ERRATA AND ADDENDA

(Part I)

page	line	
VI	12) 19) 26	for the steady state probability distribution corresponding ... read the steady state probability distribution of (3) corresponding ... .
VII	15	for $E\{ \mid k(v;[A_z]) \mid \; \mid x\} < \infty$ read $E\{ \mid k(\underline{v};[A_z]) \mid \; \mid x\} < \infty$.
VII / VIII	26) 14	for the steady state probability distribution corresponding ... read the steady state probability distribution of (1) corresponding ... .
4 / 5	11) 21) 6	for by the definition of Ω^* a point $\omega_1 \varepsilon \Omega^*$ ... read by the definition of Ω^* one and only one point $\omega_1 \varepsilon \Omega^*$ ... .
5	8	for $x_t^*(\omega)$ read $x_t^*(\omega_1)$.
6	29	for the probability measure ... read a probability measure ... .
7	1	add which depends on the past $Pa^{\omega;t'}$.
8	11	for notations read notations 1.
8	26	for Thus, if A is ... read Note that, if A is ... .
8	28	add (cf. point 2) on p. 3).
13	3	for is given by the probability measure ... read is given by a probability measure ... .
13	5	add which depends on the past $Pa_1^{\omega;t'}$.
14	12	for A in a ... If $x' \varepsilon \overline{A}$ we ... read A_γ in a ... If $x' \varepsilon \overline{A}_\gamma$ we ... .
17	1	for Property 6 read Property 6^*.
19	1	for to compare strategies ... read to compare strategies at t_o ... .

page	line	
19	7	**for** $c(Fu_1^{\omega;t}(z))$ ... **read** $c(Fu_1^{\bar{\omega};t_o}(z))$ ... .
20	7	**for** be optimal ... **read** be optimal from t_o onwards ... .
27	24	**for** $\{S_{z;x'};x' \varepsilon X\}$ **read** $\{S_{z;x'};x' \varepsilon X'\}$.
28	9	**for** We can easily verify that ... **read** For an appropriate decomposition of X' into simple ergodic sets and a set of transient states of the process (2.4) (cf. [2], p. 210) we can verify that ... .
31	1	**for** $r(z;x_1) \overset{def}{=} E\{k(\underline{u};t_o)\|x_1\}$ **read** $r(z;x_1) \overset{def}{=} \frac{1}{t_o} E\{k(\underline{u};t_o)\|x_1\}$.
31	28	**add** where $p^j_{[C]}(B;x';z)$ denotes the $j^{\underline{th}}$ order transition probability distributions of (2.17).
34	5	**for** $= \dfrac{\int_{X_1} p_{[C]}(du;x_1;z)k(u;[\hat{C}])}{\int_{X_1} p_{\|C\|}(du;x_1;z)t(u;\|\hat{C}\|)}$ **read** $= \dfrac{\int_{X_1} p_{[\hat{C}]}(du;x_1;z)k(u;[\hat{C}])}{\int_{X_1} p_{[\hat{C}]}(du;x_1;z)t(u;[\hat{C}])}$.
35	22	**for** into A ... **read** into A_γ ... .
36	7	**for** $= \int_{X_1} p_{[\hat{C}]}(du;x_1;z)k(u;[C])$ **read** $= \int_{X_1} p_{[\hat{C}]}(du;x_1;z)k(u;[\hat{C}])$.
36	9	**for** $= \int_{X_1} p_{[\hat{C}]}(du;x_1;z)t(u;[C])$ **read** $= \int_{X_1} p_{[\hat{C}]}(du;x_1;z)t(u;[\hat{C}])$.

page	line	
36	21	<u>add</u> where $k(x_1;[C]) = \int_{X_2} z(dx_2;x_1)k(x';[C])$, $t(x_1;[C]) = \int_{X_2} z(dx_2;x_1)t(x';[C])$ and $x' = (x_1,x_2)$.
37	7	<u>for</u> $k(x_1;[C]) \ldots t(x_1;[C])$ <u>read</u> $k(x_1;[\hat{C}]) \ldots t(x_1;[\hat{C}])$.
40	5	<u>for</u> $t(x;z) > 0$ <u>read</u> $t(x;[C]) > 0$.
40	7	<u>for</u> In chapter 2 of part II, (3.77) ... <u>read</u> In chapter 2 of part II, (2.77) ... .
40	8	<u>for</u> $t(x;z) = \ldots$ <u>read</u> $t(x;[A_z]) = \ldots$.
41	6	<u>for</u> 2) $p^n_{[\hat{C}]}(M_{i;n};x;z) = 1$ <u>read</u> 2) $p^{n-1}_{[\hat{C}]}(M_{i;n};x;z) = 1$.
41	22	<u>add</u> If $\hat{B} = X$, (2.55) represents the differences in expected costs between the decision processes $S_{z;x}$ and S^x_z with respect to the $S_{z;x}$-time intervals $\{[0,\underline{t}_{[\hat{C}];x} + \sum_{j=1}^{n} \underline{t}_{[\hat{C}];\underline{I}_j}); n=1,2,\ldots\}$.
42	19	<u>for</u> $\frac{1}{c_i}\sum_{j=1}^{c_i} p^{\infty c_i}_{[\hat{C}]}(B;I_1;z)$ <u>read</u> $\frac{1}{c_i}\sum_{j=1}^{c_i} p^{\infty c_i+j}_{[\hat{C}]}(B;I_1;z)$.

page line

48 14 <u>add</u> If x is the initial state, let $\underline{k}_{T;x}$ represent the random costs to incur in a period of length T. If $\hat{B} = X$, the following relation can be verified

$$c(z;x;\hat{C}') = E\underline{k}_{T;x} - Tr(z;x) + Ec(z;\underline{\xi}_T;\hat{C}') \ ,$$

where $\underline{\xi}_T$ is the state at time T.
By (2.92), we then have

$$|E\underline{k}_{T;x} - Tr(z;x)| \leq 2\kappa \ .$$

In other words, the difference in expected costs between the decision processes $S_{z;x}$ and S_z^x is uniformly bounded in x on the class of all finite time intervals.

48 16 <u>add</u> where $\underline{I}_1$ is the first future entry state in $\hat{C}$.

49 15 <u>for</u> $\{c(z;\underline{I}_1;A^z)|x;z\}$; if $x \in \bar{A}_z$
<u>read</u> $E\{c(z;\underline{I}_1;A^z)|x;z\}$; if $x \in \bar{A}_z$.

53 14) 18 <u>for</u> $p^m_{A_{z_b}}(dI_m;x;z)$ <u>read</u> $p^{m^*}_{A_{z_b}}(dI_m;x;z)$, where

$$m^* = \begin{cases} m-1, & \text{if } x \in A_{z_b} \\ m \ , & \text{if } x \in \bar{A}_{z_b} \end{cases} \ .$$

63 20 <u>add</u> $r((z_2)z;x) < r(Az';x) = r(z';x)$ with $c(Az';x) \leq c(z';x)$ or .

MATHEMATICAL CENTRE TRACTS

3

GENERALIZED MARKOVIAN DECISION PROCESSES

PART I

MODEL AND METHOD

BY

G. DE LEVE

MATHEMATISCH CENTRUM AMSTERDAM

1964

CONTENTS

PART I

Acknowledgement

The author is greatly indebted to Prof.dr. J. Hemelrijk and Prof.dr. J.Th. Runnenburg for their criticism and encouragement.

INTRODUCTION

The purpose of this book is to introduce a new approach to an extensive class of decision problems. Of the problems we have in mind, a large number refer to inventories, productions and replacements. It appears that problems of this type can be formulated as stochastic ∞-stage decision problems.

In CHAPTER 1 of PART I common characteristics of a number of decision-situations are investigated. The insight gained into the structure of decision mechanisms is used for attributing basic properties to a common mathematical model. In each situation we have a physical system of which the state can be represented by a point of an N-dimensional Cartesian space X. The space X is called the state space.

Moreover, we observe that there is a random change in the state of the system. In case no decisions are made, this evolution is called the natural process. In the mathematical model natural processes are defined by means of stationary strong Markov processes in X. In addition to this it is assumed that almost all sample functions of the natural process are continuous from the right in the time variable t and have in each finite time interval only a finite number of discontinuities. For the definitions of these concepts the reader is referred to chapter 1 in part II.

Oviously, reflections on losses and gains play a prominent part in the determination of a decision. It is no restriction to assume that only losses occur. In general two types of losses are distinguished. First, losses of which the extent changes continuously in the course of time (e.g. interest) and secondly, losses which have in- and decrements at discrete points of time (e.g. sales and costs of repair). These costs are defined in such a way that for each time interval they are completely fixed by the walk of the system (cf. section 2 in chapter 1 of part II). It will be clear that the decisionmaker, who is in charge, wants to prevent, or at least wants to make impro-

bable, the most expensive excursions through the state space.
By analysing physical decision situations we discover that decisions effect transitions in the state of the system. Moreover, it appears that in many situations decisions result in a random transition in the state of the system. For that reason decisions are defined by means of probability distributions on the states into which the system may be transferred at the moment of the decision. It is convenient to assume that at each point of time a decision is made. In this study, however, we make a distinction between interventions and null-decisions. By a null-decision the system is "transferred" with probability 1 in its present state. Both intervention and null-decision are represented by a point d of a so called decision space D.
It follows from the nature of many a decision problem that in some states certain decisions are not feasible. Consequently, to each state $x \varepsilon X$ a set of feasible decisions D(x) is assigned.
The solution of the decision problem is given in the form of a strategy. Such a strategy dictates at each point of time a feasible decision on the basis of the available information.
Obviously, because of the extra transitions, the natural process is no longer appropriate to describe the behaviour of the system if a strategy is applied. We restrict ourselves to strategies which have the property that the evolution in the state of the system can still be described by means of a stochastic process. In case the dictated decisions also depend on states assumed in the past, these processes in general are not Markovian.
In order to find out which strategy is the best one, we need a criterion. In lemma 1.2 of chapter 1 in part I we prove that, if the criterion has a number of specified properties, if certain additional conditions are imposed on the class of strategies to be considered, and if an optimal strategy exists, then there is at least one optimal strategy, which assigns to each state x in X one and only one feasible decision d in D. In case these conditions are fulfilled it is no restriction to consider only strategies z which map the state space X into the decision space D. Such strategies are represented by func-

tions $d = z(x)$ and constitute the class Z. Since we have only interventions and null-decisions the strategies $z \in Z$ divide the state space X into two disjunct sets, one denoted by A_z, comprising states in which always interventions are made, the other consisting of the states in which always null-decisions are dictated.

In chapter 2 of part II, under rather weak conditions, we prove that, if a strategy $z \in Z$ is applied, the evolution in the state of the system can still be described by means of a stationary strong Markov process (theorem 3). This stochastic process is called the decision process and is defined for each initial state x.

In chapter 2 of part II, theorem 2, we demonstrate that almost surely a finite number of interventions are made in a finite time interval. That is the reason why decision problems of this type can be formulated like stochastic ω-stage decision problems.

In chapter 2 of part II, theorem 1, we prove that the sequence of intervention states, denoted by [1])

$$\{ \underline{I}_1, \underline{I}_2, \ldots\ldots \} \quad , \tag{1}$$

constitutes a stationary Markov process with a discrete time parameter.

We now come to the explicit form of our criterion for optimality. If ω represents a realization of the decision process, let $k_T(\omega;z)$ denote the total loss incurred during the period $[0,T)$. In CHAPTER 2 of PART I, under certain conditions, it is proved that for almost all realizations ω the limit

$$\lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \tag{2}$$

exists. Note that this limit represents the mean costs per unit of time. In order to show that (2) can be expressed in a usable form we introduce the sequence of random states

$$\{ \underline{x}_{t_0;x;j} \ ; \ j=0,1,\ldots \} \tag{3}$$

---

1) Throughout this study random variables are underlined.

which represents the states of the system at the times $\{jt_o; j=0,1,\dots\}$ if x is the initial state.
Let us assume that this sequence constitutes a stationary Markov process which satisfies the Doeblin condition ([2] , p.192). We further assume that with respect to the steady state probability distribution of (3) we have

$$E\{|\,k(\underline{u};t_o)|\ \ |\,x\} < \infty, \tag{4}$$

where $k(u;t_o)$ denotes the expected costs for the interval $[0,t_o)$ if u is the initial state. Then, by using known ergodic theorems, it is proved that for all ergodic initial states x we have almost surely

$$\lim_{T\to\infty} \frac{k_T(\omega;z)}{T} = \frac{1}{t_o} E\{k(\underline{u};t_o)|\,x\}, \tag{5}$$

where $\underline{u}$ obeys the steady state probability distribution corresponding to x (cf. theorem 4 in chapter 2 of part II). If the initial state x is transient, then (2) almost surely depends on the first ergodic state assumed. Consequently, if x is transient, the mean costs per unit of time is as yet a random variable. On the set of all ergodic states x an x-function r(z;x) is defined by

$$r(z;x) \overset{\text{def}}{=} \frac{1}{t_o} E\{\,k(\underline{u};t_o)|\,x\,\}, \tag{6}$$

where $\underline{u}$ obeys the steady state probability distribution corresponding to x. Note that the x-function r(z;x) is constant on the states of a simple ergodic set [2)] and represents the mean costs per unit of time with probability 1. The domain of definition of the x-function r(z;x) is extended to the transient states and thus to X as a whole by taking

$$r(z;x) \overset{\text{def}}{=} E\{\,r(z;\underline{u})|\,x\,\}, \tag{7}$$

where $\underline{u}$ obeys the steady state probability distribution corresponding to x.

---

2) We prefer the name simple ergodic set to ergodic set, because the latter can be mixed up with the set of all ergodic states.

Note that, by (4), (5) and (6), the x-function $r(z;x)$, representing the expected mean costs per unit of time, has now been defined for all initial states x. Obviously, the z-function $r(z;x)$ is a good criterion for optimality.

In chapter 2 of part I it is also shown that there are two more ways to define the criterion function $r(z;x)$. The third one will be discussed now. To this end we consider the sequence (1) of intervention states in A_z. Let us suppose that the Markov process (1) satisfies the Doeblin condition ([2] , p.192). Let $k(v;[A_z])$ represent the expected value of the costs incurred between two successive interventions if the intervention state of the first one is v. Next let $t(v;[A_z])$ represent the expected duration of the time interval between these interventions.

We now assume for each initial state x

$$E\{|k(v;[A_z])| \; |x\} < \infty \tag{8}$$

and

$$0 < E\{t(\underline{v};[A_z])|x\} < \infty, \tag{9}$$

where $\underline{v}$ obeys the steady state probability distribution of (1) corresponding to x.

It is proved that, if x is an ergodic state of (1), we have

$$r(z;x) = \frac{E\{k(\underline{v};[A_z])|x\}}{E\{t(\underline{v};[A_z])|x\}} , \tag{10}$$

where $\underline{v}$ obeys the steady state probability distribution of (1) corresponding to x (cf. theorem 5 in chapter 2 of part I with $\hat{C} = A_z$).

Obviously, we have for all x

$$r(z;x) = E\{r(z;\underline{v})| x\} , \tag{11}$$

where $\underline{v}$ obeys the steady state probability distribution corresponding to x.

It can easily be verified that the functions $k(u;[A_z])$ and $t(u;[A_z])$ are determined by

a) the decision $d = z(u)$;

b) the natural process;

c) the set A_z.

In chapter 2 we introduce (x;d)-functions k(x;d) and t(x;d), which are determined by

a) the decision d;

b) the natural process;

c) the non-empty intersection $A_o = \bigcap_{z \in Z} A_z$.

Note that the (x;d)-functions k(x;d) and t(x;d) do not refer to any particular strategy z. In section 3 of chapter 2 (part I) we prove that for ergodic initial states x of (1) the x-function r(z;x) can also be defined by

$$r(z;x) = \frac{E\{k(\underline{v};z(\underline{v}))|\, x\}}{E\{t(\underline{v};z(\underline{v}))|\, x\}}, \tag{12}$$

where $\underline{v}$ obeys the steady state probability distribution corresponding to x. The domain of definition is extended to X by means of (11). This result implies that, if we want to compare different strategies by means of the criterion function r(z;x) we have to determine

a) the (x;d)-functions k(x;d) and t(x;d) <u>once for all</u>;

b) the steady state probability distributions of the processes (1) for each of strategies <u>individually</u>.

The introduction of the (x;d)-functions k(x;d) and t(x;d) thus leads to a considerable simplification.

We have already stated that the expected mean costs per unit of time r(z;x) is constant on a simple ergodic set of the decision process in X. In this chapter we also show that the effect of the initial state on the total expected loss is limited to a finite amount if only states of one simple ergodic set of the decision process in X are considered.

An x-function c(z;x) is introduced which, in a sense, evaluates the initial state with respect to the total expected loss. The x-functions r(z;x) and c(z;x) satisfy the functional equations

$$r(z;x) = \int_{A_z} p^1_{A_z}(dI;x;z)r(z;I) \qquad (13)$$

and

$$c(z;x) = k(x;z(x)) - r(z;x)t(x;z(x)) + $$

$$+ \int_{A_z} p^1_{A_z}(dI;x;z)c(z;I), \qquad (14)$$

where $p^1_{A_z}(B;x;z)$ represents the probability distribution of the first future intervention state if x is the initial state.

In order to be able to describe an important property of the optimal strategy we need a number of concepts which are defined now.

Let the mixed strategy $(z^*)z$ mean that all but the first intervention are made in accordance with strategy z; the first one conforms to z^*.

Let the functions $r((z^*)z;x)$ and $c((z^*)z;x)$ be defined by

$$r((z^*)z;x) \overset{\text{def}}{=} E\{r(z;\underline{I}_2)\,|\,z^*;x\} \qquad (15)$$

and

$$c((z^*)z;x) \overset{\text{def}}{=}$$

$$= E\{k(\underline{I}_1;z^*(\underline{I}_1)) - r((z^*)z;\underline{I}_1)t(\underline{I}_1;z^*(\underline{I}_1)) + c(z;\underline{I}_2)\,|\,x_1;z^*\}. \qquad (16)$$

Let the subclass Z_z of Z be defined by [3]

$$Z_z \overset{\text{def}}{=} \{z^* \mid z^* \varepsilon\ Z;\ \forall_{x\,\varepsilon X}\ r((z^*)z;x) = \inf_{\bar{z}\,\varepsilon\,Z} r((\bar{z})z;x)\}. \qquad (17)$$

In CHAPTER 3 of PART I we prove that if a strategy z_o satisfies for all $x\,\varepsilon\,X$

$$c(z_o;x) = \min_{z\,\varepsilon\,Z_{z_o}} c((z)z_o;x), \qquad (18)$$

it is an optimal strategy.

Let the mixed strategy d.z mean that after the effectuation of decision d in the initial state decisions are made in accordance with strategy z. We define x-functions r(d.z;x) and c(d.z;x) by

---

3) $\forall x\,\varepsilon\,X$ means: for all $x\,\varepsilon\,X$ we have .... .
$\exists x\,\varepsilon\,X$ means: there exists at least one $x\,\varepsilon\,X$ such that .... .

$$r(d.z;x) \overset{\text{def}}{=} E\{r(z;\underline{y})\,|\,d\} \tag{19}$$

and

$$c(d.z;x) \overset{\text{def}}{=} k(x;d) - r(d.z;x)t(x;d) + $$
$$+ E\{c(z;\underline{y})\,|\,d\} \;. \tag{20}$$

where $\underline{y}$ obeys the probability distribution corresponding to d.
The subset $D_z(x)$ of $D(x)$ in D is defined by

$$D_z(x) \overset{\text{def}}{=} \{d\,|\,d \in D(x);\; r(d;x) = \min_{d' \in D(x)} r(d'z;x)\} \;. \tag{21}$$

Let the mixed strategy A.z interdict any intervention up to the moment that the system assumes a state in A for the first time. From that time onwards decisions are made in accordance with z.
The x-functions $r(A.z;x)$ and $c(A.z;x)$ are defined by

$$r(A.z;x) \overset{\text{def}}{=} E\{r(z;\underline{y})\,|\,x;A\} \tag{22}$$

and

$$c(A.z;x) \overset{\text{def}}{=} E\{c(z;\underline{y})\,|\,x;A\} \;, \tag{23}$$

where $\underline{y}$ obeys the probability distribution of the first state in A assumed.
We further consider the class K_z of all closed sets A satisfying:

1) $A \supset A_o = \bigcap_{z \in Z} A_z$; (24)
2) $\overline{A} = \{x\,|\,r(A.z;x) < r(z;x)\} \cup \{x\,|\,r(A.z;x) = r(z;x); c(A.z;x) \leqq$
$\leqq c(z;x)\}$. (25)

Finally, we introduce the set A'_z, defined by

$$A'_z = \bigcap_{A \in K_z} A. \tag{26}$$

In chapter 3 of part I we also prove that, if a strategy z_o satisfies for all x

$$c(z_o;x) = \min_{d \in D_{z_o}(x)} c(d.z_o;x) \tag{27}$$

and

$$A'_{z_o} = A_{z_o}, \tag{28}$$

it is an optimal strategy.

In addition to this we consider two iteration procedures which may lead to an optimal strategy. The first one, called strategy improvement routine I, has reference to (18), while the second one, the strategy improvement routine II, may solve (27) and (28) in an iterative way. It is proved that, under certain conditions, these procedures are effective.

CHAPTER 4 in PART I contains a summary of the new method. Moreover, this method is compared with some known methods in this field. In section 3 the Dynamic Programming approach of RICHARD BELLMAN is discussed, while section 4 is devoted to RONALD A. HOWARD's techniques.

The new method is not a "ready-made" technique. Its final form depends heavily on the structure of the decision problem concerned. In PART III of M.C. TRACT No 3 we shall show that in several decision situations this approach leads to rather simple techniques.

The purpose of PART II is to show that there are no objections of probabilistic nature. We demonstrate that probability spaces can be constructed which cover all the requirements. Moreover, the strong Markov property of the decision process is proved in that part.

Both BELLMAN ([3], p.317 ff.) and HOWARD [4] have considered decision processes which are markovian. These processes pass under the name of MARKOVIAN DECISION PROCESSES. The state spaces concerned consist of a finite number of states, while in addition to this almost all sample functions of the decision process are step functions. In this study, however, more general state spaces and decision processes are treated. Therefore, this book comes out under the title of GENERALIZED MARKOVIAN DECISION PROCESSES.

CHAPTER 1

The mathematical model

1. State, evolution and natural process

In physical decision problems the choice of a decision depends on the state of the physical system concerned. In a replacement problem e.g. the system may be a machine, while in an inventory problem the system is the inventory or the inventory and the quantities on order. The corresponding mathematical concepts will also be called "state" and "system"; they will have the following property: [1)]

Property 1

In a mathematical model the state of the system is determined by M real-valued variables; thus by a point of an M-dimensional Cartesian space ($M < \infty$).

The set of all possible states x will be called the state space X^*. We consider physical systems which change their states if the time passes.

In the model this corresponds to a walk of the system through the state space. If the variable t runs through the time axis T,

$$T = [0, \infty) \ ,$$

each walk in the state space can be identified with a function $x = x^*(t)$.

Often an evolution in the state of the system can be described by a stochastic process. Such a stochastic process is defined by means of

1) the state space X^* with the σ-field G^* of the M-dimensional Borel sets.

---

1) Throughout this book the properties imposed on the model will be indicated by "Property", followed by a number.

2) a space Ω^* of points ω. The points ω are called realizations, while the space Ω^* is named the sample space. The space Ω^* is chosen in such a way that a 1-1 correspondence exists between points $\omega \varepsilon \Omega^*$ and elements $x=x^*(t)$ of the collection of all walks in X^*. [2)]

3) a family of ω-functions $\{x_t^*(\omega);\ t \varepsilon [0,\infty)\}$. Each ω-function $x_t^*(\omega)$ maps Ω^* into the space X^*. These functions are called sample functions and are defined as follows:
If the walk $x=x^*(t)$ corresponds to ω (cf. point 2), the t-function $x_t^*(\omega)$ satisfies

$$x_t^*(\omega) = x^*(t) \quad ; \qquad t \varepsilon [0,\infty). \qquad (1.1)$$

4) the smallest σ-field H^* of ω-sets with respect to which the ω-functions $x_t^*(\omega)$ are measurable. Consequently, if $A \varepsilon G^*$ and if $t \varepsilon [0,\infty)$, then

$$\{\ \omega \mid x_t^*(\omega) \varepsilon A\} \varepsilon H^*. \qquad (1.2)$$

5) a probability measure $P^*[\Lambda]$ of sets $\Lambda \varepsilon H^*$.

The triple $\{\Omega^*; H^*; P^*\}$ is called a Probability space.

The stochastic process is defined as the family of random variables $\{\underline{x}_t^*\ ;\ t \varepsilon [0,\infty)\}$, whose probability distributions are given by

$$\text{Prob}\ \{\underline{x}_t^* \varepsilon A\} = P^*[\{\omega | x_t^*(\omega) \varepsilon A\}] \quad ; \qquad A \varepsilon G^*. \qquad (1.3)$$

If the stochastic process $\{\underline{x}_t^*;\ t \varepsilon [0,\infty)\}$ describes the behaviour of the system, the left hand side of (1.3) denotes the probability that at time t the system will be in a state of A. Such a stochastic process is also called a random walk in X^*.

If A is any closed set in X^* and if we consider the walk represented by the point $\omega \varepsilon \Omega^*$, let $t(\omega;A)$ be the moment that the system is for the first time in A and let $t(\omega; [A])$ be the time of the first entry into A.

---

2) In this chapter only walks of infinite length are considered.

If the initial state $x_o^*(\omega)$ does not belong to A, we obviously have

$$t(\omega;A) = t(\omega;[A]) \quad . \tag{1.4}$$

Next let us introduce the states $x^*(\omega;A)$ and $x^*(\omega;[A])$, defined by

$$x^*(\omega;A) \overset{\text{def}}{=} \begin{cases} x^*_{t(\omega;A)}(\omega) ; & \text{if } t(\omega;A) < \infty \\ x_o^*(\omega) ; & \text{if } t(\omega;A) = \infty \end{cases} \tag{1.5}$$

$$x^*(\omega;[A]) \overset{\text{def}}{=} \begin{cases} x^*_{t(\omega;[A])}(\omega) ; & \text{if } t(\omega;[A]) < \infty \\ x_o^*(\omega) ; & \text{if } t(\omega;[A]) = \infty \end{cases} . \tag{1.6}$$

In this chapter we shall consider different families of probability measures. In each family $\{P^*[\Lambda ;x];\ x \in X^*\}$ we find for each initial state x of the random walk one probability measure. This probability measure has the following properties:

1) If the ω-set $\Lambda_{t;B}$ is defined by

$$\Lambda_{t;B} \overset{\text{def}}{=} \{\omega | x_t^*(\omega) \in B\} \tag{1.7}$$

and if x is the initial state of the random walk, then

$$P^* [\Lambda_{o;\{x\}} \ ; x] = 1 \ , \tag{1.8}$$

where $\{x\}$ denotes the set consisting of x only.

2) The domain of definition of $P^* [\Lambda;x]$ can be extended to a σ-field F^* with respect to which the ω-functions $t(\omega;A)$, $t(\omega;[A])$; $x^*(\omega;A)$ and $x^*(\omega;[A])$ are measurable.

3) If the ω-set $\Xi_{I;A}$ is defined by

$$\Xi_{I;A} \overset{\text{def}}{=} \{\omega | t(\omega;A) \in I \} \tag{1.9}$$

and if

$$P^* [\Xi_{[0,\infty);A} ; x] = 1 \ , \tag{1.10}$$

then almost surely $x^*(\omega; A) \in A$.

4) If the ω-set $\Xi_{I;[A]}$ is defined by

$$\Xi_{I;[A]} \stackrel{def}{=} \{\omega \mid t(\omega;[A]) \varepsilon I\} \tag{1.11}$$

and if

$$P^{*}[\Xi_{[0,\infty);[A]};x] = 1, \tag{1.12}$$

then almost surely $x^{*}(\omega;[A]) \varepsilon A$.

