(3A)	$\xi^1_h\eta^1_{h+3}$	$(h+2)0(h+3)$
(4A)	$\xi^1_h\eta^1_{h+1}$	$\overline{(h+2)(h+3)} 0(h+1)$
(5A)	$\xi^1_h 0$	$\overline{(h+2)(h+3)(h+1)} 0$

(c) For the input state $\xi^j_l\xi^2_h$ with $j = 1, 2$; $l, h = 1, 2, 3, 4$, $l \neq h$. The decision rules are the same as for the case (a) so far as they are concerned with the steps (1) to (7A). We must modify the case (7B) so as to be able to add the cases shown in table V to table III.

TABLE V

OUTPUT STATES FOR THE INPUT STATES $\xi_l^j \xi_h^2(x, y)$

Cases	Output State	Conditions for Decision
(7A)	$\xi_h^2 \eta_h^1$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)(h+3)}0h$
(8A)	$\xi_h^2 \eta_{h+1}^1$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)(h+3)h}0(h+1)$
(8B)	$\xi_h^2 0$	$\overline{(h+2)^2(h+3)^2(h+1)^2(h+2)(h+3)h(h+1)}0$

(d) For the input state $\xi_l^2 \xi_h^1$ with $l = h, h+1, h+3$.

The input state of this type implies that we have already made observations at the corresponding states $\theta_{h+2}^1(x, y)$ and $\theta_h^1(x, y)$ with the consequences that $z_{h+2}^1(x, y) \leqq z^0(x, y)$ and $z_h^1(x, y) \leqq z^0(x, y)$. Consequently our decision rules are much simplified as verified by table VI.

TABLE VI

OUTPUT STATES FOR THE INPUT STATES $\xi_l^2 \xi_h^1(x, y)$

Cases	Output State	Conditions for Decision
(2A)	$\xi_h^1 \xi_{h+3}^1$	$0(h+3)$
(3A)	$\xi_h^1 \xi_{h+1}^1$	$(h+3)0(h+1)$
(3B)	$\xi_h^1 0$	$\overline{(h+3)(h+1)}0$

(e) For the input states $\xi_l^j \xi_0^0$ and $\xi_0^0 \xi_0^0$.

An input state of this type implies that some observations on the states in a neighborhood of the state (x, y) have already been made, with the consequence that $z^0(x, y)$ is not less than any statistic corresponding to each state in the neighborhood. At first sight it seems that no further observations are needed and that we should conclude that we have reached the state which is at least locally optimal. Indeed we may make this conclusion under certain circumstances, particularly when all of the following conditions are satisfied.

(i) No change of production conditions will occur in the future.

(ii) There is some reason which assures us our local optimality is sufficient for our optimizing purpose.

(iii) Our statistics have small sampling fluctuations so that no repeated samplings will be required to ascertain our conclusions.

So far as not all of these three conditions are satisfied, our controlling procedure should continue to make observations and decisions. We here give one possible decision procedure: make a choice by a chance mechanism by which to decide an input vector from all the set of $\xi_h^1(x, y)$ with $h = 1, 2, 3, 4$.

For the input state $\xi_0^0 \xi_0^0(x, y)$, the same decision rule is used.

(f) For the input state $\xi_0^0 \xi_h^j$ with $j = 1, 2$; $h = 1, 2, 3, 4$.

We define the decision rule for the input states $\xi_0^0\xi_h^1$ and $\xi_0^0\xi_h^2$ to be the same with those of $\xi_l^1\xi_h^1$ with $l \neq h$, $h + 2$, and with those of $\xi_l^2\xi_h^2$ with $l \neq h$, $h + 2$ respectively.

Among all 64 ($= 2 \times 2 \times 4 \times 4$) combinations of the suffixes i, j, h, and l such that $i, j = 1, 2$, and $h, l = 1, 2, 3, 4$, the cases when $l = h + 2$ with $h = 1, 2, 3, 4$ are impossible, hence the number of all the possible input states of the types $\xi_l^j\xi_h^i(x, y)$ are 48 ($= 64 - 16$). Furthermore, we have to add the input states such as $\xi_0^0\xi_0^0$, $\xi_l^j\xi_0^0$, $\xi_0^0\xi_l^j$ with $j = 1, 2$; $l = 1, 2, 3, 4$ whose number is 17 ($= 1 + 8 + 8$). Consequently the total number of all the possible input states is 65.

