STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Report S 302 (VP 19)

Stochastic ∞-stage decision problems

by

G. de Leve

August 1962

0. Introduction

In many fields of human activity decisions are made and still will be made.

The number of decision problems which are solved by the aid of mathematics is increasing. Many physical decision problems seem to be translatable into mathematical problems.

It is well known that mathematical models of physical situations are often useful for describing the underlying structure. Such models are also in use in situations where decisions have to be made. Based on these models and formulated in the corresponding terminology the mathematical versions of these physical decision problems are defined. But though the mathematical decision problem is a purely mathematical problem and can be formulated by using only mathematical concepts, still in this study it will always be considered as a mathematical abstraction of a physical decision problem. For this reason we shall develop a terminology out of words reminiscent of physical concepts but with a mathematical interpretation only.

In this paper we will give an account of a study in stochastic ∞-stage decision problems.

In an ∞-stage decision problem decisions have to be made at different points of time; decisions which in general are related and can not be made independently.

By analysing some mathematical models of physical stochastic ∞-stage decision problems, we shall investigate their common characteristics. From these we shall derive some basic properties, which will be attributed to the common mathematical model.

Next we shall review some known methods for solving special decision problems.

In the third section these methods will be generalized in such a way that they apply to a larger class of decision problems.

The relation between the old methods and the new one is the topic of the 4th section.

Finally by means of two applications we shall show how the new method works.

It is not the purpose of this paper to give an exact mathematical treatise, but only to give a picture of the mathematical model, the decision problem and its solution. [See for the mathematical aspect [1]].

1) A common mathematical model

In physical ∞-stage decision problems the choice of a decision often depends on the state in which the decision has to be made. Speaking about a state, we have to specify to what that state refers. In the mathematical model we use the concept state of the system.

In a replacement problem for instance the system may be the mathematical abstraction of a machine, while in an inventory problem the system is identified with the inventory or with the inventory and the quantities on order.

Property 1

In a mathematical model the state of the system is determined by N real-valued variables; thus by a point ψ of an N-dimensional Cartesian space.

This N-dimensional Cartesian space will be called the state space Ψ. On analysing the mathematical models we discover that decisions effect transitions in the state of the system. For instance, if in a replacement problem one of the state variables is the lifetime of the machine in use, then the replacement (decision) of the old machine by a new one will change the state of the system in the mathematical model. In an inventory problem where the system comprises the quantities on order, a new order (certainly) effects the state of the system.
In addition we observe that frequently the transitions are not of a deterministic nature. In a replacement problem e.g. the initial states of new machines are not always identical and because the decision maker generally draws at random out of a set of new machines, a decision corresponds to a random transition in the mathematical model.

For this reason the following property is attributed to the mathematical model.

Property 2

In the common mathematical model a decision is a random transition which is defined by the probability distribution 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 moments of decision. They only refer to the state into which the system is transferred. A transition takes no time and 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 a small number of them effects non-degenerate transitions.

In this paper we shall make a distinction between non-degenerate decisions and degenerate decisions. By a degenerate decision the system is "transferred" with probability 1 into the state it is already in.

It is obvious that in a finite time interval only a finite number of non-degenerate decisions can be made. For this reason the following property is attributed to the mathematical model:

Property 3

In a finite time interval only a finite number of non-degenerate decisions can be made. At each point of time one and only one decision will be made.

Analogous to the state space Ψ, decisions (probability distributions) will be represented by points d of a so called decision space D.

If we restrict ourselves to probability distributions which are completely determined by their moments, the space D is an ∞-dimensional Cartesian space.

Because decisions are defined by probability distributions of states into which the system is transferred, it follows from the structure of many decision problems that in some states certain decisions are not feasible. To some extent the decision maker may be restricted in his choice of a decision.

Property 4

Whether a decision is feasible or not, depends on the state ψ of the system at the moment of decision only. Next it is stipulated that for each state ψ the set of feasible decisions $D(\psi)$ in D is a closed set.

If for an arbitrary time interval, closed on the left no non-degenerate decisions are planned and the system nevertheless changes its state during that interval, it is said to be subjected to a natural

process. In other words, if a natural process is present this process effects a walk of the system through the state space.
With respect to the natural process we have:

Property 5

In the mathematical model

(1) in each time interval closed on the left the natural process is defined by means of a Stationary Strong Markov Process.

(2) the random walk corresponding to the natural process is continuous from the right in the time parameter and has only a finite number of discontinuities in a finite time interval.

It follows from the properties of a Markov process, that it is defined for each initial state. Consequently the natural process is also defined for each initial state.

In the mathematical model we have now stipulated what happens at the moment of a certain decision, and how the behaviour of the system can be described if in a time interval closed on the left only degenerate decisions are made. We have still to state how the behaviour of the system is to be fixed if there is a non-degenerate decision at the beginning of the time interval considered.

To this end we introduce the following property:

Property 6

In the mathematical model the behaviour of the system in each time interval between two non-degenerate decisions is described by a natural process. The initial state of that process will be the state into which the system has been transferred by the decision at the beginning of the interval considered.

As we have stated already, at the moments of a non-degenerate decision the state of the system is not uniquely defined.

Let us introduce a product space Ψ' of two spaces Ψ_- and Ψ_+ both congruent to Ψ. So we have:

$$\Psi' = \Psi_- \times \Psi_+ \qquad (1.1)$$

The point $\psi_- \in \Psi_-$ fixes the state of the system at the moment of a decision before the decision is made, while $\psi_+ \in \Psi_+$ fixes the state of the system at the same point of time, but after the decision has been made. Thus if we use for fixing the state of the system the space Ψ' instead of Ψ then this state is again defined unambiguously at each point of time. At the moments of a degenerate decision we have:

$$\psi_- = \psi_+ \tag{1.2}$$

The most important features of physical decision problems are losses and gains. It will be no restriction to suppose that only losses occur.

Generally in decision problems two types of losses will be met. First, losses which increase or decrease continuously in the course of time (e.g. lack of interest or consumption of fuel), and secondly, losses which occur at discrete points of time (e.g. sales or repairments).

With respect to these losses we have:

Property 7

For losses not due to decisions the following statements hold:

a) for each walk of the system in a time interval the losses not due to decisions are defined unambiguously.

b) for each union of disjunct intervals the total losses are the sum of the losses incurred in the disjunct time intervals.

For a while let us enter into the question of how these losses are to be defined in the mathematical model. This can be done by means of two functions. The first function is a ψ_--function which represents the losses that will be suffered if the system is in the state ψ_- during one time unit. This function could be called "loss density function". The second function, which is also a ψ_--function, fixes the losses incurred by the system if it takes on a state ψ_-. This function could be called a "discrete loss function".

In [1] it will be proved that under certain conditions about the loss functions and the natural process, the mathematical model possesses this property automatically.

Property 8

If a degenerate decision is made no losses are involved. The losses due to non-degenerate decisions are defined and depend only on:

a) the state ψ_- in which the decision is made.

b) the state ψ_+ into which the system is transferred at the moment of the decision.

From properties 7 and 8 we deduce:

Statement 1

In the mathematical model the losses incurred in each time interval are unambiguously fixed by the walk of the system in the space Ψ' during that time interval.

The solution of the stochastic ∞-stage decision problem is given in the form of a strategy.

A strategy dictates a feasible decision at each point of time, on the basis of the available information, i.e.

a) the state ψ_- at the moment of decision.

b) the states $\psi' \in \Psi'$ taken on before that point of time.

It is obvious that, if a strategy is applied, because of the extra transitions the natural process is no longer appropriate to describe the behaviour of the system.

Property 9

In this study we shall restrict ourselves to the class Z_o of all strategies z, which satisfy for each point of time t_o the following properties:

(I) Even if it has not been applied before t_o, each strategy z of the class dictates feasible decisions from t_o onwards.

(II) If a strategy of the class is applied from t_o onwards then the random walk of the system in Ψ' can from that point of time be defined with the aid of the available information at t_o.

(III) For purposes of comparing strategies we introduce a real one-valued function of the available information and the strategy to be applied. For each given information this function is a criterion function defined on Z_o only with the following properties:

a) if for a particular information at t_o the criterion function for strategy z_1 takes on a smaller value than for strategy z_2, then in that situation z_1 is to be preferred to z_2.

b) if from t_o onwards the application of two different strategies will effect identical random walks in Ψ' and if one of the strategies is an element of the class considered then both strategies belong to that class and have equal criterion values, even if the informations in the initial states at t_o are different.

c) for at least one strategy the criterion function is minimal for each information at t_o. (Such strategies are called optimal).

In general, in one way or another, criterion functions are based on the losses to be incurred in the future. Consequently property a) is obvious. Property c) implies the existence of a solution of the problem and is of course indispensible. Let us consider now property b). If a realization of a random walk in Ψ' is given then the losses are fixed (statement 1). So strategies which effect identical random walks in Ψ' are also equal with respect to the losses during these random walks. Consequently there is no reason for different appreciations.

Now we shall prove that optimal decisions can be made by taking into account only the state of the system at the moment of decision.

Let z_o be an optimal strategy and let us compair two different situations.

In the first situation a system with given information about past states is at t_o in some state ψ_o.

In the second situation the system starts its random walk in that state.

The two situations are identical except for the pasts of the system.

Suppose we now apply strategy z_o from time t_o onwards in both situations, but in the second situation on the false assumption that we have the same information at t_o as in the first.

It can easily be verified that in virtue of properties 4, 6 and statement 1 systems in identical states but with different pasts are permutable with respect to their pasts. Hence under the false supposition mentioned above the behaviour of the system from t_o onwards can for both situations be described by one random walk. But if by applying z_o from t_o onwards in the second situation on the basis of false information we still obtain a feasible decision at each point of time, then this comes down to our using (unwittingly) a different strategy with reference to the right information. This strategy will be denoted by z_o'. It is uniquely determined by strategy z_o and the false information at t_o. It follows that if in the second situation strategy z_o' has been applied by using the true information at t_o, the same random walk would have been found.

According to point III[b] of property 9 the strategy z_o' belongs to the class considered and has the same criterion value as z_o in the first situation.

If the criterion functions in the first and second situations are denoted by $C_I(z)$ and $C_{II}(z)$ respectively then it follows from the optimality property of z_o that we have:

$$C_{II}(z_o) \lesseqgtr C_{II}(z_o') = C_I(z_o) \tag{1.3}$$

Let us apply again strategy z_o in both situations, but let us now proceed as if there were no information given in the first situation.

Repeating the argument given above we can state that in reality, based on true information, a strategy z_o'' is applied in the first situation.

Now the following inequality can be proved:

$$C_I(z_o) \lesseqgtr C_I(z_o'') = C_{II}(z_o) \tag{1.4}$$

and consequently we have:

$$C_I(z_o) = C_{II}(z_o) = C_I(z_o'') \tag{1.5}$$

From (1.5) it follows that information about states taken on before t_o does not effect a lower value of the criterion function. So we can always suppose the system to start its random walk at t_o.

For making a decision at t_o only the state of the system at that point of time is relevant.

Because the properties mentioned in 9 are valid for each point of time t_o and each state ψ_o we can state:

Statement 2

If the class of strategies considered satisfies property 9, then there exists a strategy which belongs to the class, is optimal and assigns unambiguously to each state $\psi \in \Psi$ a feasible decision.

In other words there exists an optimal strategy that maps the state space Ψ into the decision space D.

So it is no restriction to limit our discussion to strategies which map Ψ into D. These strategies can be represented by the relation:

$$d = z(\psi) \tag{1.6}$$

Consequently strategies z of the form (1.6) divide the state space Ψ into two disjunct sets, one denoted by A_z, comprising states in which always non-degenerate decisions will be made, the other consisting of the states in which always degenerate decisions will be made.

In order to avoid difficulties in defining the mathematical decision problem we introduce the following property:

Property 9 (continued)

(IV) Each strategy of the class assigns unambiguously a feasible decision to each state of Ψ.

(V) For each strategy of the class the set A_z will be a closed set.

(VI) Each strategy of the class dictates in a finite time interval only a finite number of non-degenerate decisions.

(VII) After each decision the system takes on a state outside A_z.

According to statement 2, point IV is no essential restriction.

The reason for introducing point V is shown by the following:

Let the system enter A_z along a continuous path. Now if A_z is a non-closed set, there may be a point of time at which the system is still outside A_z while for each later moment the system will have been in A_z for some positive time. In that case the moment and the state in which a decision is made cannot be defined.

Points VI and VII of property 9 are in conformity with property 3.

Let us consider now the sequence of states at the moments of a non-degenerate decision. These states are all elements of A_z. In [1] it will be proved that this sequence can be described by a stationary Markov process with a discrete time parameter. This process in A_z plays a prominent part in the following discussions.

In this study the set A_z will be called the intervention set.

Property 9 (continued)

(VIII) For each strategy of the class the Markov process in the set A_z has an absolute stationary probability distribution.

Our discussion of the common mathematical model will now be interrupted. It will be continued in section 3.

First we shall consider some techniques, which are already available and which have proved to be useful.

2) "Dynamic Programming" and "Markovian Decision processes"

There exists a large class of problems, which can be solved with the aid of a technique, called "Dynamic Programming".

In a dynamic programming version of a stochastic ∞-stage decision problem non-degenerate decisions can only be made at equidistant points of time. These points, denoted by Θ_k, divide the time axis into elementary time-intervals of equal length.

Fixing the total losses suffered in some time period, first the losses incurred in the composing elementary time-intervals are calculated. In order to assign costs in a unique way to time-intervals we stipulate the following: The costs which are involved in a decision d_k at Θ_k and those costs at Θ_{k+1}, which are not due to the decision d_{k+1} are included in the loss corresponding to the k^{th} interval.

The properties attributed to the mathematical model of the dynamic programming problem are more or less identical to those which we have attributed in section 1 or shall attribute to the mathematical model in section 3. In order to simplify the discussion, some of them will be reformulated here. It is assumed that the properties 1,2,3 and 4 of section 1 also apply in this section.

In addition we have:

Property 2.1

Independent of the position of an elementary time-interval for each state $I \in \Psi$ and decision $d \in D(I)$ at the beginning of the interval concerned the expected loss $h(I,d)$, assigned to the interval, is defined.

Property 2.2

If I_k is the state of the system at the beginning of the k^{th} interval and if d_k is the decision made at that point of time then the probability distribution of $\underline{I}_{k+1}$ depends on I_k and d_k but neither on k nor on states and decisions taken on or made before I_k. (Markov property).

In an ∞-stage dynamic programming problem losses are discounted. If the discounted expected value of the losses assigned to the interval $(\theta_{k+j},\theta_{k+j+1})$ has to be calculated with respect to θ_k then these losses will be multiplied by a factor α^j, where α satisfies:

$$0 \leq \alpha < 1 \tag{2.1}$$

Property 2.3

If I_k is the state of the system at the beginning of the k^{th} time-interval and if z is the strategy applied then for each k and for each state I_k the expected value of the total discounted costs from θ_k onwards can be given by a function $C(z;I_k)$.

It follows now from the definition of $C(z;I_k)$ that we have:

$$C(z;I_1) = h(I_1;z(I_1)) + \alpha\, \mathscr{E}\left\{ C(z;\underline{I}_2) \middle| I_1;z(I_1)\right\} \tag{2.2}$$

If decisions are made in order to minimize the expected value of the total discounted loss to be incurred then the function $C(z;I_1)$ will be the criterion function for the optimal strategy.

It is easily verified that the optimal strategy z_o has to satisfy the equation:

$$C(z_o;I_1) = \min_{d \in D(I_1)} \left[h(I_1;d) + \alpha\, \mathscr{E}\left\{ C(z_o;\underline{I}_2) \middle| I_1;d_1\right\}\right] \tag{2.3}$$

The solution of the stochastic ∞-stage decision problem will be obtained by solving (2.3). A detailed discussion of the existence and uniqueness of the solution can be found in [2] .