The following three situations will be of importance for our discussion:

I) The initial part of the walk $x^{*}(t)$ up to and including t_o is defined as

$$x = x^{*}(t); \qquad t \varepsilon [0,t_o] \tag{1.13}$$

Suppose we want to describe the rest of this walk from t_o onwards. Then, if $\omega \varepsilon \Omega^{*}$ corresponds to the whole walk, by the definition of Ω^{*} a point $\omega_1 \varepsilon \Omega^{*}$ can be found such that

$$x_t^{*}(\omega_1) = x_{t+t_o}^{*}(\omega); \qquad t \varepsilon [0,\infty). \tag{1.14}$$

The relation (1.14) represents a mapping of Ω^{*} on Ω^{*}. This mapping will be denoted by

$$\omega_1 = T_{t_o}(\omega). \tag{1.15}$$

II) If A is a closed set in X^{*} and if t_A is the moment that the system is for the first time in A, let the initial part of the walk up to and including $t_A < \infty$ be defined as

$$x = x^{*}(t); \qquad t \varepsilon [0;t_A]. \tag{1.16}$$

Suppose we want to describe the rest of this walk from t_A onwards. Then, if ω corresponds to the whole walk, by the definition of Ω^{*} a point $\omega_1 \varepsilon \Omega^{*}$ can be found such that

$$x_t^{*}(\omega_1) = x_{t+t(\omega;A)}^{*}(\omega); \quad t \varepsilon [0,\infty). \tag{1.17}$$

If $t_A = \infty$, we define ω_1 by

$$\omega_1 = \omega. \tag{1.18}$$

The relations (1.17) and (1.18) represent a mapping of Ω^{*} on Ω^{*}. This mapping will be denoted by

$$\omega_1 = T_A(\omega). \tag{1.19}$$

III) If A is a closed set in X^* and if $t_{[A]}$ is the time of first entry into A, let the initial part of the walk up to and including $t_{[A]} < \infty$ be defined as

$$x = x^*(t); \qquad t \in [0, t_{[A]}]. \tag{1.20}$$

Suppose we want to describe the rest of this walk from $t_{[A]}$ onwards. Then, if ω corresponds to the whole walk, by the definition of Ω^* a point $\omega_1 \in \Omega^*$ can be found such that

$$x^*_t(\omega) = x^*_{t+t(\omega;[A])}(\omega); \qquad t \in [0, \infty). \tag{1.21}$$

If $t_{[A]} = \infty$, we define ω_1 by

$$\omega_1 = \omega . \tag{1.22}$$

The relations (1.21) and (1.22) represent a mapping of Ω^* on Ω^*. This mapping will be denoted by

$$\omega_1 = T_{[A]}(\omega). \tag{1.23}$$

Convention

If the remaining part [3] of a walk has to be described in this study, it is tacitly assumed that

a) the time axis has been shifted in such a way that its origin coincides with the new starting point;

b) in accordance with the situation concerned one of the relevant point transformations (1.15), (1.19) and (1.23) has been executed.

Since each complete walk $x = x^*(t)$ in X^* can be identified with a point $\omega \in \Omega^*$, we can state that

---

3) If $t(\omega;A) = \infty$ or if $t(\omega;[A]) = \infty$, by (1.18) and (1.22) the corresponding "remaining parts" are the whole walk.

1) in the first situation the initial part of the walk $x = x^*(t)$, with $t \in [0,t_o]$, can also be represented by means of the point $(\omega,t_o) \in \Omega^* \times T$, where $T = [0,\infty)$;
2) in the second situation the initial part of the walk $x = x^*(t)$, with $t \in [0,t_A]$, can also be represented by means of the point $(\omega,t(\omega;A)) \in \Omega^* \times T$;
3) in the third situation the initial part of the walk $x = x^*(t)$, with $t \in [0,t_{[A]}]$, can also be represented by means of the point $(\omega,t(\omega;[A])) \in \Omega^* \times T$.

Since different realizations ω may have a common initial part, the representation of that initial part by points like (ω,t_o), $(\omega,t(\omega;A))$ and $(\omega,t(\omega;[A]))$ is not unique.

If the whole random walk has been defined by means of the probability space $\{\Omega^*;F^*;P^*\}$, if the initial part of that random walk is given and if the remaining part can be defined, then this part can be described by means of a probability space $\{\Omega^*;H^*;\overrightarrow{P}^*\}$, where $\overrightarrow{P}^*$ is an appropriate probability measure defined on H^*. The use of Ω^* and H^* for this description is a consequence of the convention.

We further introduce the terms past, present and future in the following sense:

a) The past $Pa^{\omega;t'}$ of the system at t' is given by

$$x = x_t^*(\omega) \;; \qquad t \in [0,t'] \;. \tag{1.24}$$

Thus the past $Pa^{\omega;t'}$ is a realized walk in X^* up to and including t'.

b) The present $Pr^{\omega;t'}$ of the system at t' is given by

$$x = x_{t'}^*(\omega). \tag{1.25}$$

Thus the present $Pr^{\omega;t'}$ is the state of the system at t'; i.e. a point of X^*.

c) The future $Fu^{\omega;t'}$ of the system at t' is given by the probability measure

$$\overline{P}^*[\Lambda] = P^*[\Lambda | Pa^{\omega;t'}]; \qquad \Lambda \in H^*. \qquad (1.26)$$

<u>Thus the future</u> $Fu^{\omega;t'}$ <u>of the system is a probability measure defined on</u> H^*.

Note that:

1) both present and future are determined by the corresponding past;
2) different pasts may generate identical presents and identical futures.

In this chapter the time t' is either a fixed time t_o or given by $t(\omega;A)$ or $t(\omega;[A])$, where A is a closed set in X^*.

If the initial part of the random walk is unknown, the future $P^*[\Lambda | Pa^{\omega;t'}]$ is an ω-function. In chapter 1 of part II this ω-function will be called a <u>conditional probability measure</u>.

A family of stochastic processes $\{S_x^*; x \in X^*\}$, one for each initial state, is called a <u>stationary Markov process</u> if the corresponding probability measures $\{P^*[\Lambda;x];\ x \in X^*\}$ have the following properties:

1) the x-function $P^*[\Lambda;x]$ is for each $\Lambda \in F^*$ measurable with respect to G^* ;
2) for each t_o and for each past $Pa^{\omega;t_o}$ we have

$$P^*[\Lambda;x \,|\, Pa^{\omega;t_o}] = P^*[\Lambda; Pr^{\omega;t_o}] ; \qquad \Lambda \in H^*,\ x \in X^*. \quad (1.27)$$

It is called, in this book, a <u>stationary strong Markov process</u> if, in addition to 1) and 2):

3) for each closed set A in X^*, satisfying for each $x \in X^*$

$$P^*[\Xi_{[0,\infty);[A]};x] = 1 , \qquad (1.28)$$

and for each past $Pa^{\omega;t(\omega;[A])}$ we have

$$P^*[\Lambda;x \,|\, Pa^{\omega;t(\omega;[A])}] = P^*[\Lambda; Pr^{\omega;t(\omega;[A])}];$$

$$\Lambda \in H^*,\ x \in X^*. \qquad (1.29)$$

(If $t(\omega;[A]) = \infty$, we take $Pr^{\omega;t(\omega;[A])} = x$.)

Physical systems will be considered, which change their states even if no decisions are made.

If no decisions are made, the evolution in the state of the system is called a natural process.

In a natural process the state of the system can at each point of time be uniquely represented by a point x of an N-dimensional Cartesian space.

Property 2 (natural process)

1) In the mathematical model a natural process is defined by means of a stochastic process.

Notations:

$$M = N,\quad X^{*} = X_{o},\quad \Omega^{*} = \Omega_{o},\quad G^{*} = G_{o},\quad x^{*}(t) = x^{o}(t),$$
$$x^{*}_{t}(\omega) = x^{o}_{t}(\omega),\ H^{*} = H_{o} \text{ and } F^{*} = F_{o}.$$

In this study we consider a family of natural processes; i.e. one for each initial state of the system. If $x_{o} = x^{o}(0)$ is the initial state of the system, the probability measure corresponding to the natural process concerned is denoted by $P_{o}[\Lambda;x_{o}]$. The natural process is defined by means of the ω-functions $\{x^{o}_{t}(\omega); t \in [0,\infty)\}$ and the probability space

$$\{\Omega_{o};H_{o};P_{o}[\Lambda;x^{o}(0)]\} \qquad ; \qquad \Lambda \in H_{o}\ .$$

Property 2 (natural process)

2) Almost all walks $x^{o}_{t}(\omega)$ are continuous from the right in t. In each finite time interval almost all t-functions $x^{o}_{t}(\omega)$ have only a finite number of discontinuities.

This property enables us to extend the domain of definition of the set-function $P_{o}[\Lambda;x]$ from H_{o} to F_{o}. (Cf. Part II, chapter 1). Thus, if A is a closed set in X_{o}, the ω-functions $t(\omega;A)$, $t(\omega;[A])$, $x(\omega;A)$ and $x(\omega;[A])$ are measurable with respect to F_{o}.

Property 2 (natural process)

3) In the mathematical model the natural processes are defined by means of stationary strong Markov processes.

2. Decisions and losses

The natural process, however, is not the only source of changes in the state of the system. For instance, if in a replacement problem at some point of time the decisionmaker decides to replace an old machine by a new one, such a decision certainly affects the state of the system. If the initial states of a new machine are not always identical, then the decision to buy anyone of a set of new machines will be represented in the mathematical model by a random transition in X_o.

Property 3 (decisions)

1) In the mathematical model a decision d is a random transition in X_o. This transition is defined by the probability distribution

$$Pd\,[A]\;; \qquad\qquad A \in G_o$$

of the state into which the system will be transferred at the moment of the decision.

So decisions are defined independently of the state at the moment of decision. They only refer to the state into which the system is transferred. A transition is assumed to take no time. Consequently, the system will be in two states at the moment of a decision.

It is convenient to assume that at each point of time a decision is made, but that only some of these decisions lead to an intervention in the natural process. In this study we shall make a distinction between "interventions" on the one hand and "null-decisions" on the other hand. In the latter case the system is transferred with probability 1 into its present state.

Decisions (probability distributions in X_o) will be represented by points d of a so called decision space D.

Property 3 (decisions)

2) In the mathematical model a decision can be represented by a point d of a K-dimensional Cartesian space D with the following properties:

a) If $p_1 > 0$, $p_2 > 0$, $p_1+p_2 = 1$ and if the probability distributions $Pd_1[A]$ and $Pd_2[A]$ correspond to the points d_1 and d_2 respectively, then the point $d = p_1 d_1 + p_2 d_2$ corresponds to the probability distribution (decision) $\sum_{i=1}^{2} p_i Pd_i[A]$.

b) If the sequence of points $\{d_i ; i=1,2,...\}$ converges to a point d and if this sequence of points corresponds to the sequence of probability distributions $\{Pd_i[A] ; i=1,2,...\}$, then for each $A \varepsilon G_o$ we have

$$\lim_{i \to \infty} Pd_i[A] = Pd[A] . \qquad (1.30)$$

Because decisions are concerned with the states into which the system is transferred, it follows from the physical structure of many decision problems that in some states certain decisions are not feasible. The decisionmaker may be restricted in his choice of a decision.

Property 3 (decisions)

3) Whether a decision is feasible or not, depends only on the state of the system at the moment of the decision.

4) For each $x \varepsilon X_o$ the set of feasible decisions denoted by D(x) is a bounded, closed and convex set in D. [4)]

5) In each state $x \varepsilon X_o$, the "null decision" is feasible.

A realization of an intervention, i.e. a random drawing from the probability distribution $Pd[A]$, will be called an intervention-transition.

---

4) Randomisations of feasible decisions are also feasible.

In the mathematical model we have now stipulated what happens at the moment of a decision and how the behaviour of the system can be described by a natural process. We have still to state how the behaviour of the system is to be represented when interventions take place. To this end we introduce the following property:

Property 4

In the mathematical model the behaviour of the system in each time interval between two interventions is described by a natural process. The initial state of that process will be the state into which the system is transferred by the intervention at the beginning of the interval concerned.

As we have stated already, the state of the system at the moments of intervention is not uniquely defined in X_o. Therefore we introduce a product space X' of two spaces X_1 and X_2, both congruent to X_o. So we have

$$X' = X_1 \times X_2 \ . \tag{1.31}$$

The points of X' are represented by $x' = (x_1, x_2)$. The space X' is a 2N-dimensional Cartesian space; thus $M = 2N$.

At each point of time the x_1-component fixes the state of the system before the decision is made, while the x_2-component describes the state at the same moment but now after the intervention-transition has been effected.

If only one decision is made at a time and if the space X' is used instead of X_o for representing the state of the system, then this state is again defined unambiguously at each point of time. At the moments of a null-decision we have $x_1 = x_2$.

Let a space Ω of points ω be chosen in such a way that a 1-1 correspondence exists between points $\omega \in \Omega$ and the elements $x' = x(t)$ of the collection of all walks in X'. Let $\{x_t(\omega);\ t \in [0,\infty)\}$ be a family of ω-functions defined on Ω .

These functions are defined as follows: If the walk $x' = x(t)$ corresponds to ω, the t-function $x_t(\omega)$ satisfies

$$x_t(\omega) = x(t) \; ; \qquad t \in [0,\infty) . \qquad (1.32)$$

Henceforth the x_1- and the x_2- component of the function $x_t(\omega)$ will be denoted by $x_{t;1}(\omega)$ and $x_{t;2}(\omega)$ respectively.

The natural process can also be defined in the space X' instead of X_o. If the system is subjected to a natural process in X', at each point of time we have $x_1 = x_2$.

For the time being we shall consider the product-state space $X' = X_1 \times X_2$ only.

<u>Notations 2</u>:

$$M = 2N, \qquad X^* = X', \quad \Omega^* = \Omega \;, \quad G^* = G',$$

$$x_t^*(\omega) = x_t(\omega) = (x_{t;1}(\omega), \, x_{t;2}(\omega)), \qquad x^*(t) = x(t),$$

$$H^* = H \quad \text{and} \quad F^* = F.$$

Let us suppose that, even if interventions are made, the behaviour of the system can still be described by a stochastic process. Then there must be a probability space $\{\Omega;H;P\}$.

For the state space X' we now give, in addition to the above definitions, special definitions of past, present and future as follows:

a') The past $Pa_1^{\omega;t'}$ of the system at t' is given by

$$\begin{aligned} x_1 &= x_{t;1}(\omega) \quad ; \qquad t \in [0,t'] \\ x_2 &= x_{t;2}(\omega) \quad ; \qquad t \in [0,t'). \end{aligned} \qquad (1.33)$$

<u>Thus the past $Pa_1^{\omega;t'}$ is a realized walk in X' up to t' and including the x_1-state at t'</u>.

b') The present $Pr_1^{\omega;t'}$ of the system at t' is given by

$$x_1 = x_{t';1}(\omega). \qquad (1.34)$$

Thus the present $Pr_1^{\omega;t'}$ is the x_1-state of the system at t'; i.e. a point of X_1.

c') The future $Fu_1^{\omega;t'}$ of the system at t' is given by the probability measure

$$\bar{P}[\Lambda] = P[\Lambda \mid Pa_1^{\omega;t'}] \quad ; \qquad \Lambda \in H \qquad (1.35)$$

Thus the future $Fu_1^{\omega;t'}$ of the system is a probability measure defined on H.

From t' onwards the remaining part of the random walk in X' is defined by means of the probability space $\{\Omega;H;\bar{P}\}$. Note that the new initial x_2-state may have an initial probability distribution. This probability distribution corresponds to the decision to be made at t'.

The most important features of physical decision problems are losses and gains. It will be no restriction to suppose that only losses occur. Gains are negative losses. Generally in decision problems three types of losses occur.

First, losses which increase or decrease continuously in the course of time; e.g. (loss of) interest or consumption of fuel.

Secondly, losses which increase or decrease at discrete points of time; e.g. owing to sales or repairs.

Finally, losses which are effected by decisions.

Let us consider how these losses are to be defined in the mathematical model. This is done by means of three functions.

The first kind of costs can be represented by an x'-function $\gamma_{cont}(x')$. It represents the losses of the first type that would be suffered if the system were in the state x' during one unit of time. The x'function $\gamma_{cont}(x')$ is called the "loss density function".

The x'-function $\gamma_{disc}(x')$ fixes the losses of the second type incurred in x', if in that state the system enters a given closed set A_γ. Each time the system enters A_γ, losses of this type will be suffered. The function $\gamma_{disc}(x')$ is called the "discrete loss function".

The third function is the (d,x')-function $\gamma_{dec}(d;x')$; it represents the costs incurred if the decision d made in x_1 leads to a transition into x_2. The function $\gamma_{dec}(d;x')$ is called the "decision cost function".

Property 5 (costs)

1) The loss density function $\gamma_{cont}(x')$ is bounded, real valued and continuous in $x' \in X'$;
2) The discrete loss function $\gamma_{disc}(x')$ is bounded, real valued and measurable with respect to the σ-field G'. The set A_γ is a closed set in X'. For each initial state x' in the natural process almost surely there will be a finite number of entries into A in a finite time interval. If $x' \in \bar{A}$, we have $\gamma_{disc}(x')=0$;
3) The decision cost function $\gamma_{dec}(d;x')$ is a bounded, real-valued function of (d,x'). Moreover, for each $x' \in X'$ it is a continuous function of d. For each $d \in D$ the x'-function $\gamma_{dec}(d;x')$ is measurable with respect to G'. For null-decisions we have $\gamma_{dec}(d;x') = 0$.

This property implies the following statement:

Statement no 1

In each time interval the losses incurred are completely fixed by

a) the walk made by the system in that time interval.
b) the interventions made by the decision maker in that period.

The losses are independent of the position of the time interval on the time axis.

3. Strategies and decision processes

The solution of the stochastic ∞-stage decision problem is given in the form of a strategy. Such a strategy dictates at each point of time a feasible decision.

If a strategy is applied, the evolution in the state of the system can still be described by a walk in X'. Thus the past of the system can be defined at each point of time.

The past does not necessarily include all information about the decisions made. It does, however, inform us about the realization of the past interventions, because at those moments $x_1 \neq x_2$. Thus the past describes everything which has really happened to the system. It is therefore reasonable to restrict ourselves to the strategies of which the decisions are based on the past $Pa_1^{\omega;t}$ only.

Consequently, each strategy z to be considered maps the space $\Omega \times T$ into the decision space D.

After introducing some properties of the strategies to be considered, we shall show that there is sense in restricting ourselves to strategies dictating decisions based on the present $Pr_1^{\omega;t}$ only. The reader who is willing to accept this statement without comment, may pass over these considerations and can take up the discussion again just after the proof of lemma 1.2.

Property 6* (strategies)[5]

1) Whether or not it has been applied before, a strategy z at any point of time t dictates a feasible decision d dependent on the past $Pa_1^{\omega;t}$ only. Notation: $d = z(Pa_1^{\omega;t})$.

If decisions are made in accordance with a given strategy, very often the evolution in the state of the system can still be described by means of a random walk.

If $\bar{\omega}$, $\omega' \in \Omega$, let the point transformation in Ω

$$\omega = T_{\bar{\omega};t_o}(\omega')$$

be defined by (cf. (1.32))

$$\begin{aligned} x_t(\omega) &= x_t(\bar{\omega}); \qquad && t \in [0,t_o) \\ x_t(\omega) &= x_{t-t_o}(\omega'); \qquad && t \in [t_o,\infty) \end{aligned} \quad . \qquad (1.36)$$

If at t_o the past $Pa_1^{\bar{\omega};t_o}$ is given and if ω' corresponds to the realization of the random walk from t_o onwards (cf. convention), then

5) This property will be reformulated at the end of this chapter.

$\omega = T_{\overline{\omega};t_o}(\omega')$ represents the complete walk.

Note that $\omega' = T_{t_o}(\omega)$ (cf. (1.15)).

Notation 3:

The remaining part of the random walk to be considered starts at t_o and can be identified with the point ω'. The past at t_o is given by $(\overline{\omega}, t_o)$ and is denoted by $Pa_1^{\overline{\omega};t_o}$. The time t_o is not always a fixed point of time.

Hence, if z is the strategy applied during the whole walk and if the past at t_o is given by $Pa_1^{\overline{\omega};t_o}$, then after t_o (convention!) the strategy $z_{\overline{\omega};t_o}$ given by

$$z_{\overline{\omega};t_o}(Pa_1^{\omega';t}) \overset{def}{=} z(Pa_1^{T_{\overline{\omega};t_o}(\omega');t+t_o}), \qquad (1.37)$$

will evidently be used.

If strategy z is applied, if $\hat{x}_1$ is the initial x_1-state of the system and if the evolution in the state of the system can be described by a random walk, the appropriate probability measure is denoted by $P[\Lambda;z;\hat{x}_1]$.

If $Pa_1^{\overline{\omega};t_o}$ is the past at t_o and if the remaining past of the random walk starting at t_o can be described by a stochastic process, the corresponding probability measure is denoted by

$$P[\Lambda;z;\hat{x}_1 | Pa_1^{\overline{\omega};t_o}].$$

Note that the future also depends on the strategy applied. Consequently, we introduce the notation $Fu_1^{\overline{\omega};t_o}(z)$.

Let us consider a class Z_o of strategies z satisfying:

Property 6* (strategies)

2) If a strategy $z \in Z_o$ is applied, for each initial x_1-state $\hat{x}_1$ the evolution in the state of the system can be described by means of a random walk. The corresponding probability measure will be denoted by $P[\Lambda;z;\hat{x}_1]$.

operty 6

3) If $z \in Z_o$, for each $\overline{\omega}$, for each t_o and hence for each $Pa_1^{\overline{\omega};t_o}$ have

a) $z_{\overline{\omega};t_o} \in Z_o$

b) $P[\Lambda;z;\hat{x}_1|Pa_1^{\overline{\omega};t_o}]$ exists and satisfies (cf. notation 3)

$$P[\Lambda;z;\hat{x}_1|Pa_1^{\overline{\omega};t_o}] = P[\Lambda;z_{\overline{\omega};t_o};Pr_1^{\overline{\omega};t_o}] =$$

$$= P[\Lambda;z_{\overline{\omega};t_o};Pr_1^{\omega';0}] \quad . \qquad (1.38)$$

4) If $z \in Z_o$, for each t and for each ω the strategy z_t, given by f. (1.15))

$$z_t(Pa_1^{\omega;t}) \overset{def}{=} z(Pa_1^{T_t(\omega);0}) \qquad (1.39)$$

an element of Z_o.

5) If z_1 and $z_2 \in Z_o$, the strategy

$$z = T^1_{z_1;t_o}(z_2),$$

fined by

$$z(Pa_1^{\omega;t}) \overset{def}{=} \begin{cases} z_1(Pa_1^{\omega;t}); & \text{if } t \in [0,t_o) \\ z_2(Pa_1^{\omega;t}); & \text{if } t \in [t_o,\infty) \end{cases} \qquad (1.40)$$

so belongs to Z_o.

6) If z_1 and $z_2 \in Z_o$, the strategy

$$z = T^2_{z_1;t_o}(z_2),$$

fined by

$$z(Pa_1^{\omega;t}) \overset{def}{=} \begin{cases} z_1(Pa_1^{\omega;t}); & \text{if } t \in [0,t_o) \\ z_2(Pa_1^{T_{t_o}(\omega);t-t_o}); & \text{if } t \in [t_o,\infty) \end{cases} \qquad (1.41)$$

so belongs to Z_o.

Point 3^b) implies for $t \geq t_o$

$$Fu_1^{\overline{\omega};t}(z) = Fu_1^{\omega';t-t_o}(z_{\overline{\omega};t_o}) \ .^{6)} \qquad (1.42)$$

The strategy $z = T^1_{z_1;t_o}(z_2)$ can be described in words by saying that z_1 is applied before t_o and z_2 from t_o onwards.

The strategy $z = T^2_{z_1;t_o}(z_2)$ can be described in words by saying that z_1 is applied before t_o and z_2 from t_o onwards <u>neglecting</u> the walk made before t_o.

If $z \in Z_o$ is the strategy applied, then by point 3^b of property 6^* from t_o onwards the evolution in the state of the system can be described by a random walk in X'. The corresponding probability measure is given by $P[\Lambda; z_{\overline{\omega};t_o}; Pr_1^{\overline{\omega};t_o}]$.

Now we shall prove the following lemma:

<u>Lemma 1.1</u>

For each strategy $z \in Z_o$ and for each past $Pa_1^{\omega;t}$ with $t \geq t_o$, the decision $d=z(Pa_1^{\omega;t})$ can be deduced from the set function

$$P[\Lambda; z_{\overline{\omega};t_o}; Pr_1^{\overline{\omega};t_o}] .$$

<u>Proof</u>.

If $z \in Z_o$, by point 3 of property 6^* the set function $P[\Lambda; z_{\overline{\omega};t_o}; Pr_1^{\overline{\omega};t_o} | Pa_1^{\omega';t-t_o}]$ is defined for each ω' and t.

If the past at t is given by $Pa_1^{\omega;t}$, the set function $P[\Lambda; z_{\overline{\omega};t_o}; Pr_1^{\overline{\omega};t_o} | Pa_1^{\omega';t-t_o}]$ determines the initial x_2-probability distribution of the random walks that start at t. This initial probability distribution represents the decision to be made at t. Hence from t_o onwards for each given $Pa_1^{\overline{\omega};t_o}$ the probability measure $P[\Lambda; z_{\overline{\omega};t_o}; Pr_1^{\overline{\omega};t_o}]$ reproduces the strategy z (and $z_{\overline{\omega};t_o}$).

This ends the proof.

---

6) This means that the past minus the present influences the future only through the strategy applied.

Suppose we want to compare strategies. Then, because of statement 1 and lemma 1.1, it is reasonable that we restrict ourselves to criterion functions which depend only on the future $Fu_1^{\bar{\omega};t_o}(z)$, given by $P[\Lambda; z_{\bar{\omega};t_o}; Pr_1^{\bar{\omega};t_o}]$.

In connection with this we introduce the following property:

Property 7* (criterion function) [7)]

1) The criterion function $c(Fu_1^{\omega;t}(z))$ is a function of probability measures, defined on H. The domain of definition is a class K of probability measures.

2) If during the first part of a random walk a strategy $z_1 \varepsilon Z_o$ has been applied, if for the remaining part that starts at t_o two strategies z_2 and z_2' are under consideration and if

$$c(Fu_1^{\bar{\omega};t_o}(z)) < c(Fu_1^{\bar{\omega};t_o}(z')), \tag{1.43}$$

where $z = T^2_{z_1;t_o}(z_2)$ and $z' = T^2_{z_1;t_o}(z_2')$, then at t_o strategy z is to be preferred.

Property 6* (strategies)

7) If $z \varepsilon Z_o$, for each $\bar{\omega}$ and for each t_o we have

$$Fu_1^{\bar{\omega};t_o}(z) \ \varepsilon \ K. \tag{1.44}$$

Definitions:

1) If $\hat{z} \varepsilon Z_o$, the class $Z_o^{\hat{z};\hat{t}}$ exists of strategies $z \varepsilon Z_o$ satisfying for each ω and $t < \hat{t}$

$$z(Pa_1^{\omega;t}) = \hat{z}(Pa_1^{\omega;t}). \tag{1.45}$$

2) A strategy z_o is called optimal from t_o onwards if, for $t \geq t_o$,

$$c(Fu_1^{\omega;t}(z_o)) = \min_{z \varepsilon Z_o^{z_o;t}} c(Fu_1^{\omega;t}(z)). \tag{1.46}$$

---

7) This property will be dropped later.

3) A <u>strategy</u> z_o is called <u>optimal</u>, if it is optimal from t=0 onwards.

4) A <u>decision</u> d, made at t_o and based on the past $Pa_1^{\bar{\omega};t_o}$, is called <u>optimal</u> if there exists a strategy z_o which is optimal from t_o onwards and which satisfies

$$d = z_o(Pa_1^{\bar{\omega};t_o}) \quad . \tag{1.47}$$

Since more than one strategy may be optimal, at t_o there may exist a number of optimal decisions.

<u>Property 7^*</u> (criterion function)

3) A strategy $z \in Z_o$ that dictates at each point of time an optimal decision, is itself optimal.

<u>Lemma 1.2</u>

If the strategy $z_o \in Z_o$ is optimal, the strategy z_{o_t} is also optimal (cf. (1.39)).

<u>Proof</u>.

Let z_o^* be defined by

$$z_o^* = T^2_{z_o;t_o}(z_o), \tag{1.48}$$

where t_o is a fixed point (cf. (1.41)).

It follows from (1.48), (1.37), (1.41) and notation 3 that

$$z^*_{o_{\bar{\omega};t_o}} = z_o \quad . \tag{1.49}$$

Consequently, by (1.42), (1.48) and point 6 of property 6^* we find

$$c(Fu_1^{\omega;t}(z_o^*)) = c(Fu_1^{\omega';t-t_o}(z^*_{o_{\bar{\omega};t_o}})) = c(Fu_1^{\omega';t-t_o}(z_o)) =$$

$$= \min_{z \in Z_o} z_{o;t-t_o}\, c(Fu_1^{\omega';t-t_o}(z)) = \min_{z \in Z_o} z^*_{o;t}\, c(Fu_1^{\omega;t}(z)),$$

if $t \geqq t_o$. (1.50)

Thus from t_o onwards the strategy z_o^* dictates optimal decisions. It follows from (1.48) that the strategy z_o^* also dictates optimal decisions before t_o. Hence, z_o^* is an optimal strategy. In particular, the decision to be dictated at t_o is optimal. This decision will also be dictated by the strategy z_{o_t}. Because (1.50) is true for each t_o, the strategy z_{o_t} always dictates optimal decisions. Hence z_{o_t} is optimal.

This ends the proof.

If a strategy z_{o_t} is applied, then the decisions to be made depend only on the present $Pr_1^{\omega;t}$. We emphasize that the proof of lemma 1.2 is based among other things on the rather complex conditions formulated in point $3^{b)}$ of property 6^* and on point 3 of property 7^*. The latter seems acceptable and agrees with our practical notion of optimality. At first sight point 3^b of property 6^* seems to be a consequence of the strong Markov property of the natural process on the one hand and of property 4 on the other hand. However, this does not seem to be true; possibly condition 3^b of property 6^* might be broken down into a number of less complex conditions, but for two reasons we have not tried to do so. First the practical implications of the condition are clear enough from its present formulation. Secondly, after introducing a reformulation of property 6^*, we shall be able to prove this condition for a smaller class of strategies (theorem 3).

