A. HORDIJK

DYNAMIC PROGRAMMING AND MARKOV POTENTIAL THEORY

DYNAMIC PROGRAMMING AND MARKOV POTENTIAL THEORY

DYNAMIC PROGRAMMING AND MARKOV POTENTIAL THEORY

ACADEMISCH PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE WISKUNDE EN NATUURWETENSCHAPPEN AAN DE UNIVERSITEIT VAN AMSTERDAM, OP GEZAG VAN DE RECTOR MAGNIFICUS, DR. A. DE FROE, HOOGLERAAR IN DE FACULTEIT DER GENEESKUNDE, IN HET OPENBAAR TE VERDEDIGEN IN DE AULA DER UNIVERSITEIT (TIJDELIJK IN DE LUTHERSE KERK, INGANG SINGEL 411, HOEK SPUI) OP WOENSDAG 16 JANUARI 1974 DES NAMIDDAGS TE 4.00 UUR

DOOR

ARIE HORDIJK

GEBOREN TE ROCKANJE

1974

MATHEMATISCH CENTRUM - AMSTERDAM

PROMOTOR: PROF.DR. J.Th. RUNNENBURG

CO-PROMOTOR: PROF.DR. J.A. BATHER

CO-REFERENT: PROF.DR. G. DE LEVE

aan Marina

CONTENTS

Acknowledgements iii

1. Introduction 1
2. Potentials and excessive functions 5
3. On the value function of an optimal control problem 20
4. Existence of optimal strategies 28
5. Semi-Markov decision processes with average return criterion 38
6. Discounted and non-discounted dynamic programming 55
7. On potentials, absorbing policies and charge structures 60
8. Recurrence for a decision process 64
9. Exponentially bounded stopping times 74
10. Sufficient conditions for the existence of an optimal policy with respect to the average return criterion 81
11. Simultaneous Doeblincondition 91
12. Connection with the work of Derman, Ross, Taylor and Veinott 101
13. Randomization and nearly optimal policies 110

Bibliography 127

List of notations 133

Summary 135

ACKNOWLEDGMENTS

The research leading to this thesis was carried out at the Mathematical Centre. The author is grateful for the splendid opportunity given to him there and wishes to thank in particular Prof.dr. J. Hemelrijk for encouraging him to study statistics and Prof.dr. G. de Leve for introducing him to the subject of Markov programming.

One of the pleasant things in life is doing mathematics together. All those with whom I had the privilege to work with I am grateful to, in particular to my promotor Prof.dr. J.Th. Runnenburg and to my co-promotor Prof.dr. J.A. Bather.

I also wish to thank my colleague Henk Tijms, who shares with me an interest in dynamic programming, for many discussions on this subject.

The stimulating interest of my co-promotor and of the members of the 1973-Colloquium on Probability Theory proved most helpful in doing the hard job of writing things down.

The author's sincere thanks go to Mr. J. Hillebrand and to Mrs. S.M.T. Hillebrand-Snijders for editing and typing the manuscript, to Mr. K.M. van Hee for proofreading, to Messrs. D. Zwarst, J. Suiker and J. Schipper for the reproduction.

1. INTRODUCTION

In this monograph we are mainly concerned with a dynamic system which at times $t = 0,1,\ldots$ is observed to be in one of a possible number of states. Let E denote the countable space of all possible states. If at time t the system is observed in state i then a decision must be chosen from a given set $\mathcal{P}(i)$. The probability that the system moves to a new state j (the so-called transition probability) is a function only of the last observed state i and the subsequently taken decision. In order to avoid an overburdened notation we shall identify the decision to be taken with the probability measure on E that is induced by it. Thus for each $i \in E$ the set $\mathcal{P}(i)$ consists of probability measures $p(i,.)$.*) Let $\mathcal{P}$ be the set of all stochastic matrices P with $p(i,.) \in \mathcal{P}(i)$ for each $i \in E$. Hence $\mathcal{P}$ has the *product property*: with P_1 and P_2 the set $\mathcal{P}$ also contains all those P with for every $i \in E$ in the i^{th} row of P either the i^{th} row of P_1, or the i^{th} row of P_2.

A policy R for controlling the system is a sequence of decision rules for the times $t = 0,1,\ldots$, where the decision rule for time t is the instruction at time t which prescribes the decision to be taken. This instruction may depend on the history i.e. the states and decisions at times $0,1,\ldots,t-1$ and the state at time t. When the decision rule is independent of the past history except for the present state then it can be identified with a $P \in \mathcal{P}$. A memoryless or Markov policy R is a sequence $P_0,P_1,\ldots \in \mathcal{P}$, where P_t denotes the decision rule at time t. P_t also gives the transition probabilities at time t.

In this monograph there are only a few places where non-memoryless policies are used. We need them to show that the value function is $c_{\mathcal{P}}$-superharmonic (see theorem 3.1). Theorem 13.2 says that when $\mathcal{P}$ contains all randomizations then the supremum over all memoryless policies equals the supremum over all policies. Hence in this case the value function may be defined as the supremum over the memoryless policies.