One of the iteration procedures which may yield the optimal strategy works as follows.

Let $z^{(1)}$ be an initial guess of the optimal strategy z_o, then the i^{th} cycle of the iteration procedure is described by

First step

Let z_i be the strategy obtained at the end of the $(i-1)^{th}$ cycle. With the aid of (2.2) the I-function $C(z_i;I)$ can be defined.

Second step

For each state I_1 minimize with respect to $d \in D(I_1)$ the d-function:

$$h(I_1;d) + \alpha\, \mathcal{E}\left\{C(z_i;\underline{I}_2) \mid I_1;d\right\} \tag{2.4}$$

If for a particular state I_1:

$$\min_{d \in D(I_1)} \left[h(I_1;d) + \alpha\, \mathcal{E}\left\{C(z_i;\underline{I}_2) \mid I_1;d\right\} = C(z_i;I_1) \right. \tag{2.5}$$

assign to I_1 the decision $z_i(I_1)$.

From this minimalization procedure a new relation between states and decisions can be derived.

This relation is the new strategy z_{i+1}.

End of the i^{th} cycle

Provided that the assumptions made are also true for the strategies $\left\{z_i;i=2,\ldots\right\}$ and provided that this sequence of strategies converges, the wanted strategy may be obtained.

For proofs the reader is still referred to [2].

Very often problems of this type can also be solved with the aid of a method, called "Markovian Decision processes". This technique is closely related to "Dynamic Programming".

In view of its generalization in section 3 the method will be presented in a somewhat unusual way.

For that purpose the following property is attributed to the mathematical model.

Property (2.4)

If a strategy z of a certain class of strategies is applied then the sequence of states $\underline{I}_k$ at the times Θ_k $(k=1,\ldots)$ can be described by a stationary Markov process with discrete time parameter, having absolute stationary probability distributions and no cyclidally moving sets.*)

----

*) If I_1 is the initial state of this Markov process and if z is the strategy applied, then the absolute stationary probability distribution mentioned above is called an $(z;I_1)$-probability distribution.

If I_1 is the initial state of the system and if z is the strategy applied, then the expected value of the loss to be assigned to a time interval in the steady state is given by:

$$r(z;I_1) = \mathcal{E}\left\{ h(\underline{I};z(\underline{I}) \middle| I_1 \right\} , \tag{2.6}$$

where the expected value is taken with respect to the $(z;I_1)$-probability distribution.

If the optimal strategy z_o exists it will be obvious, that for this strategy the z-function $r(z;I_1)$ has to be minimal.

The ∞-stage Markovian Decision problem is solved by looking at a related problem.

If z_o is the optimal strategy, let us suppose that the decision maker receives for each time interval a premium $r(z_o;I_o)$ for meeting the expenses.

If I_1 is the initial state at θ_1, if $r(z_o;I_o)$ is the promised premium and if the decision maker will apply strategy z, let $C_n(z;I_1;I_o)$ be the expected value of the amount that the decision maker has to pay out of his own pocket in n elementary time intervals.

It follows now from the definition of $C_n(z;I_1;I_o)$ that we have:

$$C_n(z;I_1;I_o) = h(I_1;z(I_1)) - r(z_o;I_o) + \mathcal{E}\left[C_{n-1}(z;\underline{I}_2;I_o) \middle| I_1;z(I_1) \right] \tag{2.7}$$

If $n \to \infty$ the I_1-function $C(z;I_1;I_o) = \lim\limits_{n \to \infty} C_n(z;I_1;I_o)$ satisfies under certain conditions:

$$C(z;I_1;I_o) = h(I_1;z(I_1)) - r(z_o;I_o) + \mathcal{E}\left[C(z;\underline{I}_2;I_o) \middle| I_1;z(I_1) \right] \tag{2.8}$$

Now it is easily verified that if

$$r(z_o;I_1) < r(z;I_1) = \mathcal{E}\left\{ h(\underline{I};z(\underline{I})) \middle| I_1 \right\} \tag{2.9}$$

we have:

$$C(z;I_1;I_o) = \lim_{n \to \infty} C_n(z;I_1;I_o) = \infty .$$

If the expected value of $C(z_o;I_1;I_o)$ is taken with respect to the $(z_o;I_o)$-probability distribution then we find:

$$\mathcal{E}\, C(z_o;\underline{I}_1;I_o) = 0 \tag{2.10}$$

It is obvious now that for each I_1 the optimal strategy z_o has to satisfy the relations:

(I) $$r(z_o;I_1) = \min_{d \in D(I_1)} \mathcal{E}\left[r(z_o;\underline{\xi}) \mid d \right] , \tag{2.11}$$

where $\underline{\xi}$ is distributed according to the probability distribution d.

(II) $$C(z_o;I_1;I_1) =$$

$$= \min_{d \in D(I_1)} \left\{ h(I_1;d) - r(z_o;I_1) + \mathcal{E}\left[C(z_o;\underline{I}_2;I_1) \mid I_1,d \right] \right\} . \tag{2.12}$$

With the aid of these two properties an iteration procedure can be developed which may yield the optimal strategy z_o.

Let $z^{(1)}$ be an initial guess of the optimal strategy z_o. The i^{th} cycle of the iteration procedure is described as follows:

<u>First step</u>

Let z_i be the strategy obtained at the end of the $(i-1)^{th}$ cycle. With the aid of (2.6) the I_o-function $r(z_i;I_o)$ can be defined.

<u>Second step</u>

For each I_1 minimize with respect to $D(I_1)$ the d-function:

$$\mathcal{E}\left[r(z_i;\underline{I}_2) \mid d \right] \tag{2.13}$$

If for a particular state I_1

$$\min_{d \in D(I_1)} \mathcal{E}\left[r(z_i;\underline{I}_2) \mid d \right] = r(z_i;I_1) \tag{2.14}$$

minimize instead of (2.13) the d-function:

$$h(I_1;d) - r(z_i;I_1) + \mathcal{E}\left[C(z_i;\underline{I}_2;I_1) \mid I_1;d \right] \tag{2.15}$$

Moreover, if

$$\min_{d \in D(I_1)} \left[h(I_1;d) - r(z_i;I_1) + \mathcal{E}\left[C(z_i;\underline{I}_2;I_1) \mid I_1;d \right] \right. =$$

$$= C(z_i;I_1;I_1), \tag{2.16}$$

assign to I_1 the decision $z_i(I_1)$.

From this minimalization procedure a new relation between states and decisions can be derived.

This relation is the new strategy z_{i+1}.

End of the i^{th} cycle

If all the assumptions needed are valid for z_i $(i=2,3,\ldots)$ and if these strategies converge then the optimal strategy may be obtained. For the case that the number of possible states is finite the reader is referred to [3].

Finally we like to remark that there exists a second type of Markovian Decision problems. In these problems the decision maker can make non-degenerate decisions only if the system changes its state.

In these problems the following properties are attributed to the mathematical model.

Property 2.5

If the strategy z is applied then the probability of a transition in a period of length dt from state I into a state of a Borel set B in Ψ is defined and given by:

$$\rho_z(B;I)dt \ . \tag{2.17}$$

The probability of two or more transitions in such a period is smaller of magnitude than $(dt)^2$.

Property 2.6

If the strategy z is applied and if the system is during a period of length dt in the state I then the losses incurred are defined and given by:

$$l(I;z)dt \ . \tag{2.18}$$

The costs, which are involved in a transition from state I into I' amount to

$$l(I';I;z)$$

In [2] and [3] it is shown that problems of this type can be solved in an analogical way as described above.

3) Generalized Markovian Decision processes

In many a stochastic ∞-stage decision problem one of the methods discussed in section 2 can be used.

However, if the decision maker is free in choosing his moment of decision it will not always be optimal to restrict himself to equidistant points of time. On the other hand in a large number of problems property (2.5) will not be satisfied. For instance such a situation can occur in inventory problems with state variables, denoting the time elapsed since the last ordering point.

In those inventory problems the state of the system changes continuously. Consequently the probability of having two or more transitions in a period of length dt will be equal to 1.

The discussion in this section is entirely based on the concepts defined and the properties attributed to the mathematical model in section 1.

In that section we have already stated that to each strategy z of the form (1.4) a set of states A_z is assigned. As soon as the system takes on a state in the set A_z a decision will be made in accordance with (1.4).

Let us suppose that Z_o is a class of strategies satisfying the properties given in property 9 of section 1 and by:

Property 9 (continued)

IX) The intersection of all sets A_z $(z \in Z_o)$, denoted by A_o, is not empty.

The set A_o will be called the stopping set. The reason why will be clear very soon.

Let us consider a sequence of random walks $\{\underline{w}_n\}$ $(n=0,1,\ldots)$ in Ψ with the following properties:

a) The initial state of each random walk $\underline{w}_n$ is given by ψ.

b) During the random walk $\underline{w}_n$ n non-degenerate decisions will be made in complete agreement with a given strategy z.

c) After the n^{th} non-degenerate decision no new non-degenerate decision will be made; on the contrary the walk will be ended as soon as a state in A_o has been taken on.

According to these properties, the random walk $\underline{w}_o$ starts in ψ and will be ended as soon as a state in A_o has been taken on.

<u>Property 9</u> (continued)

X) For each initial state ψ and for each applied strategy $z \in Z_o$ the expected loss to be incurred during the random walk $\underline{w}_n$ $(n=0,1,2,\cdots)$, denoted by $k_n(\psi;z)$, is defined and for each finite n uniformly bounded in ψ.

XI) For each initial state ψ and for each applied strategy $z \in Z_o$ the expected duration of the random walk $\underline{w}_n$, denoted by $t_n(\psi;z)$, is defined and for each finite n uniformly bounded in ψ with a positive underbound.

The expected loss, assigned to $\underline{w}_n$, includes the loss to be incurred at the end of the walk and the loss due to the decision made in the initial state ψ. The loss made in the initial state, but not due to the decision is excluded.

Let us now consider two random walks $\underline{w}^d$ and $\underline{w}_o$ in Ψ, both starting in ψ. If d is a non-degenerate decision then the system in $\underline{w}^d$ will just at the beginning of the walk be transferred from the initial state ψ into a new state in accordance with decision d. If d is a degenerate decision then the system in ψ will be "transferred" into ψ.

Anyhow, after the decision d is made, the system in $\underline{w}^d$ will be subjected to the natural process with the "transferred state" as initial state. The random walk $\underline{w}^d$ will be ended as soon as a state in A_o is taken on. In other words, after decision d is made in the initial state ψ, the remaining part of $\underline{w}^d$ is a $\underline{w}_o$-walk from the "transferred state".

As we know, the system in the random walk $\underline{w}_o$ is subjected to a natural process with initial state ψ and will also be stopped as soon as the system takes on a state in A_o.

If the decision d in the initial state of $\underline{w}^d$ is degenerate, then it follows from the definitions of the walks considered, that they are identical.

A function $k(\psi,d)$ can now be defined as being the difference in expected losses between the $\underline{w}^d$- and the $\underline{w}_o$-walk.

A function $t(\psi;d)$ can now be defined as being the difference in expected durations between the $\underline{w}^d$-and the $\underline{w}_o$- walk.

From the definitions of $k(\psi;d)$ and $t(\psi;d)$ it follows:

a) If d is degenerate then:

$$k(\psi;d) = 0 \tag{3.1}$$

$$t(\psi;d) = 0 \tag{3.2}$$

b) If d is non-degenerate then:

$$k(\psi;d) = k_1(\psi;z) - k_o(\psi;z) \tag{3.3}$$

$$t(\psi;d) = t_1(\psi;z) - t_o(\psi;z) \tag{3.4}$$

for each z satisfying:

$$d = z(\psi) \tag{3.5}$$

Note that in spite of (3.3) and (3.4) the structures of the functions $k(\psi;d)$ and $t(\psi;d)$ are independent of any particular strategy z.

For each pair (ψ,d) they are uniquely defined.

Now it can be proved that the following relation is true [1])

$$k_n(\psi;z) = k_m(\psi;z) + \mathcal{E}\left[k_{n-m}(\underline{I}_m;z) - k_o(\underline{I}_m;z) \middle| \psi\right], \tag{3.6}$$

where $\underline{I}_m$ is the state of the system at the time of the m^{th} non-degenerate decision.

It follows now from (3.3) and (3.6) that we have:

$$k_n(\psi;z) = k_o(\psi;z) + \sum_{j=1}^{n} \left[k_j(\psi;z) - k_{j-1}(\psi;z)\right] =$$

$$= k_o(\psi;z) + \sum_{j=1}^{n} \mathcal{E}\left[k(\underline{I}_j;z(\underline{I}_j)) \middle| \psi\right]. \tag{3.7}$$

----

1) At a first glance this relation seems to be obvious. If only <u>one</u> probability space is used then, because of the fact that in $\underline{I}_m$ the costs assigned to the alternative random walks $\underline{w}_{n-m}$ and $\underline{w}_o$ are considered simultaneously, some probabilistic difficulties have to be conquerred. In proving (3.6) we need also the strong Markov property of the natural process (see [1]).

In the same way we find:

$$t_n(\psi;z) = t_o(\psi;z) + \sum_{j=1}^{n} \mathcal{E}\left[t(\underline{I}_j;z(\underline{I}_j))\middle|\psi\right] . \tag{3.8}$$

The solution of the stochastic ∞-stage decision problem will be obtained by solving a more complicated decision problem.

Extension of the problem

In the state ψ_o at the beginning of the random walk $\underline{w}_n$ the decision maker can also decide to put out to contract the control of the system. If he calls in the aid of a controller and if the system is during Δt time units in some state ψ, then for that period he has to pay a fee equal to $r(z;\psi). \Delta t$. (Note that the fee depends on the state of the system at a particular point of time.) According to the contract in return for the fee the hired controller is obliged to apply strategy z from the initial state ψ_o and to pay all the costs, which may occur. Furthermore it is agreed that the fee $r(z;\psi)$ satisfies the following property:

If ψ_1 and ψ_2 belong to the same simple ergodic set of the z-process, then we have [2]:

$$r(z;\psi_1) = r(z;\psi_2) \tag{3.9}$$

In other words, as soon as the system enters a simple ergodic set then the fee per unit of time $r(z;\psi)$ becomes constant.

Now it follows from (3.9) that, if in ψ_o the controller's aid is called in, the expected value of the costs in the steady state per unit of time is given by:

$$\mathcal{E}\left[r(z;\underline{I})\middle|\ \psi_o\right] \tag{3.10}$$

The expected value in (3.10) is taken with respect to the ψ_o-probability distribution in A_z.[3]

----

2) If the strategy z is applied the z-process describes the behaviour of the system in Ψ (thus not only in A_z!!). Simple ergodic sets can not be divided in more than one ergodic set.

3) The ψ_o-probability distribution in A_z is the absolute stationary probability distribution of the Markov process with discrete time parameter in A_z, that corresponds to ψ_o.

Suppose that the strategy z ought to be applied. Let us consider first the situation in which the initial state ψ_o belongs to a simple ergodic set of the z-process.

In accordance with (3.9) the fee $r(z;\psi)$ per unit of time is constant and is given by $r(z;\psi_o)$.

If the decision maker makes decisions in order to minimize the expected future costs and if strategy z ought to be applied during a random walk $\underline{w}_n$ with initial state ψ_o, then he will take the control in his own hands if:

$$k_n(\psi_o;z) \leqq r(z;\psi_o)t_n(\psi_o;z) \tag{3.11}$$

or

$$\frac{k_n(\psi_o;z)}{t_n(\psi_o;z)} \leqq r(z;\psi_o) \quad . \tag{3.12}$$

It follows now from (3.7) and (3.8) that (3.12) is equivalent with:

$$\frac{\frac{1}{n}\, k_o(\psi_o;z) + \frac{1}{n}\sum_{j=1}^{n} \mathcal{E}\left[k(\underline{I}_j;z(\underline{I}_j))\,\middle|\,\psi_o\right]}{\frac{1}{n}\, t_o(\psi_o;z) + \frac{1}{n}\sum_{j=1}^{n} \mathcal{E}\left[t(\underline{I}_j;z(\underline{I}_j))\,\middle|\,\psi_o\right]} \leqq r(z;\psi_o). \tag{3.13}$$

<u>Property 9</u> (continued)

XII) For each strategy $z \in Z_o$ the Markov process in A_z with discrete time parameter,

a) satisfies the Doeblin condition [4)]

b) has no cyclically moving sets.