Because of the lemma just proved, it is reasonable to restrict ourselves to the class Z of strategies z with the following property:

Property 6 (strategies)

1) Each strategy $z \in Z$ maps the state space X_1 into the decision space D.

This relation between x_1-states and decisions will be denoted by

$$d = z(x_1). \tag{1.51}$$

Consequently, strategies $z \in Z$ divide the state space X_1 into two disjunct sets, one denoted by A_z and comprising states in which always interventions are made, the other consisting of the states in which always null-decisions are made. The set A_z is called the <u>intervention set</u>.

<u>Property 6</u> (strategies)

2) The x_1-function $d = z(x_1)$ is measurable with respect to G_1.

3) The x'-function $\gamma_{dec}(z(x_1);x')$ is measurable with respect to G'.

4) For each strategy $z \varepsilon Z$ the intervention set A_z is a closed set.

5) The intersection A_o of all sets A_z $(z \varepsilon Z)$ is not empty and satisfies for $x_1 \varepsilon X_1$ $\quad P_o\left[\Xi_{[0,\infty);A_o};x_1\right] = 1$ and

$$\int_0^\infty t\,P_o\left[\Xi_{dt;A_o};x_1\right] < \infty. \qquad (1.52)$$

By point 1 of property 6 the strategies $z \varepsilon Z$ map the state space X_1 into the decision space D. We have stipulated that decisions shall correspond to probability distributions in the space X_2. So we can state that each strategy z corresponds to a family of transition probabilities

$$\{\, z(B,x_1);\ x_1 \varepsilon X_1,\ B \varepsilon G_2 \,\}$$

which, for $d = z(x_1)$, are given by

$$z(B;x_1) = P_{z(x_1)}[B] \ ; \qquad B \varepsilon G_2 \quad (1.53)$$

where G_2 is the σ-field of Borelsets in X_2.

<u>Property 6</u> (strategies)

6) If strategy $z \varepsilon Z$, then

a) for each $x_1 \varepsilon A_z$ we have $z(A_z;x_1) = 0$.

b) for each $B \varepsilon G_2$ the x_1-function $z(B;x_1)$ is measurable with respect to G_1.

If one of the strategies $z \varepsilon Z$ is applied, the system is said to be subjected to a decision process.

By lemma 2.24 of part II the decision process can be defined by a stochastic process. Each decision process effects a walk of the system through the state space $X' = X_1 \times X_2$. If z is the strategy applied and if x' is the initial state, then the decision process is denoted by $S_{z;x'}$.

In part II, ch.2-4, we prove that the domain of definition of $P[\Lambda;x_1;z]$ can be extended to a σ-field F, with respect to which the ω-functions $t(\omega;A)$, $t(\omega;[A])$, $x(\omega;A)$ and $x(\omega;[A])$ are measurable. From now on we shall use the probability space $\{\Omega;F;P\}$.

Let us consider the sequence of stochastic x_1-states $\{\underline{I}_j;\ j=1,2,\dots\}$ at the moments of intervention. Since the set A_z is a closed set in X_1, it follows from point 2 of property 2, property 4 and (1.52) that these states almost surely belong to A_z.

In theorem 1 of chapter 2 in Part II we prove

Theorem 1

If a strategy $z \varepsilon Z$ is applied, the sequence of states $\{\underline{I}_j;\ j=1,2,\dots\}$ at the moments of intervention can be described by a stationary Markov process in A_z with a discrete time parameter.

Property 6 (strategies)

7) For each strategy $z \varepsilon Z$ the Markov process in A_z with discrete time parameter satisfies the Doeblin condition (cf.[2], p.192).

Point 7 of property 6 implies that the stationary Markov process in A_z has for each initial state a stationary absolute probability distribution (cf.[2], p.192 ff.).

In theorem 2 of chapter 2 in part II we prove

Theorem 2.

If a strategy $z \varepsilon Z$ is applied, then in each finite time interval almost surely only a finite number of interventions will be made.

We also prove in that chapter

<u>Theorem 3</u>

If a strategy $z \in Z$ is applied, the decision process in X' can be described by means of a stationary strong Markov process.

From theorem 3 we can deduce

$$P[\Lambda;z;\hat{x}_1 | Pa_1^{\overline{\omega};t_o}] = P[\Lambda;z;Pr_1^{\overline{\omega};t_o}]. \qquad (1.54)$$

If $z \in Z$, then we can easily verify that $z = z_{\overline{\omega};t_o}$ (cf. (1.37)). Hence, if $z \in Z$, it follows from (1.54) that

$$P[\Lambda;z;\hat{x}_1 | Pa_1^{\overline{\omega};t_o}] = P[\Lambda;z_{\overline{\omega};t_o};Pr_1^{\overline{\omega};t_o}] \quad . \qquad (1.55)$$

Thus, the strategies $z \in Z$ satisfy point 3^b of property 6^*.

CHAPTER 2

A criterion for optimality

1. Introduction

In this chapter we shall construct a criterion function for strategies.

If $z \in Z$ is the strategy applied and if in the initial x_1-state the decision transition has not been effected, then the x_2-component obeys the initial distribution $z(B;x_1)$. Decision processes with such an initial distribution are denoted by $\{S_{z;x_1};x_1 \in X_1\}$. They are defined by means of stationary strong Markov processes in X' with initial probability distributions (cf. theorem 3).

If only the x_1-states of the decision process $S_{z;x_1}$ are recorded, the random walk in X_1 also is a stationary strong Markov process. This process is called the decision process $S_{z;x_1}$ in X_1.

In chapter 1 we have determined the way in which losses enter the model. Obviously, the choice of a strategy depends on these losses in one way or another.

If $z \in Z$ is the strategy applied and if ω denotes a realization of the decision process $S_{z;x_1}$, let $k_T(\omega;z)$ be the costs incurred during the period $[0,T)$.

Using certain additional properties of the mathematical model, we shall prove that for almost all $\omega \in \Omega$

$$\lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \tag{2.1}$$

exists.

Moreover, if the set E in X_1 is a simple ergodic set [1] of the decision process $S_{z;x_1}$ in X_1 and if $\omega_1,\omega_2 \in \Omega$ satisfy

$$x_{o;1}(\omega_i) \in E;\ i=1,2,$$

---

1) Simple ergodic sets can not be divided into more than one ergodic set; $x_{o;1}(\omega)$ is the x_1-component of $x_o(\omega)$.

then almost surely

$$\lim_{T \to \infty} \frac{k_T(\omega_1;z)}{T} = \lim_{T \to \infty} \frac{k_T(\omega_2;z)}{T} .$$

Consequently, an x_1-function $r(z;x_1)$ can be defined on the set Y of all ergodic states [2] of $S_{z;x_1}$ in X_1 such that, if ω satisfies $x_{o;1}(\omega) = x_1$,

$$r(z;x_1) = \lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \qquad (2.2)$$

for almost all $\omega \in \Omega$.

Note that the x_1-function $r(z;x_1)$ is constant on a simple ergodic set of $S_{z;x_1}$ in X_1. The domain of definition of $r(z;x_1)$ is extended to the whole space X_1 by

$$r(z;x_1) \overset{\text{def}}{=} E\{r(z;\underline{y})|x_1\} \quad , \qquad (2.3)$$

where the random state $\underline{y}$ is the first ergodic state in Y taken on in the decision process $S_{z;x_1}$ in X_1.
Roughly speaking, the x_1-function $r(z;x_1)$ represents for almost all realizations the mean costs per unit of time if the initial x_1-state is ergodic. In case that the initial x_1-state is not ergodic the mean costs depend on the first ergodic state assumed, and, therefore, they are random.

Hence, by (2.3), the function $r(z;x_1)$ determines the expected mean costs per unit of time for all initial x_1-states.

Obviously, the z-function $r(z;x_1)$ is a good criterion for optimality. It will be demonstrated that the function $r(z;x_1)$ can also be expressed in a more usable form.

If C is a closed set in X' of the form

$$C = \hat{C} \times X_2,$$

it can be proved that, under obvious conditions, the sequence of x_1-components of the successive entry states in C constitutes a stationary Markov process with a discrete time parameter. This process is called

---

2) An ergodic state is a state of a simple ergodic set.

the Markov process in $[\hat{C}]$.

Although almost all entry states belong to C, the state space of the Markov process in $[\hat{C}]$ is not $\hat{C}$ but X_1. Thus the first and higher order transition probabilities are defined for each initial x_1-state.

If the Markov process in $[\hat{C}]$ satisfies the Doeblin condition (cf. [2], p.192), then for each initial x_1-state x_1 a stationary absolute probability distribution $p_{[\hat{C}]}(B;x_1;z)$ exists.

Let $S_z^{x_1}$ now denote the decision process in X_1 which starts in a random x_1-state obeying the steady state probability distribution $p_{[\hat{C}]}(B;x_1;z)$.

In section 4 we shall prove that on the class of all finite time intervals the difference in expected costs between the decision processes $S_{z;x_1}$ and $S_z^{x_1}$ is uniformly bounded in x_1.

This result implies that the effect of the initial state on the expected loss is limited to a finite amount if only states of one simple ergodic set in X_1 are considered.

Finally, an x_1-function $c(z;x_1)$ is introduced which in a sense enables us to value the initial state with respect to the total expected loss.

The x_1-functions $c(z;x_1)$ and $r(z;x_1)$ which jointly satisfy a pair of functional equations, are used in an iteration procedure for obtaining optimal strategies (chapter 3).

2. The criterion function

We first consider the decision processes $\{S_{z;x'};x' \varepsilon X\}$. For the time being the state of the system in X' is only recorded at the points of time $\{jt_o;j=0,1,\ldots\}$. They are represented by the stochastic variables

$$\{\underline{x}'_{t_o;x';j};\ j=0,1,\ldots\}\ , \tag{2.4}$$

where $\underline{x}'_{t_o;x';0}$ stands for the initial state x'. Since the decision process $S_{z;x'}$ is a stationary Markov process, the sequence (2.4) constitues a stationary Markov process with a discrete time parameter (cf. part II, chapter 1, lemma 1.37). The first and higher order transition probabilities are denoted by $\{p^j_{t_o}(B;x';z);j=1,2,\ldots\}$.

We now assume:

Assumption 1*

The Markov processes $\{\underline{x}'_{t_o;x';j};j=0,1,\ldots\}$ with $x' \in X'$ satisfy the Doeblin condition (cf. [2], p.192).

Consequently, for each $x' \in X'$ there exists a stationary absolute probability distribution $p_{t_o}(B;x';z)$ that satisfies (cf. [2], p.214)

$$p_{t_o}(B;x';z) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} p^j_{t_o}(B;x';z). \tag{2.5}$$

We can easily verify that, if x' is an ergodic state of the $S_{z;x'}$-process, it is also an ergodic state of the process (2.4) and conversely.

Next let the x'-set A be given by

$$A = A_Y \cup (A_z \times X_2). \tag{2.6}$$

With respect to the decision processes $\{S_{z;x'};x' \in X'\}$ we assume:

Property 6 (strategies)

8) If $z \in Z$, a finite number of entries in A almost surely occur in a finite time interval.

According to lemma 1.31 in chapter 1 of part II:

a) the random losses, incurred during $[0,t_o)$, can be represented by a stochastic variable $\underline{k}'_{t_o;x'}$ with mean $k'(x';t_o)$;

b) the number of entries into A during $[0,t_o)$ can be represented by a stochastic variable $\underline{n}'_{t_o;x'}$ with mean $n'(x';t_o)$.

Assumption 2*

For each initial state $x' \in X'$ we have

$$E\{n'(\underline{y};t_o)|\ x'\} < \infty , \tag{2.7}$$

where $\underline{y}$ obeys the stationary absolute probability distribution of the Markov process (2.4).

The following theorem is an immediate consequence of lemma 1.49 in chapter 1 of part II.

Theorem 4.1

Under the assumptions 1^* and 2^*, for almost all realizations ω of the decision process $S_{z;x'}$, the limit

$$\lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \tag{2.1}$$

exists and is

$$\frac{1}{t_o} E \{k'(\underline{y};t_o) | x'\} \tag{2.8}$$

if the initial state x' is an ergodic state of the process (2.4) and if $\underline{y}$ obeys the corresponding stationary absolute probability distribution.

Let the x_1-components of (2.4) be denoted by

$$\{\underline{x}_{t_o;x_1;j}; j=0,1,\ldots\} \quad . \tag{2.9}$$

If only the x_1-components of (2.4) are recorded at $\{jt_o; j=0,1,\ldots\}$ and if the x_2-component of the initial state obeys the probability distribution $z(B;x_1)$, then the sequence (2.9) also constitutes a stationary Markov process with a discrete time parameter (in X_1). The first and higher order transition probabilities of this process are then given by

$$p^j_{t_o}(B;x_1;z) = \int_{X_2} z(dx_2;x_1) p^j_{t_o}(B \times X_2;x';z); \qquad j=1,2,\ldots, \tag{2.10}$$

where $x' = (x_1,x_2)$ and $B \in G_1$.

The stationary absolute probability distribution $p_{t_o}(B;x_1;z)$ of this discrete time process in X_1 satisfies

$$p_{t_o}(B;x_1;z) = \int_{X_2} z(dx_2;x_1) p_{t_o}(B \times X_2;x';z), \tag{2.11}$$

where $x' = (x_1,x_2)$ and $B \in G_1$.

Next we define the x_1-function $k(x_1;t_o)$ by

$$k(x_1;t_o) \overset{\text{def}}{=} \int_{X_2} z(dx_2;x_1)k'(x';t_o), \qquad (2.12)$$

where $x' = (x_1,x_2)$.

Obviously, if x_1 is an ergodic state of the stochastic process (2.9), the state $x' = (x_1,x_2)$ is an ergodic state of the process (2.4) with probability 1.

Provided that the initial state x' belongs to one given simple ergodic state with probability 1, it follows from (2.8) that theorem 4.1 remains true if x' has an initial probability distribution. We now consider the decision process $S_{z;x_1}$. As we know this process has such an initial distribution if x_1 is ergodic.

Since

$$\frac{1}{t_o} \int_{X_2} z(dx_2;x_1) \int_{X'} p_{t_o}(dy;x';z)k'(y;t_o) = \qquad (2.13)$$

$$= \frac{1}{t_o} \int_{X_2} p_{t_o}(du;x_1;z)k(u;t_o),$$

the following theorem has been proved:

Theorem 4

Under the assumptions 1^* and 2^* for almost all realizations ω of the decision process $S_{z;x_1}$ the limit

$$\lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \qquad (2.1)$$

exists and is

$$\frac{1}{t_o} E\{k(\underline{u};t_o)|\ x_1\} \qquad (2.14)$$

if the initial x_1-state x_1 is an ergodic state of the process (2.9) and if $\underline{u}$ obeys the corresponding stationary absolute probability distribution.

On the set Y_{t_o} of all ergodic states of the process (2.9), the criterion function $r(z;x_1)$ is now defined by

$$r(z;x_1) \overset{\text{def}}{=} E\{k(\underline{u};t_o)|x_1\} \quad . \qquad (2.15)$$

The domain of definition of the x_1-function $r(z;x_1)$ is extended to X_1 by taking

$$r(z;x_1) \overset{\text{def}}{=} E\{r(z;\underline{u})|x_1\} \quad , \qquad (2.16)$$

where $\underline{u}$ obeys the stationary absolute probability distribution of the process (2.9).

Let us return to the decision processes $\{S_{z;x'};x' \varepsilon X'\}$.

A similar result can be obtained by means of a closed set C in X' satisfying:

Assumption 1

The length of the period preceding the first entry into C is finite with probability 1.

If x' is the initial state of the decision process $S_{z;x'}$, the length of the period preceding the first entry into C is represented by a stochastic variable $\underline{t}'_{[C];x'}$ (cf. lemma 2.30 in chapter 2 of part II). The expected duration of this period is denoted by $t'(x';[C])$. Since the decision process is a stationary strong Markov process, the sequence of entry states

$$\{\underline{x}'_{[C];x';j};j=0,1,\ldots\} \qquad (2.17)$$

in C constitutes a stationary Markov process with a discrete time parameter (cf. lemma 1.50 in chapter 1 of part II).

We further impose:

Assumption 2

The Markov processes $\{\underline{x}'_{[C];x';j};j=0,1,\ldots\}$ with $x' \varepsilon X'$ satisfy the Doeblin condition (cf. [2] ,p.192).

Consequently, for each $x' \varepsilon X'$ there exists a stationary absolute probability distribution $p_{[C]}(B;x';z)$ that satisfies (cf.[2] ,p.214)

$$p_{[C]}(B;x';z) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p^{j}_{[C]}(B;x';z). \qquad (2.18)$$

Assumption 3

During the period $\underline{t}'_{[C];x'}$ a finite number of entries into A occur with probability 1.

According to lemma 1.31 in chapter 1 of part II:

a) the random losses, incurred during $[0,\underline{t}'_{[C];x'})$, can be represented by a stochastic variable $\underline{k}'_{[C];x'}$ with mean $k'(x';[C])$;

b) the number of entries into A during $[0,\underline{t}'_{[C];x'})$ can be represented by a stochastic variable $\underline{n}'_{[C];x'}$ with mean $n'(x';[C])$.

Assumption 4

For each initial state $x' \in X'$

$$E\{n'(\underline{y};[C])|x'\} < \infty$$

and

$$0 < E\{t'(\underline{y};[C])|x'\} < \infty,$$

where $\underline{y}$ obeys the corresponding stationary absolute probability distribution of the Markov process (2.17).

The following theorem follows at once from lemma 1.57 in chapter 1 of part II.

Theorem 5.1

Under the assumptions 1,2,3 and 4 for almost all realizations ω of the decision process $S_{z;x'}$ the limit

$$\lim_{T\to\infty} \frac{k_T(\omega;z)}{T} \tag{2.1}$$

exists and is

$$\frac{E\{k'(\underline{y};[C])|x'\}}{E\{t'(\underline{y};[C])|x'\}} \tag{2.19}$$

if the initial state x' is an ergodic state of the process (2.17) and if $\underline{y}$ obeys the corresponding stationary absolute probability distribution.

We now consider a closed set C of the form

$$C = \hat{C} \times X_2. \tag{2.20}$$

The x_1-components of the entry states in C are denoted by

$$\{\underline{x}_{[\hat{C}];x_1;j}; j=0,1,\ldots\} . \tag{2.21}$$

If only the x_1-components of (2.17) are recorded at the successive entry states in C and if the x_2-component of the initialstate obeys the probability distribution $z(B;x_1)$, then the sequence (2.21) also constitutes a stationary Markov process with a discrete time parameter. This process is called the Markov process in $[\hat{C}]$. The first and higher order transition probabilities of this process are then given by

$$p^j_{[\hat{C}]}(B;x_1;z) = \int_{X_2} z(dx_2;x_1)p^j_{[C]}(B \times X_2;x';z); j=1,2,\ldots , \tag{2.22}$$

where $x' = (x_1,x_2)$ and $B \in G_1$.

The stationary absolute probability distribution $p_{[\hat{C}]}(B;x_1;z)$ of the Markov process in $[\hat{C}]$ satisfies

$$p_{[\hat{C}]}(B;x_1;z) = \int_{X_2} z(dx_2;x_1)p_{[C]}(B \times X_2;x';z), \tag{2.23}$$

where $x' = (x_1,x_2)$ and $B \in G_1$.

Next we define the x_1-functions $k(x_1;[\hat{C}])$ and $t(x_1;[\hat{C}])$ by

$$k(x_1;[\hat{C}]) \stackrel{\text{def}}{=} \int_{X_2} z(dx_2;x_1)k'(x';[C]) \tag{2.24}$$

and

$$t(x_1;[\hat{C}]) \stackrel{\text{def}}{=} \int_{X_2} z(dx_2;x_1)t'(x';[\hat{C}]), \tag{2.25}$$

where $x' = (x_1,x_2)$.

Obviously, if x_1 is an ergodic state of the stochastic process (2.21), $x'= (x_1,x_2)$ is an ergodic state of the process (2.17) with probability 1. Provided that the initial state x' belongs to one given simple ergodic set with probability 1, it follows from (2.19) that theorem 5.1 remains true if x' has an initial probability distribution.

We now consider the decision process $S_{z;x_1}$. Since for each ergodic x_1-state of the Markov process in $[\hat{C}]$

$$\int_{X_2} z(dx_2;x_1) \frac{\int_{X'} p_{[C]}(dy;x';z)k'(y;[C])}{\int_{X'} p_{[C]}(dy;x';z)t'(y;[C])} =$$

$$= \frac{\int_{X_1} p_{[C]}(du;x_1;z)k(u;[\hat{C}])}{\int_{X_1} p_{[C]}(du;x_1;z)t(u;[\hat{C}])} \tag{2.26}$$

the following theorem has been proved:

Theorem 5

Under the assumptions 1,2,3 and 4 for almost all realizations ω of the decision process $S_{z;x_1}$ the limit

$$\lim_{T \to \infty} \frac{k_T(\omega;z)}{T} \tag{2.1}$$

exists and is

$$\frac{E\{k(\underline{u};[\hat{C}])\,|\,x_1\}}{E\{t(\underline{u};[\hat{C}])\,|\,x_1\}} \tag{2.27}$$

if the initial state x_1 is an ergodic state of the process (2.21) and if $\underline{u}$ obeys the corresponding stationary absolute probability distribution.

On the set $Y_{[\hat{C}]}$ of all ergodic states of the process (2.21), the criterion function $r(z;x_1)$ can also be defined by

$$r(z;x_1) \overset{def}{=} \frac{E\{k(\underline{u};[\hat{C}])\,|\,x_1\}}{E\{t(\underline{u};[\hat{C}])\,|\,x_1\}}\ , \tag{2.28}$$

where $\underline{u}$ obeys the stationary absolute probability distribution corresponding to x_1.

The domain of definition of the x_1-function $r(z;x_1)$ is extended to X_1 by taking

$$r(z;x_1) \overset{def}{=} E\,\{r(z;\underline{u})\,|\,x_1\}\ , \tag{2.29}$$

where $\underline{u}$ obeys the stationary absolute probability distribution corresponding to x_1 of the Markov process in $[\hat{C}]$.

Finally, we shall give a third expression for the function $r(z;x_1)$.

To this end we consider the natural processes $\{S_{o;x_1};x_1 \varepsilon X_1\}$. As we know the natural processes can also be described by means of the state space X'. In that case we always have $x_1 = x_2$.

Let B be a closed set in X' satisfying:

Assumption 5 (natural process)

The length of the period, preceding the moment at which the system first assumes a state of B, is finite with probability 1.

By lemma 1.5 in part II,this moment can be represented by a stochastic variable $\underline{t}^o_{B;x'}$, with mean $t^o(x_1;B)$ if x' is the initial state.

Assumption 6 (natural process)

During the period $\underline{t}^o_{B;x'}$, a finite number of entries in A_γ occur with probability 1.

Note that during a natural process no decision costs are incurred (cf. point 8 of property 6).

According to lemma 1.31 in chapter 1 of part II:

a) the random losses incurred during $[0,\underline{t}^o_{B;x'})$ can be represented by a stochastic variable $\underline{k}^o_{B;x'}$, with mean $k^o(x_1;B)$;

b) the number of entries into A during $[0,\underline{t}^o_{B;x'})$ can be represented by a stochastic variable $\underline{n}^o_{B;x'}$, with mean $n^o(x_1;B)$.

Assumption 7 (natural process)

For each initial state x' of the natural process

$$n^o(x_1;B) < \infty$$

$$\text{and} \qquad t^o(x_1;B) < \infty .$$

From assumption 7 it follows that $k^o(x_1;B) < \infty$ for each $x_1 \varepsilon X_1$.

We now return to the decision processes $\{S_{z;x_1};x_1 \varepsilon X_1\}$ and consider the x_1-functions $k(x_1;z)$ and $t(x_1;z)$, defined by

$$k(x_1;z) \overset{def}{=} k(x_1;[\hat{C}]) + \int_{X_1} p^1_{[\hat{C}]}(dv;x_1;z)k^o(v;B) - k^o(x_1;B) \tag{2.30}$$

and

$$t(x_1;z) \overset{def}{=} t(x_1;[\hat{C}]) + \int_{X_1} p^1_{[\hat{C}]}(dv;x_1;z)t^o(v;B) - t^o(x_1;B) \tag{2.31}$$

respectively.

Obviously, we have for each $x_1 \varepsilon X_1$

$$\int_{X_1} p_{[\hat{C}]}(du;x_1;z)k(u;z) = \int_{X_1} p_{[\hat{C}]}(du;x_1;z)k(u;[C]) \tag{2.32}$$

and

$$\int_{X_1} p_{[\hat{C}]}(du;x_1;z)t(u;z) = \int_{X_1} p_{[\hat{C}]}(du;x_1;z)t(u;[C]). \tag{2.33}$$

The following theorem is an immediate consequence of (2.28), (2.32) and (2.33).

Theorem 6

Under the assumptions 1 through 7, the criterion function $r(z;x_1)$ can be defined by

$$r(z;x_1) = \frac{E\{k(\underline{u};z)\,|\,x_1\}}{E\{t(\underline{u};z)\,|\,x_1\}} \tag{2.34}$$

on the set of all ergodic states of the Markov process in $[\hat{C}]$.

3. Two important choices of the sets C and B

1. Sometimes it is possible to choose the set C in such a way that the entry states in C always have the same x_1-component, say $\hat{x}_1$. In that case the criterion function $r(z;x_1)$ becomes

$$r(z;x_1) = \frac{k(\hat{x}_1;[C])}{t(\hat{x}_1;[C])} \quad . \tag{2.35}$$

2. In most cases, however, we choose

$$C = A_z \times X_2 \qquad B = A_o \times X_2, \tag{2.36}$$

where A_o is defined by $A_o \overset{\text{def}}{=} \bigcap_{z \varepsilon Z} A_z$ (cf.(1.52)).
Since the decision process between two entries into $A_z \times X_2$ can be described by means of a natural process, it follows from the definitions of $k(x_1;[C])$, $k^o(x_1;B)$, $t(x_1;[C])$ and $t^o(x_1;B)$ that (2.36) implies:

$$k(x_1;[A_z]) = \int_{X_2} z(dx_2;x_1) \{\gamma_{dec}(x';z(x_1)) + \gamma_{disc}(x') + k_+^o(x_2;A_z)\} \tag{2.37}$$

and

$$t(x_1;[A_z]) = \int_{X_2} z(dx_2;x_1) t^o(x_2;A_z), \tag{2.38}$$

where $x' = (x_1,x_2)$ and $k_+^o(x_2;A_z) = k^o(x_2;A_z) - \gamma_{disc}(x_2,x_2))$.
Since $A_o \subset A_z$, by (2.30), (2.31), (2.37) and (2.38),

$$\begin{aligned} k(x_1;z) &= \int_{X_2} z(dx_2;x_1) \{\gamma_{dec}(x';z(x_1)) + \gamma_{disc}(x') + \\ &\quad + k_+^o(x_2;A_z)\} + \int_{X_1} p^1_{[A_z]}(du;x_1;A_z) k^o(u;A_o) - k^o(x_1;A_o) = \\ &= \int_{X_2} z(dx_2;x_1) \{\gamma_{dec}(x';z(x_1)) + \gamma_{disc}(x') + k_+^o(x_2;A_o)\} - k^o(x_1;A_o). \end{aligned} \tag{2.39}$$

$$\begin{aligned} t(x_1;z) &= \int_{X_2} z(dx_2;x_1) t(x_2;A_z) + \\ &\quad + \int_{X_1} p^1_{[A_z]}(du;x_1;A_z) t^o(u;A_o) - t^o(x_1;A_o) = \\ &= \int_{X_2} z(dx_2;x_1) t^o(x_2;A_o) - t^o(x_1;A_o). \end{aligned} \tag{2.40}$$

We now introduce the $(x_1;d)$-functions $k(x_1;d)$ and $t(x_1;d)$, defined by (cf. p.88)

$$k(x_1;d) = \int_{X_2} P_d\,[dx_2]\ \{\gamma_{dec}(x';d) + \gamma_{disc}(x') + k_+^o(x_2;A_o)\} - k^o(x_1;A_o) \tag{2.41}$$

and

$$t(x_1;d) = \int_{X_2} P_d\,[dx_2]\ t^o(x_2;A_o) - t^o(x_1;A_o) \tag{2.42}$$

respectively.