Now let us turn to the output states. To each point there are associated the output vectors $\eta_p^k(x, y)$ with $k = 1, 2$; $p = 1, 2, 3, 4$ and $\eta_0^0(x, y)$. We define our output state by a combination of each one of the first vectors of $\xi_h^i(x, y)$ and $\xi_0^0(x, y)$ with each one of the second vectors $\eta_p^k(x, y)$ and $\eta_0^0(x, y)$. Some combinations are impossible.

But our principal device in dealing with output states associated with a point is to think of each output state as an input state associated with its endpoint.

For instance, let our output state be the connection of the end vector $\xi_h^i(x, y)$ of the input state $\xi_l^j\xi_h^i(x, y)$ with $\eta_p^k(x, y)$. As an output vector $\eta_p^k(x, y)$ has its origin (starting point) at the point (x, y) and as its endpoint $\psi_p^k(x, y)$. Our device is first to notice that $\eta_p^k(x, y) = \xi^k[\psi_p^k(x, y)]$ and then to think of the connection as $\xi_h^i\xi_{p+2}^k[\psi_p^k(x, y)]$. This device makes it possible to give the matrix of the transition probabilities among the input states. Let the coordinate of a point be denoted by (x, y) in $1 \leqq x \leqq m$, $1 \leqq y \leqq n$. With each point (x, y) let there be in general associated 65 input states. Hence we can define now a matrix of the transition probabilities among the total input states. Each input state can be given by the coordinate (x, y, u), where (x, y) is the coordinate of the point and u is the order number of the input state within 65 input states associated with the point (x, y).

Let our ordering of the states be such that an input state (x, y, u) is ranked in the $[65(x - 1)(y - 1) + 65(x - 1) + u]$th order. It is noted that each of all the possible output states associated with the point (x, y) is transformed to an input state associated with a certain point (x', y') belonging to the set of points $(x \pm 1, y)$, $(x \pm 2, y)$, $(x, y \pm 1)$, and $(x, y \pm 2)$.

Consequently it is obvious that our matrix of the transition probabilities has the form of the matrix (4.13) with the specifications of (4.14) and with the difference that $a_j^{(i)}$, $b_j^{(i)}$, $b_j'^{(i)}$, $c_j^{(i)}$, $c_j'^{(i)}$, $f_j^{(i)}$, $f_j'^{(i)}$, $g_j^{(i)}$ and $g_j'^{(i)}$ are now not the constants but in general the 65×65 matrices respectively.

9. Simplified aspects of global behaviors of controlling procedures

In section 8 a transition matrix is given which will enable us to describe the whole aspect of certain controlling procedures in their global behaviors as well as in their local behaviors. However, these complete descriptions do not give us direct observations on the global behaviors of our controlling procedures, and some device must be introduced to facilitate us in obtaining a simplified and sufficiently approximative picture of controlling procedures. A few techniques are indicated in this section for this purpose.

(1) *Simplified and approximate transition matrix regarding transitions among the vector directions.* In the previous sections detailed descriptions of decision functions appealing to statistics based upon observations at different controlled states were given in which both translations and rotations were taken into consideration. These considerations lead us to neurons governing translations and rotations without memory or with memory and to a net system of such neurons located at each point in the two-dimensional lattice domain. In their simpler cases transitions are concerned with the points in that domain, and in their more complicated cases we have to discuss transitions among the input states each of which can be characterized by the location of the input point, the directions and the lengths of two component vectors. Two input states having the same direction and the same length of two corresponding vectors have to be distinguished from each other provided they have different input points. Similar remarks are valid for the formulation in section 7. These distinctions are required in general because two such states may have in general different matrices of transition probabilities, so far as their input points are not identical. In spite of these general assertions, a device may sometimes give a substantial simplification of our descriptions of controlling procedures.