It can now be proved (see [1]) that for $n \to \infty$ the left hand side of (3.10) converges to:

$$\frac{\mathcal{E}\left[k(\underline{I};z(\underline{I}))\,\middle|\,\psi_o\right]}{\mathcal{E}\left[t(\underline{I};z(\underline{I}))\,\middle|\,\psi_o\right]}, \tag{3.14}$$

where the expected value is taken with respect to the ψ_o-probability distribution in A_z.

----

4) See J.L. Doob: Stochastic Processes p.192.

If $n \gg 0$ and if strategy z ought to be applied then the decision maker will control the system himself if:

$$\frac{\mathcal{E}\left[k(\underline{I};z(\underline{I}))\mid\psi_o\right]}{\mathcal{E}\left[t(\underline{I};z(\underline{I}))\mid\psi_o\right]} < r(z;\psi_o) \tag{3.15}$$

If ψ_o belongs to one of the simple ergodic sets of the z-process, let $r(z;\psi_o)$ from now on be given by:

$$r(z;\psi_o) = \frac{\mathcal{E}\left[k(\underline{I};z(\underline{I}))\mid\psi_o\right]}{\mathcal{E}\left[t(\underline{I};z(\underline{I}))\mid\psi_o\right]} \tag{3.16}$$

Let us now consider the general case and define $r(z;\psi_o)$ by:

$$r(z;\psi_o) = \mathcal{E}\left[\frac{\mathcal{E}\left[k(\underline{I};z(\underline{I}))\mid I_o\right]}{\mathcal{E}\left[t(\underline{I};z(\underline{I}))\mid I_o\right]}\;\middle|\;\psi_o\right] \tag{3.17}$$

where $\underline{I}_o$ is distributed according to the ψ_o-probability distribution in A_z.

In other words, if in ψ_o the controller's aid is called in then $r(z;\psi_o)$ is equal to the expected value of the costs in the steady state per unit of time.

Property 9 (continued)

XIII) For each initial state ψ and for each applied strategy $z \in Z_o$ the expected value of the fee to be paid to the controller during the random walk $\underline{w}_n$, denoted by $k_n^*(\psi;z)$, is defined and uniformly bounded in ψ.

It is easily verified (cf. 3.9) that, if ψ belongs to a simple ergodic set, we have:

$$k_n^*(\psi;z) = r(z;\psi)\,t_n(\psi;z) \tag{3.18}$$

Analogous to the function $k(\psi;z)$ we can also define the function $k^*(\psi;z)$.

For a random walk $\underline{w}_n$ with initial state ψ_o the difference in expected costs between both types of control is given by:

$$g_n(z;\psi_o) = k_n(\psi_o;z) - k_n^*(\psi_o;z) \tag{3.19}$$

Now it can be proved with the aid of the points X, XI, XII and XIII of property 9 that the limit

$$g(z;\psi_o) = \lim_{n\to\infty} g_n(z;\psi_o)$$

exists and can be given by:

$$g(z;\psi_o) = k_o(\psi_o;z) - k_o^*(\psi_o;z) +$$

$$+ \sum_{j=1}^{\infty} \mathcal{E}\left[k(\underline{I}_j;z(\underline{I}_j)) - k^*(\underline{I}_j;z(\underline{I}_j)) \,\Big|\, \psi_o\right] \tag{3.20}$$

where $\underline{I}_1 = \psi_o$ if $\psi_o \in A_z$.

If the controller's aid is called in and if two different strategies z_1 and z_2 are considered, which satisfy

$$r(z_1;\psi_o) > r(z_2;\psi_o) \tag{3.21}$$

then for an infinite period of time the difference in expected costs between the strategy z_1 and the strategy z_2 will be $+\infty$.

If the decision maker takes the control in his own hands, and if both strategies may be applied during an infinite period of time, then, because $g(z;\psi)$ is finite, it follows that the difference in expected costs is also $+\infty$.

Consequently, in case of own control the optimal strategy will also minimize:

$$r(z;\psi_o) = \mathcal{E}\left[\frac{\mathcal{E}\{k(\underline{I};z(\underline{I}))|\underline{I}_o\}}{\mathcal{E}\{t(\underline{I};z(\underline{I}))|\underline{I}_o\}} \,\Big|\, \psi_o\right] \tag{3.22}$$

where $\underline{I}_o$ is distributed according to the ψ_o-probability distribution in A_z.

Now we shall mention four properties of the optimal strategy z_o:

(I) Let us assume that for a state $\psi_o \in \Psi$ a feasible decision d can be found, such that:

$$\mathcal{E}\left[r(z;\underline{\xi})\,\middle|\,d\right] < r(z;\psi_o) \tag{3.23}$$

If (3.23) is true for the optimal strategy then it should not be profitable to apply z_o in ψ_o.
But by definition $z_o(\psi_o)$ is the optimal decision and so we have:

$$r(z_o;\psi_o) = \min_{d\in D(\psi_o)} \mathcal{E}\left[r(z_o;\underline{\xi})\,\middle|\,d\right] \tag{3.24}$$

(II) Let ψ_o be the initial state of a random walk $\underline{w}_{\infty}$ in Ψ. If the system is in ψ and if the strategy z is applied let us suppose that, for meeting the expenses during $\underline{w}_{\infty}$, the decision maker receives a premium equal to $r(z;\psi)$ per unit of time. (Note, that the premium is equal to the controller's fee.) In addition, if he does not call in controller's aid, he gets in the initial state an amount equal to $k_o(\psi_o;z) - k_o^*(\psi_o;z)$.

So if he takes the control in his own hands and if he applies strategy z the expected value of the amount he has to pay out of his own pocket is given by:

$$C(z;\psi_o) = g(z;\psi_o) - k_o(\psi_o;z) + k_o^*(\psi_o;z) \tag{3.25}$$

or, according to (3.20)

$$C(z;\psi_o) = \sum_{j=1}^{\infty} \mathcal{E}\left[k(\underline{I}_j;z(\underline{I}_j)) - k^*(\underline{I}_j;z(\underline{I}_j))\,\middle|\,\psi_o\right]^{5)} \tag{3.26}$$

where $\underline{I}_1 = \psi_o$ if $\psi_o \in A_z$.

Hence (cf. (3.1))

$$C(z;\psi_o) = k(\psi_o;z(\psi_o)) - k^*(\psi_o;z(\psi_o)) +$$

$$+ \sum_{j=1}^{\infty} \mathcal{E}\left[k(\underline{I}_j;z(\underline{I}_j)) - k^*(\underline{I}_j;z(\underline{I}_j))\,\middle|\,z(\psi_o)\right] \tag{3.27}$$

or

$$C(z;\psi_o) = k(\psi_o;z(\psi_o)) - k^*(\psi_o;z(\psi_o)) + \mathcal{E}\left[C(z_o;\underline{\xi})\,\middle|\,z(\psi_o)\right] \tag{3.28}$$

----

5) Note that in $C(z;\psi_o)$ no costs are worked up which will be incurred before the first decision point. The first decision, however, has to be made before or at the moment the system enters A_o.

It will be clear that the decision maker wants to minimize the expected value of his own share.

Now let us assume that the decision maker in ψ_o is allowed to minimize that amount, unless he effects an increase of the expected costs in the steady state per unit of time. In other words (cf.(3.24)), if the decision maker applies z_o, he is authorized in ψ_o to make a decision only if this decision does not change the expected value of future premiums.

Thus:

$$\mathcal{E}\left[r(z_o;\underline{\xi})\middle| d\right] = r(z_o;\psi_o) \tag{3.29}$$

Now it is easily verified that the optimal strategy has to satisfy (3.24) and

$$C(z_o;\psi_o) = \min_{d\in D^*(\psi_o)} \left[k(\psi_o;d) - k^*(\psi_o;d) + \mathcal{E}\left\{C(z_o;\underline{\eta})\middle| d\right\}\right], \tag{3.30}$$

where $D^*(\psi_o)$ is the intersection of $D(\psi_o)$ and the set of decisions, which satisfy (3.29).

(III) Let us suppose that the decision maker in state $\psi\in \bar{A}_o$ is wondering, whether he will make a non-degenerate decision according to strategy z now or he will wait T time units, provided that the system takes on no state in A_o during these T time units. If in the second reflection no state in A_o is taken on then in accordance with z a decision will be made at the end of the period of T time units. But, if the system takes on a state in A_o during that period then a decision will be made earlier. In that case in accordance with z the decision is made in the entering state of A_o. (Note: A_o is the intersection of all intervention sets, so the decision maker cannot wait longer.) It follows from the definition of the optimal strategy z_o that in states in A_{z_o} a postponement of a non-degenerate decision is not profitable. Or in other words, if $\underline{\xi}$ is the state in which in the second reflection the decision is made, then

$$\min_{T} \mathcal{E}\left[r(z_o;\underline{\xi}) \,\middle|\, \psi, T\right] = r(z_o;\psi) \qquad (3.31)$$

(IV) Let ψ_o be the initial state of a random walk $\underline{w}_{\infty}$ in state space. If the system is in ψ and if the strategy z is applied suppose that for meeting the expenses during $\underline{w}_{\infty}$ the decision maker receives a premium equal to $r(z;\psi)$ per unit of time. In addition if he does not call in controller's aid he gets in the initial state an amount equal to $k_o(\psi_o;z) - k_o^*(\psi_o;z)$.
So if he takes the control in his own hands and if he applies strategy z the expected value of the amount he has to pay out of his own pocket is given by:

$$C(z;\psi_o) = k(\psi_o;z(\psi_o) - k^*(\psi_o;z(\psi_o) + \\ + \mathcal{E}\left[C(z;\underline{\xi}) \,\middle|\, z(\psi_o)\right] \qquad (3.28)$$

It will be clear that the decsion maker wants to minimize the expected value of his own share.

Now let us assume that the decision maker is allowed to minimize that amount unless he effects an increase of the expected value of the costs in the steady state per unit of time (i.e. premium). In other words in ψ_o he is authorized to postpone decisions, if (cf. (III))

$$r(z_o;\psi) = \mathcal{E}\left[r(z_o;\underline{\xi}) \,\middle|\, \psi, T\right] \qquad (3.32)$$

Now it can easily be verified that the optimal strategy has to satisfy (3.31) and

$$C(z_o;\psi) = \min_{T \in T(\psi)} \mathcal{E}\left[C(z_o;\underline{\xi}) \,\middle|\, \psi; T\right] \text{ 6)} \qquad (3.33)$$

where $T(\psi)$ is the set of T values, which satisfy (3.32).

The functional equations (3.31) and (3.33) fix the shape of the intervention set of the optimal strategy. In many decision problems the boundary of that intervention set is given by:

----

6) cf. 5)

$$\left[\frac{d}{dT}\ \mathcal{E}\left\{r(z_o;\underline{\xi})\,\middle|\,\psi\ ;T\right\}\right]_{T=0} = 0 \qquad (3.34)$$

and

$$\left[\frac{d}{dT}\ \mathcal{E}\left\{C(z_o;\underline{\xi})\,\middle|\,\psi\ ;T\right\}\right]_{T=0} = 0 \qquad (3.35)$$

With the aid of these four properties an iteration procedure can now be developed which may yield the optimal strategy.

Let $z_1 \in Z_o$ be an initial guess of the optimal strategy.

The i^{th} cycle of the iteration procedure is now described by:

<u>First step</u>

Let $z_i \in Z_o$ be the strategy obtained at the end of the $(i-1)^{th}$ cycle. With the aid of (3.22) and (3.26) the functions $r(z_i;\psi)$ and $C(z_i;\psi)$ are defined.

<u>Second step</u>

The intervention set A_{z_i} will first be purified in such a way that only states in A_{z_i} remain, which satisfy:

$$\min_{T}\ \mathcal{E}\left\{r(z_i;\underline{\xi})\,\middle|\,\psi;T\right\} = r(z_i;\psi) \qquad (3.36)$$

The meaning of $\underline{\xi}$ is given in property III of the optimal strategy.

The set of states $\psi \in \Psi$ which satisfy (3.36) will be denoted by B_i.

Next we shall remove those states in A_{z_i} which belong to B_i and satisfy:

$$\min_{T \in T(\psi)}\ \mathcal{E}\left[C(z_i;\underline{\xi})\,\middle|\,\psi,T\right] < C(z_i;\psi), \qquad (3.37)$$

where $T(\psi)$ is the set of T values which satisfy:

$$\mathcal{E}\left[r(z_i;\underline{\xi})\,\middle|\,\psi\ ,T\right] = r(z_i;\psi) \qquad (3.38)$$

The remaining states I in A_{z_i} and their corresponding decisions form a new strategy $z_i^{(1)}$. It is assumed that $z_i^{(1)} \in Z_o$.

<u>Third step</u>:

Repeat the first step but now with $z_i^{(1)}$.

Fourth step:

For each $\psi \in \Psi$ minimize with respect to $d \in D(\psi)$ the d-function:

$$\mathcal{E}\left\{ r(z_i^{(1)};\underline{I}) \mid d \right\}. \tag{3.39}$$

The set of states $\psi \in \Psi$ which satisfy

$$\min_{d \in D(\psi)} \mathcal{E}\left\{ r(z_i^{(1)};\underline{I}) \mid d \right\} = r(z_i^{(1)};\psi) \tag{3.40}$$

will be denoted by B_i'.

Next minimize for each $\psi \in B_i'$ with respect to $d \in D^*(\psi)$ the d-function:

$$k(\psi;d) - r(z_i^{(1)};\psi)t(\psi;d) + \mathcal{E}\left\{ C(z_i^{(1)};\underline{I}) \mid d \right\}, \tag{3.41}$$

where $D^*(\psi)$ is the intersection of $D(\psi)$ and the set of decisions d which satisfy

$$\mathcal{E}\left\{ r(z_i^{(1)};\underline{I}) \mid d \right\} = r(z_i^{(1)};\psi). \tag{3.42}$$

From this minimalization procedure a new relation between states and decisions can be derived.

If for some state $\psi \in B_i'$ we have:

$$C(z_i^{(1)};\psi) = \min_{d \in D^*(\psi)} \left[k(\psi;d) - r(z_i^{(1)};\psi)t(\psi;d) + \mathcal{E}\left\{ C(z_i^{(1)};\underline{I}) \mid d \right\} \right] \tag{3.43}$$

assign to that state ψ the decision $z_i^{(1)}(\psi)$.

The relation between states and decision which will be obtained in this way, is denoted by z_{i+1}. It is assumed that this strategy is an element of Z_o.

End of the i^{th} cycle

Without more details about the structure of the functions, probability distributions etc. used, it seems difficult to give a rigorous proof of the effectiveness of the iteration procedure given above.

In [1] we shall stipulate sufficient conditions for obtaining strategies z_i and $z_i^{(1)}$ which belong to Z_o.

Assuming that these conditions are fulfilled, we shall now show how the effectiveness of the procedure in a special case can be proved.

For that purpose we introduce mixed strategies of the form $(z_a)^n(z_b)^m z_c$ $(z_j \in Z_o,\ j=a,b,c)$.

In succession such a strategy prescribes:

1) n decisions in accordance with z_a
2) m decisions with the aid of z_b
3) an infinite number of decisions in conformity with strategy z_c.