Clearly, if d is a null decision

$$k(x_1;d) = 0 \qquad (2.43)$$
$$t(x_1;d) = 0.$$

Moreover, if $x_1 \varepsilon A_z$,

$$k(x_1;z(x_1)) = k(x_1;z) \qquad (2.45)$$
$$t(x_1;z(x_1)) = t(x_1;z). \qquad (2.46)$$

If x_1 is an ergodic state of the Markov process in $[A_z]$, then by (2.34), (2.45) and (2.46) the criterion function $r(z;x_1)$ is also given by

$$r(z;x_1) = \frac{\int_{X_1} p_{[A_z]}(du;x_1;z)k(u;z(u))}{\int_{X_1} p_{[A_z]}(du;x_1;z)t(u;z(u))} \qquad (2.47)$$

Note that the $(x_1;d)$-functions $k(x_1;d)$ and $t(x_1;d)$ depend only on

a) the structure of the natural process

b) the stopping set A_o

and not on a particular strategy z.
Thus, if we have to compare different strategies z, the $(x_1;d)$-functions $k(x_1;d)$ and $t(x_1;d)$ used in (2.47) can be determined once for all. In order to determine the criterion function $r(z;x_1)$, we then only need to know the stationary absolute probability distributions of the Markov process in $[A_z]$. This stochastic process has a discrete time parameter. This justifies the introduction of the rather complex functions $k(x_1;z)$ and $t(x_1;z)$.

By point $6^{a)}$ of property 6 the Markov process in $[A_z]$ and the Markov process in A_z (cf. theorem 1) are identical.

4. An additional property of the decision process

Let us return to the sets C and B, given by $C = \hat{C} \times X_2$ and $B = \hat{B} \times X_2$, with $\hat{B}$ and $\hat{C}$ unspecified.

It follows from (2.47) that the criterion function is a function of z and the x_1-component of the initial state only. Therefore we state:

Statement 2

If the x_1-functions $k(x_1;z)$ and $t(x_1;z)$ are known, then to find out the optimal strategy only the x_1-states of the system need to be considered.

For that reason our future discussions will refer to the state space X_1 only.

Notation 4:

In order to simplify the notation we write X ($x \varepsilon X$) instead of X_1 and G instead of G_1.

Let $S_{z;x}$ denote the decision process with initial state x and let $S_z^{\,x}$ denote the decision process which starts in a random state obeying the steady state probability distribution $p_{[\hat{C}]}(B;x;z)$.

In theorem 5 we have proved that, if the state x belongs to a simple ergodic set, then for the decision processes $S_{z;x}$ and $S_z^{\,x}$ the mean costs per unit of time over an infinite period are equal.

In this section we shall prove a result which will enable us to state the following:

"If a strategy $z \varepsilon Z$ is applied the difference in expected costs between the decision processes $S_{z;x}$ and $S_z^{\,x}$ is uniformly bounded in x on the class of all finite time intervals".

Let us consider the Markov process in $[\hat{C}]$ and let us introduce the following notations:

Notation 5:

1) The entry states in $\hat{C}$ will be denoted by $\{\underline{I}_j;\ j=1,2,\ldots\}$.
2) If z is the strategy applied and if x is the initial state, then

$$p^{\,n}_{[\hat{C}]}(B;x;z) \overset{def}{=} \text{Prob}\,\{\underline{I}_n \varepsilon B\,|\,x;z\}\ ;\ n=1,2,\ldots;\quad B \varepsilon G. \quad (2.48)$$

Let us reformulate assumptions 2 through 7.

Assumption 2'

The Markov process in $[\hat{C}]$ satisfies the Doeblin condition.

Assumption 3'

For each $z \varepsilon Z$ the x-functions $t(x;z)$ and $k(x;z)$ are bounded. Moreover, we have for each x $t(x;z) > 0$.

If $\hat{C} = A_z$, by point 7 of property 6 assumption 2' is satisfied. In chapter 2 of part II, (3.77) and ff., we prove for each $x \varepsilon X$

$$t(x;z) = \int_X z(d\hat{x};x) \int_0^\infty t \; P_0[\Xi_{dt;A_z};\hat{x}] > 0. \qquad (2.49)$$

In order to satisfy assumption 3' completely for $\hat{C} = A_z$ we shall introduce the following property:

Property 5 (costs)

4) For $d \varepsilon D(x)$ the (x;d)-functions $k(x;d)$ and $t(x;d)$, defined by (2.41) and (2.42), are bounded.

If $z \varepsilon Z$ is applied, there exists a decomposition of $\hat{C}$ into s disjunct simple ergodic sets $E_i \varepsilon G$ $(i=1,2,\ldots,s)$ and one set of transient states $O \varepsilon G$.

The i^{th} simple ergodic set E_i can be subdidived into c_i cyclically moving subsets $M_{ij} \varepsilon G$ $(j=1,2,\ldots,c_i)$ (cf.[2], p.206 ff.).

We shall now prove that from assumption 2' it follows that

a) the number of disjunct simple ergodic sets is finite;
b) for each simple ergodic set the number of cyclically moving subsets is finite.

Using Doob's notation, if $s = \infty$ then for each $\varepsilon > 0$ a simple ergodic set E_k can be found such that (cf. [2], p.192)

$$1) \quad \phi(E_k) \leqq \varepsilon \; ; \qquad (2.50)$$

$$2) \quad p^n_{[\hat{C}]}(E_k;x;z) = 1; \qquad x \varepsilon E_k;\; n=1,2,\ldots \qquad (2.51)$$

This contradicts the Doeblin condition and thus $s<\infty$.

Using Doob's notation again, if $c_i=\infty$ then for each $\varepsilon>0$ and for each $n_o \geqq 0$ an integer $n \geqq n_o$, a cyclically moving subset $M_{i;n}$ can be found such that (cf. [2], p.192)

$$1)\quad \phi(M_{i;n}) \leqq \varepsilon; \qquad (2.52)$$

$$2)\quad p^n_{[\hat{C}]}(M_{i;n};x;z) = 1; \qquad x \varepsilon M_{i;1}. \qquad (2.53)$$

This contradicts the Doeblin condition and thus $c_i<\infty$.

Since there are only a finite number of simple ergodic sets it follows from assumption 3' that the x-function $r(z;x)$ is bounded.

Let us now introduce the x-function $v(x;z)$ defined by

$$v(x;z) \overset{\text{def}}{=} k(x;z) - r(z;x)t(x;z). \qquad (2.54)$$

By assumption 3' the x-function $v(x;z)$ is bounded. For later reference we state

Lemma 2.1

If $z \varepsilon Z$, the x-functions $r(z;x)$ and $v(x;z)$ are bounded.

In the remainder of this section we shall prove that, for $n\rightarrow\infty$, the sequences

$$\{\, v(x;z) + \sum_{j=1}^{n}\int_{\hat{C}} p^j_{[\hat{C}]}(dI;x;z)v(I;z);\ n=1,2,\ldots \qquad (2.55)$$

oscillate between finite bounds.

To this end we introduce an x-set $\hat{C}'$ which is a union of cyclically moving subsets; i.e. one for each simple ergodic set of the Markov process in $[\hat{C}]$.

If the initial state $I_1 \varepsilon \hat{C}$ is a state of a simple ergodic set E_i and if $m(I_1)$ is the number of entries to be made before a state of $\hat{C}'$ will be taken on, let $c_n(z;I_1;\hat{C}')$ be defined by

$$c_n(z;I_1;\hat{C}') \overset{\text{def}}{=} v(I_1;z) + \sum_{j=1}^{nc_i+m(I_1)} \int_{\hat{C}} p^j_{[\hat{C}]}(dI;I_1;z)v(I;z). \qquad (2.56)$$

Lemma 2.2

If I_1 is an ergodic state, the sequences

$$\{ c_n(z;I_1;\hat{C}'); \quad n=1,2,\dots \} \tag{2.57}$$

converge uniformly in I_1 to a limit denoted by $c(z;I_1;\hat{C}')$.

Proof:

It can be proved that (cf. [2], p.208 ff.) for $B \varepsilon G$ the sequences

$$\{ p_{[\hat{C}]}^{nc_i+j}(B;I_1;z); \; n=1,2,\dots \}; \qquad j=1,2,\dots,c_i \tag{2.58}$$

converge, exponentially fast, and uniformly in B and I_1 to a probability distribution $p_{[\hat{C}]}^{\infty c_i+j}(B;I_1;z)$.

The stationary absolute probability distribution of the Markov process in $[\hat{C}]$ corresponding to the initial state I_1 satisfies (cf. [2], p.192 ff.)

$$p_{[\hat{C}]}(B;I_1;z) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p_{[\hat{C}]}^{j}(B;I_1;z). \tag{2.59}$$

From (2.59) we deduce for $I_1 \varepsilon E_i$

$$p_{[\hat{C}]}(B;I_1;z) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p_{[\hat{C}]}^{j}(B;I_1;z) =$$

$$= \lim_{n\to\infty} \frac{1}{nc_i} \sum_{j=1}^{nc_i} p_{[\hat{C}]}^{j}(B;I_1;z) =$$

$$= \lim_{n\to\infty} \frac{1}{nc_i} \sum_{k=0}^{n-1} \sum_{j=1}^{c_i} p_{[\hat{C}]}^{kc_i+j}(B;I_1;z) =$$

$$= \frac{1}{c_i} \sum_{j=1}^{c_i} \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{[\hat{C}]}^{kc_i+j}(B;I_1;z)\right) = \frac{1}{c_i} \sum_{j=1}^{c_i} p_{[\hat{C}]}^{\infty c_i}(B;I_1;z). \tag{2.60}$$

Consequently, by the definitions of $r(z;x)$ and $v(x;z)$ we have for $I_1 \varepsilon E_i$

$$\sum_{j=1}^{c_i} \int_{\hat{C}} p_{[\hat{C}]}^{\infty c_i + j}(dI;I_1;z)v(I;z) = c_i \int_{\hat{C}} p_{[\hat{C}]}(dI;I_1;z)v(I;z) = 0. \tag{2.61}$$

Now let

a) $$\partial \overset{\text{def}}{=} \sup_{I \in \hat{C}} |v(I;z)| < \infty \tag{2.62}$$

b) $$A_{\mu;\nu} \overset{\text{def}}{=} \{I \mid \frac{\mu\partial}{2^\nu} \leqq v(I;z) < \frac{(\mu+1)\partial}{2^\nu}\} \quad ; \tag{2.63}$$

$$; \ \mu = -2^\nu \ldots +2^\nu \ ; \ \nu = 1,2,\ldots$$

and let us consider the integral

$$\int_{\hat{C}} p_{[\hat{C}]}^{nc_i + j}(dI;I_1;z)v(I;z) \tag{2.64}$$

and the sum

$$\sum_{\mu=-2^\nu}^{+2^\nu} \frac{\mu\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j}(A_{\mu;\nu};I_1;z). \tag{2.65}$$

Obviously,

$$\left| \int_{\hat{C}} p_{[\hat{C}]}^{nc_i+j}(dI;I_1;z)v(I;z) - \sum_{\mu=-2^\nu}^{2^\nu} \frac{\mu\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j}(A_{\mu;\nu};I_1;z) \right| \leqq \frac{\partial}{2^\nu} . \tag{2.66}$$

Let the x-set $B_{\mu;\nu}$ be defined by

$$B_{\mu;\nu} \overset{\text{def}}{=} \{ I \mid v(I;z) \geqq \frac{\mu\partial}{2^\nu} \} \quad . \tag{2.67}$$

From (2.63), (2.65) and (2.67) it follows that

$$\sum_{\mu=-2^\nu}^{+2^\nu} \frac{\mu\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j}(A_{\mu;\nu};I_1;z) =$$

$$= \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\mu\partial}{2^\nu} \left[p_{[\hat{C}]}^{nc_i+j}(B_{\mu;\nu};I_1;z) - p_{[\hat{C}]}^{nc_i+j}(B_{\mu+1;\nu};I_1;z) \right] =$$

$$= - \partial + \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j} (B_{\mu+1;\nu}; I_1; z). \tag{2.68}$$

The relations (2.66) and (2.68) imply that the corresponding elements of the sequences

$$\{ \int_{\hat{C}} p_{[\hat{C}]}^{nc_i+j} (dI; I_1; z) v(I;z);\ n=1,2,\dots \} \tag{2.69}$$

and

$$\{ \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j} (B_{\mu+1;\nu}; I_1; z) - \partial;\ n=1,2,\dots \} \tag{2.70}$$

differ $\frac{\partial}{2^\nu}$ at the most.

Because the sequences (2.58) converge exponentially fast and uniformly in B and I_1 to $p_{[\hat{C}]}^{\infty c_i+j}(B;I_1;z)$ we can find for each j a triple $(N_o;\rho;\kappa)$ with $\rho<1$ and $\kappa<\infty$ such that for $n \geq N_o$

$$\left| p_{[\hat{C}]}^{nc_i+j} (B;I_1;z) - p_{[\hat{C}]}^{\infty c_i+j} (B;I_1;z) \right| \leq \kappa \cdot \rho^n . \tag{2.71}$$

This implies

$$\left| \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} p_{[\hat{C}]}^{nc_i+j} (B_{\mu+1;\nu}; I_1; z) - \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} p_{[\hat{C}]}^{\infty c_i+j} (B_{\mu+1;\nu}; I_1; z) \right|$$

$$\leq \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} \kappa \rho^n = \kappa^* \rho^n, \tag{2.72}$$

where $\kappa^* = 2 \partial \kappa$.

Consequently, the sequences (2.70) converge exponentially fast and uniformly in I_1 to

$$\sum_{\mu=-2^\nu}^{+2^\nu} \frac{\partial}{2^\nu} p_{[\hat{C}]}^{\infty c_i+j} (B_{\mu+1;\nu}; I_1; z) - \partial;\ j=1,2,\dots,c_i . \tag{2.73}$$

Obviously, (2.73) differs from

$$\int_{\hat{C}} p_{[\hat{C}]}^{\infty c_i+j} (dI; I_1; z) v(I;z) \tag{2.74}$$

at the most $\frac{\delta}{2^\nu}$.

Now we have for each j

$$\left| \int_{\hat{C}} P_{[\hat{C}]}^{\infty c_i + j}(dI;I_1;z)v(I;z) - \int_{\hat{C}} P_{[\hat{C}]}^{nc_i+j}(dI;I_1;z)v(I;z) \right| =$$

$$\leq \left| \int_{\hat{C}} P_{[\hat{C}]}^{\infty c_i + j}(dI;I_1;z)v(I;z) - \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\delta}{2^\nu} P_{[\hat{C}]}^{\infty c_i + j}(B_{\mu+1;\nu};I_1;z) + \delta \right|$$

$$+ \left| \sum_{\mu=-2^\nu}^{2^\nu} \frac{\delta}{2^\nu} P_{[\hat{C}]}^{\infty c_i + j}(B_{\mu+1;\nu};I_1;z) - \sum_{\mu=-2^\nu}^{2^\nu} \frac{\delta}{2^\nu} P_{[\hat{C}]}^{nc_i + j}(B_{\mu+1;\nu};I_1;z) \right| +$$

$$+ \left| \int_{\hat{C}} P_{[\hat{C}]}^{nc_i + j}(dI;I_1;z)v(I;z) - \sum_{\mu=-2^\nu}^{+2^\nu} \frac{\delta}{2^\nu} P_{[\hat{C}]}^{nc_i + j}(B_{\mu+1;\nu};I_1;z) + \delta \right|$$

$$\leq \frac{2\delta}{2^\nu} + \kappa^* \rho^n = \frac{\delta}{2^{\nu-1}} + \kappa^* \rho^n \; ; \qquad \nu=1,2,\ldots \; . \tag{2.75}$$

Consequently, for each j there exists a triple (N_o,ρ,κ^*) with $\rho < 1$ and $\kappa^* < \infty$ such that for $n \geq N_o$

$$\left| \int_{\hat{C}} P_{[\hat{C}]}^{\infty c_i + j}(dI;I_1;z)v(I;z) - \int_{\hat{C}} P_{[\hat{C}]}^{nc_i+j}(dI;I_1;z)v(I;z) \right| \leq \kappa^* \rho^n . \tag{2.76}$$

Thus the sequences

$$\left\{ \int_{\hat{C}} P_{[\hat{C}]}^{nc_i+j}(dI;I_1;z)v(I;z) \; ; \; n=1,2,\ldots \right\} \tag{2.77}$$

converge exponentially fast and uniformly in I_1 to

$$\int_{\hat{C}} P_{[\hat{C}]}^{\infty c_i+j}(dI;I_1;z)v(I;z) \; . \tag{2.78}$$

Obviously, by (2.61) the sequences

$$\left\{ \sum_{j=1}^{c_i} \int_{\hat{C}} P_{[\hat{C}]}^{nc_i+j}(dI;I_1;z)v(I;z) \; ; \; n=1,2,\ldots \right\} \tag{2.79}$$

converge exponentially fast and uniformly in I_1 to

$$\sum_{j=1}^{c_i} \int_{\hat{C}} P_{[\hat{C}]}^{\infty c_i + j}(dI;I_1;z)v(I;z) = 0 \ . \tag{2.80}$$

Now we are able to prove that the sequences (2.58) converge uniformly in I_1 to an I_1-function $c(z;I_1;\hat{C}')$.

Let us consider the limit

$$\lim_{n\to\infty} \sum_{N_o c_i + m(I_1)+1}^{nc_i+m(I_1)+1} \int_{\hat{C}} P_{[\hat{C}]}^{j}(dI;I_1;z)v(I;z) \ , \tag{2.81}$$

which is equal to

$$\lim_{n\to\infty} \sum_{h=N_o}^{n} \sum_{j=1}^{c_i} \int_{\hat{C}} P_{[\hat{C}]}^{hc_i+m(I_1)+j}(dI;I_1;z)v(I;z) \ . \tag{2.82}$$

Since the sequences (2.79) converge exponentially fast and uniformly in I_1 to zero we can find a triple $(N_o;\rho;\kappa)$ with $\rho < 1$ and $\kappa < \infty$ such that for $h \geq N_o$

$$\left| \sum_{j=1}^{c_i} \int_{\hat{C}} P_{[\hat{C}]}^{hc_i+m(I_1)+j}(dI;I_1;z)v(I;z) \right| \leq \kappa\, \rho^h \tag{2.83}$$

and hence, uniformly in I_1,

$$\lim_{n\to\infty} \left| \sum_{h=N_o}^{n} \sum_{j=1}^{c_i} \int_{\hat{C}} P_{[\hat{C}]}^{hc_i+m(I_1)+j}(dI;I_1;z)v(I;z) \right| \leq \frac{\kappa\rho^{N_o}}{1-\rho}. \tag{2.84}$$

By (2.58) and (2.62) uniformly $I_1 \varepsilon E_i$ we have

$$\left| c_{N_o}(z;I_1;\hat{C}') \right| \leq (N_o+1)c_i\delta < \infty. \tag{2.85}$$

Hence, by (2.84) and (2.85) the sequences (2.57) converge uniformly in $I_1 \varepsilon \bigcup_{i=1}^{s} E_i$ to $c(z;I_1;\hat{C}')$. This function satisfies

$$\left| c(z;I_1;\hat{c}') \right| \leq (N_o+1)c\delta + \frac{\kappa\rho^{N_o}}{1-\rho} < \infty, \tag{2.86}$$

where $c = \max_i c_i$. (2.87)

This ends the proof.

Finally, we consider the case that I_1 is a transient state. It can be proved (cf. [2] ,p.207) that for some $\rho < 1$

$$\sum_{j=N}^{\infty} p^{j}_{[\hat{C}]}(O;I_1;z) \leq \text{const}\ \frac{\rho^N}{1-\rho}\ , \qquad (2.88)$$

where O is the set of transient states.

Consequently,

$$| \sum_{j=N}^{\infty} \int_O p^{j}_{[\hat{C}]}dI;I_1;z)v(I;z)| \leq$$

$$| \sum_{j=N}^{\infty} \int_O p^{j}_{[\hat{C}]}(dI;I_1;z)\ v(I;z)| \leq \text{const}\ \frac{\rho^N}{1-\rho} \qquad (2.89)$$

and thus a positive value $k' < \infty$ can be found such that

$$| v(I_1;z) + \sum_{j=1}^{\infty} \int_O p^{j}_{[\hat{C}]}(dI;I_1;z)v(I;z)| \leq \kappa' < \infty\ . \qquad (2.90)$$

It can easily be verified that the probability distribution of the first ergodic state reached in $\hat{C}$ can be defined. Let us denote this probability distribution by $p^{E}_{[\hat{C}]}(B;x;z)$.
So far the function $c(z;I_1;\hat{C}')$ has only been defined for ergodic states I_1. The domain of definition will now be extended to X by taking

$$c(z;x;\hat{C}') = v(x;z) + \sum_{j=1} \int_O p^{j}_{[\hat{C}]}(dI;x;z)v(I;z) +$$

$$+ \int_{\hat{C}} p^{E}_{[\hat{C}]}(dI;x;z)c(z;I;\hat{C}')\ . \qquad (2.91)$$

By means of (2.86) and (2.90) we can easily find a positive $\kappa'' < \infty$ such that for each $x \in X$

$$| c(z;x;\hat{C}') | \leq \kappa'' < \infty\ . \qquad (2.92)$$

For later reference we state:

Lemma 2.3

The domain of definition of the x-function $c(z;x;\hat{C}')$, defined on the set of all ergodic states as the limit of (2.58) can be extended to X by means of (2.91).

Note that for each choice of $\hat{C}'$ we find an x-function $c(z;x;\hat{C}')$.

Obviously, for $n \to \infty$, the sequences

$$\{ \ v(x;z) + \sum_{j=1}^{n} \int_{\hat{C}} p^{j}_{[\hat{C}]}(dI;x;z)v(I;z); \ n=1,2,\dots \} \tag{2.55}$$

oscillate between the bounds

$$c(z;x;\hat{C}') \pm (c+1)\,\delta \ . \tag{2.93}$$

So we have proved the following lemma:

Lemma 2.4

If a strategy $z \varepsilon Z$ is applied, for $n \to \infty$ the sequences

$$v(x;z) + \sum_{j=1}^{n} \int_{\hat{C}} p^{j}_{[\hat{C}]}(dI;x;z)v(I;z); \ n=1,2,\dots \tag{2.55}$$

oscillate uniformly in x between finite bounds.

From (2.90) we deduce

$$c(z;x;\hat{C}') = v(x;z) + E\{ c(z;\underline{I}_1;\hat{C}') \ x;z\} \ . \tag{2.94}$$

5. The x-functions r(z;x) and c(z;x)

In chapter 3 we shall only consider the case $\hat{C} = A_z$ and $\hat{B} = A_o$.

By point $6^{a)}$ of property 6 the Markov process in $[A_z]$ and the Markov process in A_z (cf. theorem 1) are identical.

From now on the intervention states are denoted by $\underline{I}_j$ (j=1,2,...).

By point $6^{a)}$ of property 6 the intervention states $\{ \underline{I}_j; j=2,\dots \}$ are almost surely entry states in $[A_z]$. If the initial state $x \varepsilon A_z$, then the first intervention state I_1 is x itself and thus is <u>no</u> entry state. On the other hand, if $x \varepsilon \bar{A}_z$, $\underline{I}_1$ is an entry state.

It follows from (2.94) that, if

$$\hat{C} = A_z \qquad \hat{B} = A_o \qquad \hat{C}' = A^z , \tag{2.95}$$

we have

$$c(z;x;A^z) = k(x;z(x)) - r(z;x)t(x;z(x)) + E\{c(z;\underline{I};A^z)|x;z\} , \tag{2.96}$$

where $\underline{I}$ is the first intervention state after the state x. (Thus, if $x \in \bar{A}_z$, $\underline{I}=\underline{I}_1$. Otherwise: $\underline{I}=\underline{I}_2$.)

Since for null decisions we have

$$k(x;d) = 0 \tag{2.97}$$

$$t(x;d) = 0 \tag{2.98}$$

we find

$$c(z;x;A^z) =$$

$$= \begin{cases} k(I_1;z(I_1))-r(z;I_1)t(I_1;z(I_1))+ E\{c(z;\underline{I}_2;A^z)|I_1;z\} \ ; \text{ if } x=I_1 \varepsilon A_z . \\ \{c(z;\underline{I}_1;A^z)|x;z\} \ ; \text{ if } x \varepsilon \bar{A}_z . \end{cases} \tag{2.99}$$

Further, we have for $I_1 \varepsilon A_z$ (cf.(2.33))

$$r(z;I_1) = \int_{A_z} p_{[A_z]}(dI;I_1;z)r(z;I) =$$

$$= \int_{A_z} p^1_{[A_z]}(dI_2;I_1;z) \int_{A_z} p_{[A_z]}(dI;I_2;z)r(z;I) =$$

$$= \int_{A_z} p^1_{[A_z]}(dI_2;I_1;z)r(z;I_2). \tag{2.100}$$

<u>Notation 6</u>:

If z is the strategy applied, and if x is the initial state, then

$$p^n_{A_z}(B;x;z) \overset{\text{def}}{=} \text{Prob}\{\underline{I}^*_n \varepsilon B|x;z\} \ ; \ n=1,2,\ldots; \qquad B \varepsilon G, \tag{2.101}$$

where $\underline{I}^*_n$ stands for the n^{th} <u>future</u> intervention state.

Consequently, we have for $I_1 \varepsilon A_z$

$$r(z;I_1) = \int_{A_z} p^1_{A_z}(dI_2;I_1;z)\, r(z;I_2), \tag{2.102}$$

$$c(z;I_1;A^z) = k(I_1;z(I_1)) - r(z;I_1)\,t(I_1;z(I_1)) + \\ + \int_{A_z} p^1_{A_z}(dI_2;I_1;z)\, c(z;I_2;A^z). \tag{2.103}$$

It follows from (2.102) and (2.103) that the set of functional equations

$$r(z;I_1) = \int_{A_z} p^1_{A_z}(dI_2;I_1;z)\, r(z;I_2), \tag{2.102}$$

$$c(z;I_1) = k(I_1;z(I_1)) - r(z;I_1)\,t(I_1;z(I_1)) + \\ + \int_{A_z} p^1_{A_z}(dI_2;I_1;z)\, c(z;I_2) \tag{2.104}$$

has at least as many solutions as different choices of A^z exist.

Let us consider the functional equations (2.102) and (2.104) more carefully.

Suppose that the Markov process in A_z has m simple ergodic sets E_i $(i=1,2,\ldots,m)$. If δ_i $(i=1,2,\ldots)$ are arbitrary real numbers, let the x-function $c^*(z;x)$ be defined by (cf. (2.90))

$$c^*(z;x) \overset{\text{def}}{=} \begin{cases} c(z;x;A^z) + \delta_i \;; & \text{if } x \,\varepsilon\, E_i \\ v(x;z) + \sum_{j=1}^{\infty} \int_O p^j_{[A_z]}(dI;x;z)\, v(I;z) + \\ \quad + \int_{A_z} p^E_{[A_z]}(dI;x;z)\, c^*(z;I)\,; & \text{if } x \,\varepsilon\, O \\ E\{ c^*(z;\underline{I}_1) | x;z \} \;; & \text{if } x \,\varepsilon\, \bar{A}_z \,. \end{cases} \tag{2.105}$$

We can now easily verify that the I_1-functions $r(z;I_1)$ and $c^*(z;I_1)$ also constitute a solution of the set of functional equations (2.102) and (2.104).

In chapter 3 an x-function $c(z;x)$ will be considered that satisfies (2.104). This function needs not to be one of the functions

$c(z;x;A^z)$. The functions $c(z;x;A^z)$ have already proved their usefulness in proving the existence of a solution of the set of functional equations (2.102) and (2.104).

CHAPTER 3

Optimal strategies

1. Introduction and definitions

In this chapter we only consider the case

$$\hat{C} = A_z \tag{3.1}$$

$$\hat{B} = A_o . \tag{3.2}$$

Let the x-function $c(z;x)$ be one of the functions which satisfy the equation

$$c(z;x) = k(x;z(x)) - r(z;x)t(x;z(x)) + + \int_{A_z} p^1_{A_z}(dI;x;z)c(z;I) \tag{3.3}$$

with

$$r(z;x) = \int_{A_z} p^1_{A_z}(dI;x;z)r(z;I). \tag{3.4}$$

We shall discuss three decision problems. The solutions of these problems enable us to formulate properties of the optimal strategy.

Next we shall construct an iteration procedure that yields a sequence of strategies $\{z^i; i=1,2,\ldots\}$ satisfying for each x

$$r(z^i;x) \geqq r(z^j;x); \qquad j \geqq i. \tag{3.5}$$

Under conditions (properties of the mathematical model) to be stated below we prove for each $x \in X$

$$\lim_{i \to \infty} r(z^i;x) = \inf_{z \in Z} r(z;x). \tag{3.6}$$

Let us first introduce some tools needed for the description of the three decision problems mentioned above.

<u>Tool</u> a) The <u>extended class</u> $\hat{Z} \supset Z$ of all strategies z satisfying $A_z \supset A_o$ and points 1,2,3,4 and 6[b)] of property 6.

<u>Tool</u> b) The <u>mixed strategies</u> of the form $(z_a)^n(z_b)^m z_c$ with $z_a \varepsilon \hat{Z}$, $z_b \varepsilon \hat{Z}$ and $Z_c \varepsilon Z$. Such a mixed strategy dictates successively

1) n interventions in accordance with z_a;
2) m interventions in accordance with z_b;
3) ∞ interventions in accordance with z_c.