Let us now be concerned with transitions among vector directions. For instance, we now have the identity relations to the effect that for any pair of (x', y') and (x, y)

$$\xi_h^1(x, y) = \xi_h^1(x', y'), \tag{9.1}$$

$$\eta_h^1(x, y) = \eta_h^1(x', y'), \tag{9.2}$$

for $h = 1, 2, 3, 4$. In addition to these, the conventions used in (7.1) to (7.4) simplify our situation so that our concern with vectors can now be reduced to four vector directions: $(0, 0) \to (1, 0)$; $(0, 0) \to (0, 1)$; $(0, 0) \to (-1, 0)$; $(0, 0) \to (0, -1)$, which we denote tentatively by 1, 2, 3, and 4. Vector directions will not give us, however, any substantial contributions unless the corresponding elements of the transition matrices associated with every point of the interior of a certain domain are all equal to each other, that is, matrices of transition probabilities have space homogeneity. This condition is rather too severe to be satisfied throughout the whole domain of the state points under consideration. Nevertheless, a restriction of our state points to some suitable subdomain may secure us the condition for an approximate space homogeneity. Under this assumption we may and shall confine our attention to the transitions among the four vector directions.

Let us assume that an approximate space homogeneity is valid within a certain subdomain, that is, as our reference state point moves within the interior of the subdomain, very small numerical changes of elements of the transition matrices are expected so that a common constant transition matrix (independent of its reference state point) can be used within the interior of the subdomain, in order to give an approximate picture of our controlling procedures in its sufficient accuracy of approximation.

Let us begin with examples for which our device will serve to simplify observations on the successive behavior of our controlling procedures.

EXAMPLE 8. Prevailing flows with rotations. Let the matrix of transitions among four vector directions be given by

$$\text{(9.3)} \qquad T = \begin{matrix} \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} \begin{matrix} 1 & 2 & 3 & 4 \\ \end{matrix} \begin{bmatrix} a & b+u & c & d \\ a & b & c+u & d \\ a & b & c & d+u \\ a+u & b & c & d \end{bmatrix}$$

One of the possible interpretations may be given as follows. There is a prevailing flow of transitions with a fixed set of probabilities a, b, c, d valid for all points within the interior of the subdomain where (a, b, c, d) are independent of the previous direction along which we have proceeded just before we reach the point.

On the other hand, there is one additional possibility of changing the previous direction into another one such that our particle will move along the vector direction $(h + 1)$ with a certain probability u when it has come along the vector direction h. This possibility may imply in some instances the admissibility of rotations. We have $a, b, c, d, u \geqq 0$ and $a + b + c + d + u = 1$. By virtue of the direct applications of simple Markov chains we can handle the various types of controlling problems. For instance, the evaluation of the mean number of steps for our particle to attain the boundary of our subdomain for the first time is fundamentally important for controlling procedures. It is noted that since we have

$$\text{(9.4)} \qquad A_T(\lambda) \equiv |\lambda E - T| = (\lambda - 1)(\lambda^3 + \lambda^2 u + \lambda u^2 + u^3),$$

we can write out by the method in our previous paper [20] the iteration T^n by means of the characteristic values of T, which will give us direct and explicit evaluations of the probabilities of absorption at the boundary. Our example includes the particular cases when $u = 0$ and $u = 1$ respectively. When $u = 1$, there happens a route of cyclic rotations around four states.