For the mixed strategies we shall define the functions

$r((z_a)^n(z_b)^m z_c;\psi)$ and $C((z_a)^n(z_b)^m z_c;\psi)$ by:

$$r((z_a)^n(z_b)^m z_c;\psi) = \mathcal{E}\left\{r((z_b)^m z_c;\underline{\xi})\ \middle|\ \psi\right\} = \mathcal{E}\left\{r(z_c;\underline{\eta})\middle|\psi\right\} \tag{3.44}$$

and

$$C((z_a)^n(z_b)^m z_c;\psi) = \sum_{j=1}^{n} \mathcal{E}\left[k(\underline{I}_j;z_a(\underline{I}_j)) - k^*(\underline{I}_j;z_a(\underline{I}_j))\ \middle|\ \psi\right]$$

$$+ \sum_{j=n+1}^{n+m} \mathcal{E}\left\{k(\underline{I}_j;z_b(\underline{I}_j)) - k^*(\underline{I}_j;z_b(\underline{I}_j))\middle|\underline{\xi}\right\} + \mathcal{E}\,C(z_c;\underline{\eta}) \tag{3.45}$$

respectively, where $\underline{\xi}$ is distributed according to $z_a(\underline{I}_m)$ and $\underline{\eta}$ is distributed according to $z_b(\underline{I}_{n+m})$.

Let us consider now the second step of the iteration procedure. According to (3.36) we find for each $\psi \in \Psi$:

$$r(z_i^{(1)};z_i;\psi) \leqq r(z_i;\psi) \tag{3.46}$$

and by induction:

$$r((z_i^{(1)})^n z_i;\psi) \leqq r(z_i;\psi) \quad . \tag{3.47}$$

If $n \rightarrow \infty$ it can be proved with the aid of property 9 that:

$$\lim_{n\rightarrow\infty} r((z_i^{(1)})^n z_i;\psi) = \mathcal{E}\left[r(z_i;\underline{I})\ \middle|\psi\right] \leqq r(z_i;\psi) \quad , \tag{3.48}$$

where $\underline{I}$ is distributed according to the ψ probability distribution in $A_{z_i^{(1)}}$.

From $\lim_{n\to\infty} r((z_i^{(1)})^n z_i;\psi) = \lim_{n\to\infty} r((z_i^{(1)})^{n-k}(z_i^{(1)})^k z_i;\psi)$ (3.49)

we can deduce for each finite integral number k:

$$\mathcal{E}\ r((z_i^{(1)})^k z_i;\underline{I}) = \mathcal{E}\ r(z_i;\underline{I}) \tag{3.50}$$

and thus with respect to the ψ-probability distribution in $A_{z_i^{(1)}}$:

$$r((z_i^{(1)})^k z_i;\underline{I}) = r(z_i;\underline{I}) \text{ spr } O . \tag{3.51}$$

It follows from the definition of B_i that:

$$\underline{I} \in B_i \qquad \text{spr } O . \tag{3.52}$$

It can now easily be verified that the simple ergodic sets of the $z_i^{(1)}$-process are also inside B_i.

If ψ_1 and ψ_2 belong to the same simple ergodic set of the $z_i^{(1)}$-process then we have:

$$r(z_i^{(1)} z_i;\psi_j) = r(z_i;\psi_j) \qquad j=1,2 \tag{3.53}$$

and by induction:

$$r((z_i^{(1)})^n z_i;\psi_j) = r(z_i;\psi_j) \qquad j=1,2 \tag{3.54}$$

If $n\to\infty$

$$\lim_{n\to\infty} r((z_i^{(1)})^n z_i;\psi_j) = \mathcal{E}\ r(z_i;\underline{I}) = r(z_i;\psi_j) \tag{3.55}$$

$j=1,2$

where $\underline{I}$ is distributed according to the identical ψ_j-probability distributions in $A_{z_i^{(1)}}$.

Hence $\quad r(z_i;\psi_1) = r(z_i;\psi_2)$ (3.56)

It follows from (3.37) that for each ψ' belonging to a simple ergodic set of the $z_i^{(1)}$-process we have:

$$C(z_i^{(1)} z_i;\psi') \leqq C(z_i;\psi') \tag{3.57}$$

Using this inequality just after the first decision we find:

$$C((z_i^{(1)})^2 z_i; \mathcal{Y}') \leqq C(z_i^{(1)} z_i; \mathcal{Y}') \tag{3.58}$$

It follows now from (3.45) that (3.58) is equivalent with:

$$\mathcal{E}\left\{k(\underline{I}_1; z_i^{(1)}(\underline{I}_1)) - k^*(\underline{I}_1; z_i^{(1)}(\underline{I}_1)) \,\middle|\, \mathcal{Y}'\right\} +$$

$$+ \mathcal{E}\left\{C(z_i^{(1)} z_i; \underline{I}_2) \,\middle|\, \mathcal{Y}'\right\} \leqq C(z_i^{(1)} z_i; \mathcal{Y}') \tag{3.59}$$

Let us suppose that, if the decision maker applies strategy $(z_i^{(1)})^2 z_i$ from $\mathcal{Y}'$ onwards, he gets a premium per unit of time, which is equal to the expected value of the costs in the steady state per unit of time. In the initial state $\mathcal{Y}'$ the premium is given by $r((z_i^{(1)})^2 z_i; \mathcal{Y}')$. It follows now from (3.54) and (3.56) that, if $\mathcal{Y}$ is a state in a simple ergodic set of the $z_i^{(1)}$-process, the premium will always be equal to $r(z_i; \mathcal{Y}')$.

Consequently (3.59) is equivalent with (cf. (3.18)):

$$\mathcal{E}\left\{k(\underline{I}_1; z_i^{(1)}(\underline{I}_1)) - r(z_i; \mathcal{Y}')t(\underline{I}_1; z_i^{(1)}(\underline{I}_1)) \,\middle|\, \mathcal{Y}'\right\} +$$

$$+ \mathcal{E}\left\{C(z_i^{(1)} z_i; \underline{I}_2) \,\middle|\, \mathcal{Y}'\right\} \leqq C(z_i^{(1)} z_i; \mathcal{Y}') \tag{3.60}$$

If $\mathcal{Y}''$ belongs to one of the simple ergodic sets let $\mathcal{Y}'$ in (3.60) now be distributed according to the $\mathcal{Y}''$-distribution in $A_{z_i^{(1)}}$.

It is easily verified that by taking the expectation of both sides of (3.60) we find:

$$\mathcal{E}\left\{k(\underline{I}; z_i^{(1)}(\underline{I})) - r(z_i; \mathcal{Y}'')t(\underline{I}; z_i^{(1)}(\underline{I})) \,\middle|\, \mathcal{Y}''\right\} \leqq 0 \tag{3.61}$$

where $\underline{I}$ is distributed according to the $\mathcal{Y}''$-distribution in $A_{z_i^{(1)}}$.

From (3.61) it follows:

$$r(z_i^{(1)}; \mathcal{Y}'') = \frac{\mathcal{E}\left\{k(\underline{I}; z_i^{(1)}(\underline{I})) \,\middle|\, \mathcal{Y}''\right\}}{\mathcal{E}\left\{t(\underline{I}; z_i^{(1)}(\underline{I})) \,\middle|\, \mathcal{Y}''\right\}} \leqq r(z_i; \mathcal{Y}'') \tag{3.62}$$

Substituting $\psi''=\underline{I}$ in (3.62) and taking expectations with respect to some ψ -probability distribution in $A_{z_i^{(1)}}$ then we find with the aid of (3.48):

$$\mathcal{E}\left\{ r(z_i^{(1)};\underline{I}) \middle| \psi \right\} \leqq \mathcal{E}\left\{ r(z_i;\underline{I}) \middle| \psi \right\} \leqq r(z_i;\psi), \tag{3.63}$$

where $\underline{I}$ is distributed according to the ψ-distribution in $A_{z_i^{(1)}}$.

Hence
$$r(z_i^{(1)};\psi) \leqq r(z_i;\psi) \tag{3.64}$$

This proves the effectiveness of the second step in the iteration procedure.

In the same way we can prove that z_{i+1} is to prefer to $z_i^{(1)}$ (fourth step).

Now we have shown that for each $\psi \in \Psi$ the sequence of strategies $\left\{ z_i \right\}$ satisfies:

$$r(z_i;\psi) \leqq r(z_{i-1};\psi) \quad . \tag{3.65}$$

Except the speed of convergence, in practice the only point of interest is whether:

$$r_o(\psi) = \lim_{n \to \infty} r(z_n;\psi) \text{ is equal to } r(z_o;\psi)$$

or not.

Let us introduce the following assumption:

a) after a finite number of cycles M_o the sets B_i and $B_i{}'$ are always identical with Ψ.

b) the sequence of ψ-functions defined on $B_i{}'$ by:

$$k(\psi;z_{i+1}(\psi)) - k^*(\psi;z_{i+1}(\psi)) + \mathcal{E}\left\{ C(z_i^{(1)};\underline{I}) \,\middle|\, z_{i+1}, \psi) \right\} + \\ -C(z_i^{(1)};\psi) \tag{3.66}$$

converges uniformly in ψ to zero.

Consequently if z_o is the optimal strategy and if $i \geqq M_o$ we have for each ψ :

$$r(z_o z_i^{(1)};\psi) \geqq r(z_i^{(1)};\psi) \qquad (3.67)$$

and by induction:

$$r((z_o)^n z_i^{(1)};\psi) \geqq r(z_i^{(1)};\psi) \ . \qquad (3.68)$$

It can now be proved with the aid of property 9 that:

$$\lim_{n\to\infty} r((z_o)^n z_i^{(1)};\psi) = \mathcal{E}\left\{r(z_i^{(1)};\underline{I})\middle|\psi\right\} \geqq r(z_i^{(1)};\psi), \qquad (3.69)$$

where $\underline{I}$ is distributed according to the ψ-distribution in A_{z_o}. The convergence in (3.69) is uniformly in ψ.

It follows now from (3.69) that if ψ_1 and ψ_2 belong to some simple ergodic set of the z_o-process we have:

$$\lim_{n\to\infty} r((z_o)^n z_i^{(1)};\psi_1) = \lim_{n\to\infty} r((z_o)^n z_i^{(1)};\psi_2) \qquad (3.70)$$

Let us consider the strategy $(z_o)^n z_i^{(1)}$. According to (3.45) the following relation is true:

$$C((z_o)^n z_i^{(1)};\psi) = \mathcal{E}\left\{k(\underline{I}_1;z_o(\underline{I}_1)) - k^*(\underline{I}_1;z_o(\underline{I}_1))\middle|\psi\right\} + \\ + \mathcal{E}\left\{C((z_o)^{n-1} z_i^{(1)};\underline{I}_2)\middle|\psi\right\}. \qquad (3.71)$$

By induction it follows from the definition of z_{i+1}

$$C((z_o)^{n-1} z_{i+1} z_i^{(1)};\psi) \leqq C((z_o)^n z_i^{(1)};\psi) \ . \qquad (3.72)$$

According to b) for a given $\varepsilon > 0$ an integral value $N_o(\varepsilon)$ can be found such that for $i \geqq N_o(\varepsilon)$

$$C((z_o)^{n-1} z_i^{(1)};\psi) - \varepsilon \leqq C((z_o)^{n-1} z_{i+1} z_i^{(1)};\psi) \qquad (3.73)$$

Take $i \geqq \max(N_o(\varepsilon), M_o)$ then it follows from (3.71), (3.72) and (3.73) that:

$$C((z_o)^{n-1} z_i^{(1)};\psi) - \varepsilon \leqq \mathcal{E}\left\{k(\underline{I}_1;z_o(\underline{I}_1)) - k^*(\underline{I}_1;z_o(\underline{I}_1))\middle|\psi\right\} + \\ + \mathcal{E}\left\{C((z_o)^{n-1} z_i^{(1)};\underline{I}_2)\middle|\psi\right\} \ . \qquad (3.74)$$

Let ψ_o be a state in a simple ergodic set of the z_o-process and let ψ in (3.74) be distributed according to the ψ_o-distribution in A_{z_o}.

Then after taking expectations of both sides of (3.74) we get:

$$-\varepsilon \leqq \mathcal{E}\left\{k(\underline{I};z_o(\underline{I})) - k^*(\underline{I};z_o(\underline{I})) \mid \psi_o\right\} \tag{3.75}$$

where $\underline{I}$ is distributed according to the ψ_o-probability distribution in A_{z_o}.

According to (3.70) if $n \to \infty$ then with respect to the ψ_o-probability distribution in A_{z_o}

$$\lim_{n\to\infty} r((z_o)^n z_i^{(1)};\underline{I}) = \mathcal{E}\left\{r(z_i^{(1)};\underline{\xi}) \mid \psi_o\right\} \quad \text{spr } 0 \tag{3.76}$$

where $\underline{\xi}$ is distributed according to the ψ_o-probability distribution in A_{z_o}. The convergence is uniformly in $\underline{I}$.

Thus if $n \to \infty$ we have:

$$-\varepsilon \leqq \mathcal{E}\left[k(\underline{I};z_o(\underline{I})) - \mathcal{E}\left\{r(z_i^{(1)};\underline{\xi}) \mid \psi_o\right\} . t(\underline{I};z_o(\underline{I})) \mid \psi_o\right] \tag{3.77}$$

or

$$\frac{-\varepsilon}{\mathcal{E}\left\{t(\underline{I};z_o(\underline{I})) \mid \psi_o\right\}} \leqq r(z_o;\psi_o) - \mathcal{E}\left\{r(z_i^{(1)};\underline{I}) \mid \psi_o\right\} \tag{3.78}$$

Rewriting (3.78) we obtain with the aid of (3.69):

$$r(z_o;\psi_o) \geqq r(z_i^{(1)};\psi_o) - \frac{\varepsilon}{\mathcal{E}\left\{t(\underline{I};z_o(\underline{I})) \mid \psi_o\right\}} \tag{3.79}$$

and thus if ψ_o belongs to a simple ergodic set of the z_o-process:

$$r(z_o;\psi_o) = \lim_{n\to\infty} r(z_n^{(1)};\psi_o) \quad . \tag{3.80}$$

Finally, if ψ_o is a transient state of the z_o-process, we have:

$$r(z_i^{(1)};\psi_o) \leqq \lim_{n\to\infty} r((z_o)^n z_i^{(1)};\psi_o) = \mathcal{E}\left\{r(z_i^{(1)};\underline{I}) \mid \psi_o\right\}, \tag{3.81}$$

where $\underline{I}$ is distributed according to the ψ-distribution in A_{z_o} (cf. (3.69)).

Because $\underline{I}$ is a state of some simple ergodic set we find with the aid of (3.79) and (3.81):

$$r(z_i^{(1)};\psi_o) \leqq \mathcal{E}\left\{(r(z_i^{(1)};\underline{I}) \mid \psi_o\right\} \leqq \mathcal{E}\left\{r(z_o;\underline{I}) \mid \psi_o\right\} + \\ + \mathcal{E}\left[\frac{\varepsilon}{\mathcal{E}\left\{t(\underline{I}';z_o(\underline{I}'))\middle|\,\underline{I}\right\}} \,\middle|\, \psi_o\right], \qquad (3.82)$$

where $\underline{I}'$ is distributed according to the $\underline{I}$-probability distribution in A_{z_o} and $\underline{I}$ is distributed according to the ψ_o-distribution in A_{z_o}.

Hence

$$\mathcal{E}\left\{r(z_o;\underline{I}))\middle|\psi_o\right\} = r(z_o;\psi_o) = r_o(\psi_o) \quad . \qquad (3.83)$$

Starting from certain assumptions the effectiveness of the iteration procedure is proved.

4) The relation between the new and the old method

In this section we will discuss problems which can be formulated as Markovian Decision problems, but which are also solvable with the aid of the method discussed in section 3.

Moreover, the freedom of formulation we meet in the new method, can be utilized in such a way that both approaches are identical.