If $m > 0$ and if $z_b \neq z_c$ or if $n > 0$ and if $z_a \neq z_b$, the mixed strategy $z' = (z_a)^n(z_b)^m z_c$ does not belong to $\hat{Z}$. Note that, conversely, each strategy $z \varepsilon Z$ is a mixed strategy (n=m=0). Let us consider the case n=0; i.e. the strategy $z' = (z_b)^m z_c$. We then define the functions $r(z';x)$ and $c(z';x)$ by [1)]

$$r(z';x) = r((z_b)^m z_c;x) \overset{def}{=}$$

$$\overset{def}{=} \int_{A_{z_b}} p^m_{A_{z_b}}(dI_m;x;z_b) \int_X z_b(dy^b_m;I_m) r(z_c;y^b_m) =$$

$$= E\{r(z_c;\underline{y}^b_m) \mid x;z_b\} \ , \qquad (3.7)$$

$$c(z';x) = c((z_b)^m z_c;x) \overset{def}{=}$$

$$\overset{def}{=} \sum_{j=1}^{m} E\{k(\underline{I}_j;z_b(\underline{I}_j)) - r((z_b)^{m-j+1}z_c;\underline{I}_j)t(\underline{I}_j;z_b(\underline{I}_j)) \mid x;z_b\} +$$

$$+ \int_{A_{z_b}} p^m_{A_{z_b}}(dI_m;x;z_b) \int_X z_b(dy^b_m;I_m) c(z_c;y^b_m) =$$

$$= \sum_{j=1}^{m} E\{k(\underline{I}_j;z_b(\underline{I}_j)) - r((z_b)^{m-j+1}z_c;\underline{I}_j)t(\underline{I}_j;z_b(\underline{I}_j)) \mid x;z_b\} +$$

$$+ \ E\{c(z_c;\underline{y}^b_m) \mid x;z_b\} \ , \qquad (3.8)$$

where $\underline{y}^b_m$ is the state into which the system is transferred by the decision $z_b(\underline{I}_m)$.

---

1) The states $\underline{y}^b_m$ (i=1,2,...) belong to X.

Next we consider the case $n \neq 0$ and define the functions $r(z';x)$ and $c(z';x)$ by

$$r(z';x) = r((z_a)^n(z_b)^m z_c;x) \overset{\text{def}}{=}$$

$$\overset{\text{def}}{=} E\{r((z_b)^m z_c;\underline{y}_n^a) \mid x;z_a\} , \tag{3.9}$$

$$c(z';x) = c((z_a)^n(z_b)^m z_c;x) \overset{\text{def}}{=}$$

$$\overset{\text{def}}{=} \sum_{j=1}^{n} E\{k(\underline{I}_j;z_a(\underline{I}_j))-r((z_a)^{n-j+1}(z_b)^m z_c;\underline{I}_j)\ t(\underline{I}_j;z_a(\underline{I}_j)) \mid x;z_a\} +$$

$$+ \quad E\{c((z_b)^m z_c;\underline{y}_n^a) \mid x;z_a\} , \tag{3.10}$$

where $\underline{y}_n^a$ is the state into which the system is transferred by the decision $z_a(\underline{I}_n)$.

<u>Tool</u> c) The <u>mixed strategies of the form dz'</u>, where z' is a mixed strategy of the form $(z_a)^n(z_b)^m z_c$ (cf. tool b)). The mixed strategy dz' dictates

1) the decision d in the initial state and after that
2) decisions in accordance with z'.

We now define for $d \in D(x)$

$$r(dz';x) \overset{\text{def}}{=} \int_X P_d(dy) r(z';y) \tag{3.11}$$

$$c(dz';x) \overset{\text{def}}{=} k(x;d) - r(dz';x)t(x;d) + \int_X P_d(dy)c(z';y) . \tag{3.12}$$

<u>Tool</u> d) The <u>mixed strategies</u> of the form Az', where z' is a mixed strategy of the form $(z_a)^n(z_b)^m z_c$ and A is a closed set in X. The mixed strategy Az' interdicts any intervention up to the moment that for the first time a state of A is taken on by the system. From that time onwards decisions are made in accordance with z'.

We now define

$$r(Az';x) \overset{\text{def}}{=} \int_A p_A^1(dy;x)r(z';y) \tag{3.13}$$

$$c(Az';x) \overset{\text{def}}{=} \int_A p_A^1(dy;x)c(z';y), \tag{3.14}$$

where $p_A^1(B;x)$ denotes the probability distribution of the first state of A which is taken on if x is the initial state. Note that the probability distribution $p_A^1(B;x)$ depends only on the natural process.

Notation 7:

If z' is a mixed strategy $(z_a)^n(z_b)^m z_c$ and if $n > 0$, then $A_{z'}$ is given by

$$A_{z'} \overset{\text{def}}{=} A_{z_a} . \tag{3.15}$$

Tool e) Minimizing subset of $A_{z'}$. If z' is a mixed strategy $(z_a)^n(z_b)^m z_c$, consider the class $K_{z'}$ of all closed sets A satisfying

1) $A \supset A_o$;

2) $\overline{A} = \{x|r(Az';x) < r(z';x)\} \cup \{x|r(Az';x) = r(z';x);$

$$c(Az';x) \leqq c(z';x)\} . \tag{3.16}$$

Obviously, we have X and $A_{z'} \in K_{z'}$.

The minimizing subset of $A_{z'}$ is now defined by the intersection $A'_{z'}$ of all sets $A \in K_{z'}$.

Thus,

$$A'_{z'} = \bigcap_{A \in K_{z'}} A. \tag{3.17}$$

Property 2 (natural process)

4) If A is a closed set in X, a finite number of entries into A occur with probability 1 in a finite time interval.

Lemma 3.1

If the x-sets A_1 and A_2 are closed and contain the set A_o, then for each bounded measurable x-function $q(x)$

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)q(x_1) =$$

$$=\int_{A_1} p^1_{A_1}(dx_1;x)q(x_1) + \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \cdot$$

$$\cdot \left[\int_{A_2} p^1_{A_2}(dx_2;x_1)q(x_2) - q(x_1)\right] +$$

$$+\ldots \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \int_{\bar{A}_1} p^1_{A_2}(dx_2;x_1) \ldots \int_{\bar{A}^{(k)}} p^1_{A^{(k-1)}}(dx_k;x_{k-1}) \times$$

$$\times \left[\int_{A^{(k)}} p^1_{A^{(k)}}(dx_{k+1};x_k)q(x_{k+1}) - q(x_k)\right] \tag{3.18}$$

$+ \ldots,$

where

$$A^{(k)} \overset{\text{def}}{=} \begin{cases} A_1 & ; \text{ if } k = \text{even} \\ A_2 & ; \text{ if } k = \text{odd} \end{cases} . \tag{3.19}$$

<u>Proof</u>:

Point 4 of property 2) and point 5 of property 6) imply

$$p^1_{A_1 \cap A_2}(B;x) = p^1_{A_1}(B \cap A_2;x) + \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) p^1_{A_2}(B \cap A_1;x_1)$$

$$+ \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \int_{\bar{A}_1} p^1_{A_2}(dx_2;x_1)\, p^1_{A_1}(B \cap A_2;x_2) + \ldots\ldots \tag{3.20}$$

Consequently,

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)q(x_1) = \int_{A_2} p^1_{A_1}(dx_1;x)q(x_1) +$$

$$+ \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \int_{A_1} p^1_{A_2}(dx_2;x_1)q(x_2) +$$

$$+ \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \int_{\bar{A}_1} p^1_{A_2}(dx_2;x_1) \int_{A_2} p^1_{A_1}(dx_3;x_2)q(x_3) + \ldots . \tag{3.21}$$

Obviously, the right hand sides of (3.18) and (3.21) are equal.

Lemma 3.2

If $A_1, A_2 \varepsilon K_{z'}$, then

$$A_1 \cap A_2 \varepsilon K_{z'} \quad . \tag{3.22}$$

Proof:

In lemma 3.1 the substitution $q(x) = r(z';x)$ yields for $x \varepsilon \bar{A}_1$:

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)r(z';x_1) = \int_{A_1} p^1_{A_1}(dx_1;x)r(z';x_1) +$$
$$+ \int_{\bar{A}_2} p^1_{A_1}(dx_1;x)\Big[\int_{A_2} p^1_{A_2}(dx_2;x_1)r(z';x_2) - r(z';x_1)\Big] +$$
$$+ \int_{\bar{A}_2} p^1_{A_1}(dx_1;x)\int_{\bar{A}_1} p^1_{A_2}(dx_2;x_1)\Big[\int_{A_1} p^1_{A_1}(dx_3;x_2)r(z';x_3) - r(z';x_2)\Big] +$$
$$+ \ldots\ldots \tag{3.23}$$

By the definition of the class $K_{z'}$ we find that

a) $\int_{A_1} p^1_{A_1}(dx_1;x)r(z';x_1) \leq r(z';x); \qquad x \varepsilon \bar{A}_1 \quad ;$ (3.24)

b) the remaining terms of the right hand side of (3.23) are smaller than or equal to zero.

Hence, for $x \varepsilon \bar{A}_1$

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)r(z';x_1) \leq \int_{A_1} p^1_{A_1}(dx_1;x)r(z';x_1) \leq r(z';x). \tag{3.25}$$

The proof of (3.25) for $x \varepsilon \bar{A}_2$ is similar and is therefore omitted. Thus, for $x \varepsilon \overline{A_1 \cap A_2}$ (3.25) is true.

Now let us consider the case that $x \varepsilon \overline{A_1 \cap A_2}$ and

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)r(z';x_1) = r(z';x). \tag{3.26}$$

By (3.23) and (3.26) we find that

a) $\int_{A_1} p^1_{A_1}(dx_1;x)r(z';x_1) = r(z';x)$ (3.27)

b) the remaining terms of the right hand side of (3.23) are equal to zero.

Consequently, almost surely (cf. 3.23)

$$r(z';\underline{x}_k) = \begin{cases} \int_{A_1} p^1_{A_1}(dx_{k+1};\underline{x}_k)r(z';x_{k+1}); & k = \text{even} \\ \int_{A_2} p^1_{A_2}(dx_{k+1};\underline{x}_k)r(z';x_{k+1}); & k = \text{odd}. \end{cases} \quad (3.28)$$

If $\underline{x}_k \varepsilon \bar{A}_1$ (k=even) or if $\underline{x}_k \varepsilon \bar{A}_2$ (k=odd), then since $A_1 \varepsilon K_{z'}$ and $A_2 \varepsilon K_{z'}$, we have almost surely

$$\underline{x}_k \varepsilon \{x \mid \int_{A_1} p^1_{A_1}(dy;x)r(z';y) = r(z';x); \; \int_{A_1} p^1_{A_1}(dy;x)c(z';y) \leqq c(z';x)\} \;; \quad k = \text{even} \quad (3.29)$$

$$\underline{x}_k \varepsilon \{x \mid \int_{A_2} p^1_{A_2}(dy;x)r(z';y) = r(z';x); \; \int_{A_2} p^1_{A_2}(dy;x)c(z';y) \leqq c(z';x)\}; \quad k = \text{odd}. \quad (3.30)$$

By substituting $q(x) = c(z';x)$ in (3.18) we find for $x \varepsilon \bar{A}_1$

$$\int_{A_1 \cap A_2} p^1_{A_1 \cap A_2}(dx_1;x)c(z';x_1) = \int_{A_1} p^1_{A_1}(dx_1;x)c(z';x_1) + \int_{\bar{A}_2} p^1_{A_1}(dx_1;x)\left[\int_{A_2} p^1_{A_2}(dx_2;x_1)c(z';x_2) - c(z';x_1)\right] + \int_{\bar{A}_2} p^1_{A_1}(dx_1;x) \int_{\bar{A}_1} p^1_{A_2}(dx_2;x_1)\left[\int_{A_1} p^1_{A_1}(dx_3;x_2)c(z';x_3) - c(z';x_2)\right] + \dots \quad (3.31)$$

If (3.26) is true, by (3.29) and (3.30) we find for $x \varepsilon \bar{A}_1$

$$\int_{A_1\cap A_2} p^1_{A_1\cap A_2}(dx_1;x)c(z';x_1) \leq \int_{A_1} p^1_{A_1}(dx_1;x)c(z';x_1) \leq c(z';x). \tag{3.32}$$

If $x \in \bar{A}_2$ and if x satisfies (3.26), then the proof of (3.32) is similar.

Thus, if $x \in \overline{A_1\cap A_2}$, we have either

$$\int_{A_1\cap A_2} p^1_{A_1\cap A_2}(dx_1;x)r(z';x_1) < r(z';x) \tag{3.33}$$

or

$$\int_{A_1\cap A_2} p^1_{A_1\cap A_2}(dx_1;x)r(z';x_1) = r(z';x) \tag{3.34}$$

and

$$\int_{A_1\cap A_2} p^1_{A_1\cap A_2}(dx_1;x)c(z';x_1) \leq c(z';x). \tag{3.35}$$

Hence, $A_1\cap A_2 \in K_{z'}$.

This ends the proof.

Note that lemma 3.2 does <u>not</u> imply $A'_{z'} \in K_{z'}$.

<u>Tool</u> f) If $z \in Z$, <u>the minimizing subset $D_z(x)$</u> of D(x) is defined by

$$D_z(x) \overset{def}{=} \{d \mid d \in D(x);\ r(dz;x) = \min_{d' \in D(x)} r(d'z;x)\}. \tag{3.36}$$

<u>Lemma 3.3</u>

If $z \in Z$, the minimizing subset $D_z(x)$ is a closed non empty set.

<u>Proof</u>:

From lemma 2.1 and point 2^b of property 3, it follows that the d-function r(dz;x) is bounded and continuous in d. Since the set D(x) is closed, the assertion is obvious.

<u>Tool</u> g) If $z \in Z$, the <u>A-compressed strategy of</u> $[A]z$ <u>of z</u> is defined by

$$[A]z \overset{\text{def}}{=} \begin{cases} z(x), & \text{if } x \in A \\ \text{null decisions}, & \text{if } x \in \bar{A} \end{cases} \quad (3.37)$$

Tool h) If $z \in Z$ and if $x \in X$, the minimizing subset $Z_z(x)$ of $\hat{Z}$ is defined by

$$Z_z(x) \overset{\text{def}}{=} \{z^* \mid z^* \in \hat{Z}; r((z^*)z;x) = \inf_{\bar{z} \in \hat{Z}} r((\bar{z})z;x)\} \ . \quad (3.38)$$

2. The basic problems and theorems

In this section we consider three decision problems.

First problem

To minimize the d-function $c(dz;x)$ for each x subject to the constraint $d \in D_z(x)$.

By lemma 3.3 the set $D_z(x)$ is a closed set. By point 2^b of property 3 and by the points 3) and 4) of property 5 it follows from the definitions that the d-functions $k(x;d)$ and $t(x;d)$ are bounded and continuous in d. Now it can easily be verified (cf.(2.94),(2.105),(3.12)) that the d-function $c(dz;x)$ is also bounded and continuous in d. This implies that for each x at least one decision $d \in D_z(x)$ can be found that minimizes $c(dz;x)$. Such a decision is called a minimizing decision.

The solution of the first problem is not necessarily unique. Therefore, we introduce the following property:

Property 6 (strategies)

9) If $z \in Z$, a selection procedure exists such that
   a) to each x one and only one minimizing decision $d_{z;x}$ is assigned;
   b) if for some x the decision $d=z(x)$ is a minimizing decision, then $d_{z;x} = z(x)$;
   c) the strategy z_1 , defined by

$$z_1(x) = d_{z;x} \quad (3.39)$$

belongs to $\hat{Z}$.

Lemma 3.4

If $z \in Z$ and if for some x the null decision is a minimizing decision, then the decision z(x) is also a minimizing decision.

Proof:

If d is a null decision, then

$$r(dz;x) = r(z;x) = r(z(x)z;x) \quad (3.40)$$

$$c(dz;x) = c(z;x) = c(z(x)z;x). \quad (3.41)$$

The assertion follows now at once.

Lemma 3.4 and point 9[b)] of property 6 imply

$$A_{z_1} \supset A_z \ . \quad (3.42)$$

Lemma 3.5

If z_1 is the solution of the first problem, then for each x we have either

$$r((z_1)z;x) < r(z;x) \quad (3.43)$$

or

$$r((z_1)z;x) = r(z;x) \quad (3.44)$$

and

$$c((z_1)z;x) \leq c(z;x). \quad (3.45)$$

Proof:

Let the set B be defined by

$$B \overset{def}{=} \{x \mid r((z_1)z;x) < r(z;x)\} \cup$$

$$\cup \{x \mid r((z_1)z;x) = r(z;x); c((z_1)z;x) \leq c(z;x)\} \ . \quad (3.46)$$

If follows from the definition of the strategy z_1 that

$$B \supset A_{z_1} \supset A_z. \quad (3.47)$$

If $x \in \bar{B}$, then by (3.47)

$$r(z;x) = \int_{A_{z_1}} p^1_{A_{z_1}}(dy;x)\,r(z;y) \;\geqq$$

$$\geqq \int_{A_{z_1}} p^1_{A_{z_1}}(dy;x)\,r((z_1)z;y) = r((z_1)z;x). \tag{3.48}$$

If $x \in \bar{B}$ and if $r(z;x) = r((z_1)z;x)$, then

$$c(z;x) = \int_{A_{z_1}} p^1_{A_{z_1}}(dy;x)\,c(z;y) \;\geqq$$

$$\geqq \int_{A_{z_1}} p^1_{A_{z_1}}(dy;x)\,c((z_1)z;y) = c((z_1)z;x). \tag{3.49}$$

Consequently, if $x \in \bar{B}$, we have either

$$r(z;x) > r((z_1)z;x) \tag{3.50}$$

or

$$r(z;x) = r((z_1)z;x) \tag{3.51}$$

and

$$c(z;x) \geqq c((z_1)z;x). \tag{3.52}$$

It follows from (3.46), (3.50), (3.51) and (3.52) that $\bar{B} = \emptyset$. This ends the proof.

Second problem

To determine for the strategy $z' = (z_1)z$ the subset $A'_{z'}$ of A_{z_1}.

Property 6 (strategies)

10) If $z \in Z$, if z_1 is the solution of the first problem with z and if $z' = (z_1)z$, then (cf. tool e)

$$z_2 \overset{\text{def}}{=} [A'_{z'}]\; z_1 \in Z. \tag{3.53}$$

Lemma 3.6

If z_1 is the solution of the first problem with z and if the sets A_1 and A_2 belong to $K_{z'}$, $(z'=(z_1)z)$ then we have either

$$r(A_1 \cap A_2 z';x) < r(A_i z';x) \;;\quad i=1,2 \tag{3.54}$$

or

$$r(A_1 \cap A_2 z';x) = r(A_i z';x) \quad ; \quad i=1,2 \tag{3.55}$$

and

$$c(A_1 \cap A_2 z';x) \leq c(A_i z';x) \quad ; \quad i=1,2 \tag{3.56}$$

<u>Proof:</u>

Let us first consider the case i=1.

If $x \in A_1$, then

$$r(A_1 z';x) = r(z';x) \tag{3.57}$$

$$c(A_1 z';x) = c(z';x) \; . \tag{3.58}$$

The assertion is a direct consequence of $A_1 \cap A_2 \in K_{z'}$ (lemma 3.2). If $x \in \overline{A}_1$, the assertion follows at once from (3.25) and (3.32).
The proof for i=2 is similar and is therefore omitted.

This ends the proof.

<u>Property 6</u> (strategies)

11) If $z \in Z$, if z_1 is the solution of the first problem with z and if $z' = (z_1)z$, then

$$A'_{z'} \in K_{z'} \, . \tag{3.59}$$

<u>Lemma 3.7</u>

If z_1 is the solution of the first problem with z, if z_2 is the solution of the second problem with $z'=(z_1)z$ and if $A \in K_{z'}$, then we have either

$$r((z_2)z;x) \leq r(Az';x) < r(z';x) \tag{3.60}$$

or

$$r((z_2)z;x) = r(Az';x) = r(z';x) \tag{3.61}$$

and

$$c((z_2)z;x) \leq c(Az';x) \leq c(z';x) \, . \tag{3.62}$$

<u>Proof:</u>

Since for each $A \in K_{z'}$ we have

$$z' = Xz' \tag{3.63}$$

$$Az' = A \cap Xz' \tag{3.64}$$

$$(z_2)z = A'_{z'}, z' = A'_{z'} \cap Az' = A'_{z'} \cap A \cap Xz' \tag{3.65}$$

the assertion follows at once from lemma 3.6.

This ends the proof.

Lemma 3.8

If z_1 is the solution of the first problem with z and if z_2 is the solution of the second problem with $z'=(z_1)z$, then we have either

$$r((z_2)z;x) < r(z;x) \tag{3.66}$$

or

$$r((z_2)z;x) = r(z;x) \tag{3.67}$$

and

$$c((z_2)z;x) \leqq c(z;x). \tag{3.68}$$

Proof :

The assertion follows at once from lemmas (3.5) and (3.7).

Lemma 3.9

If z_1 is the solution of the first problem with z and if z_2 is the solution of the second problem with $z'=(z_1)z$, then for each x we have

$$r(z_2;x) \leqq r(z;x). \tag{3.69}$$

Proof:

It follows from (3.66) and (3.67) that for all x we have

$$r((z_2)z;x) \leqq r(z;x). \tag{3.70}$$

Consequently, (cf.(3.7)) for all x

$$r((z_2)^k z;x) = E\{r((z_2)z;\underline{y}_{k-1})\,|\,x;z_2\} \leqq$$

$$\leqq E\{r(z;\underline{y}_{k-1})\,|\,x;z_2\} = r((z_2)^{k-1}z;x), \tag{3.71}$$

where $\underline{y}_{k-1}$ is the state into which the system is transferred by the $(k-1)^{th}$ intervention according to z_2. Hence, for all x and $k \geqq 1$

$$r((z_2)^k z;x) \leqq r((z_2)z;x) \leqq r(z;x). \tag{3.72}$$

First we shall prove that, if for some $x_1 \varepsilon A_z$ we have

$$r((z_2)^k z;x_1) = r((z_2)z;x_1); \quad k=1,2,\ldots \tag{3.73}$$

we also have

$$c((z_2)^2 z;x_1) \leqq c((z_2)z;x_1) \tag{3.74}$$

and

$$r((z_2)^2 z;x_1) = E\{r((z_2)z;\underline{I}_\infty)|x_1;z_2\} , \tag{3.75}$$

where $\underline{I}_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_{z_2} corresponding to x_1.

Let us suppose that (3.73) holds, then

$$\int_X z_2(dy;x_1)r((z_2)^{k-1}z;y) = \\ = \int_X z_2(dy;x_1)r(z;y). \tag{3.76}$$

It follows from (3.72) and (3.76) that with respect to the probability distribution $z_2(x_1)$ we have almost surely

$$r((z_2)^{k-1}z;\underline{y}) = r(z;\underline{y}) \ ; \quad k=2,3,\ldots \ . \tag{3.77}$$

By (3.67) and (3.68) we find with respect to the probability distribution $z_2(x_1)$ almost surely

$$c((z_2)z;\underline{y}) \leqq c(z;\underline{y}) \tag{3.78}$$

and thus

$$E\{c((z_2)z;\underline{y})|x_1;z_2\} \leqq E\{c(z;\underline{y})|x_1;z_2\} \ . \tag{3.79}$$

By (3.73) we have for k=2

$$r((z_2)^2 z;x_1) = r((z_2)z;x_1) \tag{3.80}$$

and thus

$$k(x_1;z_2(x_1)) - r((z_2)^2z;x_1)t(x_1;z_2(x_1)) =$$

$$= k(x_1;z_2(x_1)) - r((z_2)z;x_1)t(x_1;z_2(x_1)). \qquad (3.81)$$

By adding (3.79) and (3.81) we find (cf. (3.10))

$$c((z_2)^2z;x_1) = c((z_2)(z_2)z;x_1) \leqq c((z_2)z;x_1). \qquad (3.82)$$

Consequently, (3.74) is true.

Further it follows from (3.73) that (cf. lemma 2.1)

$$r((z_2)^2z;x_1) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} r((z_2)^jz;x_1) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \int_{A_{z_2}} p^j_{A_{z_2}}(dI;x_1;z_2)r((z_2)z;I) =$$

$$= \int_{A_{z_2}} p_{A_{z_2}}(dI;x_1;z)r((z_2)z;I) = E\{r((z_2)z;\underline{I}_\infty)|x_1;z_2\}, \qquad (3.83)$$

which gives us (3.75).

Let us now return to (3.72). Obviously we have for all x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z_2)^iz;x) \leqq r((z_2)z;x). \qquad (3.84)$$

For each integer $j \geqq 1$ and for all x we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z_2)^iz;x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z_2)^{i+j-1}z;x) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{A_{z_2}} p^i_{A_{z_2}}(dI;x;z_2)r((z_2)^jz;I) =$$

$$= \int_{A_{z_2}} p_{A_{z_2}}(dI;x;z_2)r((z_2)^jz;I) = E\{r((z_2)^jz;\underline{I}_\infty)|x;z_2\}, \qquad (3.85)$$

where $\underline{I}_\infty$ now obeys the stationary absolute probability distribution in A_{z_2} corresponding to x.

From (3.72), (3.84) and (3.85) we can easily deduce for all x

$$E\{r((z_2)z;\underline{I}_\infty)|x;z_2\} \leq r(z;x). \qquad (3.86)$$

By taking j=k and j=1 in (3.85) for all x

$$E\{r((z_2)^k z;\underline{I}_\infty)|x;z_2\} = E\{r((z_2)z;\underline{I}_\infty)|x;z_2\}. \qquad (3.87)$$

By (3.72) and (3.87) with respect to the probability distribution of $\underline{I}_\infty$ almost surely

$$r((z_2)^k z;\underline{I}_\infty) = r((z_2)z;\underline{I}_\infty). \qquad (3.88)$$

Since (3.73) implies (3.74) and (3.75) we obtain from (3.88) almost surely

$$c((z_2)^2 z;\underline{I}_\infty) \leq c((z_2)z;\underline{I}_\infty) \qquad (3.89)$$

$$r((z_2)^2 z;\underline{I}_\infty) = E\{r((z_2)z;\underline{I}'_\infty)|\underline{I}_\infty;z_2\}, \qquad (3.90)$$

where $\underline{I}'_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_{z_2} with initial state I_∞. The expected value of $r((z_2)z;\underline{I}'_\infty)$ is then a function of I_∞ and if we substitute the random state $\underline{I}_\infty$ in this function the random variable in the right hand side of (3.90) is obtained.

If the original initial state of the Markov process in A_{z_2} is an <u>ergodic</u> state, then $\underline{I}'_\infty$ has the same distribution as $\underline{I}_\infty$ for almost all $I_\infty \in A_{z_2}$ and thus we have almost surely

$$r((z_2)^2 z;\underline{I}_\infty) = E\{r((z_2)z;\underline{I}'_\infty)|\underline{I}_\infty;z_2\} =$$

$$= E\{r((z_2)z;\underline{I}_\infty)|x;z_2\}. \qquad (3.91)$$

Now it follows from (3.89) that for all x almost surely

$$c((z_2)^2 z;\underline{I}_\infty) =$$

$$= k(\underline{I}_\infty;z_2(\underline{I}_\infty)) - r((z_2)^2 z;\underline{I}_\infty)t((\underline{I}_\infty;z_2(\underline{I}_\infty)) +$$

$$+ E\{c((z_2)z;\underline{I}'_2)|\underline{I}_\infty;z_2\} \leq c((z_2)z;\underline{I}_\infty), \qquad (3.92)$$

where $\underline{I}_2'$ obeys the probability distribution of the second intervention state in A_{z_2} if I_∞ is the initial (intervention) state.

If the original initial state x of the Markov process in A_{z_2} is an ergodic state, then (3.92) becomes almost surely (cf.(3.91))

$$k(\underline{I}_\infty;z_2(\underline{I}_\infty)) - E\{r((z_2)z;\underline{I}_\infty)|x;z_2\}\ t(\underline{I}_\infty;z_2(\underline{I}_\infty)) + \\ + E\{c((z_2)z;\underline{I}_2')|\underline{I}_\infty;z_2\} \leqq c((z_2)z;\underline{I}_\infty). \qquad (3.93)$$

It can easily be verified that for x ergodic

$$E_{\underline{I}_\infty}\{E_{\underline{I}_2'}\{c((z_2)z;\underline{I}_2')|\underline{I}_\infty;z_2\}\ |x;z_2\} = \\ = E_{\underline{I}_\infty}\{c((z_2)z;\underline{I}_\infty)|x;z_2\}\ . \qquad (3.94)$$

Consequently, by taking expectations in (3.93) we find for x ergodic

$$E\{k(\underline{I}_\infty;z_2(\underline{I}_\infty))|x;z_2\} + \\ - E\{r((z_2)z;\underline{I}_\infty)|x;z_2\}\ E\{t(\underline{I}_\infty;z_2(\underline{I}_\infty))|x;z_2\} \leqq 0. \qquad (3.95)$$

By (2.46), (2.47), (2.49) and (3.53) we have

$$E\{t(\underline{I}_\infty;z_2(\underline{I}_\infty))|x;z_2\} > 0 \qquad (3.96)$$

and thus for x ergodic

$$r(z_2;x) = \frac{E\{k(\underline{I}_\infty;z_2(\underline{I}_\infty))|x;z_2\}}{E\{t(\underline{I}_\infty;z_2(\underline{I}_\infty))|x;z_2\}} \leqq E\{r((z_2)z;\underline{I}_\infty)|x;z_2\}\ . \qquad (3.97)$$

Now let x be an arbitrary state. Then the state $\underline{I}_\infty$ will nevertheless almost surely be ergodic. Hence, by (3.97) with respect to the stationary absolute probability distribution of the Markov process in A_{z_2} with initial state x we find almost surely

$$r(z_2;\underline{I}_\infty) \leqq E\{r((z_2)z;\underline{I}_\infty')|\underline{I}_\infty;z_2\}\ , \qquad (3.98)$$

where $\underline{I}'_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_{z_2} with initial state I_∞.