EXAMPLE 9. Prevailing flows with rotations under sequential experiments. Let the translation matrix of transitions among four vector directions and stopping of steps be given by

$$\text{(9.5)} \qquad T = \begin{matrix} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} \begin{matrix} 0 & 1 & 2 & 3 & 4 \\ \end{matrix} \begin{bmatrix} e+u & a & b & c & d \\ e & a & b+u & c & d \\ e & a & b & c+u & d \\ e & a & b & c & d+u \\ e & a+u & b & c & d \end{bmatrix}$$

The main difference between the present example and the previous example 8 is the introduction of the new degenerate vector 0 which means no step in any direction. The transition probabilities ($e + u$, a, b, c, d) may be based upon our decision rule regarding our experiments to decide (i) whether we proceed in any vector direction or we remain at the same point and (ii) to which direction we proceed in the former case. Consequently, the probability $e + u$ can be understood as implying the possibility that we appeal to a new experiment when we are placed in the situation 0. Other aspects of our interpretations of the matrix are similar to those in example 8. We have now

$$a + b + c + d + e + u = 1. \tag{9.6}$$

It is to be noted that

$$A_T(\lambda) \equiv |\lambda E - T| = (\lambda - 1)(\lambda^4 - u^4) \tag{9.7}$$

is different from (9.4). Besides any answers to the questions to be discussed about example 8, we may have to discuss the mean numbers of replicated experiments for our particle to reach the boundary of the subdomain for the first time.

(2) *Transitions among global subdomains.* In order to have information about the global behavior of transitions of our states, it is sufficient to separate the set of all states into a system of subsets of states. For instance the states corresponding to the mn points in the domain $1 \leqq x \leqq m$, $1 \leqq y \leqq n$ are divided into the system of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{K}$, as shown in table VII.

TABLE VII

SUBDIVISION OF THE WHOLE DOMAIN
$1 \leqq x \leqq m, \quad 1 \leqq y \leqq m$
INTO NINE SUBDOMAINS

$\mathcal{B}$	$\mathcal{C}$	$\mathcal{D}$
$\mathcal{E}$	$\mathcal{A}$	$\mathcal{F}$
$\mathcal{G}$	$\mathcal{H}$	$\mathcal{K}$

As an illustration let us consider example 3 with the matrix of transition probabilities given in (4.4) to (4.7), which gives us its expression with reference to the present separation as shown in table VIII. Here the 0 denote the matrices whose elements are all zero, while the * mean certain matrices which are not 0. In table VIII the matrix corresponding to the combination of $\mathcal{H}$ in the column and $\mathcal{Y}$ in the row will be written by $\mathcal{H}\mathcal{Y}$; such as $\mathcal{A}\mathcal{A}$, $\mathcal{A}\mathcal{D}$, $\mathcal{F}\mathcal{D}$, $\mathcal{F}\mathcal{K}$, $\mathcal{F}\mathcal{F}$ and so on. The contents of each of these matrices $\mathcal{A}\mathcal{A}$, $\mathcal{A}\mathcal{D}$. $\cdots$. $\mathcal{F}\mathcal{K}$ and $\mathcal{F}\mathcal{K}$ can be readily obtained by means of description in example 3. Let us now discuss the transition from the set $\mathcal{B} + \mathcal{D} + \cdots + \mathcal{G} + \mathcal{K}$ to the set $\mathcal{A}$. For this purpose the absorbing

TABLE VIII

TRANSITION MATRIX AMONG NINE SUBDOMAINS

	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{D}$	$\mathcal{G}$	$\mathcal{H}$	$\mathcal{K}$	$\mathcal{E}$	$\mathcal{F}$
$\mathcal{A}$	*	0	*	0	0	*	0	*	*
$\mathcal{B}$	0	*	*	0	0	0	0	*	0
$\mathcal{C}$	*	*	*	*	0	0	0	0	0
$\mathcal{D}$	0	0	*	*	0	0	0	0	*
$\mathcal{G}$	0	0	0	0	*	*	0	*	0
$\mathcal{H}$	*	0	0	0	*	*	*	0	0
$\mathcal{K}$	0	0	0	0	0	*	*	0	*
$\mathcal{E}$	*	*	0	0	*	0	0	*	0
$\mathcal{F}$	*	0	0	*	0	0	*	0	*