Let us suppose that the decision problem is formulated as an "old" Markovian Decision problem. In "old" Markovian Decision problems, only the states of the system, either at equidistant points of time or at moments at which the system changes its state, are of interest. In section 2 we have always marked these states with the symbol I.

Let us restrict ourselves to problems in which only at equidistant points of time non-degenerate decisions can be made.

Let the state space and the decision space in the old version be given by Ψ and D respectively.

Now we add to the state space Ψ a new state variable s, that measures the length of the period elapsed since the last non-degenerate decision.

The states in the extended state space Ψ^* are indicated by $\psi^* = (\psi; s)$.

At each point of time the state of the system can be given uniquely by a point $\psi^* \in \Psi^*$.

If the state of the system is presented by such a point then in the problem concerned non-degenerate decisions are allowed to be made only in states, in which the state variable s takes on an integral value. In other words, if the s-component of ψ^* is not an integer then $D(\psi^*)$ contains only the probability distribution concentrated in ψ^*.

Non-degenerate decisions defined in the original model can also be presented by probability distributions of states ψ^* in Ψ^*. These probability distributions are determined by the old ones with the aid of the relation: $\underline{\psi}^* = (\underline{\psi}; 0)$.

Now we have shown that the restrictions inherent in the "old" Markovian Decision problem can also be worked into the more general model. This model will be used in the new method.

However, if we compare the Markov process in A_z with that of the old version, then we observe that the two processes are not identical. This is obvious, because in the new version we limit ourselves to states which correspond to non-degenerate decisions, whereas in the old version states at equidistant points of time are considered, whose related decisions may be degenerate.

Consequently the new method introduces a new technique for solving "old" Markovian decision problems. In section 5 we shall use this technique for solving the well-known (S,s)-inventory problem.

In order to get identical techniques we stipulate for the more general model:

a) In $(\psi;1)$ the degenerate decision given by the probability distribution concentrated in $(\psi;1)$ is not feasible.

b) In $(\psi;1)$ the decision given by the probability distribution concentrated in $(\psi;0)$ is feasible and does not effect costs.

As a consequence of this regulation in the new model at each of the equidistant points of time a non-degenerate decision is made.

In other words for each strategy z the intervention set A_z contains all states of the form $(\psi;1)$ and no other states.

Consequently we have:

$$A_z = A_o \tag{4.1}$$

Now it is easily verified that the Markov processes in the two formulations are identical. Both processes describe the states of the system at the equidistant points of time.

As we know the functions $k(\psi^*;d)$ and $t(\psi^*;d)$ in the general model are defined with the aid of the two random walks $\underline{w}^d$ and $\underline{w}_o$.

From (4.1) it follows that the random walk $\underline{w}_o$ consists of the initial state only, while $\underline{w}^d$ is a random walk between two equidistant points of time.

Now it follows from the definitions of the functions $k(\psi^*;z)$, $h(\psi;z)$, $t(\psi^*;z)$, $r(\psi^*;z)$ and $r(\psi;z)$ that we have:

$$k((\psi;1);z) = h(\psi;z) \tag{4.2}$$

$$t((\psi;1);z) = 1 \tag{4.3}$$

and consequently (cf.(2.6) and (3.17))

$$r(z;(\psi;1)) = r(z;\psi) \tag{4.4}$$

Finally the identity (4.1) also implies that the second step in the iteration procedure of the new method always fails (delay is not permitted). For this reason the two iteration procedures are identical too.

5) Examples

I) The (S,s)-inventory problem

The (S,s)-inventory problem is often used for illustrating dynamic programming techniques [2].

In [4] this inventory problem is solved with the aid of Markov processes.

In this section, however, we shall show that the (S,s)-inventory problem can also be solved with the aid of the method discussed in section 3.

Let us suppose that a wholesaledealer can replenish without lead time a stock of some commodity at equidistant points of time Θ_k (k=1,2,...). The intervals between these points of time are of unit length and are called elementary time intervals.

In addition we assume the following:

a) the purchase price of q units is expressed by a function Q(q).
b) the inventory costs are C_1 per unit a unit of time in stock.
c) emergency purchases will be made if the inventory is run out between two equidistant points of time. The purchase price is then C_2 per unit.
d) the probability distribution of the number of clients, arriving in T elementary time intervals is given by a Poisson distribution with parameter λT.
e) the demand $\underline{x}$ per client is distributed according to the distribution function F(x). It is assumed that there is no dependence between the demands of successive clients.

In the solution of the (S,s)-inventory problem we restrict ourselves to those strategies which prescribe only purchases at points of time Θ_k, whereupon the stock is equal or below a certain level s(s $\geqq$0). At such points of time the stock will be supplied till S. (S$\geqq$s).

In other words the strategies considered differ in the corresponding levels s and S only.

In fig.1 a state space Ψ is suggested.

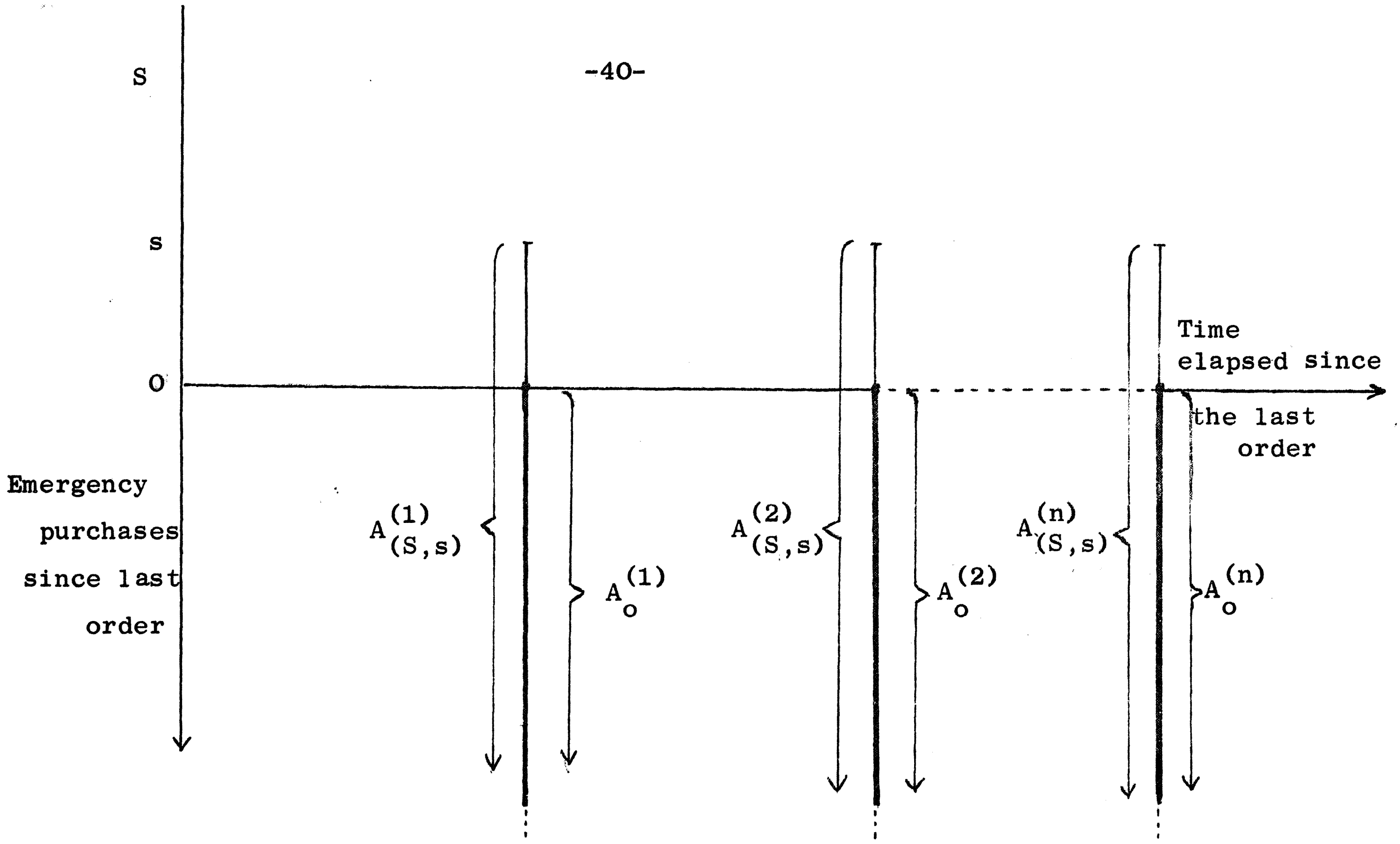

fig.1 The state space ψ

On the horizontal axis the state variable "time since the last order" is plotted, while on the vertical axis the stock and the emergency purchases can be fixed. Because the strategies concerned are completely determined by the two parameters S and s the intervention sets will be denoted by $A_{(S,s)}$ instead of A_z.

In fig.1 an intervention set $A_{(S,s)}$ and the stopping set A_o are marked. Note that the two sets consist of an infinite number of non-connected subsets (half lines), $A^{(k)}_{(S,s)}$ and $A^{(k)}_{o}$ respectively.

It can now easily be verified that the probability distribution of the demand in T time units is given by the distribution function:

$$G_T(y) = \sum_{j=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^j}{j!} \quad F_j(y) \tag{5.1}$$

where $F_j(y)$ is the distribution function of the sum of j equally and independently distributed stochastic variables $\underline{x}_i$ and where $F_o(y)=1$.

Of interest is also the distribution function of the time needed for selling a quantity x. Let us denote this time by the stochastic variable $\underline{t}_x$ and let the distribution function be given by $K_x(t)$.

It is easily verified that we have:

$$G_T(x) = 1 - K_x(T) \quad . \tag{5.2}$$

Hence

$$\mathcal{E}\, \underline{t}_x = \int_0^\infty t dK_x(t) = \int_0^\infty (1-K_x(t))dt =$$

$$= \sum_{j=0}^\infty F_j(x) \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^j}{j!} dt = \frac{1}{\lambda} \sum_{j=0}^\infty F_j(x) \tag{5.3}$$

If the strategy (S,s) is applied then the distribution of the seize of the stock $\underline{u}$ left at the time of a new supply (non degenerate decision) is given by:

$$H_{S,s}(u) = \left[1-G_1(S-u)\right] + \sum_{n=1}^\infty \int_0^{S-s} \left[1-G_1(S-u-y)\right] g_n(y)dy +$$

$$+ \sum_{n=1}^\infty e^{-\lambda n}\left[1-G_1(S-u)\right] = \tag{5.4}$$

$$= \sum_{n=0}^\infty \int_0^{S-s} \left[1-G_1(S-u-y)\right] dG_n(y) .$$

In order to be able to apply the method discussed in section 3 we have to determine the functions $k(\psi;S,s)$ and $t(\psi;S,s)$.

If $\psi \in \bar{A}_{(S,s)}$ we have by definition

$$\begin{aligned} k(\psi;S,s) &= 0 \\ t(\psi;S,s) &= 0 \end{aligned} \tag{5.5}$$

Now let us suppose that at θ_1 the stock $\underline{u}$ is smaller than s. Thus $\psi \in A_{(S,s)}$.

As we know the function $k(1,u;S,s)$ is defined with the aid of two random walks.

In the initial state of the first walk $\underline{w}^d$ the stock is supplied till S (decision d) and after that no purchases (non-degenerate decisions) will be made. The stock is controlled until at some point

θ_{r_1} the stock is run out.

In the second walk $\underline{w}_o$ no purchases will be made and the stock will be controlled until at some point θ_{r_2} the stock is run out.

The expected costs to be incurred in the first walk are:

<u>purchase price</u>: a) $Q(S)$ if $u < 0$ (5.6)
b) $Q(S-u)$ if $u \geqq 0$

<u>inventory costs</u>: [1] $C_1 \int_0^S (\mathcal{E}\underline{t}_x)dx = C_1 \int_0^S (\sum_{j=0}^{\infty} F_j(x))dx$. (5.7)

For determining the <u>emergency purchase price</u> we need to know the probability distribution of the quantitity $\underline{v}$, ordered in emergency purchases during the random walk $\underline{w}^d$. It is easily verified that the distribution function of v is given by:

$$H_{S;0}(v) = (1-G_1(S-v)) + \sum_{n=1}^{\infty} \int_{+0}^{S} \left[1-G_1(S-v-y)\right] g_n(y)dy +$$

$$+ \sum_{n=1}^{\infty} e^{-\lambda n} \left[1-G_1(S-v)\right] \tag{5.8}$$

<u>emergency purchase price</u>:

$$-C_2 \int_{-\infty}^{0} u \, dH_{S;0}(u) = C_2 \int_{-\infty}^{0} H_{S;0}(u)du \tag{5.9}$$

The expected costs to be incurred in the second walk are zero for $u \leqq 0$ and for $u > 0$ given by:

<u>inventory costs</u>: $C_1 \int_0^u (\mathcal{E}\underline{t}_x)dx = C_1 \int_0^u (\sum_{j=0}^{\infty} F_j(x))dx$. (5.10)

For determining the <u>emergency purchase price</u> we need to know the probability distribution of the quantity $\underline{v}$, ordered in emergency purchases during the random walk $\underline{w}_o$. It is easily verified that the distribution function of $\underline{v}$ is given by:

---

1) The expected inventory costs of a quantity dx to be sold between $\underline{t}_x$ and $\underline{t}_{x+dx}$ are $C_1.\mathcal{E}\underline{t}_x.dx$.

$$H_{u;0}(v) = (1-G_1(u-v)) + \sum_{n=1}^{\infty} \int_{+0}^{u} \left[1-G_1(u-v-y)\right] g_n(y)\,dy + \sum_{n=1}^{\infty} e^{-\lambda n} \left[1-G_1(u-v)\right] \tag{5.11}$$

<u>emergency purchase price</u>:

$$-C_2 \int_{-\infty}^{0} v\,dH_{u;0}(v) = C_2 \int_{-\infty}^{0} H_{u,0}(v)\,dv \tag{5.12}$$

With the aid of (5.6) up to and including (5.12) we find for $u \geqq 0$:

$$k(1,u;S,s) = Q(S-u) + C_1 \int_{u}^{S} \left(\sum_{j=0}^{\infty} F_j(x)\right) dx + C_2 \int_{-\infty}^{0} (H_{S;0}(v) - H_{u;0}(v))\,dv \tag{5.13}$$

and for $u < 0$:

$$k(1;u;S,s) = Q(S) + C_1 \int_{0}^{S} \left(\sum_{j=0}^{\infty} F_j(x)\right) dx + C_2 \int_{-\infty}^{0} H_{0;s}(v)\,dv\ . \tag{5.14}$$

The function $t(\psi;S,s)$ is also defined with the aid of the two random walks $\underline{w}^d$ and $\underline{w}_0$.

The probability that a random walk $\underline{w}^d$ takes n elementary time intervals is given by:

$$p(n) = e^{-\lambda(n-1)} \left[1-G_1(S)\right] + \int_{+0}^{S} \left[1-G_1(S-y)\right] g_{n-1}(y)\,dy =$$

$$= \int_{0}^{S} \left[1-G_1(S-y)\right] dG_{n-1}(y)\ . \tag{5.15}$$

So we find for the expected duration of the first random walk:

$$\sum_{n=1}^{\infty} n\,p_n = \sum_{n=1}^{\infty} n \int_{0}^{S} \left[1-G_1(S-y)\right] dG_{n-1}(y) =$$

$$\sum_{n=1}^{\infty} n \left[G_{n-1}(S) - G_n(S)\right] = \sum_{n=0}^{\infty} G_n(y) \tag{5.16}$$

and in the same way for the expected duration of the second random walk:

$$\text{a)} \quad \sum_{n=0}^{\infty} G_n(u) \qquad \text{if} \quad u > 0. \tag{5.17}$$

$$\text{b)} \quad 0 \qquad \text{if} \quad u \leqq 0.$$

Consequently $t(1,u;S,s)$ is given by:

$$\text{a)} \quad \sum_{n=0}^{\infty} G_n(S) - \sum_{n=0}^{\infty} G_n(u) \qquad \text{if} \quad u > 0. \tag{5.18}$$

$$\text{b)} \quad \sum_{n=0}^{\infty} G_n(S) \qquad \text{if} \quad u = 0.$$

Now it is easily verified that the Markov process in A_z with discrete time parameter has no cyclically moving sets and only one ergodic set.