By taking expectations in (3.98) we obtain

$$r(z_2;x) = E\{r(z_2;\underline{I}_\infty)|x;z_2\} \leq$$
$$\leq E_{\underline{I}_\infty}\{E_{\underline{I}'_\infty}\{r((z_2)z;\underline{I}'_\infty)|\underline{I}_\infty;z_2\}|x;z_2\} = E\{r((z_2)z;\underline{I}_\infty)|x;z_2\} ; \tag{3.99}$$

thus (3.97) also holds for non-ergodic states x. Hence by (3.86) and (3.99) for all x

$$r(z_2;x) \leq E\{r((z_2)z;\underline{I}_\infty)|x;z_2\} \leq r(z;x). \tag{3.100}$$

This ends the proof.

As a third problem we consider the following:

<u>Third problem</u>

To minimize the z^*-function $c((z^*)z;x)$ for each x subject to the constraint $z^* \varepsilon Z_z(x)$.

After introducing an additional property of the mathematical model we shall demonstrate that the solution of the second problem also solves the third problem.

<u>Property 6</u> (strategies)

12) There exists a strategy $z_3 \varepsilon \hat{Z}$, which is for <u>each</u> x a solution of the third problem.

<u>Lemma 3.9</u>

If z_1 is the solution of the first problem with z and if z_2 is the solution of the second problem with $z' = (z_1)z$, then z_2 is for each x also a solution of the third problem.

<u>Proof</u>:

If z_3 satisfies point 12) of property 6 and if for some x the decision $z_3(x)$ is an intervention, then we have

$$r((z_3)z;x) = r((z_1)z;x) = r((z_2)z;x) \qquad (3.101)$$
$$c((z_3)z;x) = c((z_1)z;x) = c((z_2)z;x). \qquad (3.102)$$

If $z_3(x)$ is a null decision we have either

$$r((z_3)z;x) < r((z_1)z;x) \qquad (3.103)$$

or

$$r((z_3)z;x) = r((z_1)z;x) \qquad (3.104)$$

and

$$c((z_3)z;x) \leqq c((z_1)z;x). \qquad (3.105)$$

If (3.103) is true, then by (3.101)

$$r(z';x) > r((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r((z_3)z;y) =$$
$$= \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y). \qquad (3.106)$$

If

$$r(z';x) = r((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y), \qquad (3.107)$$

then by (3.102), (3.104) and (3.105)

$$c(z';x) \geqq c((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)c((z_3)z;y) =$$
$$= \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)c(z';y). \qquad (3.108)$$

Consequently, if $x \in \bar{A}_{z_3}$, then we have either

$$r(z';x) > \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y) \qquad (3.109)$$

or

$$r(z';x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y) \qquad (3.110)$$

and

$$c(z';x) \geqq \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)c(z';y). \qquad (3.111)$$

Thus $A_{z_3} \varepsilon K_{z'}$.

Hence, we have either (cf. lemma 3.7)

$$r((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y) > r((z_2)z;x) \qquad (3.112)$$

or

$$r((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)r(z';y) = r((z_2)z;x) \qquad (3.113)$$

and $$c((z_3)z;x) = \int_{A_{z_3}} p^1_{A_{z_3}}(dy;x)c(z';y) \geqq c((z_2)z;x). \qquad (3.114)$$

Thus $$r((z_3)z;x) = r((z_2)z;x) \qquad (3.115)$$

$$c((z_3)z;x) = c((z_2)z;x). \qquad (3.116)$$

This ends the proof.

Since only <u>one</u> solution of the second problem exists the converse of lemma 3.9 is <u>not</u> true in general.

<u>Theorem 8</u>

If $z_{o;3}$ is a solution of the third problem with $z=z_o$ and if $z_{o;3}=z_o$, then for each x

$$r(z_o;x) = \min_{z \,\varepsilon\, Z} r(z;x). \qquad (3.117)$$

<u>Proof</u>:

Since $z_{o;3}=z_o$ we find for each $z \,\varepsilon\, Z$ and $x \,\varepsilon\, X$ either

$$r((z)z_o;x) > r(z_o;x) \qquad (3.118)$$

or $$r((z)z_o;x) = r(z_o;x) \qquad (3.119)$$

and $$c((z)z_o;x) \geqq c(z_o;x). \qquad (3.120)$$

By (3.118) and (3.119) we have

$$r((z)^k z_o;x) = E\{r(z_o;\underline{y}_k)|x;z\} \leqq$$

$$\leqq E\{r((z)z_o;\underline{y}_k)|x;z\} = r((z)^{k+1}z_o;x), \qquad (3.121)$$

where $\underline{y}_k$ is the state into which the system is transferred by the k^{th} intervention. Thus

$$r(z_o;x) \leqq r((z)z_o;x) \leqq r((z)^k z_o;x), \qquad (3.122)$$

$$r(z_o;x) \leqq \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} r((z)^k z_o;x) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \int_{A_z} p^k_{A_z}(dI;x;z)r((z)z_o;I) =$$

$$= E\{r((z)z_o;\underline{I}_\infty)|x;z\} , \qquad (3.123)$$

where $\underline{I}_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_z corresponding to x.

First we shall prove that, if for some $x_1 \varepsilon A_z$ we have

$$r((z)^k z_o;x_1) = r((z)z_o;x_1); \quad k=1,2,\dots , \qquad (3.124)$$

we also have

$$c((z)^2 z_o;x_1) \geqq c((z)z_o;x_1) \qquad (3.125)$$

and

$$r((z)^2 z_o;x_1) = E\{r((z)z_o;\underline{I}_\infty)|x_1;z\} , \qquad (3.126)$$

where $\underline{I}_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_z corresponding to x_1.

Let us suppose that (3.124) holds, then

$$\int_X z(dy;x_1)r((z)^{k-1}z_o;y) = \int_X z(dy;x_1)r(z_o;y). \quad (3.127)$$

It follows from (3.122) and (3.127) that with respect to the probability distribution $z(x_1)$ we have almost surely

$$r((z)^{k-1}z_o;\underline{y}) = r(z_o;\underline{y}); \quad k=2,3,\dots . \qquad (3.128)$$

Since $z_{0;3} = z_o$ it follows from (3.128) with k=2 that almost surely

$$c((z)z_o;\underline{y}) \geqq c(z_o;\underline{y}) \tag{3.129}$$

and thus

$$E\{c((z)z_o;\underline{y})|x_1;z\} \geqq E\{c(z_o;\underline{y})|x_1;z\} . \tag{3.130}$$

By (3.124) we have for k=2

$$r((z)^2 z_o;x_1) = r((z)z_o;x_1) \tag{3.131}$$

and thus

$$k(x_1;z(x_1)) - r((z)^2 z_o;x_1)t(x_1;z(x_1)) =$$

$$= k(x_1;z(x_1)) - r((z)z_o;x_1)t(x_1;z(x_1)). \tag{3.132}$$

By adding (3.130) and (3.132) we find (cf. 3.10)

$$c((z)^2 z_o;x_1) \geqq c((z)z_o;x_1). \tag{3.133}$$

Consequently, (3.125) is true.

Further it follows from (3.124) that (cf. lemma 2.2)

$$r((z)^2 z_o;x_1) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} r((z)^j z_o;x_1) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \int_{A_z} p^j_{A_z}(dI;x_1;z)r((z)z_o;I) =$$

$$= \int_{A_z} p_{A_z}(dI;x_1;z)r((z)z_o;I) = E\{r((z)z_o;\underline{I}_\infty)|x_1;z\} , \tag{3.134}$$

which gives us (3.126).

Let us return to (3.120). Obviously we have for all x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z)^i z_o;x) \geqq r((z)z_o;x). \tag{3.135}$$

For each integer $j \geqq 1$ and for all x we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z)^i z_o;x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r((z)^{i+j-1} z_o;x) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{A_z} p^i_{A_z}(dI;x;z) r((z)^j z_o;I) =$$

$$= \int_{A_z} p_{A_z}(dI;x;z) r((z)^j z_o;I) = E\{r((z)^j z_o;\underline{I}_\infty)|x;z\}, \qquad (3.136)$$

where $\underline{I}_\infty$ now obeys the stationary absolute probability distribution in A_z corresponding to x.

By taking j=k and j=1 for all x

$$E\{r((z)^k z_o;\underline{I}_\infty)|x;z\} = E\{r((z)z_o;\underline{I}_\infty)|x;z\} . \qquad (3.137)$$

By (3.122) and (3.137) with respect to the probability distribution of $\underline{I}_\infty$ almost surely

$$r((z)^k z_o;\underline{I}_\infty) = r((z)z_o;\underline{I}_\infty). \qquad (3.138)$$

Since (3.124) implies (3.125), we obtain from (3.138) almost surely

$$c((z)^2 z_o;\underline{I}_\infty) \geq c((z)z_o;\underline{I}_\infty) \qquad (3.139)$$

$$r((z)^2 z_o;\underline{I}_\infty) = E\{r((z)z_o;\underline{I}'_\infty)|\underline{I}_\infty;z\} , \qquad (3.140)$$

where $\underline{I}'_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_z, with initial state I_∞ . The expected value of $r((z)z_o;\underline{I}'_\infty)$ is then a function of I_∞ and if we substitute the random state $\underline{I}_\infty$ in this function the random variable in the right hand side of (3.140) is obtained.

If the original initial state x of the Markov process in A_z is an ergodic state, then $\underline{I}'_\infty$ has the same distribution as $\underline{I}_\infty$ for almost all $I_\infty \in A_z$ and thus we have almost surely

$$r((z)^2 z_o;\underline{I}_\infty) = E\{r((z)z_o;\underline{I}'_\infty)|\underline{I}_\infty;z\} =$$

$$= E\{r((z)z_o;\underline{I}_\infty)|x;z\} . \qquad (3.141)$$

Now it follows from (3.139) that for all x almost surely

$$c((z)^2 z_o;\underline{I}_\infty) = k(\underline{I}_\infty;z(\underline{I}_\infty))-r((z)^2 z_o;\underline{I}_\infty)t(\underline{I}_\infty;z(\underline{I}_\infty)) +$$

$$+ E\{c((z)z_o;\underline{I}_2')|\underline{I}_\infty;z\} \geqq c((z)z_o;\underline{I}_\infty), \qquad (3.142)$$

where $\underline{I}_2'$ obeys the probability distribution of the second intervention state in A_z if I_∞ is theinitial (intervention) state.

If the original initial state x of the Markov process in A_z is an ergodic state, then (3.142) becomes almost surely (cf. (3.141))

$$k(\underline{I}_\infty;z(\underline{I}_\infty)) - E\{r((z)z_o;\underline{I}_\infty)|x;z\}\ t(\underline{I}_\infty;z(\underline{I}_\infty)) +$$

$$+ E\{c((z)z_o;\underline{I}_2')|\underline{I}_\infty;z\} \geqq c((z)z_o;\underline{I}_\infty). \qquad (3.143)$$

It can be easily verified that for x ergodic

$$E_{\underline{I}_\infty}\{E_{\underline{I}_2'}\{c((z)z_o;\underline{I}_2')|\underline{I}_\infty;z\}|x;z\} =$$

$$= E_{\underline{I}_\infty}\{c((z)z_o;\underline{I}_\infty)|x;z\} . \qquad (3.144)$$

Consequently, by taking expectations in (3.143) we find for x ergodic

$$E\{k(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\} +$$

$$- E\{r((z)z_o;\underline{I}_\infty)|x;z\} . E\{t(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\} \geqq 0 \qquad (3.145)$$

or

$$r(z;x) = \frac{E\{k(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\}}{E\{t(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\}} \geqq$$

$$\geqq E\{r((z)z_o;\underline{I}_\infty)|x;z\} . \qquad (3.146)$$

Now let x be an arbitrary state. Then the state $\underline{I}_\infty$ will nevertheless almost surely be ergodic. Hence, by (3.146) with respect to the absolute stationary probability distribution of the Markov process in A_z with initial state x we find almost surely

$$r(z;\underline{I}_{\infty}) \geq E\{r((z)z_o;\underline{I}'_{\infty})|\underline{I}_{\infty};z\} \ , \qquad (3.147)$$

where $\underline{I}'_{\infty}$ obeys the stationary absolute probability distribution of the Markov process in A_z with initial state I_{∞}.

By taking expectations in (3.147) we obtain

$$\begin{aligned} r(z;x) &= E\{r(z;\underline{I}_{\infty})|x;z\} \geq \\ &\geq E_{\underline{I}_{\infty}}\{E_{\underline{I}'_{\infty}}\{r((z)z_o;\underline{I}'_{\infty})|\underline{I}_{\infty};z\}|x;z\} = \\ &= E\{r((z)z_o;\underline{I}_{\infty})|x;z\}. \end{aligned} \qquad (3.148)$$

Thus (3.146) also holds for non-ergodic states x. It follows from (3.123) and (3.148) that for all x

$$r(z_o;x) \leq E\{r((z)z_o;\underline{I}_{\infty})|x;z\} \leq r(z;x). \qquad (3.149)$$

This ends the proof.

By lemma 3.9 and theorem 8 we also have :

Theorem 9

If $z_{0;1}$ is the solution of the first problem with z_o, if $z_{0;2}$ is the solution of the second problem with $z'_o = (z_{0;1})z_o$ and if $z_{0;2}=z_o$, then for each x

$$r(z_o;x) = \min_{z \in Z} r(z;x). \qquad (3.150)$$

3. Optimal strategies and the strategy improvement routines

The most plausible definition of an optimal strategy is the following:

A strategy z_o is called optimal if it minimizes for each x the z-function $r(z;x)$. In other words:

$$r(z_o;x) = \min_{z \in Z} r(z;x). \qquad (3.151)$$

By theorem 8 a strategy z'_o is optimal if it satisfies

$$c(z_o';x) = \min_{z \,\varepsilon\, Z_{z_o'}(x)} c((z)z_o';x). \qquad (3.152)$$

By theorem 9 a strategy z_o'' is optimal if it satisfies

$$c(z_o'';x) = \min_{d \,\varepsilon\, D_{z_o}(x)} c((d)z_o'';x) \qquad (3.153)$$

and

$$A'_{z_o''} = A_{z_o''} . \qquad (3.154)$$

Note that z_o'' also satisfies (3.152) and (3.151), while in general z_o' does not satisfy (3.153) and (3.154).

Without more detailed information about the mathematical model we cannot prove the existence and the uniqueness of the solution of the above equations.

If an optimal strategy z_o can be obtained neither from (3.151), nor from (3.152) and nor from (3.153) and (3.154), for practical purposes an iteration procedure is required, that yields a sequence of strategies $\{z^i;i=1,2,...\}$ such that for each x

$$\lim_{n \to \infty} r(z^n;x) = \inf_{z \,\varepsilon\, Z} r(z;x). \qquad (3.155)$$

The functional equation (3.152) suggests the following iteration procedure. This iteration procedure, called the strategy improvement method I, starts with an arbitrary strategy $z^o \,\varepsilon\, Z$. If z^{i-1} is the strategy obtained at the end of the $(i-1)^{th}$ cycle, the i^{th} cycle runs as follows:

First step

Determine a solution of the functional equations (3.3) and (3.4) with $z=z^{i-1}$.

Second step

Solve the third problem with $z=z^{i-1}$ and determine a strategy $z_3 = z_3^{i-1}$.

The strategy z^i is given by z_3^{i-1}.

End of the i^{th} cycle.

The equations (3.153) and (3.154) suggest the following iteration procedure. This iteration procedure, called the strategy improvement method II, starts with an arbitrary strategy $z^o \varepsilon Z$. If z^{i-1} is the strategy obtained at the end of the $(i-1)^{th}$ cycle the i^{th} cycle runs as follows:

First step

Determine a solution of the functional equations (3.3) and (3.4) with $z=z^{i-1}$.

Second step

Solve the first problem with $z=z^{i-1}$ and determine the strategy $z_1 = z_1^{i-1}$.

Third step

Determine the x-functions $r((z_1^{i-1})z^{i-1};x)$ and $c((z_1^{i-1})z^{i-1};x)$.

Fourth step

Solve the second problem with $z'=(z_1^{i-1})z^{i-1}$ and determine the strategy $z_2=z_2^{i-1}$.

The strategy z^i is given by z_2^{i-1}.

End of the i^{th} cycle.

Note that the strategy improvement method II is a special case of strategy improvement method I. In the second routine the third problem is solved in a prescribed way.

In order to prove the effectiveness of these iteration routines we introduce the following property :

Property 6 (strategies)

13) There is at least one initial strategy $z^o \varepsilon Z$ and an integer M_o such that for each $z \varepsilon Z$, for each x and for each $i \geqq M_o$ we have

$$r((z)z^{i};x) \geq r(z^{i};x). \qquad (3.156)$$

If for all $z \in Z$ the decision process has only one simple ergodic set, point 13) of property 6 is always satisfied with the equality sign. In that case the x-function $r(z^{i};x)$ is a constant. This holds for every $z^{o} \in Z$.

Theorem 10

If $z \in Z$ is an arbitrary strategy and if z^{o} obeys point 13) of property 6, the strategy improvement method starting from z^{o} generates a sequence of strategies $\{z^{i};i=1,2,\ldots\}$, satisfying for each x

$$r(z;x) \geq \lim_{n \to \infty} r(z^{n};x). \qquad (3.157)$$

Proof :

If $i \geq M_{o}$, by point 13) of property 6 we have for each x

$$r((z)z^{i};x) \geq r(z^{i};x). \qquad (3.158)$$

Consequently, for all x and $k \geq 2$

$$r((z)^{k}z^{i};x) = E\{r((z)z^{i};\underline{y}_{k-1})|x;z\} \geq$$

$$\geq E\{r(z^{i};\underline{y}_{k-1})|x;z\} = r((z)^{k-1}z^{i};x), \qquad (3.159)$$

where $\underline{y}_{k-1}$ is the state into which the system is transferred owing to the $(k-1)^{th}$ intervention according to z.

Hence, for all x and $k \geq 1$

$$r((z)^{k}z^{i};x) \geq r((z)z^{i};x) \geq r(z^{i};x). \qquad (3.160)$$

Consequently,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} r((z)^{k}z^{i};x) =$$

$$= E\{r((z)z^{i};\underline{I}_{\infty})|x;z\} \geq r(z^{i};x), \qquad (3.161)$$

where $\underline{I}_{\infty}$ obeys the stationary absolute probability distribution in A_{z}

with initial state x.

First we shall prove that, if for some $x_1 \varepsilon A_z$

$$r((z)^k z^i;x_1) = r((z)z^i;x_1); \quad k=1,2,\ldots , \quad (3.162)$$

we also have

$$c((z)^2 z^i;x_1) \geqq c((z)(z^{i+1})z^i;x_1) \quad (3.163)$$

and

$$r((z)^2 z^i;x_1) = E\{r((z)z^i;\underline{I}_\infty)|x_1;z\} , \quad (3.164)$$

where $\underline{I}_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_z corresponding to x_1.

Let us suppose that (3.162) holds, then

$$\int_X z(dy;x_1)r((z)^{k-1}z^i;y) = \int_X z(dy;x_1)r(z^i;y). \quad (3.165)$$

It follows from (3.160) and (3.165) that with respect to the probability distribution $z(x_1)$ we have almost surely

$$r((z)^{k-1}z^i;\underline{y}) = r(z^i;\underline{y}); \quad k=2,3,\ldots \quad (3.166)$$

and thus almost surely $z \varepsilon Z_{z^i}(\underline{y})$.

It follows from the solution of the third problem that we have almost surely ($z^{i+1}=z_3^i$)

$$c((z)z^i;\underline{y}) \geqq c((z^{i+1})z^i;\underline{y}) \quad (3.167)$$

and thus

$$E\{c((z)z^i;\underline{y})|x_1;z\} \geqq E\{c((z^{i+1})z^i;\underline{y})|x_1;z\} . \quad (3.168)$$

By point 13 of property 6 we have for all x

$$r((z^{i+1})z^i;x) = r(z^i;x) \quad (3.169)$$

and thus by (3.162) and (3.169)

$$r((z)^2 z^i;x_1) = r((z)(z^{i+1})z^i;x_1). \tag{3.170}$$

So we find

$$k(x_1;z(x_1)) - r((z)^2 z^i;x_1)t(x_1;z(x_1)) =$$
$$= k(x_1;z(x_1)) - r((z)(z^{i+1})z^i;x_1)t(x_1;z(x_1)). \tag{3.171}$$

By adding (3.168) and (3.171) we find

$$c((z)^2 z^i;x_1) \geqq c((z)(z^{i+1})z^i;x_1). \tag{3.172}$$

Consequently, (3.163) is true.

Further it follows from (3.162) that

$$r((z)^2 z^i;x_1) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} r((z)^j z^i;x_1) =$$
$$= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \int_{A_z} p^j_{A_z}(dI;x_1;z)r((z)z^i;I) =$$
$$= \int_{A_z} p_{A_z}(dI;x_1;z)r((z)z^i;I) = E\{r((z)z^i;\underline{I}_\infty)|x_1;z\}, \tag{3.173}$$

which gives us (3.164).

Let us return to (3.160). Obviously we have for all x

$$\lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} r((z)^h z^i;x) \geqq r((z)z^i;x). \tag{3.174}$$

For each integer $j \geqq 1$ and for all x we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} r((z)^h z^i;x) = \lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} r((z)^{h+j-1} z^i;x) =$$
$$= \lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} \int_{A_z} p^h_{A_z}(dI;x;z)r((z)^j z^i;I) =$$
$$= \int_{A_z} p_{A_z}(dI;x;z)r((z)^j z^i;I) = E\{r((z)^j z^i;\underline{I}_\infty)|x;z\}, \tag{3.175}$$

where $\underline{I}_\infty$ now obeys the stationary absolute probability distribution in A_z corresponding to x.

By taking j=k and j=1 for all x

$$E\{r((z)^k z^i;\underline{I}_\infty)|x;z\} = E\{r((z)z^i;\underline{I}_\infty)|x;z\} \quad . \tag{3.176}$$

By (3.160) and (3.176) with respect to the probability distribution of $\underline{I}_\infty$ almost surely

$$r((z)^k z^i;\underline{I}_\infty) = r((z)z^i;\underline{I}_\infty). \tag{3.177}$$

Since (3.162) implies (3.163) we obtain from (3.177) almost surely

$$c((z)^2 z^i;\underline{I}_\infty) \geqq c((z)(z^{i+1})z^i;\underline{I}_\infty) \tag{3.178}$$

$$r((z)^2 z^i;\underline{I}_\infty) = E\{r((z)z^i;\underline{I}'_\infty)|\underline{I}_\infty;z\} \quad , \tag{3.179}$$

where $\underline{I}'_\infty$ obeys the stationary absolute probability distribution of the Markov process in A_z with initial state I_∞ . The expected value of $r((z)z^i;\underline{I}'_\infty)$ is then a function of I_∞ and if we substitute the random state $\underline{I}_\infty$ in this function the random variable in the right hand side of (3.179) is obtained. If the original initial state of the Markov process in A_z is an ergodic state, then $\underline{I}'_\infty$ has the same distribution as $\underline{I}_\infty$ for almost all $I_\infty \varepsilon A_z$ and thus we have almost surely

$$r((z)^2 z^i;\underline{I}_\infty) = E\{r((z)z^i;\underline{I}'_\infty)|\underline{I}_\infty;z\} =$$
$$= E\{r((z)z^i;\underline{I}_\infty)|x;z\} \quad . \tag{3.180}$$

Now it follows from (3.178) that for all x almost surely

$$c((z)^2 z^i;\underline{I}_\infty) = k(\underline{I}_\infty;z(\underline{I}_\infty))-r((z)^2 z^i;\underline{I}_\infty)t(\underline{I}_\infty;z(\underline{I}_\infty)) +$$
$$+ E\{c((z)z^i;\underline{I}'_2)|\underline{I}_\infty;z\} \geqq c((z)(z^{i+1})z^i;\underline{I}_\infty), \tag{3.181}$$

where $\underline{I}'_2$ obeys the probability distribution of the second intervention state in A_z if I_∞ is the initial (intervention) state.

If the original initial state x of the Markov process in A_z is an ergodic state then (3.181) becomes almost surely (cf. (3.180))

$$k(\underline{I}_\infty;\ z(\underline{I}_\infty)) - E\{r((z)z^i;\underline{I}_\infty)\,|\,x;z\}\,t(\underline{I}_\infty;\ z(\underline{I}_\infty))$$

$$+\ E\{c((z)z^i;\underline{I}_2')\,|\,\underline{I}_\infty;z\} \geq c((z)(z^{i+1})z^i;\underline{I}_\infty). \qquad (3.182)$$

By taking expectations in (3.182) we find for x ergodic

$$E\{k(\underline{I}_\infty;z(\underline{I}_\infty))\,|\,x;z\}\ +$$

$$-\ E\{r((z)z^i;\underline{I}_\infty)\,|\,x;z\}\ E\{t(\underline{I}_\infty;z(\underline{I}_\infty))\,|\,x;z\}\ \geq$$

$$\geq E\{c((z)(z^{i+1})z^i;\underline{I}_\infty) - c((z)z^i;\underline{I}_\infty)\,|\,x;z\} \qquad (3.183)$$

or

$$r(z;x) = \frac{E\{k(\underline{I}_\infty;z(\underline{I}_\infty))\,|\,x;z\}}{E\{t(\underline{I}_\infty;z(\underline{I}_\infty))\,|\,x;z\}} \geq E\{r((z)z^i;\underline{I}_\infty)\,|\,x;z\}\ +$$

$$+\ \frac{E\{c((z)(z^{i+1})z^i;\underline{I}_\infty) - c((z)z^i;\underline{I}_\infty)\,|\,x;z\}}{E\{t(\underline{I}_\infty;z(\underline{I}_\infty))\,|\,x;z\}}. \qquad (3.184)$$

Now let x be an arbitrary state. Then the state $\underline{I}_\infty$ will nevertheless almost surely be ergodic. Hence, by (3.184) with respect to the stationary absolute probability distribution of the Markov process in A_z with initial state x we find almost surely

$$r(z;\underline{I}_\infty) \geq E\{r((z)z^i;\underline{I}_\infty')\,|\,\underline{I}_\infty;z\}\ +$$

$$+\ \frac{E\{c((z)(z^{i+1})z^i;\underline{I}_\infty') - c((z)z^i;\underline{I}_\infty')\,|\,\underline{I}_\infty;z\}}{E\{t(\underline{I}_\infty';z(\underline{I}_\infty'))\,|\,\underline{I}_\infty;z\}}, \qquad (3.185)$$

where $\underline{I}_\infty'$ obeys the stationary absolute probability distribution of the Markov process in A_z with initial state I_∞.

By taking expectations in (3.185) we find

$$r(z;x) = E\{r(z;\underline{I}_\infty)\,|\,x;z\} \geq E\{E\{r((z)z^i;\underline{I}_\infty')\,|\,\underline{I}_\infty;z\}\,|\,x;z\}\ +$$

$$+\ E\left\{\frac{E\{c((z)(z^{i+1})z^i;\underline{I}_\infty') - c((z)z^i;\underline{I}_\infty')\,|\,\underline{I}_\infty;z\}}{E\{t(\underline{I}_\infty';z(\underline{I}_\infty'))\,|\,\underline{I}_\infty;z\}}\,|\,x;z\right\}$$

$$= E\{r((z)z^i;\underline{I}_\infty)\,|\,x;z\}\ +$$

$$+ E\left\{ \frac{E\{c((z)(z^{i+1})z^{i};\underline{I}'_{\infty})-c((z)z^{i};\underline{I}'_{\infty})|\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))|\underline{I}_{\infty};z\}} \mid x;z \right\} \cdot \quad (3.186)$$

By (3.161) and (3.186) we have for any x' and $i \geq M_o$

$$r(z;x') \geq r(z^{i};x') +$$

$$+ E\left\{ \frac{E\{c((z)(z^{i+1})z^{i};\underline{I}'_{\infty})-c((z)z^{i};\underline{I}'_{\infty})|\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))|\underline{I}_{\infty};z\}} \mid x;z \right\} \quad (3.187)$$

and thus

$$r(z;x') \geq \frac{1}{n} \sum_{i=M_o}^{M_o+n-1} r(z^{i};x') +$$

$$\frac{1}{n} \sum_{i=M_o}^{M_o+n-1} E\left\{ \frac{E\{c((z)(z^{i+1})z^{i};\underline{I}'_{\infty})-c((z)z^{i};\underline{I}'_{\infty})|\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))|\underline{I}_{\infty};z\}} \mid x';z \right\}. \quad (3.188)$$

$$r(z;x') \geq \frac{1}{n} \sum_{i=M_o}^{M_o+n-1} r(z^{i};x') +$$

$$+ \frac{1}{n} \sum_{i=M_o+1}^{M_o+n-1} E\left\{ \frac{E\{c((z)(z^{i})z^{i-1};\underline{I}'_{\infty})-c((z)z^{i};\underline{I}'_{\infty})|\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))|\underline{I}_{\infty};z\}} \mid x';z \right\} +$$

$$+ \frac{1}{n} E\left\{ \frac{E\{c((z)(z^{M_o+n})z^{M_o+n-1};\underline{I}'_{\infty})-c((z)z^{M_o};\underline{I}'_{\infty}) \mid \underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))|\underline{I}_{\infty};z\}} \mid x';z \right\}. \quad (3.189)$$

If the x-functions $\{c(z^{i};x);\ i=1,2,\dots\}$ are any set of functions satisfying :

$$r(z^{i};x) = \int_{A_{z^{i}}} p^{1}_{A_{z^{i}}}(dI;x;z^{i})\, r(z^{i};I) \quad (3.190)$$

$$c(z^{i};x) = k(x;z^{i}(x))-r(z^{i};x)t(x;z^{i}(x)) +$$

$$+ \int_{A_{z^{i}}} p^{1}_{A_{z^{i}}}(dI;x;z^{i})\, c(z^{i};I), \quad (3.191)$$

if k_i (i=1,2,...) are arbitrary constants and if the x-functions $c'(z^i;x)$ are defined by

$$c'(z^i;x) \overset{\text{def}}{=} c(z^i;x) + k_i , \qquad (3.192)$$

then the set of x-functions $\{c'(z^i;x);\ i=1,2,...\}$ also satisfies (3.190) and (3.191).