Markov chain is introduced in which $\mathcal{A}$ is defined to be an absorbing boundary. Its transition matrix therefore has the form

$$T_{\mathcal{A}} = \begin{bmatrix} I & O \\ R & Q \end{bmatrix}, \tag{9.8}$$

where I is the identity matrix,

$$R = \begin{bmatrix} 0 \\ \mathcal{C}\mathcal{A} \\ 0 \\ 0 \\ \mathcal{G}\mathcal{A} \\ 0 \\ \mathcal{E}\mathcal{A} \\ \mathcal{F}\mathcal{A} \end{bmatrix} \tag{9.9}$$

and

$$Q = \begin{bmatrix} \mathcal{B}\mathcal{B} & \mathcal{B}\mathcal{C} & 0 & 0 & 0 & 0 & \mathcal{B}\mathcal{E} & 0 \\ \mathcal{C}\mathcal{B} & \mathcal{C}\mathcal{C} & \mathcal{C}\mathcal{D} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathcal{D}\mathcal{C} & \mathcal{D}\mathcal{D} & 0 & 0 & 0 & 0 & \mathcal{D}\mathcal{F} \\ 0 & 0 & 0 & \mathcal{G}\mathcal{G} & \mathcal{G}\mathcal{H} & 0 & \mathcal{G}\mathcal{E} & 0 \\ 0 & 0 & 0 & \mathcal{H}\mathcal{G} & \mathcal{H}\mathcal{H} & \mathcal{H}\mathcal{K} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathcal{K}\mathcal{H} & \mathcal{K}\mathcal{K} & 0 & \mathcal{K}\mathcal{F} \\ \mathcal{E}\mathcal{B} & 0 & 0 & \mathcal{E}\mathcal{G} & 0 & 0 & \mathcal{E}\mathcal{E} & 0 \\ 0 & 0 & \mathcal{F}\mathcal{D} & 0 & 0 & \mathcal{F}\mathcal{K} & 0 & \mathcal{F}\mathcal{F} \end{bmatrix} \tag{9.10}$$

Then the fundamental matrix for our absorbing chain is defined by

$$N = (I - Q)^{-1}. \tag{9.11}$$

In order to have an expression for the inverse of $I - Q$, let us write

$$I - Q = \left[\begin{array}{cccc} K_1 & 0 & L_1 & L_3 \\ 0 & K_2 & L_2 & L_4 \\ \hdashline M_1 & M_3 & I - \mathcal{E}\mathcal{E} & 0 \\ M_2 & M_4 & 0 & I - \mathcal{F}\mathcal{F} \end{array}\right] = \begin{bmatrix} K & L \\ M & N \end{bmatrix} \tag{9.12}$$

where

$$
(9.13)\qquad
\begin{aligned}
K_1 &= \begin{bmatrix} I - \mathfrak{B}\mathfrak{B} & -\mathfrak{B}\mathfrak{C} & 0 \\ -\mathfrak{E}\mathfrak{B} & I - \mathfrak{C}\mathfrak{C} & -\mathfrak{C}\mathfrak{D} \\ 0 & -\mathfrak{D}\mathfrak{C} & I - \mathfrak{D}\mathfrak{D} \end{bmatrix}, \\
K_2 &= \begin{bmatrix} I - \mathfrak{G}\mathfrak{G} & -\mathfrak{G}\mathfrak{H} & 0 \\ -\mathfrak{H}\mathfrak{G} & I - \mathfrak{H}\mathfrak{H} & -\mathfrak{H}\mathfrak{K} \\ 0 & \mathfrak{K}\mathfrak{H} & I - \mathfrak{K}\mathfrak{K} \end{bmatrix}
\end{aligned}
$$

$$
(9.14)\qquad L_1 = \begin{bmatrix} -\mathfrak{B}\mathfrak{E} \\ 0 \\ 0 \end{bmatrix}, \quad L_2 = \begin{bmatrix} -\mathfrak{G}\mathfrak{E} \\ 0 \\ 0 \end{bmatrix}, \quad L_3 = \begin{bmatrix} 0 \\ 0 \\ -\mathfrak{D}\mathfrak{F} \end{bmatrix}, \quad L_4 = \begin{bmatrix} 0 \\ 0 \\ -\mathfrak{K}\mathfrak{F} \end{bmatrix}
$$