Because of the fact that the functions $k(1,u;S,s)$ and $t(1,u;S,s)$ do not depend on 1 we need only to know the probability distribution of $\underline{u}$ in the steady state. This probability distribution, however, is given by (5.4) and so we find (cf.(3.22)):

$$r(S,s;\psi) =$$

$$= \frac{H_{S',s}(0)\left[\Phi(S) + C_1 \int_0^{S'} \left(\sum_{j=0}^{\infty} F_j(x)\right) dx + C_2 \int_{-\infty}^{0} H_{S',0}(v)\, dv\right] + \int_0^{s} \left[\Phi(s-u) + C_1 \int_u^{S'} \left(\sum_{j=0}^{\infty} F_j(x)\right) dx + C_2 \int_{-\infty}^{0} \{H_{S',0}(v) - H_{u,0}(v)\}\, dv\right] d\, H_{S',s}(u)}{H_{S',s}(0) \sum_{n=0}^{\infty} G_n(S) - \int_0^{s} \left[\sum_{n=1}^{\infty} \{G_n(S') - G_n(u)\}\right] d\, H_{S',s}(u)} \tag{5.19}$$

The right hand side of (5.19) can be rewritten in the form:

$$\frac{C_1 \int_0^{S'} \left\{\sum_{j=0}^{\infty} F_j(x)\right\} dx + C_2 \int_{-\infty}^{0} H_{S',0}(v)\, dv + H_{S',s}(0) . \Phi(S) + \int_0^{s} \left[\Phi(s-u) - C_1 \int_0^{u} \left\{\sum_{j=0}^{\infty} F_j(x)\right\} dx - C_2 \int_{-\infty}^{0} H_{u,0}(v)\, dv\right] d\, H_{S',s}(u)}{\sum_{n=0}^{\infty} G_n(S' - s)} \tag{5.20}$$

The optimal choice of S and s can now be obtained directly from the minimalization of (5.11) with respect to S and s.

II) The motorist's problem

A motorist has effected an accident insurance. In the insurance-policy, belonging to it, among others the following conditions are stated:

1) The insurance will run for one year. At the end of each year it can be continued. The premium has to be paid at the beginning of each period of one year.

2) The premium amounts to E_1, unless
   a) in the preceding period of one year no damages have been claimed. In that case the premium will be E_2, unless
   b) in the preceding period of two years no damages have been claimed. In that case the premium will be E_3, unless
   c) in the preceding period of three or more years no damages have been claimed. In that case the premium will be E_4.

3) Damages have to be claimed immediately. Only the difference between the damage and a fixed amount a_o, the socalled own risk, is covered by insurance.

It will be obvious that our motorist will claim no damages smaller than a_o. But in view of point 2 of the insurance-policy we may expect that it will also be unprofitable to claim damages, which are not much larger than a_o, unless a loss is already claimed that year. In addition the optimal lower bound of the amount to claim depends on the time of the accident and on the last premium paid.

Let us suppose that the following data are available. The motorist association has established that for motorists like our driver the number of accidents in a period of length T is distributed by a Poisson distribution with mean λT. Next it is observed that the losses $\underline{s}$ are independently distributed according to the distribution function F(s).

The motorist's problem is how to establish an optimal strategy for claiming losses.

Let us try to solve this problem with the aid of the method described in section 3.

First we have to create a state space, next to define a natural process and finally to state what decisions imply in the model to be constructed.

Let us approach the problem from the point of view of a motorist, who is a fanatical opponent of insurances. He will certainly advise not to pay premiums. If a premium is already paid then he will suggest to claim all losses to be incurred and not to continue the insurance at the end of the year.

But, one year after the last premium has been paid, claims will be rejected by the insurance company.

In accordance with the opinion of that motorist in the natural process no premiums will be paid and all losses will be claimed.

If we accept this picture of the natural process, then the payment of a premium and the suppression of a claim are effected by decisions.

Analysing the motorist's problem in this way we observe that at each point of time the following facts are of interest:

a) the fact, whether an accident happens at the moment considered or not.
b) the extent of the damage.
c) the fact, whether the damage is covered by insurance or not.
d) the fact, whether a damage is claimed since the last payment of a premium or not.
e) the amount of the last premium paid.
f) the point of time considered.

This information has to be worked into the state of the system.

In addition we have to see that the random walk effected by the natural process is continuous from the right in the time parameter (property 5). In fig.2 a 4-dimensional state space is suggested that comes up to these requirements.

The s-coordinate of a state point denotes the extent of the last damage claimed after the last payment of a premium. The u-coordinate denotes the length of the period elapsed since the first claim after the last payment of a premium. Both coordinates are obvious and can be determined at each point of time.

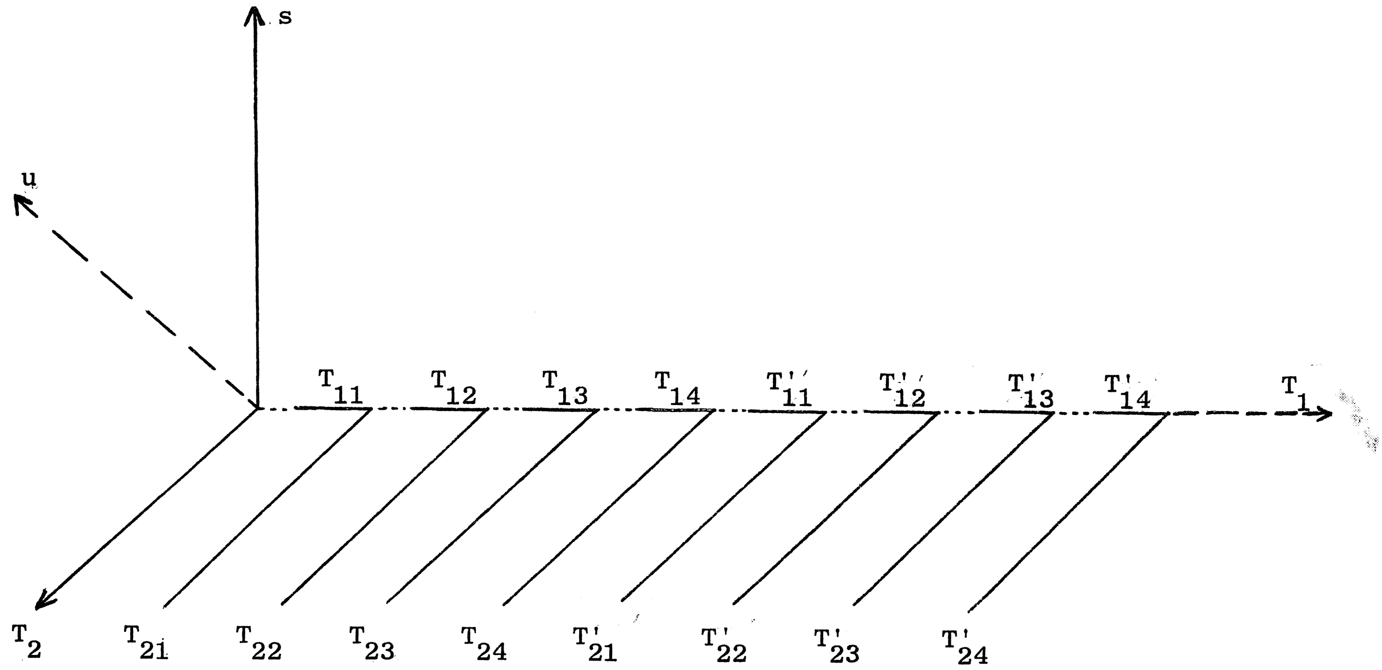

fig.2 State space

The T_1-axis is a time axis and as long as damages are still covered by insurance the points of time are fixed on this axis.

The T_2-axis is also a time axis, whose points fix the points of time from the moment whereupon damages are no longer covered by insurance.

In the (T_1,T_2)-plane one can find eight broken time intervals.

The intervals T_{11},T'_{11},T_{12}, etc. are time intervals of one year. The half lines T_2,T_{21},T'_{21} etc. are also time intervals with a same time scale.

In determining the (T_1,T_2)-component of a state point only points of the lines $T_{11},T_2,T'_{11},T_{21}$ etc. will be used.

If the orthogonal projection of a state point on the (T_1,T_2)-plane is:

1) a point of T_{1i} (i=1,2,3,4) then:

   a) the last premium paid is E_i.

   b) the damages are still covered.

   c) no damage has been claimed since the last premium is paid.

d) date and point of time considered are fixed by the location of the projection on T_{1i}.

2) a point of T_{2i} (i=1,2,3,4) then:

a) the last premium paid is E_i.

b) the damages are not longer covered.

c) during the period they were covered no damage was claimed

d) date and time considered are fixed by the location of the projection on T_{2i}.

3) a point of T'_{1i} (i=1,2,3,4) then:

a) the last premium paid is E_i.

b) the damages are still covered.

c) one or more damages have been claimed since the last premium has been paid.

d) date and time considered are fixed by the location of the projection on T'_{1i}.

4) a point of T'_{2i} (i=1,2,3,4) then:

a) the last premium paid is E_i.

b) the damages are not longer covered.

c) during the period they were covered one or more damages have been claimed.

d) date and time considered are fixed by the location of the projection on T'_{2i}.

5) a point of T_2 then:

a) never a premium has been paid.

b) no damages are or have been covered by insurance.

c) date and time considered are fixed by the location of the projection on T_2.

From this construction of the state space we can deduce for the random walk of the system in the natural process the following statements:

(1) If never a premium has been paid the projection of the state point on the (T_1,T_2)-plane will run along the T_2-axis.

(2) If the last premium paid is E_i, then, as long as no damage is claimed and damages are still covered, the projection of the state point will run along the line T_{1i} (i=1,2,3,4). On the moment upon which damages are no longer covered by insurance it turns round an angle of 90 degrees and after that the projection will follow the half line T_{2i} (i=1,2,3,4).

(3) If the last premium paid is E_i and if at the time, whereupon the first damage since the last payment of a premium has been claimed, the damage was still covered, then from that time onwards the projection of the state point runs along the lines T'_{1i} and T'_{2i}.

(4) The u-and s-coordinate of the moving state point are unambiguously determined by their definitions.

So the random walk of the system, effected by the natural process can now be defined for different initial states. The natural process itself is determined by the Poisson distribution of the number of accidents and by the probability distribution of the extent of the damage.

It can now easily be verified that these random walks are continuous from the right in the time parameter. The proof that the natural process is strong Markovian will not be given here.

As we have stated before payments of premiums and suppressions of claims are effected by decisions.

Let us first discuss the payments of premiums and their effects on the natural process. Premiums can be paid at the ends of the periods in which damages are covered by insurance and at the moments, whereupon no damages are covered.

In other words non-degenerate decisions of this type can be made at the moments that the projection of the state point on the (T_1,T_2) plane takes on either one of the end points of T_{1i},T'_{1i} or one of the points of T_2,T_{2i},T'_{2i}.

If at the end of T_{1i} (i=1,2,3) a premium E_{i+1} is paid then the state point will be transferred into the point of which the projection on the (T_1, T_2)-plane is given by the initial state of T_{1i+1} and for which holds: u=0 and s=0.

If at the end of T_{14} a premium E_4 is paid then the state point will be transferred into the point of which the projection on the (T_1, T_2)-plane is given by the initial state of T_{14} and for which holds: u=0 and s=0.

If at the end of T'_{1i} (i=1,2,3,4) or in a point of T_2, T_{2i} and T'_{1i} (i=1,2,3,4) a premium E_1 is paid then the state point will be transferred into the point of which the projection on the (T_1, T_2)-plane is the initial state of T_{11} and for which holds: u=0 and s=0.

Now the effects and the costs of this type of decisions are stipulated.

In the sequel our considerations are based on the assumption that at the end of each period premiums will be paid in which the damages are covered by insurance.

So if a strategy is applied then the points of T_2, T'_{21} etc. will not be taken on by the projection of the state point on the (T_1, T_2)-plane. Consequently we can restrict ourselves to the state space given in fig.3.

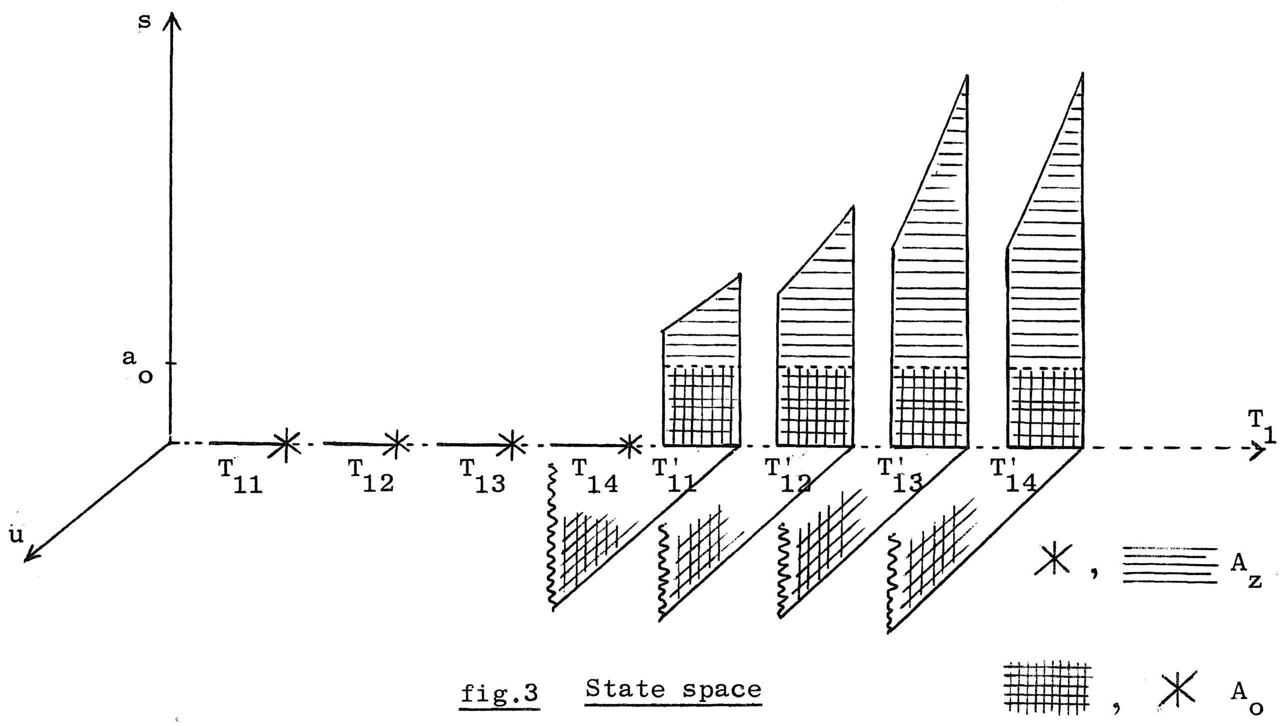

fig.3 State space

From the assumption about the payment of the premium made above it follows that for each strategy to be considered at least the intervention set consists of

a) twodimensional planes through the endpoints of T'_{1i} (i=1,2,3,4) and perpendicular to T_1.

b) the endpoints of the intervals T_{1i} (i=1,2,3,4).

Now we will discuss decisions, which effect suppressions of claims.

Let us assume that just before an accident the projection of the state point on T_1 is a point of T_{1i}.

If s' is the extent of the loss incurred, then the system is at the moment of the accident in a state of which the projection on T_1 is a point of T'_{1i}, that corresponds to the time considered and for which holds $s=s'$ and $u=0$.