Using the functions $c'(z^i;x)$ instead of $c(z^i;x)$, the strategies z^{i+1} $(i \geq 0)$ are still solutions of the third problem with z^i.

Consequently,

$$r(z;x') \geq \frac{1}{n} \sum_{i=M_o}^{M_o+n-1} r(z^i;x') +$$

$$\frac{1}{n} \sum_{i=M_o+1}^{M_o+n-1} E\left\{ \frac{E\{c'((z)(z^i)z^{i-1};\underline{I}'_{\infty})-c'((z)z^i;\underline{I}'_{\infty})\,|\,\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))\,|\,\underline{I}_{\infty};z\}} \,\Big|\, x';z \right\} +$$

$$+ \frac{1}{n} E\left\{ \frac{E\{c'((z)(z^{M_o+n})z^{M_o+n-1};\underline{I}'_{\infty})-c'((z)z^{M_o};\underline{I}'_{\infty})\,|\,\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))\,|\,\underline{I}_{\infty};z\}} \,\Big|\, x';z \right\}. \qquad (3.193)$$

Now let the constants k_i $(i=M_o,...,M_o+n-1)$ be chosen in such a way that

$$E\left\{ \frac{E\{c'((z)(z^i)z^{i-1};\underline{I}'_{\infty})-c'((z)z^i;\underline{I}'_{\infty})\,|\,\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))\,|\,\underline{I}_{\infty};z\}} \,\Big|\, x';z \right\} = 0;$$

$$;\ i=M_o+1,...,M_o+n-1. \qquad (3.194)$$

This implies

$$E\left\{ \frac{k_i-k_{i-1}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))\,|\,\underline{I}_{\infty};z\}} \,\Big|\, x';z \right\} =$$

$$= E\left\{ \frac{E\{c((z)(z^i)z^{i-1};\underline{I}'_{\infty})-c((z)z^i;\underline{I}'_{\infty})\,|\,\underline{I}_{\infty};z\}}{E\{t(\underline{I}'_{\infty};z(\underline{I}'_{\infty}))\,|\,\underline{I}_{\infty};z\}} \,\Big|\, x';z \right\};$$

$$;\ i=M_o+1,...,M_o+n-1. \qquad (3.195)$$

If we choose $k_{M_o} = 0$, then the remaining constants k_i $(i=M_o+1,\dots,M_o+n-1)$ are unambiguously determined by (3.195).

Hence, by this choice of the constants k_i the relation (3.193) becomes

$$r(z;x') \geq \frac{1}{n} \sum_{i=M_o}^{M_o+n-1} r(z^i;x') +$$

$$+ \frac{1}{n} E\left\{ \frac{E\{c'((z)(z^{M_o+n})z^{M_o+n-1};\underline{I}'_{-\infty}) - c'((z)z^{M_o};\underline{I}'_{-\infty}) \mid \underline{I}_{-\infty};z\}}{E\{t(\underline{I}'_{-\infty};z(\underline{I}'_{-\infty})) \mid \underline{I}_{-\infty};z\}} \right| . \qquad$$

$$\left. . \mid x';z \right\} . \qquad (3.196)$$

Finally, if $n \to \infty$, we find

$$r(z;x') \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=M_o}^{M_o+n-1} r(z^i;x') = \lim_{n\to\infty} r(z^n;x'). \qquad (3.197)$$

This ends the proof.

CHAPTER 4

A new method and some related techniques

1. Introduction

The results obtained in the preceding chapters furnish us with a new method for solving stochastic ∞-stage decision problems. Two formulations of this method will be described in section 2.

In some books the theory dealing with problems of this type is considered as a part of dynamic programming. In this book, however, the term dynamic programming is reserved for techniques treated by RICHARD BELLMAN in his book "Dynamic Programming" [3] . In section 3 of this chapter we shall give a brief description of a dynamic programming approach to the problems considered in this book.

Two decision problems of this type have been solved by RONALD A. HOWARD [4] . In section 4 we shall show that his first policy improvement technique follows from the second formulation of the new method. With respect to his second problem, however, our method leads to a technique different from his.

2. Summary of the new method

In this section we shall outline the new method for solving stochastic ∞-stage decision problems. We do not pretend that this method never fails. But, if the mathematical model has the properties mentioned in chapter 1,2 and 3, the method will be effective. These properties are sufficient, but not always necessary.

Now we merely stipulate the properties as stated below.
The state of the system can be represented by a point of a so-called state space. The state space, denoted by X, consists of points of an M-dimensional Cartesian space. The decisions can be represented by points d of a decision space D. We differentiate between decisions called null-decisions and decisions called interventions.

An intervention corresponds to a random transition in the state of the system. A null-decision "transfers" the system with probability 1 into its present state. Decisions are defined by the probability distributions of the state into which the system is transferred.

For each state x of the system we have a set D(x) of feasible decisions d in D.

A class of stationary Markov processes in X is defined; i.e. one for each initial state. This class of processes is called the natural process. In each particular problem a class Z of strategies z with corresponding intervention sets A_z in X is given. Each strategy $z \in Z$ dictates interventions in states of A_z and null-decisions elsewhere (d=z(x)). The stopping set A_o is defined by

$$A_o \stackrel{\text{def}}{=} \bigcap_{z \in Z} A_z \ . \tag{4.1}$$

To each state x and decision $d \in D(x)$ two random walks, denoted by $\underline{w}^o$ and $\underline{w}^d$, can be assigned. During the random walk $\underline{w}^o$ the system is subjected to the natural process having x as initial state. The walk ends as soon as the system takes on a state of the stopping set A_o. At the start of the walk $\underline{w}^d$ the system is transferred into a random state $\underline{y}$ with the probability distribution of d.
After this transition the system is subjected to the natural process having $\underline{y}$ as initial state and will end the random walk $\underline{w}^d$ as soon as it assumes a state of the stopping set A_o.

The (x;d)-function k(x;d) represents the difference in expected losses incurred during $\underline{w}^d$ and $\underline{w}^o$. The costs of the decision d are included in k(x;d).

The (x;d)-function t(x;d) represents the difference in expected durations of $\underline{w}^d$ and $\underline{w}^o$. If d is a null-decision, we obviously have

$$k(x;d) = 0 \tag{4.2}$$

$$t(x;d) = 0 \ . \tag{4.3}$$

If z is the strategy applied and if x is the current state, the probability distribution of the first future intervention state I_1 is denoted by $p^1_{A_z}(A;x;z)$.

Further we consider the functional equations

$$r(z;x) = \int_{A_z} p^1_{A_z}(dI;x;z)\ r(z;I) \qquad (4.4)$$

$$(= E\{r(z;\underline{I}_1)\,|\,x,z\})$$

and

$$c(z;x) = k(x;z(x)) - r(z;x)t(x;z(x)) + \\ + \int_{A_z} p^1_{A_z}(dI;x;z)\ c(z;I). \qquad (4.5)$$

$$(= k(x;z(x)) - r(z;x)\ t(x;z(x)) + E\{c(z;\underline{I}_1)\,|\,x,z\})$$

In the sequel the x-functions $r(z;x)$ and $c(z;x)$ denote some solution of the equations (4.4) and (4.5).

We now introduce <u>mixed strategies</u> of the following types:

a) <u>The mixed strategy of the form $(z_a)z_b$</u> with z_a and $z_b \in Z$, dictating
   1) first an intervention in accordance with z_a;
   2) then interventions in accordance with z_b.

If $z' = (z_a)z_b$, the x-functions $r(z';x)$ and $c(z';x)$ are defined by

$$r(z';x) \overset{def}{=} E\{r(z_b;\underline{y})\,|\,x;z_a\} \qquad (4.6)$$

and

$$c(z';x) \overset{def}{=} E\{k(\underline{I}_1;z_a(\underline{I}_1)) - r(z';\underline{I}_1)t(\underline{I}_1;z_a(\underline{I}_1))|x;z_a\} + \\ + E\{c(z_b;\underline{y})\,|\,x;z_a\}\ , \qquad (4.7)$$

where $\underline{y}$ is the state into which the system is transferred by the decision $z_a(\underline{I}_1)$.

b) <u>The mixed strategy of the form $d\cdot z$</u> with $z \in Z$, dictating
   1) the decision d in the initial state and
   2) then decisions in accordance with z.

We define $r(d\cdot z;x)$ and $c(d\cdot z;x)$ by

$$r(d \cdot z;x) \overset{def}{=} E\{r(z;\underline{y})|d\} \qquad (4.8)$$

and

$$c(d \cdot z;x) \overset{def}{=} k(x;d) - r(d \cdot z;x)\ t(x;d) + + E\{c(z;\underline{y})|d\}\ , \qquad (4.9)$$

where now $\underline{y}$ obeys the probability distribution corresponding to d.

c) The mixed strategy of the form Az', where z' is a mixed strategy of the form $(z_a)z_b$ and A is a closed set in X.
This strategy interdicts any intervention up to (but not including) the moment that for the first time a state of A is taken on by the system.

From that time onwards decisions are made in accordance with z'. The x-functions r(Az';x) and c(Az';x) are defined by

$$r(Az';x) \overset{def}{=} E\{r(z';\underline{y})|x;A\} \qquad (4.10)$$

and

$$c(Az';x) \overset{def}{=} E\{c(z';\underline{y})|x;A\}\ , \qquad (4.11)$$

where $\underline{y}$ obeys the probability distribution of the first state of A taken on if x is the initial state. Note that this probability distribution depends only on the natural process.

If z' is a mixed strategy $(z_a)z_b$, the set $A_{z'}$ is defined by

$$A_{z'} \overset{def}{=} A_{z_a} \qquad (4.12)$$

Further we consider the class $K_{z'}$ of all intervention sets A satisfying:

$$\bar{A} = \{x|r(Az';x) < r(z';x)\} \cup \{x|r(Az';x) = r(z';x);c(Az';x) \leqq c(z';x)\}\ . \qquad (4.13)$$

Obviously, we have $A_{z'} \in K_{z'}$.

Next we consider minimizing subsets of the following types:

a) If $z \in Z$, the minimizing subset $D_z(x)$ of $D(x)$ is defined by

$$D_z(x) \overset{def}{=} \{d|d \in D(x);\ r(d \cdot z;x) = \min_{d' \in D(x)} r(d' \cdot z;x)\}\ . \qquad (4.14)$$

b) The minimizing subset $A'_{z'}$ of $A_{z'}$ is defined by the intersection of all sets $A \varepsilon K_{z'}$.

Thus

$$A'_{z'} = \bigcap_{A \varepsilon K_{z'}} A. \tag{4.15}$$

c) If $z \varepsilon Z$ and if $x \varepsilon X$, the minimizing subset $Z_z(x)$ of Z is defined by

$$Z_z(x) \overset{def}{=} \{z^* \mid z^* \varepsilon Z; r((z^*)z;x) = \inf_{\bar{z} \varepsilon Z} r((\bar{z})z;x)\} . \tag{4.16}$$

d) If $z \varepsilon Z$, the minimizing subset Z_z of Z is defined by

$$Z_z \overset{def}{=} \bigcap_{x \varepsilon X} Z_z(x). \tag{4.17}$$

If $z \varepsilon Z$, the A-compressed strategy $[A]\, z$ of z is defined by

$$[A]\, z(x) \overset{def}{=} \begin{cases} z(x), & \text{if } x \varepsilon A. \\ \text{null-decisions}, & \text{if } x \varepsilon \bar{A}. \end{cases} \tag{4.18}$$

A strategy $z_o \varepsilon Z$ is called optimal if it satisfies for each x

$$r(z_o;x) = \min_{z \varepsilon Z} r(z;x). \tag{4.19}$$

We now consider the following problem:

"To determine an optimal strategy z_o of Z".

Several approaches are possible for solving this problem. The seemingly more simple approaches are in general only practicable in the most simple cases, but lead to impracticable problems in more complicated cases. Therefore a number of alternative approaches is described, of increasing complexity but also applicable to cases of increasing difficulty.

First formulation

I. Preparatory part

Determine the $(x;d)$-functions $k(x;d)$ and $t(x;d)$.

II. Determination of the optimal strategy

A. Direct approach

a) Determine a strategy z_o that satisfies for each x

$$r(z_o;x) = \min_{z \in Z} r(z;x). \tag{4.19}$$

b) Determine a strategy $z_o \in Z_{z_o}$ that satisfies for each x

$$c(z_o;x) = \min_{z \in Z_{z_o}} c((z)z_o;x), \tag{4.20}$$

where $c(z_o;x)$ satisfies (4.4) and (4.5), while $c((z)z_o;x)$ and Z_{z_o} are given by (4.7) and (4.17) respectively.

B. Iterative approach

If $z^o \in Z$ is an arbitrary initial strategy and if z^{i-1} is the strategy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) function-determination operation

Determine a solution of the functional equations (cf. (4.4) and (4.5))

$$r(z^{i-1};x) = \int_{A_{z^{i-1}}} p^1_{A_{z^{i-1}}}(dI;x;z^{i-1})\; r(z^{i-1};I) \tag{4.21}$$

and

$$c(z^{i-1};x) = k(x;z^{i-1}(x)) - r(z^{i-1};x)\; t(x;z^{i-1}(x)) + \\ + \int_{A_z^{i-1}} p^1_{A_{z^{i-1}}}(dI;x;z^{i-1})c(z^{i-1};I). \tag{4.22}$$

b) strategy improvement routine

1) Determine the minimizing subset $Z_{z^{i-1}}$ of Z.
2) Minimize uniformly in x the z-function $c((z)z^{i-1};x)$ subject to the constraint $z \in Z_{z^{i-1}}$. Select one of the solutions. The selected strategy is denoted by z^i.

End of the i^{th} cycle.

The second part of the new method can also be formulated in a more specified way; the complete formulation then becomes as follows:

Second formulation

I. Preparatory part

Determine the (x;d)-functions $k(x;d)$ and $t(x;d)$.

II. Determination of the optimal strategy

A. Direct approach

Determine a strategy z_o that satisfies for each x $z_o(x) \varepsilon D_{z_o}(x)$,

$$c(z_o;x) = \min_{d \,\varepsilon\, D_{z_o}(x)} c(d\cdot z_o;x) \qquad (4.23)$$

and

$$A'_{z_o} = A_{z_o}, \qquad (4.24)$$

where $c(z_o;x)$ satisfies (4.4) and (4.5), while $c(d\cdot z_o;x)$, $D_{z_o}(x)$ and A'_{z_o} are given by (4.9), (4.14) and (4.15) respectively.

B. Iterative approach

If $z^o \varepsilon Z$ is an arbitrary initial strategy and if z^{i-1} is the strategy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) function-determination operation

Determine a solution of the functional equations

$$r(z^{i-1};x) = \int_{A_{z^{i-1}}} p^1_{A_{z^{i-1}}}(dI;x;z^{i-1})\, r(z^{i-1};I) \qquad (4.25)$$

and

$$c(z^{i-1};x) = k(x;z^{i-1}(x)) - r(z^{i-1};x)\, t(x;z^{i-1}(x))$$

$$+ \int_{A_{z^{i-1}}} p^1_{A_{z^{i-1}}}(dI;x;z^{i-1})\, c(z^{i-1};I). \qquad (4.26)$$

b) strategy improvement routine

1) Determine for each x the minimizing subset $D_{z^{i-1}}(x)$.

2) Minimize the d-function $c(d \cdot z;x)$ for each x subject to the constraint $d \in D_{z^{i-1}}(x)$. Select for each x one of the minimizing decisions. The selected decision, $d_{z^{i-1};x}$, is chosen equal to $z^{i-1}(x)$ if $z^{i-1}(x)$ is a minimizing decision. The strategy z_1^{i-1} is defined by

$$z_1^{i-1}(x) \overset{def}{=} d_{z^{i-1};x} \quad . \tag{4.27}$$

3) Determine for the strategy $z' = (z_1^{i-1})z^{i-1}$ the minimizing subset $A'_{z'}$ of $A_{z'}$. The strategy z^i is the $A'_{z'}$-compressed strategy of z_1^{i-1}.

End of the i^{th} cycle.

Finally, some remarks about the usefulness of this method.

Let $p^k_{A_z}(B;x;z)$ denote the probability distribution of the k^{th} future intervention state in A_z and let the strategies $z \in Z$ have the following properties:

Property A

1) The functional equations (4.4) and (4.5) have at least one solution;

2) The Markov process in A_z has stationary absolute probability distributions given by

$$p_{A_z}(B;x;z) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p^k_{A_z}(B;x;z); \tag{4.28}$$

3) The function $r(z;x)$ is bounded and the expected values of $c(z;\underline{I}_\infty)$ with respect to the stationary absolute probability distributions $p_{A_z}(B;x;z)$ exist.

It follows from (4.4) that

$$r(z;x) = \frac{1}{n} \sum_{k=1}^{n} \int_{A_z} p^k_{A_z}(dI;x;z)\, r(z;I). \tag{4.29}$$

Hence, by point 3) and (4.28)

$$r(z;x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \int_{A_z} p^k_{A_z}(dI;x;z)\ r(z;I) =$$

$$= \int_{A_z} p_{A_z}(dI;x;z)\ r(z;I). \tag{4.30}$$

Consequently, $r(z;x)$ is constant on simple ergodic sets in A_z. If x is an ergodic state,it follows from (4.5) and (4.30) that

$$E\{c(z;\underline{I}_\infty)|x;z\} = E\{k(\underline{I}_\infty;z(\underline{I}_\infty)) - r(z;x)\ t(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\} +$$

$$+ E\{\int_{A_z} p^1_{A_z}(dI;\underline{I}_\infty;z)\ c(z;I)|x;z\}\ , \tag{4.31}$$

where $\underline{I}_\infty$ has the absolute stationary probability distribution corresponding to x.

We can easily verify that

$$E\{c(z;\underline{I}_\infty)|x;z\} = E\{\int_{A_z} p^1_{A_z}(dI;\underline{I}_\infty;z)\ c(z;I)|x;z\}\ . \tag{4.32}$$

By (4.31) and (4.32)

$$E\{k(\underline{I}_\infty;z(\underline{I}_\infty)) - r(z;x)\ t(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\} = 0 \tag{4.33}$$

and thus if x is a state of a simple ergodic set

$$r(z;x) = \frac{E\{k(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\}}{E\{t(\underline{I}_\infty;z(\underline{I}_\infty))|x;z\}}\ . \tag{4.34}$$

By (4.30) and (4.34) we find for an arbitrary x

$$r(z;x) = E\{\frac{E\{k(\underline{I}'_\infty;z(\underline{I}'_\infty))|\underline{I}_\infty;z\}}{E\{t(\underline{I}'_\infty;z(\underline{I}'_\infty))|\underline{I}_\infty;z\}}\ |\ x;z\}\ , \tag{4.35}$$

where $\underline{I}_\infty$ has the absolute stationary probability distribution in A_z corresponding to x and z. The distribution of $\underline{I}'_\infty$ is obtained as follows: let $\underline{I}''_\infty$ have the absolute stationary distribution in A_z corresponding to I_∞ and z; substituting $\underline{I}_\infty$ for I_∞ then $\underline{I}''_\infty$ becomes $\underline{I}'_\infty$.

Repeating the arguments made in the proof of theorem 8 in chapter 3 we can demonstrate that, if z_o satisfies either (4.20) or (4.23) and (4.24), we have for each x

$$r(z_o;x) = \min_{z \in Z} r(z;x). \tag{4.36}$$

In addition to property A we also introduce:

<u>Property B</u>

1) The strategies z^i $(i=1,2,\ldots)$, yielded by one of the iteration procedures, belong to Z;
2) The minimizing subsets $A'_{(z_1^{i-1})z^{i-1}} \varepsilon K_{(z_1^{i-1})z^{i-1}}$ $(i=1,2,\ldots)$;
3) An integer M_o can be found such that for each $z \varepsilon Z$, for each $x \varepsilon X$ and for each $i \geqq M_o$ we have

$$r((z)z^i;x) \geq r(z^i;x). \tag{4.37}$$

Repeating the arguments made in the proof of theorem 10 in chapter 3 we can demonstrate that

$$\lim_{i \to \infty} r(z^i;x) = \inf_{z \varepsilon Z} r(z;x). \tag{4.38}$$

By means of properties A and B, however, we cannot prove that the z-function r(z;x) is a good criterion for optimality (cf. section 2 of chapter 2). Unlike the properties of the mathematical model introduced in chapters 1,2 and 3, the properties A and B do not provide us with a deeper understanding of the structure of the decision process. But for practical purposes this need not always be necessary.

3. Dynamic programming

Many stochastic ∞-stage decision problems can be solved by means of dynamic programming. According to BELLMAN ([3], p.81) the corresponding decision processes have the following features in common:

a) In each case we have a physical system characterized at any stage by a small set of parameters, the state variables;

b) At each stage of either process we have a choice of a number of decisions; [1)]

c) The effect of a decision is a transformation of the state variables;

d) The past history of the system is of no importance in determining future actions;

e) The purpose of the process is to maximize some function of the state variables.

BELLMAN considers an infinite sequence of stages. The state of the system at the k^{th} stage is denoted by $\underline{I}_k$. At every stage a decision is taken. The properties mentioned above imply that at each stage the decisionmaker can make as if the process starts at this stage in its actual state.

The set of feasible decisions corresponding to the state I is indicated by D(I). If at a stage the state is I and if the decision is d, then a loss k(I;d) will be incurred.

The solution of the stochastic ∞-stage dynamic programming problem is given by a <u>policy</u>. Quoting BELLMAN (cf. [3] ,p.82): "A policy is any rule for making decisions which yields an allowable sequence of decisions; and an optimal policy is a policy which maximizes a preassigned function of the final state variables."

In this formulation one of the state variables is the total gain.

The method is based on the following principle (cf. [3] ,p.83):

<u>Principle of optimality</u>

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

In BELLMAN's dynamic programming approach of ∞-stage problems losses are discounted. Therefore, if a policy z is applied, the expected value of the loss to be incurred at the n^{th} stage with regard to the k^{th} stage is given by

1) Including the null-decision.

$$\alpha^{n-k}\ E\{k(\underline{I}_n;z(\underline{I}_n))\,|\,I_k;z\}\ ;\ n=k,k+1,\ldots, \qquad (4.39)$$

where the discount factor α satisfies $0 \leq \alpha < 1$ and $z(\underline{I}_n)$ is the decision to be made at the n^{th} stage.

If a policy z is applied, let the total expected loss to be incurred since the k^{th} decision and discounted with regard to the k^{th} stage be given by $f(z;I_k)$. Hence, for each I_1 and for each z to be considered we find

$$f(z;I_1) = k(I_1;z(I_1)) + \alpha E\{f(z;\underline{I}_2)\,|\,I_1;z(I_1)\}. \qquad (4.40)$$

Now it follows from the principle of optimality that the optimal policy z_o satisfies for each I_1

$$f(z_o;I_1) = \min_{d\,\varepsilon\, D(I_1)} \left[k(I_1;d) + \alpha\ E\{f(z_o;\underline{I}_2)\,|\,I_1;d\}\right]\ . \qquad (4.41)$$

The optimal policy z_o can be determined from the solution of the functional equation (4.41).

A detailed discussion of the existence and the uniqueness of the solution can be found in [3].

One of the iteration procedures which may yield the optimal policy is closely related to the strategy improvement routines considered in section 2. This iteration procedure will now be described.

<u>Determination of the optimal policy</u> (BELLMAN). <u>Iterative approach</u>

If $z^o \varepsilon Z$ is an arbitrary initial strategy and if z^{i-1} is the strategy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) <u>function-determination operation</u>

Determine the solution of the functional equation

$$f(z^{i-1};I_1) = k(I_1;z^{i-1}(I_1)) + \alpha\ E\{\,f(z^{i-1};\underline{I}_2)\,|\,I_1;z^{i-1}(I_1)\,\}. \qquad (4.42)$$

b) policy improvement routine

Minimize the d-function

$$k(I_1;d) + \alpha E\{f(z^{i-1};\underline{I}_2)|I_1;d\} \qquad (4.43)$$

for each I_1 subject to the constraint $d \in D(I_1)$. In this way we find for each I_1 at least one decision d. From these decisions we select an arbitrary one with this restriction: if for a particular state I_1 we have

$$f(z^{i-1};I_1) = \min_{d \in D(I_1)} \left[k(I_1;d) + \alpha E\{f(z^{i-1};\underline{I}_2)|I_1;d\}\right], \qquad (4.44)$$

then the decision $z^{i-1}(I_1)$ must be selected. The selected decisions constitute a policy z^i.

End of the i^{th} cycle.

Under certain conditions, the effectiveness of this iteration procedure can be proved (cf. [3]).

4. HOWARD's policy improvement methods

In [4] R.A. HOWARD considers two different types of stochastic ∞-stage decision problems. Using HOWARD's terminology, we shall formulate these problems and the corresponding solutions. Then, we shall show that with respect to these problems methods for solution can also be derived from the strategy improvement routines of section 2.

HOWARD's first problem

Let us consider a system that can be in N different states. These states are numbered from 1 to N. At equidistant points of time θ (θ=1,2,...) decisions can be made. For each state a finite number of feasible decisions are given. Thus, the decisions d^* which are feasible in at least one state, can also be numbered (d^*= 1,...,M). If the system at θ is in the state j and if at that time decision d^* is made, then

a) a loss $q_j^{d^*}$ will be incurred;

b) at $\theta+1$ the system will be in the state h with probability $p_{jh}^{d^*}$ $\left(\sum_{h=1}^{N} p_{jh}^{d^*} = 1\right)$.

If to each state of the system one and only one feasible decision has been assigned, this relation between states and decisions is called a <u>policy</u>.

It is assumed that, if decisions are made in accordance with a policy, the behaviour of the system can be described by means of a Markov chain without cyclically moving subsets.

If j is the initial state of the system and if a given policy is applied, then

a) we drop the index d^* in $p_{jh}^{d^*}$ and $q_j^{d^*}$;

b) the expected loss per stage in the steady state is denoted by g_j;

c) the expected loss to be incurred in the first n stages is denoted by $v_j(n)$ $(v_j(1) = q_j)$;

d) the value v_j is defined by

$$v_j \overset{\text{def}}{=} \lim_{n \to \infty} (v_j(n) - ng_j). \tag{4.45}$$

A policy is called <u>optimal</u>, if it minimizes for each initial state j the expected loss g_j in the steady state.

The problem is how to determine an optimal policy.

HOWARD proves in [4] that an optimal policy can be found by means of the following iteration procedure (cf. [4] p.64):

<u>Determination of the optimal policy</u> (HOWARD I). <u>Iterative approach</u>

If z^o is an arbitrary initial policy and if z^{i-1} is the policy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) <u>value-determination operation</u>

Use p_{jh} and q_j for a given policy z^{i-1} to solve the double set of equations

$$g_j = \sum_{h=1}^{N} p_{jh}\, g_h ; \qquad j=1,2,\ldots,N \tag{4.46}$$

and

$$v_j + g_j = q_j + \sum_{h=1}^{N} p_{jh}\, v_h ; \qquad j=1,2,\ldots,N \tag{4.47}$$

for all v_j and g_j, by setting the value of one v_j in each simple ergodic set zero.

b) <u>policy improvement routine</u>

1) Determine for each j the set D_j^* of decisions d^* which minimize

$$\sum_{h=1}^{N} p_{jh}^{d^*} g_h \ . \tag{4.48}$$

2) Minimize

$$q_j^{d^*} + \sum_{h=1}^{N} p_{jh}^{d^*} v_h \tag{4.49}$$

for each j subject to the constraint $d^* \in D_j^*$.
For each j this yields at least one decision d^*. From these decisions we select an arbitrary one with this restriction: if for a particular state j the decision dictated by the given policy belongs to D_j^* and if in addition to this we have

$$v_j + g_j = \min_{d^* \in D_j^*} \left(q_j^{d^*} + \sum_{h=1}^{N} p_{jh}^{d^*} v_h\right), \tag{4.50}$$

then this decision is chosen. The selected decisions, one for each state, constitute a new policy. In the next cycle this policy will be the given policy z^i.