$$
(9.15)\qquad
\begin{aligned}
M_1 &= (-\mathfrak{E}\mathfrak{B} \quad 0 \quad 0), & M_3 &= (-\mathfrak{E}\mathfrak{G} \quad 0 \quad 0) \\
M_2 &= (0 \quad 0 \quad -\mathfrak{F}\mathfrak{D}), & M_4 &= (0 \quad 0 \quad -\mathfrak{F}\mathfrak{K})
\end{aligned}
$$

$$
(9.16)\qquad
\begin{aligned}
K &= \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}, & L &= \begin{bmatrix} L_1 & L_3 \\ L_2 & L_4 \end{bmatrix}, \\
M &= \begin{bmatrix} M_1 & M_3 \\ M_2 & M_4 \end{bmatrix}, & N &= \begin{bmatrix} I - \mathfrak{E}\mathfrak{E} & 0 \\ 0 & I - \mathfrak{F}\mathfrak{F} \end{bmatrix}.
\end{aligned}
$$

In consequence, we have

$$
(9.17)\qquad N = (I - Q)^{-1} = \begin{bmatrix} K^{-1} + XZ^{-1}Y & -XZ^{-1} \\ -Z^{-1}Y & Z^{-1} \end{bmatrix}
$$

where $X = K^{-1}$, $Y = MK^{-1}$, and $Z = N - MK^{-1}L$, provided that Z^{-1} exists. Consequently, the problem of finding the inverse of the matrix $I - Q$ is now reduced to that of evaluating Z^{-1}, K_1^{-1}, and K_2^{-1}. We can proceed to the further reduction of our problem to the evaluation of the inverses of matrices of smaller orders. These procedures may be said to be intolerably tedious and laborious for a general transition matrix. It is however to be expected that, for our particular transition matrix, tremendous simplification may sometimes be given in calculating its minor submatrices.

10. Game-theoretic approach and application of Monte Carlo method

Our formulation of successive processes of statistical optimizing procedures in terms of the Markov chains implies the necessity for establishing some probabilistic theorems by which to make clear not only the local but also the global behavior of iterated products of the transition matrix P. Results obtained in our previous paper [20] which are based on reduction theorems and analysis of routes may serve to facilitate the calculation of the characteristic roots of $A(x) \equiv |\lambda E - P| = 0$. However, in addition to probabilistic approaches there is a possibility of attacking our problem from game-theoretic approaches. This derives from the fact that the transition matrices now under consideration are determined by two conditions. The first condition is concerned with our process of statistical optimizing procedures, that is to say, with our chosen strategy. The second condition is determined by a set of populations associated with each point of the one- or two-dimensional lattice set on which we have incomplete information. This situation may permit us to formulate a game-theoretic approach in which one player is the control engineer while the other is nature, who chooses a setup of populations. For instance, in the case of a two-dimensional lattice let $\alpha(x, y)$ be the population mean of a population associated with the lattice point (x, y). Under our formulation we have incomplete information about $\alpha(x, y)$, irrespective of whether the assumption $\alpha(x, y)$ may be a deterministic function or a sample function from some stochastic processes. In this sense our solution to the problem of giving a strategy for choosing a process of statistical control among the set of possible processes may be adequately formulated under a decision function approach which takes into consideration game-theoretic aspects and which may belong to stochastically approximative analysis in the sense of [18].

Experimental approach also will be suggestive in developing our theoretic considerations for the comparison of the merits and deficiencies of various optimizing procedures. Some experimental results are given by Hirai, Asai, and Kitajime [11] which treat the case when $\alpha(x, y) = (32^2 - x^2 - y^2)^{1/2}$ and no observational errors are involved. The introduction of observational errors in such model experiments may lead us to an application of the Monte Carlo method, which is probably indispensable to our theoretic considerations in giving us empirical grounds leading to conjectures as to the asymptotic behavior of the powers of P.

REFERENCES

[1] G. E. P. Box and G. A. Coutie, "Application of digital computers in the exploration of function relationships," *Proc. Inst. Elec. Eng.*, Vol. 103 (1956), pp. 100–107.