If the claim is suppressed then the system will be transferred into the state of which the projection on T_1 is a point of T_{1i}, that corresponds to the time considered and for which holds $s=0$ and $u=0$.

If just before an accident the projection of the state point on T_1 is a point of T'_{1i} then a loss has been claimed already and it has no sense to suppress the claim.

Consequently suppressions of claims take only place if $u=0$ and if the projections of the state points on T_1 are points of T'_{1i} (i=1,2,3,4). Now it follows from point 3 of the policy of insurance that if $u=0$ and if the projection of the state point is a point of T'_{1i} a claim will certainly be suppressed if $s \leqq a_o$. It is obvious that if it is profitable to claim a loss of the extent s', a loss $s > s'$ has to be claimed too.

For this and other heuristic reasons that part of the intervention set A_z which corresponds to the suppressions of claims, is in the (s,T_1)-plane and has a form as drawn in fig.3.

It is easily verified that the stopping set A_o, according to its definition must have the form as drawn in fig.2.

It can now easily be proved that the Markov-process in the intervention set A_z with discrete time parameter has only one ergodic set and no cyclically moving sets, while in addition it satisfies the Doeblin condition.

Now the strategies and their implications have been discussed in full extent.

Our next task is to determine the functions $k(\psi;d)$ and $t(\psi;d)$.[1)]

As we know the function $k(\psi;d)$ is determined by the difference in expected values of the loss to be incurred in the random walk $\underline{w}^d$ and in the random walk $\underline{w}_o$, both from the state ψ.

The function $t(\psi;d)$ is defined in a similar way.

Let us consider now the random walk $\underline{w}_o$. If the projection of the initial state on the T_1-axis is a point t of T'_{1i} then it follows from the construction of the state space that the random walk $\underline{w}_o$ will take (1-t) units of time, unless u=0 and $s \leqq a_o$. In that case $\underline{w}_o$ consists of the initial state only and by definition no costs are involved in that walk. If the random walk takes (1-t) units of time then the expected number of accidents in that period is $\lambda(1-t)$. Because of the fact that in the natural process all damages will be claimed the expected loss per accident is given by:

$$K(a_o) = \int_0^{a_o} s \, dF(s) + a_o \int_{a_o}^{\infty} dF(s) \qquad (5.21)$$

Thus we find for the expected loss to be incurred in $\underline{w}_o$:

$$\begin{array}{ll} 0 & \text{if } u=0 \text{ and } s \leqq a_o \\ \lambda(1-t)\ K(a_o) & \text{otherwise} \end{array} \qquad (5.22)$$

Now we discuss the case that the projection of the initial state on the T_1-axis is a point t of T_{1i}.

The probability of no accident during the period $[t,1]$ is $e^{-\lambda(1-t)}$. If no accident occurs then the walk will end at t=1, thus at the end of T_{1i}. In such a walk no losses are incurred.

The probability of one or more accidents during the period $[t,1]$ is $1-e^{-\lambda(1-t)}$. Suppose that the first accident happens at $\underline{t}_1$, then one of the following eventualities may arise:

----

1) The reader, who is only interested in the solution of this problem can omit the determination of $k(\psi;d)$ and $t(\psi;d)$.

a) the damage s_1 at $\underline{t}_1$ is smaller than a_o. In that case the walk is ended at $\underline{t}_1$ and the total loss to be incurred is s_1.

b) the damage s_1 at $\underline{t}_1$ is larger than a_o. In that case the walk will go on till t=1. The probability of a continuation after the first accident is $(1-F(a_o))$. The loss at t_1 is equal to a_o. Because of the fact that after the accident the projection of the state point on T_1 is a point of T'_{1i} we find for the expected value of the loss to be incurred in the remaining part of the walk:

$$\lambda (1-\underline{t}_1)K(a_o) \tag{5.23}$$

Combining these results we can state that, if the projection of the initial state on the T_1-axis is a point of T_{1i}, the expected loss to be incurred in $\underline{w}_o$ is given by:

$$\begin{aligned}(1-e^{-\lambda(1-t)})K(a_o) + \left[1-F(a_o)\right] K(a_o) \int_0^{1-t} \lambda^2(1-t-r)e^{-\lambda r}dr = \\ =F(a_o)K(a_o)(1-e^{-\lambda(1-t)}) + \left[1-F(a_o)\right] K(a_o) \lambda(1-t).\end{aligned} \tag{5.24}$$

In a similar way we will find for the expected duration of the random walk $\underline{w}_o$ in this case:

$$\begin{aligned}e^{-\lambda(1-t)}(1-t)+F(a_o) \int_0^{1-t} \lambda te^{-\lambda t}dt + \left[1-F(a_o)\right] (1-e^{-\lambda(1-t)})(1-t) = \\ (1-t)-F(a_o)\left[1-t- \frac{1}{\lambda} (1-e^{-\lambda(1-t)})\right]\end{aligned} \tag{5.25}$$

Now we have determined the expected loss to be incurred in- and the expected durations of - the random walks $\underline{w}_o$, starting from different points of the state space. With the aid of these results we are in the position to determine the functions $k(\psi;d)$ and $t(\psi;d)$ too.

It is convenient to introduce here a new notation for the state of the system.

With

$$\psi = \left[i,t;s;u\right] \tag{5.26}$$

we indicate that state point, which projection on T_1 is in t of T_{1i}.

With $$\psi = [i',t;s;u] \tag{5.27}$$

we indicate that state point, which projection on T_1 is in t of T'_{1i}.

As we know the endpoints of T_{1i}, given by $[i,1;0;0]$, belong to A_o.

So for determining $k(\psi;z(\psi))$ we have only to calculate the expected value of the loss to be incurred during the random walk $\underline{w}^{z(\psi)}$. After the premium E_{i+1} is paid the random walk $\underline{w}^{z(\psi)}$ is identical with a random walk $\underline{w}_o$ from the initial point of $T_{1,i+1}$ $(i+1,0;0;0)$. So if ψ is the endpoint of T_{1i} then we find with the aid of (5.24) for $k(\psi;z(\psi))$:

$$k(\psi;z(\psi)) = \begin{cases} E_{i+1} + F(a_o)K(a_o)(1-e^{-\lambda}) + \lambda K(a_o)(1-F(a_o)) \\ \qquad\qquad \text{if } \psi = [i,1;0;0] \text{ with } i=1,2,3 \\ E_4 + F(a_o)K(a_o)(1-e^{-\lambda}) + \lambda K(a_o)(1-F(a_o)) \\ \qquad\qquad \text{if } \psi = [4,1;0;0] \end{cases} \tag{5.28}$$

In a similar way we find with the aid of (5.25):

$$t(\psi;z(\psi)) = 1-F(a_o)\left[1-\frac{1}{\lambda}(1-e^{-\lambda})\right] \qquad \text{if } \psi = [i,1;0;0] \tag{5.29}$$

Now it can easily be verified that for $\psi = [i',1;0;0]$ we have:

$$k(\psi;z(\psi)) = E_1+F(a_o)K(a_o)(1-e^{-\lambda})+\lambda(1-F(a_o))K(a_o) \tag{5.30}$$

and

$$t(\psi;z(\psi)) = 1-F(a_o)\left[1-\frac{1}{\lambda}(1-e^{-\lambda})\right] \tag{5.31}$$

If a strategy z is applied let the boundary of the intervention set A_z in the $(T_1;s)$-plane be given by the curve:

$$s = S(i',t;z) \qquad \begin{matrix} i'=1,2,3,4 \\ t \in [0,1] \end{matrix} \tag{5.32}$$

If according to a strategy z we have to suppress a claim in $\psi = [i',t;s;0]$ then one of the following eventualities may arise:

a) The damage s at t is smaller than a_o. In that case a random walk $\underline{w}_o$ starting in ψ consists of the initial state ψ only. Because of the suppression of the claim the system is transferred into $[i,t;0;0]$. The random walk $\underline{w}^{z(\psi)}$ from ψ is after the suppression identical to a $\underline{w}_o$-walk from $[i,t;0;0]$. Because of the fact that $s \leqq a_o$ the losses involved in the suppression of the claim (non-degenerate decision) are zero. So we find with the aid of (5.24) and (5.25) for the functions $k(\psi;z(\psi))$ and $t(\psi;z(\psi))$,

$$k(\psi;z(\psi)) = F(a_o)K(a_o)(1-e^{-\lambda(1-t)})+ \left[1-F(a_o)\right] K(a_o)\lambda(1-t) \tag{5.33}$$

$$t(\psi;z(\psi)) = (1-t)-F(a_o)\left[1-t-\frac{1}{\lambda}(1-e^{-\lambda(1-t)})\right] \tag{5.34}$$

respectively.

b) The damage s at t is larger than a_o, but smaller than or equal to $S(i',t;z)$. In this case a random walk $\underline{w}_o$ starting in ψ takes (1-t) time units. The expected loss to be incurred in this walk is given by (cf. (5.23)):

$$\lambda(1-t)K(a_o) \tag{5.35}$$

Because of the suppression of the claim the system is transferred into $[i,t;0;0]$. The random walk $\underline{w}^{z(\psi)}$ from ψ is after the suppression identical to a $\underline{w}_o$-walk from $[i,t;0;0]$. Because of the fact that $s > a_o$ the losses involved in the suppression of the claim (non-degenerate decision) are $s-a_o$. So we find with the aid of (3.3), (3.4), (5.24), (5.25) and (5.35) for the functions $k(\psi;z(\psi))$ and $t(\psi;z(\psi))$:

$$\begin{aligned} k(\psi;z(\psi)) &= k_1(\psi;z) - k_o(\psi;z) = \\ &= \left\{ s-a_o+F(a_o)K(a_o)(1-e^{-\lambda(1-t)}) + \left[1-F(a_o)\right] K(a_o)\lambda(1-t)\right\} \\ &\quad - \lambda(1-t)K(a_o) = \\ &= s-a_o + K(a_o)F(a_o)\left\{1-e^{-\lambda(1-t)}-\lambda(1-t)\right\} \end{aligned} \tag{5.36}$$

$$t(\psi;z(\psi)) = t_1(\psi;z) - t_o(\psi;z) =$$
$$= \left[1-t-F(a_o)\left\{1-t-\frac{1}{\lambda}(1-e^{-\lambda(1-t)})\right\}\right] - (1-t) =$$
$$= F(a_o)\left[\frac{1}{\lambda}(1-e^{-\lambda(1-t)})-(1-t)\right] \quad . \qquad (5.37)$$

Recapitulating:

$$k(\psi;z(\psi)) = \begin{cases} E_{i+1}+F(a_o)K(a_o)(1-e^{-\lambda})+\lambda K(a_o)\left[1-F(a_o)\right] & ;\text{if } \psi=[i,1;0;0] \\ & \quad i\neq 4 \\ E_4+F(a_o)K(a_o)(1-e^{-\lambda})+\lambda K(a_o)\left[1-F(a_o)\right] & ;\text{if } \psi=[4,1;0;0] \\ E_1+F(a_o)K(a_o)(1-e^{-\lambda})+\lambda K(a_o)\left[1-F(a_o)\right] & ;\text{if } \psi=[i',1,s;u] \\ F(a_o)K(a_o)(1-e^{-\lambda(1-t)})+\left[1-F(a_o)\right]K(a_o)\lambda(1-t); & \text{if } \psi=[i',t;s;0] \\ & \quad s\leqq a_o \\ S-a_o+K(a_o)F(a_o)\left[1-e^{-\lambda(1-t)}-\lambda(1-t)\right] & \text{if } \psi=[i',t;s;0 \\ & \quad a_o\leqq s\leqq S(i',t;z) \end{cases}$$

(5.38)

$$t(\psi;z(\psi)) = \begin{cases} 1-F(a_o)\left[1-\frac{1}{\lambda}(1-e^{-\lambda})\right] & ;\text{ if } \psi=[i,1;0;0] \text{ or } [i',1;s;u] \\ (1-t)-F(a_o)\left[1-t-\frac{1}{\lambda}(1-e^{-\lambda(1-t)})\right] & \text{if } \psi=[i',t;s;0] \quad s\leqq a_o \\ F(a_o)\left[\frac{1}{\lambda}(1-e^{-\lambda(1-t)})-(1-t)\right] & ;\text{ if } \psi=[i',t;s;0] \\ & \quad a_o\leqq s\leqq S(i',t;z) \end{cases}$$

(5.39)

In section 3 we have stated that the boundary of the intervention set of the optimal strategy can be determined with the aid of the relations (3.34) and (3.35).

Because of the fact that for each strategy z to be considered the z-process has only one ergodic set the relation (3.34) is satisfied for each $\psi \in \Psi$.

Let us consider the relation (3.35). According to (3.35) each point of the boundary of the intervention set A_{z_o} has to satisfy:

$$\left[\frac{d}{dt}\,\mathcal{E}\left\{C(z_o;\underline{\xi})\,\middle|\,\psi;T\right\}\right]_{T=0} = 0 \quad . \qquad (3.35)$$

In this section we will use (3.35) in a somewhat different form.

In section 3 we have introduced the function $g(z;\psi)$. This function expresses the expected value of the amount that the decision maker has to pay out of his own pocket, if he applies strategy z from the initial state ψ onwards.

Let us consider the following states:

$$\begin{aligned}
\psi_1 &= \left[i', t; S(i,t;z); 0\right] \\
\psi_2 &= \left[i', t+\Delta t; S(i,t+\Delta t;z); 0\right] \\
\psi_3 &= \left[i, t; 0; 0\right] \\
\psi_4 &= \left[i, t+\Delta t; 0; 0\right] \\
\psi_5 &= \left[i', t+\Delta t; s'; u'\right]
\end{aligned} \tag{5.40}$$

Now the following identities are easily verified:

$$g(z;\psi_1) = S(i,t;z) - a_o + g(z;\psi_3) \tag{5.41}$$

$$g(z;\psi_2) = S(i,t+\Delta t;z) - a_o + g(z;\psi_4) \tag{5.42}$$

and for each s' and u'

$$g(z;\psi_2) = g(z;\psi_5) \tag{5.43}$$

As we know the probability of one accident in a period of length Δt is equal to $\lambda.\Delta t$, while the probability of two or more accidents is of magnitude $(\Delta t)^2$.

Following the definition of $g(z;\psi)$ we find:

$$g(z;\psi_1) = \lambda.\Delta t\, K(a_o) - \Delta t . r(z;\psi_1) + g(z;\psi_2) + O(\Delta t^2) \tag{5.44}$$

and, if $K\left[S(i,t;z)\right]$ is given by:

$$K\left[S(i,t;z)\right] = \int_0^{S(i,t;z)} x \, dF(x) + a_o \int_{S(i,t;z)}^{\infty} dF(x), \tag{5.45}$$

$$g(z;\psi_3) = \lambda.\Delta t\, K[S(i,t;z)] - \Delta t\, r(z;\psi_1) +$$
$$+ \left\{ e^{-\lambda.\Delta t} + \lambda.\Delta t\, F[S(i,t;z)] \right\} g(z;\psi_4) +$$
$$+ \lambda.\Delta t [1-F[S(i,t;z]] \; g(z;\psi_2) + O(\Delta t^2) =$$
$$= g(z;\psi_4) + \lambda\, \Delta t\, K[S(i,t;z)] - \Delta t\, r(z;\psi_1) +$$
$$+ \lambda.\Delta t [1-F[S(i,t;z)]] \left\{ g(z;\psi_2) - g(z;\psi_4) \right\} + O(\Delta t^2). \qquad (5.46)$$

From (5.41), (5.42) and (5.44) we can deduce:

$$\lambda.\Delta t\, K(a_o) - \Delta t\, r(z;\psi_1) + O(\Delta t^2) = S(i,t;z) - S(i,t+\Delta t;z) +$$
$$+ \lambda\, \Delta t\, K[S(i,t;z)] - \Delta t\, r(z;\psi_1) + \lambda\, \Delta t [1-F[S(i,t;z)]] \cdot$$
$$\cdot [S(i,t+\Delta t;z) - a_o] \qquad (5.47)$$

or

$$S(i,t+\Delta t;z) - S(i,t;z) =$$
$$= \lambda\, \Delta t \left\{ K[S(i,t;z)] - K(a_o) + [1-F[S(i,t;z)]] \, [S(i,t+\Delta t;z) - a_o] \right\}$$
$$(5.48)$$

Consequently:

$$\frac{d\, S(i,t;z)}{dt} = \lim_{\Delta t \downarrow 0} \frac{S(i,t+\ \ t;z) - S(i,t;z)}{\Delta t} =$$

$$= \lambda \left\{ K[S(i,t;z)] - K(a_o) + [1-F[S(i,t;z)]] \, [S(i,t;z) - a_o] \right\} =$$

$$= \lambda \left\{ \int_{a_o}^{\infty} (x-a_o)\, dF(x) + \int_{S(i,t;z)}^{\infty} [x-S(i,t;z)]\, dF(x) \right\} . \qquad (5.49)$$

From (5.49) it follows that the boundary of the intervention set consists of parts of that curve S(t) which satisfies the equation:

$$\frac{d\, S(t)}{dt} = \lambda \int_{a_o}^{\infty} (x-a_o)\, dF(x) - \lambda \int_{S(t)}^{\infty} [x-S(t)]\, dF(x) . \qquad (5.50)$$

The curve S(t) is determined by (5.50) except for a translation in t.