End of the cycle.

It can be proved that an optimal policy will be found after a finite number of iterations. For proofs the reader is referred to [4] .

<u>HOWARD's second problem</u>

Again a system that can be in N different states (1,...,N) is considered. Suppose that the decision maker can influence the evolution in the state of the system by means of M different actions.

If the system is at t in the state j and if the action d^* is undertaken, then

a) for each unit of time the system is in the state j a loss $r_{jj}^{d^*}$ will be incurred;

b) for a transition from state j into state h a loss $r_{jh}^{d^*}$ will be incurred;

c) at t+ Δt the system will be in state h with probability $a_{jh}^{d^*} \Delta t + o(\Delta t^2)$ $(h \neq j)$.

If to each state of the system one and only one action has been assigned, the relation between states and actions is called a policy. It is assumed that, if actions are carried out in accordance with a policy, the behaviour of the system can be described by means of a Markov process with a continuous time parameter.

If j is the initial state of the system and if a <u>given</u> policy is applied, then

a) we drop the index d^* in $a_{jh}^{d^*}$, $r_{jj}^{d^*}$ and $r_{jh}^{d^*}$;

b) the expected loss per unit of time in the steady state is denoted by g_j;

c) the expected loss for the period $(0,t]$ is denoted by $v_j(t)$;

d) the value v_j is now defined by

$$v_j \overset{\text{def}}{=} \lim_{t \to \infty} (v_j(t) - tg_j). \qquad (4.51)$$

A policy is called <u>optimal</u>, if it minimizes for each initial state j the expected loss g_j per unit of time in the steady state.

The problem is how to determine the optimal policy.

Let us try to solve this problem along the same lines as above. To this end we split up the time axis into periods of length Δt. If at the beginning of the interval $(t', t' + \Delta t]$ the system is in the state j and if during that interval the decision maker will carry out the action d^*, then

a) we say that he makes the decision d^* at t';

b) the corresponding loss $q_j^{d^*}$ is given by [2)]

$$q_j^{d^*} = (r_{jj}^{d^*} + \sum_{h\neq j} a_{jh}^{d^*} r_{jh}^{d^*})\ \Delta t + o(\Delta t^2). \tag{4.52}$$

Notation 8:

We now define $a_{jj}^{d^*}$ and $\hat{q}_j^{d^*}$ by

$$a_{jj}^{d^*} \overset{\text{def}}{=} - \sum_{h\neq j} a_{jh}^{d^*}\ ;\quad j=1,2,\dots,N \tag{4.53}$$

and

$$\hat{q}_j^{d^*} \overset{\text{def}}{=} r_{jj}^{d^*} + \sum_{h\neq j} a_{jh}^{d^*} r_{jh}^{d^*};\quad j=1,2,\dots,N \tag{4.54}$$

respectively.

If a given policy is applied, the following relation is obvious (cf. (4.46)):

$$g_j = \sum_{h\neq j} a_{jh}\ \Delta t g_h + (1- \sum_{h\neq j} a_{jh}\ \Delta t) g_j + o(\Delta t^2), \tag{4.55}$$

and thus by (4.53)

$$\begin{aligned} 0 &= \sum_{h\neq j} a_{jh}\ \Delta t g_h - \sum_{h\neq j} a_{jh}\ \Delta t g_j + o(\Delta t^2) = \\ &= \sum_{h=1}^{N} a_{jh}\ g_h\ \Delta t + o(\Delta t^2). \end{aligned} \tag{4.56}$$

Consequently,

$$\sum_{h=1}^{N} a_{jh}\ g_h = 0. \tag{4.57}$$

Further we have (cf. (4.47))

$$v_j = (\hat{q}_j - g_j)\ \Delta t + \sum_{h\neq j} a_{jh}\ \Delta t\ v_h + (1- \sum_{h\neq j} a_{jh}\ \Delta t) v_j + o(\Delta t^2), \tag{4.58}$$

and thus by (4.53)

$$g_j - \hat{q}_j = \sum_{h=1}^{N} a_{jh}\ v_h. \tag{4.59}$$

---

2) The summation $\sum_{h\neq j}$ means that h runs through the numbers 1,...,N with the exception of j.

In the first step of the policy improvement routine corresponding to HOWARD's <u>first</u> problem we had to determine for each j the set D_j^* of decisions d^* which minimize

$$\sum_{h=1}^{N} p_{jh}^{d^*} q_h \qquad . \qquad (4.48)$$

This corresponds to seeking out the decisions d^* which minimize

$$\{\sum_{h\neq j} a_{jh}^{d^*} g_h + (1- \sum_{h\neq j} a_{jh}^{d^*}) g_j\} \Delta t + o(\Delta t^2) \qquad (4.60)$$

or

$$\{\sum_{h=1}^{N} a_{jh}^{d^*} g_h + g_j\} \Delta t + o(\Delta t^2) \qquad (4.61)$$

and thus

$$\sum_{h=1}^{N} a_{jh}^{d^*} g_h + o(\Delta t). \qquad (4.62)$$

In the second step of the policy improvement routine corresponding to HOWARD's first problem we had to minimize for each j

$$q_j^{d^*} + \sum_{h=1}^{N} p_{jh}^{d^*} v_h \qquad (4.49)$$

subject to the constraint $d^* \varepsilon D_j^*$.

This corresponds to minimizing, with respect to $d^* \varepsilon D_j^*$,

$$\hat{q}_j^{d^*} \Delta t + \sum_{h\neq j} a_{jh}^{d^*} \Delta t\, v_h + (1- \sum_{h\neq j} a_{jh}^{d^*} \Delta t) v_j + o(\Delta t^2) \qquad (4.63)$$

or

$$\hat{q}_j^{d^*} \Delta t + \sum_{h=1}^{N} a_{jh}^{d^*} v_h \Delta t + v_j + o(\Delta t^2) \qquad (4.64)$$

and thus the minimization of (4.63) is equivalent with that of

$$\hat{q}_j^{d^*} + \sum_{h=1}^{N} a_{jh}^{d^*} v_h + o(\Delta t). \qquad (4.65)$$

Obviously, the following iteration procedure yields an optimal strategy (cf. [4], p.108):

Determination of the optimal policy (HOWARD II). Iterative approach

If z^o is an arbitrary initial policy and if z^{i-1} is the policy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) value-determination operation

Use a_{jh} and $\hat{q}_j$ for a given policy to solve the double set of equations (cf. (4.57) and (4.59))

$$\sum_{h=1}^{N} a_{jh}\, g_h = 0 \qquad ; \quad j=1,2,\ldots,N \tag{4.66}$$

and

$$g_j = \hat{q}_j + \sum_{h=1}^{N} a_{jh}\, v_h \qquad ; \quad j=1,2,\ldots,N \tag{4.67}$$

for all v_j and g_j, by setting the value of one v_j in each simple ergodic set to zero.

b) policy improvement routine

1) Determine for each j the set D_j^* of actions d^* which minimize

$$\sum_{h=1}^{N} a_{jh}^{d^*}\, g_h . \tag{4.68}$$

2) Minimize

$$\hat{q}_j^{d^*} + \sum_{h=1}^{N} a_{jh}^{d^*}\, v_h \tag{4.69}$$

for each j subject to the constraint $d^* \varepsilon D_j^*$. For each j this yields at least one action d^*. From these actions we select an arbitrary one with this restriction: if for a particular state j the action dictated by the given policy belongs to D_j^* and if in addition to this we have

$$v_j + g_j = \min_{d^* \varepsilon D_j^*} \left(\hat{q}_j^{d^*} + \sum_{h=1}^{N} a_{jh}^{d^*}\, v_h\right) , \tag{4.70}$$

this action is chosen. The selected actions, one for each state, constitute a new policy z^i. In the next cycle this policy will be the given policy.

End of the cycle.

It can be proved that an optimal policy will be found after a finite number of iterations. For proofs the reader is referred to [4] .

We shall now demonstrate that HOWARD's first technique can also be derived from the second routine of section 2. To this end we introduce a new state variable e. If at HOWARD's equidistant points of time θ a ("HOWARD-") decision still has to be made, then e=0. In all other cases e denotes the last ("HOWARD-") decision made (e=1,...,M).

In our model the state space X consists of points x=(j,e) of a twodimensional lattice.

Because in our model decisions are defined by means of the probability distributions of the state into which the system is transferred, it follows from the construction of X that in our model a "HOWARD"-decision d^* made in j can be denoted by (j,d^*). The (decision-) probability distribution is now concentrated in the state (j,d^*). In our model the decision space D consists of points $d = (j,d^*)$ of a two dimensional lattice. [3)]

We now stipulate that

a) the sets D(j,e) of feasible decisions d only consist of null-decisions if $e \neq 0$;

b) the strategies z to be considered dictate interventions if e=0.

Hence, for all $z \in Z$ the intervention sets A_z consist of the states $\{(j,0);\ j=1,\dots,N\}$. Thus, for all $z \in Z$

$$A_z = A_o. \tag{4.71}$$

Using HOWARD's notations, but with the applied strategy indicated by z, we easily verify that $p^1_{A_z}(B;x;z)$, $k(x;d)$ and $t(x;d)$ (cf. section 2) are given by

$$p^1_{A_z}((h,0);(j,0);z) = p_{jh}(z), \tag{4.72}$$

3) This decision space does not satisfy points 2^a and 4 of property 3 in chapter 1. With respect to these properties an $M \times N$-dimensional Cartesian space is needed. In that space the end points of the MN unit vectors correspond to the decisions (j,d^*).

$$k((j,0);(j,d^*)) = q_j^{d^*} \tag{4.73}$$

and

$$t((j,0);(j,d^*)) = 1. \tag{4.74}$$

It follows from (4.4), (4.5), (4.46) and (4.47) that the functions $r(z;x)$ and $c(z;x)$ (cf. section 2) satisfy the equations

$$r(z;(j,0)) = g_j(z) = \sum_{h=1}^{N} p_{jh}(z)g_h(z) \tag{4.75}$$

and

$$c(z;(j,0)) = v_j(z) = q_j(z)-q_j(z) + \sum_{h=1}^{N} p_{jh}(z)v_h(z). \tag{4.76}$$

Further it follows from (4.8), (4.9), (4.72), (4.73) and (4.74) that the functions $r(d\cdot z;x)$ and $c(d\cdot z;x)$ satisfy the equations

$$r((j,d^*)z;(j,0)) = \sum_{h=1}^{N} p_{jh}^{d^*} g_h(z) \tag{4.77}$$

and

$$c((j,d^*)z;(j,0)) = q_j^{d^*} - \sum_{h=1}^{N} p_{jh}^{d^*} g_h(z) + \sum_{h=1}^{N} p_{jh}^{d^*} v_h(z). \tag{4.78}$$

The second formulation of the new method (p.93) provides us with the following iteration procedure:

I) Preparatory part

Determine the $(x;d)$-functions

$k(x;d)$ and $t(x;d)$.

These functions are given by (cf.(4.73) and (4.74))

$$k((j,e);(j,d^*)) = \begin{cases} 0 & ;\ \text{if } e\neq 0 \\ q_j^{d^*} & ;\ \text{if } e=0 \end{cases} \tag{4.79}$$

and

$$t((j,e);(j,d^*)) = \begin{cases} 0 & ;\ \text{if } e\neq 0 \\ 1 & ;\ \text{if } e=0. \end{cases} \tag{4.80}$$

II) Determination of the optimal strategy

B. Iterative approach

If $z^o \in Z$ is an arbitrary initial strategy and if z^{i-1} is the strategy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) function-determination operation

Determine a solution of the functional equations (cf. (4.75) and (4.76)) [4]

$$g_j(z^{i-1}) = \sum_{h=1}^{N} p_{jh}(z^{i-1}) g_h(z^{i-1}) \tag{4.81}$$

and

$$v_j(z^{i-1}) = q_j(z^{i-1}) - g_j(z^{i-1}) + \sum_{h=1}^{N} p_{jh}(z^{i-1}) v_h(z^{i-1}). \tag{4.82}$$

b) strategy-improvement routine

1) Determine for each $x=(j,e)$ the minimizing subset $D_{z^{i-1}}(j,e)$ of decisions $d=(j,d^*)$ which minimize (cf. (4.77))

$$\sum_{h=1}^{N} p_{jh}^{d^*} g_h(z^{i-1}) \tag{4.83}$$

2) Minimize the (j,d^*)-function (cf. (4.78))

$$q_j^{d^*} - \sum_{h=1}^{N} p_{jh}^{d^*} g_h(z) + \sum_{h=1}^{N} p_{jh}^{d^*} v_h(z) \tag{4.84}$$

for each j subject to the constraint $(j,d^*) \in D_{z^{i-1}}(j,e)$. Since (4.83) is constant for $(j,d^*) \in D_{z^{i-1}}(j,e)$ we may replace (4.84) by

$$q_j^{d^*} + \sum_{h=1}^{N} p_{jh}^{d^*} v_h(z). \tag{4.85}$$

Select for each (j,e) one of the minimizing decisions. If $z^{i-1}(j,e)$ is a minimizing decision, then the selected

---

4) $g_j(z)$ and $v_j(z)$ are "j-functions" defined on $(1,2,\dots,N)$.

decision $d_{z^{i-1};(j,e)}$ is chosen equal to $z^{i-1}(j,e)$.

The strategy z_1^{i-1} is defined by

$$z_1^{i-1}(j,e) \overset{def}{=} d_{z^{i-1};(j,e)} \quad . \tag{4.86}$$

3) Determine for the strategy $z' = (z_1^{i-1})z^{i-1}$ the minimizing subset $A'_{z'}$ of $A_{z_1^{i-1}}$.

   Since the class $K_{z'}$ only consists of the set A_o (cf.(4.71)) we find

$$A'_{z'} = A_{z^{i-1}} = A_o. \tag{4.87}$$

   Consequently,

$$z^i = z_1^{i-1}. \tag{4.88}$$

Comparing this routine with that of HOWARD for the first problem we easily verify that the two techniques are identical.

We now return to Howard's second problem. First we remark that at each point of time an action is going on.

Without restricting the generality we now add to the description of the problem the following two points:

1) After each alteration in the j-state of the system the decision maker decides whether the running action will be continued or not;
2) If j is the actual state of the system, if d^* is the running action and if $\sum_{h\neq j} a_{jh}^{d^*}=0$, then the decision maker decides after each unit of time whether the running action will be continued or not.

In order to incorporate these points in the mathematical model we introduce an additional state-variable e.

The e-component of the state is equal to d^*

a) if with respect to the present action d^* and state j we have $\sum_{h\neq j} a_{jh}^{d^*} = 0$ and if in addition to this the length of the period elapsed since the last decision (cf. points 1 and 2) is less than one unit of time;

b) if with respect to the present action d^* and state j we have $\sum_{h \neq j} a_{jh}^{d^*} > 0$ and if in addition to this the present action has been started or continued after the last alteration in the j-state of the system.

The e-component of the state is equal to 0

c) if with respect to the present action d^* and state j we have $\sum_{h \neq j} a_{jh}^{d^*} = 0$ and if in addition to this the length of the period elapsed since the last decision (cf. points 1 and 2) is larger than or equal to one unit of time;

b) if no decision has been made since the last alteration in the state of the system (cf. point 1).

In our model (cf. section 2) the state space X consists of points x=(j,e) of a two-dimensional lattice.
Since in our model decisions d are defined by means of probability distributions of the state into which the system is transferred, it follows from the construction of X that a ("Howard"-) action d^* undertaken in j can be denoted by (j,d^*). In our model the decision space D consists of a two-dimensional lattice.

We now stipulate that

a) the sets D(j,e) of feasible decisions only concist of null-decisions if $e \neq 0$;

b) the strategies z to be considered dictate only interventions if e=0 (cf. points 1 and 2).

By a) and b) all intervention sets A_z consist of the states $x \in \{(j,0);j=1,\ldots,N\}$.

Thus for all $z \in Z$

$$A_z = A_o. \tag{4.89}$$

Using HOWARD's notation but with the applied strategy indicated by z, we easily verify that $p^1_{A_z}((j,0);(k,0);z)$ (cf. section 2) satisfies

$$p^1_{A_z}((j,0);(k,0);z) = \begin{cases} a_{jk}(z)(\sum_{h\neq j} a_{jh}(z))^{-1}, \text{ if } \sum_{h\neq j} a_{jh}(z) > 0 \\ 1, \text{ if } \sum_{h\neq j} a_{jh}(z) = 0 \text{ and if } k=j. \\ 0, \text{ if } \sum_{h\neq j} a_{jh}(z) = 0 \text{ and if } k\neq j. \end{cases} \tag{4.90}$$

It follows from the formulation of Howard's second problem that the length of the period between two successive alterations in the j-state of the system obeys the negative exponential distribution. If j is the initial ("HOWARD"-) state and if d^* is the action to carry out, then the parameter of this distribution is given by $\sum_{h\neq j} a^{d^*}_{jh}$.
Consequently, if j is the initial ("Howard"-) state and if $\sum_{h\neq j} a^{d^*}_{jh} > 0$, the expected duration of the walk $\underline{w}^d$ (cf. section 2) is equal to $(\sum_{h\neq j} a^{d^*}_{jh})^{-1}$.

If the initial state of the walk $\underline{w}^o$ is given by (j,0), then by (4.89) the expected duration of this walk is equal to zero (cf.section 2). Therefore, we find (cf. section 2)

$$t((j,0);(j,d^*)) = \begin{cases} (\sum_{h\neq j} a^{d^*}_{jh})^{-1}, \text{ if } \sum_{h\neq j} a^{d^*}_{jh} > 0. \\ 1, \text{ if } \sum_{h\neq j} a^{d^*}_{jh} = 0. \end{cases} \tag{4.91}$$

Obviously, we have $t((j,e);(j,d^*)) = 0$ if $e\neq 0$. Hence, the function $t(x;d)$ has been determined.

Next we consider the function $k(x;d)$. By (4.89) the expected loss to incur during $\underline{w}^o$ is zero if $e=0$. With regard to the walk $\underline{w}^d$ the expected loss is given by $(\sum_{h\neq j} a^{d^*}_{jh} r^{d^*}_{jh} + r^{d^*}_{jj})(\sum_{h\neq j} a^{d^*}_{jh})^{-1}$ if $\sum_{h\neq j} a^{d}_{jh} > 0$ and otherwise by r^d_{jj}.
Therefore, we find (cf. section 2)

$$k((j,0);(j,d^*)) = \begin{cases} (\sum_{h\neq j} a^{d^*}_{jh} r^{d^*}_{jh} + r^{d^*}_{jj})(\sum_{h\neq j} a^{d^*}_{jh})^{-1}, \text{ if } \sum_{h\neq j} a^{d^*}_{jh} > 0. \\ r^{d^*}_{jj}, \text{ if } \sum_{h\neq j} a^{d^*}_{jh} = 0. \end{cases} \tag{4.92}$$

Obviously, we have $k((j,e);(j,d^*)) = 0$ if $e \neq 0$. This determines the function $k(x;d)$.

It follows from (4.4), (4.5), (4.90), (4.91) and (4.92) that the functions $r(z;x)$ and $c(z;x)$ satisfy

$$r(z;(j,0)) = \begin{cases} \left[\sum_{h\neq j} a_{jh}(z) r(z;(h,0))\right] \left(\sum_{h\neq j} a_{jh}(z)\right)^{-1}, \\ \qquad\qquad \text{if } \sum_{h\neq j} a_{jh}(z) \neq 0. \\ r(z;(j,0)), \text{ if } \sum_{h\neq j} a_{jh}(z) = 0. \end{cases} \tag{4.93}$$

and

$$c(z;(j,0)) = \begin{cases} \left[\sum_{h\neq j} a_{jh}(z)\{r_{jh}(z)+c(z;(h,0))\}+r_{jj}(z) - r(z;(j,0))\right]\cdot \\ \qquad\qquad \cdot\left(\sum_{h\neq j} a_{jh}(z)\right)^{-1}, \text{ if } \sum_{h\neq j} a_{jh}(z) > 0. \\ r_{jj} - r(z;(j,0)) + c(z;(j,0)), \text{ if } \sum_{h\neq j} a_{jh}(z) = 0. \end{cases} \tag{4.94}$$

respectively.

From (4.53) and (4.93) we obtain

$$\sum_{h=1}^{N} a_{jh}(z) r(z;(h,0)) = 0, \tag{4.95}$$

while by means of (4.54) the equation (4.94) becomes

$$r(z;(j,0)) = \hat{q}_j(z) + \sum_{h=1}^{N} a_{jh}(z) c(z;(h,0)). \tag{4.96}$$

It follows from (4.8) and (4.9) that the functions $r(d\cdot z;x)$ and $c(d\cdot z;x)$ satisfy

$$r((j,d^*)z;(j,0)) = \begin{cases} \left[\sum_{h\neq j} a_{jh}^{d^*} r(z;(h,0))\right]\left(\sum_{h\neq j} a_{jh}^{d^*}\right)^{-1}, \quad \text{if } \sum_{h\neq j} a_{jh}^{d^*} > 0. \\ r(z;(j,0)), \text{ if } \sum_{h\neq j} a_{jh}^{d^*} = 0. \end{cases} \tag{4.97}$$

and

$$c((j,d^*)z;(j,0)) = \begin{cases} \left[\sum_{h\neq j} a_{jh}^{d^*}\{r_{jh}^{d^*}+c(z;(h,0))\}+ r_{jj}^{d^*} - r(z;(j,0))\right]\cdot \\ \qquad\qquad \cdot\left(\sum_{h\neq j} a_{jh}^{d^*}\right)^{-1}, \text{ if } \\ \qquad\qquad \sum_{h\neq j} a_{jh}^{d^*} > 0. \\ r_{jj}^{d^*} - r(z;(j,0)), \text{ if } \sum_{h\neq j} a_{jh}^{d^*} = 0. \end{cases} \tag{4.98}$$

It follows from (4.57) and (4.59) on the one hand and from (4.95) and (4.96) on the other hand that the functions $r(z;(j,0))$ and $c(z;(j,0))$ can also be represented by $g_j(z)$ and $v_j(z)$ respectively.
So we write

$$r(z;(j,0)) = g_j(z) \tag{4.99}$$

and

$$c(z;(j,0)) = v_j(z). \tag{4.100}$$

In order to simplify the notations let us introduce the notations

$$r((j,d^*)z;(j,0)) = g_j^{d^*}(z) \tag{4.101}$$

and

$$c((j,d^*)z;(j,0)) = v_j^{d^*}(z). \tag{4.102}$$

By (4.99) through (4.102) the equations (4.95) through (4.98) become

$$\sum_{h=1}^{N} a_{jh}(z)g_h(z) = 0, \tag{4.103}$$

$$g_j(z) = \hat{q}_j(z) + \sum_{h=1}^{N} a_{jh}(z)v_h(z), \tag{4.104}$$

$$g_j^{d^*}(z) = \begin{cases} \left[\sum_{h\neq j} a_{jh}^{d^*} g_h(z)\right] \left(\sum_{h\neq j} a_{jh}^{d^*}\right)^{-1}, \text{ if } \sum_{h\neq j} a_{jh}^{d^*} > 0. \\ g_j(z), \text{ if } \sum_{h\neq j} a_{jh}^{d^*} = 0. \end{cases} \tag{4.105}$$

and

$$v_j^{d^*}(z) = \begin{cases} \left[\hat{q}_j^{d^*} + \sum_{h\neq j} a_{jh}^{d^*} v_h(z) - g_j^{d^*}(z)\right] \left(\sum_{h\neq j} a_{jh}^{d^*}\right)^{-1}, \\ \qquad \text{if } \sum_{h\neq j} a_{jh}^{d^*} > 0. \\ r_{jj}^{d^*} - g_j(z), \text{ if } \sum_{h\neq j} a_{jh}^{d^*} = 0. \end{cases} \tag{4.106}$$

The method, discussed in section 2, provides us with the following iteration procedure:

I. <u>Preparatory part</u>

Determine the $(x;d)$-functions $k(x;d)$ and $t(x;d)$.

These functions are given by (4.92) and (4.91) respectively.

Hence,

$$t((j,e);(j,d^{*})) = \begin{cases} 0, \text{ if } e\neq 0. \\ (\sum_{h\neq j} a_{jh}^{d^{*}})^{-1}, \text{ if } e=0 \text{ and if } \sum_{h\neq j} a_{jh}^{d^{*}} > 0. \\ 1, \text{ if } e=0 \text{ and if } \sum_{h\neq j} a_{jh}^{d^{*}} = 0. \end{cases} \quad (4.107)$$

and

$$k((j,e);(j,d^{*})) = \begin{cases} 0, \text{ if } e\neq 0. \\ (\sum_{h\neq j} a_{jh}^{d^{*}} r_{jh}^{d^{*}} + r_{jj}^{d^{*}})(\sum_{h\neq j} a_{jh}^{d^{*}})^{-1}, \text{ if } e=0 \\ \qquad \text{and if } \sum_{h\neq j} a_{jh}^{d^{*}} > 0 \\ r_{jj}^{d^{*}}, \text{ if } e=0 \text{ and if } \sum_{h\neq j} a_{jh}^{d^{*}} = 0. \end{cases} \quad (4.108)$$

II. Determination of the optimal strategy

B. Iterative approach

If $z^{o} \varepsilon Z$ is an arbitrary initial strategy and if z^{i-1} is the strategy obtained at the end of the $(i-1)^{st}$ cycle, the i^{th} cycle runs as follows:

a) function determination operation

Determine a solution of the functional equations (cf. (4.103) and (4.104))

$$\sum_{h=1}^{N} a_{jh}(z^{i-1})g_{h}(z^{i-1}) = 0 \quad (4.109)$$

and

$$g_{j}(z^{i-1}) = \hat{q}_{j}(z^{i-1}) + \sum_{h=1}^{N} a_{jh}(z^{i-1})v_{h}(z^{i-1}) \quad (4.110)$$

b) strategy improvement routine

1) Determine for each (j,0) the minimizing subset $D_{z^{i-1}}(j,0)$ of decisions $d=(j,d^{*})$ which minimize (cf. (4.105))

$$g_{j}^{d^{*}}(z^{i-1}) = \begin{cases} \left[\sum_{h\neq j} a_{jh}^{d^{*}} g_{h}(z^{i-1})\right](\sum_{h\neq j} a_{jh}^{d^{*}})^{-1}, \\ \qquad \text{if } \sum_{h\neq j} a_{jh}^{d^{*}} > 0. \\ g_{j}(z^{i-1}), \text{ if } \sum_{h\neq j} a_{jh}^{d^{*}} = 0. \end{cases} \quad (4.111)$$

2) Minimize for each (j,0) the function

$$v_j^{d^*}(z) = \begin{cases} \left[\hat{q}_j^{d^*} + \sum_{h \neq j} a_{jh}^{d^*} v_h(z^{i-1}) - g_j^{d^*}(z^{i-1})\right] \cdot \\ \qquad \cdot \left(\sum_{h \neq j} a_{jh}^{d^*}\right)^{-1}, \text{ if } \sum_{h \neq j} a_{jh}^{d^*} > 0. \\ r_{jj}^{d^*} - g_j(z), \text{ if } \sum_{h \neq j} a_{jh}^{d^*} = 0. \end{cases} \qquad (4.112)$$

subject to the constraint $d \in D_{z^{i-1}}(j,0)$. Select for each (j,e) one of the minimizing decisions. The selected decision $d_{z^{i-1};(j,e)}$ is equal to $z^{i-1}(j,e)$ if $z^{i-1}(j,e)$ is a minimizing decision.

The strategy z_1^{i-1} is defined by

$$z_1^{i-1}(j,e) \stackrel{def}{=} d_{z^{i-1};(j,e)}. \qquad (4.113)$$

3) Determine for the strategy $z' = (z_1^{i-1})z^{i-1}$ the minimizing subset $A'_{z'}$, of $A_{z_1^{i-1}}$.
Since the class $K_{z'}$ only consists of the set A_0 (cf. (4.89)) we find

$$A'_{z'} = A_{z^{i-1}} = A_0 . \qquad (4.114)$$

Consequently,

$$z^i = z_1^{i-1}. \qquad (4.115)$$

End of the i^{th} cycle.

By comparing the object functions (4.111) and (4.112) with (4.68) and (4.69) we can easily verify that this routine is closely related to that of HOWARD for the second problem. They lead to the same minimum of the expected costs per unit of time in the steady state. Each optimum solution of the one is an optimum solution of the other and vice versa.

REFERENCES :

1. M. LOÈVE, Probability theory. D. van Nostrand Company, New York, 1960.

2. J.L. DOOB, Stochastic Processes. John Wiley and Sons, New York, 1953.

3. R. BELLMAN, Dynamic Programming. Princeton University Press, New Jersey, 1957.

4. R.A. HOWARD, Dynamic Programming and Markov Processes. The Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, New York, 1960.