[2] G. E. P. Box and J. S. Hunter, "Multifactor experimental designs for exploring response surfaces," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 195–241.

[3] G. E. P. Box, "Evolutionary operation: A method for increasing industrial productivity," *Appl. Statist.*, Vol. 6 (1957), pp. 3–23.

[4] ———, "Some general considerations in process optimization," *J. Basic Eng., Trans. ASME, Ser. D*, Vol. 82 (1960), pp. 113–119.

[5] D. B. Brandon, "Developing mathematical models for computer control," *Instrument Soc. Amer. J.*, Vol. 6 (1959), pp. 70–74.

[6] R. R. Brown, "A generalized computer procedure for the design of optimum systems. Parts I and II," *Commun. Electronics*, Vol. 43 (1959), pp. 285–289 and 289–293.

[7] R. R. Bush and W. K. Estes, "Studies in mathematical learning theory," *Stanford Mathematical Studies in the Social Sciences*, Vol. 3, Part I, "Stimulus-sampling models," Stanford, Stanford University Press, 1959.

[8] R. L. Cosgriff and R. A. Emerling, "Optimizing control system," *Appl. Ind.*, Vol. 35 (1958), pp. 13–16.

[9] A. Dvoretzky, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 39–55.

[10] S. GORN, "On the mechanical simulation of habit-forming and learning," *Information and Control*, Vol. 2 (1959), pp. 226–259.

[11] H. HIRAI, K. ASSAI, and S. KITAZIMA, "Optimizing control by logical circuits," report in *Nippon Denki Tsushin Gakukai*, 1960. (In Japanese.)

[12] R. HOOKE and R. I. VAN NICE, "Optimizing control by automatic experimentation," *Instrument Soc. Amer. J.*, Vol. 6 (1959), pp. 78–79.

[13] S. KANO, "Linear controlled stochastic processes," *Sci. Rep. Kagoshima Univ.*, Vol. 8 (1959), pp. 1–14.

[14] ———, "Linear controlled stochastic processes with continuous parameter," *Bull. Math. Statist.*, Vol. 9 (1960), pp. 43–49.

[15] S. KARLIN AND J. L. MCGREGOR, "The differential equations of simple birth and death processes," *Trans. Amer. Math. Soc.*, Vol. 85 (1957), pp. 489–549.

[16] J. G. KEMENY and J. L. SNELL, *Finite Markov Chains*, New York, Van Nostrand, 1959.

[17] T. KITAGAWA, "Successive processes of statistical controls (1)," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 7 (1952), pp. 13–28.

[18] ———, "Some aspects of stochastically approximative analysis," *Bull. Math. Statist.*, Vol. 6 (1956), pp. 109–129.

[19] ———, "Successive processes of statistical controls (2)," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 13 (1959), pp. 1–16.

[20] ———, "Successive processes of statistical controls (3)," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 14 (1960), pp. 1–33.

[21] ———, "Successive process of statistical inferences applied to linear regression analysis and its specializations to response surface analysis," *Bull. Math. Statist.*, Vol. 8 (1959), pp. 80–114.

[22] ———, "A mathematical formulation for evolutionary operation programs," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 15 (1960), in preparation.

[23] ———, "The logical aspects of successive processes of statistical inferences and controls," *The 32nd Session of the International Statistical Institute at Tokyo*, 1960, No. 20, pp. 1–14.

[24] W. LEDERMANN and G. E. H. RUETER, "Spectral theory for the differential equations of simple birth and death processes," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 246 (1954), pp. 321–369.

[25] I. LEFKOWITZ and D. P. ECKMAN, "Optimizing control by model methods," *Instrument Soc. Amer. J.*, Vol. 6 (1959), pp. 74–77.

[26] H. ROBBINS and S. MONRO, "A stochastic approximation method," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 400–407.

[27] T. SEGUCHI, "Some statistical controls (II)—controls of finite Markov chains," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 15 (1960), in preparation.