It can be proved that for this example the relations (5.49) and (3.35) are equivalent.

Next we will consider the relation (3.30). If the optimal strategy z_o is applied, then according to (3.30) each point $\psi \in \Psi$ has to satisfy:

$$C(z_o;\psi) = \min_{d \in D^*(\psi)} \left[k(\psi;d) - k^*(\psi;d) + \mathcal{E}\left\{ C(z_o;\underline{\eta}) \mid d \right\} \right] \tag{3.30}$$

Because of the fact that for each strategy z to be considered the z-process has only one ergodic set the relation (3.30) is equivalent with:

$$C(z_o;\psi) = \min_{d \in D(\psi)} \left[k(\psi;d) - r(z_o;\psi)t(\psi;d) + \mathcal{E}\left\{ C(z_o;\underline{\eta}) \mid d \right\} \right] \tag{5.51}$$

In this section the relation (5.51) will be applied in a somewhat different form.

Let us suppose that just after the payment of the premium E_1 an accident happens. If the extent of the damage is s then at the moment of the accident the system will be in the state $\psi = [1',0;s;0]$.

If the decision maker claims this damage and if he will apply the optimal strategy z_o afterwards, then the expected value of the amount he has to pay out of his own pocket is given by:

$$\lim_{\delta \to +0} g(z_o;[1',\delta;s;\delta]) \quad . \tag{5.52}$$

If he suppresses the claim and if he will apply the optimal strategy z_o afterwards, then the expected value of the amount he has to pay out of his own pocket is given by:

$$s - a_o + g(z_o;[1,0;0;0]) \quad . \tag{5.53}$$

Now it is easily verified that:

$$\lim_{\delta \to +0} g(z_o,[1',\delta;s;\delta]) = \lambda K(a_o) - r(z_o;\psi) + E_1 + g(z_o,[1,0;0;0]) \quad . \tag{5.54}$$

Now it follows from (5.53) and (5.54) that the suppression of the claim is only profitable, if

$$s-a_o+g(z_o;[1,0;0;0]) \leqq \lambda K(a_o)-r(z_o;\psi) + E_1 + g(z_o;[1,0;0;0]) \tag{5.55}$$

or:

$$s \leqq a_o + \lambda K(a_o) - r(z_o;\psi) + E_1 \tag{5.56}$$

Thus:

$$S(1,0;z_o) = a_o + \lambda K(a_o) - r(z_o;\psi) + E_1 \tag{5.57}$$

Let us suppose that just after the payment of the premium E_1 an accident happens. If the extent of the damage is s then at the moment of the accident the system will be in the state $\psi = [i',0;s;0]$.

If the decision maker claims this damage and if he will apply the optimal strategy z_o afterwards, then the expected value of the amount he has to pay out of his own pocket is given by:

$$\lim_{\delta \to +0} g(z_o,[i',\delta;s;\delta]) \tag{5.58}$$

If he suppresses the claim and if he will apply the optimal strategy z_o afterwards, then the expected value of the amount he has to pay out of his own pocket is

$$s-a_o+g(z_o,[i,0;0;0]) \tag{5.59}$$

Now it follows from the definition of $g(z;\psi)$:

a) $$\lim_{\delta \to +0} g(z_o;[i'-1,1-\delta;S(i-1,1-\delta;z_o);0]) =$$

$$= \lim_{\delta \to +0} \left\{ S(i-1,1-\delta;z_o)-a_o \right\} +g(z_o;[i,0;0;0]) + E_i \tag{5.60}$$

b) $$\lim_{\delta \to +0} g(z_o;[i'-1,1-\delta;S(i-1,1-\delta;z_o);0]) = E_1+g(z_o;[1,0;0;0]) \tag{5.61}$$

c) $$\lim_{\delta \to +0} g(z_o;[i',\delta;s;\delta]) = \lambda K(a_o)-r(z_o;\psi) + E_1 + g(z_o;[1,0;0;0]) \tag{5.62}$$

With the aid of (5.58) up to and including (5.62) we find that the suppression of the claim is only profitable if:

$$s-a_o+g(z_o;\ [i,0;0;0]\ ) \leqq \lim_{\delta\to +0}\ g(z_o;[i',\delta;s;\delta]\ ) \tag{5.63}$$

or

$$s-a_o \leqq \lambda\, K(a_o)+E_i-r(z_o;\psi)+ \lim_{\delta\to +0}\left\{ S(i-1,1-\delta;z_o)-a_o\right\} \tag{5.64}$$

Thus:

$$S(i,0;z_o) = \lambda\ K(a_o) + E_i-r(z_o;\psi)+ \lim_{\delta\to +0}\ S(i-1,1-\delta;z_o) \tag{5.65}$$

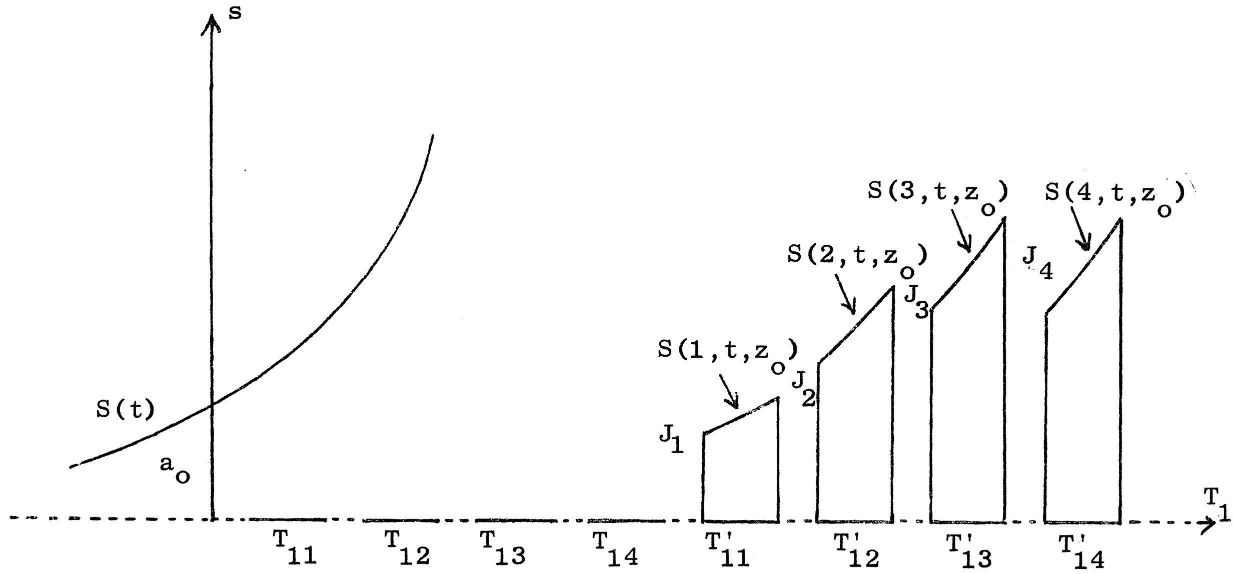

fig.4 A detail of the state space

Recapitulation: (cf.fig.4)

$$\begin{aligned} J_1 &= K(a_o) + E_1 - r(z_o;\psi) \\ J_2 &= K(a_o) + E_2 - r(z_o;\psi) \\ J_3 &= K(a_o) + E_3 - r(z_o;\psi) \\ J_4 &= K(a_o) + E_4 - r(z_o;\psi) \end{aligned} \tag{5.66}$$

We have already stated that the boundaries $S(i,t;z_o)$ are parts of a curve $S(t)$, which is also plotted in fig.4.

From the structure of the problem it can easily be deduced that for the optimal strategy z_o the boundaries $S(3,t;z_o)$ and $S(4,t;z_o)$ are identical.

In other words the following relation has to be true:

$$S(3,1;z_o) - S(3,0;z_o) = \lambda\, K(a_o) + E_4 - r(z_o;\psi) \qquad (5.67)$$

With the aid of (5.66), (5.67) and of one of the solutions $S(t)$ of (5.50) the motorist's problem can be solved. This will be done in the following numerical example.

Let the following data be given:

$$\begin{array}{ll} 1) & \lambda = 2 \\ 2) & F(s) = 1-e^{-s} \\ 3) & a_o = 0.4 \\ 4) & E_1 = 1.6 \quad E_3 = 1.2 \\ & E_2 = 1.4 \quad E_4 = 1.1 \end{array} \qquad (5.68)$$

Now by using (5.68) we find for (5.50):

$$\frac{d\,S(t)}{dt} = 2\int_{0.4}^{\infty} \left[s-0.4\right] e^{-s} ds - 2\int_{S(t)}^{\infty} \left[s-S(t)\right] e^{-s} ds \qquad (5.69)$$

and thus:

$$\frac{d\,S(t)}{dt} = 2e^{-0.4} - 2e^{-S(t)} \qquad (5.70)$$

Hence:

$$\tfrac{1}{2}\int \frac{d\,S(t)}{e^{-0.4} - e^{-S(t)}} = t+c \qquad (5.71)$$

and consequently:

$$S(t) = 0.4 + \ln\left\{1+\tfrac{1}{2}e^{2e^{-0.4}(t+c)}\right\} \qquad (5.73)$$

In the sequel we shall limit ourselves to the solution

$$S(t) = 0.4 + \ln(1+e^{bt}) \ , \tag{5.74}$$

where

$$b = 2e^{-0.4} \ .$$

From (5.68) we derive:

1) $K(0.4) = 0.3297$

2)
$$\begin{aligned} J_1 &= 0.6594 + 1{,}6 - r(z_o;\psi) = 2.2594 - r(z_o;\psi) \\ J_2 &= 0.6594 + 1.4 - r(z_o;\psi) = 2.0594 - r(z_o;\psi) \\ J_3 &= 0.6594 + 1.2 - r(z_o;\psi) = 1.8594 - r(z_o;\psi) \\ J_4 &= 0.6594 + 1.1 - r(z_o;\psi) = 1.7594 - r(z_o;\psi) \end{aligned} \tag{5.75}$$

Now let t_i be given by:

$$S(t_i) = S(i,0;z_o) \tag{5.76}$$

From the definitions of J_i (i=1,2,3,4) it follows that the following relations are true:

$$S(t_1) = 0.4 + J_1 \tag{5.77}$$

$$S(t_2) - S(t_1+1) = J_2 \tag{5.78}$$

$$S(t_3) - S(t_2+1) = J_3 \tag{5.79}$$

$$S(t_3+1) - S(t_3) = J_4 \tag{5.80}$$

Now t_1 can be determined from (5.77). It will be a function of the unknown $r(z_o;\psi)$. If the $r(z_o;\psi)$-function (t_1+1) is substituted in (5.78) then from this substitution the $r(z_o;\psi)$-functions t_2 and (t_2+1) can be obtained. In a similar way we can find the $r(z_o;\psi)$-functions t_3 and (t_3+1).

Finally the relation (5.80) will turn out to be equivalent to:

$$e^{-3r(z_o;\psi)} - 0{,}122134e^{-2r(z_o;\psi)} + 0.0009e^{-r(z_o;\psi)} + 0.000011 = 0. \tag{5.81}$$

By solving (5.81) we find:

$$r(z_o;\psi) = 2{,}177 \tag{5.82}$$

and consequently:

$$
\begin{aligned}
J_1 &= 0{,}0824 \\
J_2 &= -0{,}1176 \\
J_3 &= -0{,}3176 \\
J_4 &= -0{,}4176
\end{aligned}
\qquad (5.83)
$$

The solution of this numerical example is given in fig.5.

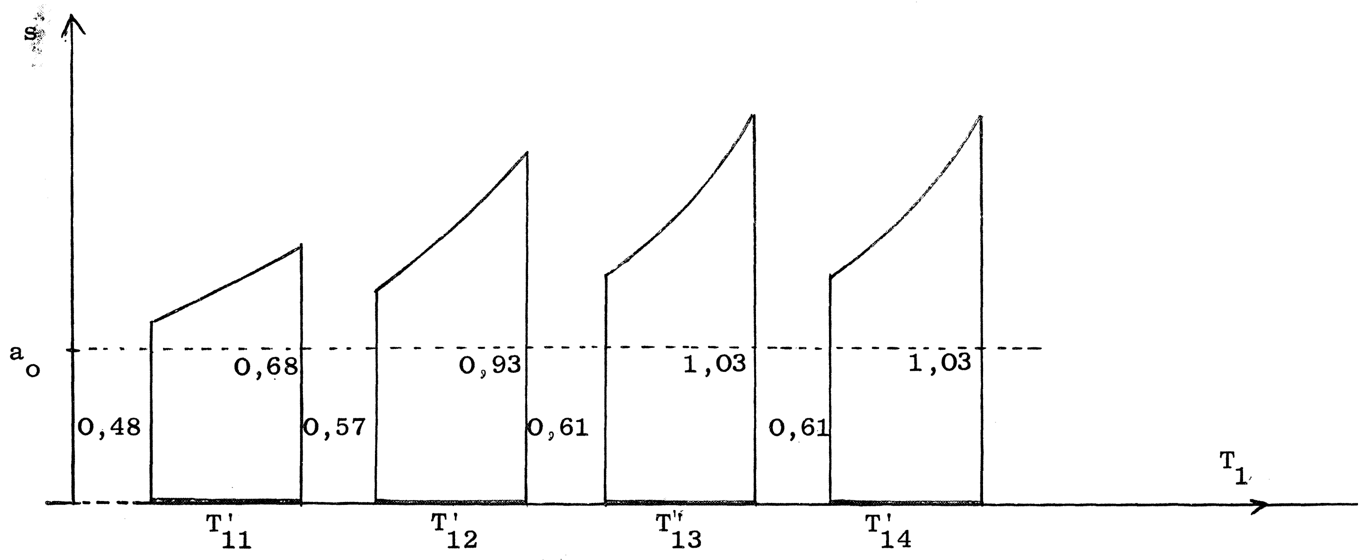

fig.5 Solution of the motorist's problem

References

1 G. de Leve: Generalized Markovian Decision Processes (in preparation).

2 R. Bellman: Dynamic Programming, Princeton U.P.'57.

3 R.A. Howard: Dynamic Programming and Markov processes, The Technology Press of the M.I T. and John Wiley '60.

4 K.J. Arrow, S.Karlin and H. Scarf: Studies in the theory of inventory and production: Stanford U.P.'58.