Since the law of motion of the dynamic system can be described by a non-stationary Markov chain when a memoryless policy is used, we prefer to

---

*) We allow that with positive probability the system "breaks down" or "disappears", so $p(i,j) \geq 0$, $i,j \in E$ and $p(i,E) := \sum_{j \in E} p(i,j) \leq 1$, $i \in E$.

introduce a decision process as a collection of non-stationary Markov chains (for a more general foundation of decision processes see [Hinderer]). A memoryless policy which takes at all times the same decision rule i.e. $P^\infty := (P,P,\ldots)$, $P \in \mathcal{P}$ is called a stationary policy and induces a stationary Markov chain.

One of the features of this monograph is the generalization of well-known results for one Markov chain to a collection of Markov chains. We give some examples. In theorem 8.6 it is proved that the maximal average expected reward does not depend on the initial state given that the system is recurrent. This is a direct generalization of the well-known theorem that each excessive function on a recurrent chain is constant.

The main assumption in theorem 5.1 (relation 5.1.1) is nothing else than a condition guaranteeing that all Markov chains are uniformly positive recurrent. This condition is a direct generalization to a collection of Markov chains of a so-called Foster criterion or a Liapunov function criterion as it is called elsewhere (see subsection 2.7).

Finally the simultaneous Doeblin condition (see section 11) is a straightforward extension to a collection of Markov chains of the well-known Doeblin condition.

Nowadays potential theory for Markov chains is well developed. A systematic treatment of potential theory for dynamic systems would in our opinion be desirable. Although the second part of the title of this monograph suggests more, our contribution to potential theory for dynamic systems consists only in the introduction of some useful terminology and the derivation of some interesting results (sections 2 and 7). The reason is that we were mainly interested in dynamic programming. It seems that many interesting questions were left untouched.

When in state i decision $p(i,.)$ is taken then an immediate cost depending on i and $p(i,.)$ is incurred *). Let $c_P(i)$ be the immediate cost when taking decision $p(i,.)$ (the i^{th} row of matrix P) in state i and write c_P for the vector with i^{th} component $c_P(i)$. Note that if $P,Q \in \mathcal{P}$ with $p(i,.) = q(i,.)$ then $c_P(i) = c_Q(i)$.

The expectation of the cost at time n when starting in state i at time

*) It is common to minimize when speaking of costs. We shall always maximize. The reason is that along with a cost structure also a reward function shall be used (see section 3).

zero and using policy $R = (P_0, P_1, \ldots)$ will be denoted by $\mathbb{E}_{i,R}\, c(\underline{x}_n)$, where $\underline{x}_n$ *) is the state at time n. $\mathbb{E}_R\, c(\underline{x}_n)$ denotes the vector with i^{th} component $\mathbb{E}_{i,R}\, c(\underline{x}_n)$ (for stationary policy P^∞ we write $\mathbb{E}_P[\ldots]$ instead of $\mathbb{E}_{P^\infty}[\ldots]$). It is easily seen that

$$\mathbb{E}_R\, c(\underline{x}_n) = P_0\, P_1\, \ldots\, P_{n-1}\, c_{P_n}\,.$$

In some of the following sections it is assumed that the cost function is a charge structure (see definition 2.12). In dynamic programming a weaker assumption like "all relevant expectations do not attain the value plus infinity" could be used. Our gain is a greater simplicity in the statements of the results. Also a nice implication is that the well-known theorem in optimal stopping remains valid: the value function is the minimum of the excessive functions that majorize the reward function.

The basic reason for taking the state space a countable set was that many of the problems which arise in general state spaces already appear in the countable state space. The countable state space does not have the "compactness" properties of the finite state space and with the countable state space one avoids the "measurability" questions of more general state spaces. As to the generalization of the results of this monograph, some can be generalized in a straightforward way, some results cannot be generalized and for the other results we do not know.

In an important part of the literature on Markovian decision processes it is assumed that for each state the set of available decisions in that state is a finite set. Usually randomized decisions i.e. convex combinations of the available decisions with a corresponding convex combination of the costs as the immediate cost, are allowed. We prefer to start with general sets of decisions $\mathcal{P}(i)$, $i \in E$, which may contain all randomizations. As long as there are no constraints introduced the distinction between randomized and non-randomized decisions is in our opinion not very important (cf. section 13).

In several places we need a notion of convergence on $\mathcal{P}$. A sequence

*) Random variables are underlined.

P_n, $n = 1,2,\ldots$ is convergent to P if $\lim_{n\to\infty} p_n(i,j) = p(i,j)$ for all i and j. In this case, we shall say that $\lim_{n\to\infty} P_n = P$. $\mathcal{P}$ with this topology is a metric space (see section 13).

The identification of the set of actions with the set of probability measures and several notations are adopted from [Bather].

The number of papers on dynamic programming is overwhelming. Only those books or papers referred to in this monograph, or those that proved important for the author's study of these topics are included in the bibliography.

It is difficult to provide a readable and consequent notation for the topics studied. The list of notations may be helpful to overcome possible notational shortcomings.

2. POTENTIALS AND EXCESSIVE FUNCTIONS

The aim of this section is twofold. First to generalize some well-known theorems in Markov potential theory (theorems 2.9 and 2.20 to 2.23). The second intention of this section is to introduce notions which, in our opinion, are basic in the study of discrete time dynamic programming problems. Further we collect in this section definitions and results which play an important role throughout this monograph.

Each function used in this monograph is assumed to be a finite and real valued function. Moreover when writing $\mathbb{E}_P\, f(\underline{x}_n)$ or $P^n f$ it is tacitly assumed that

$$\sum_j p^n(i,j)|f(j)| < \infty \text{ for all } i \in E.$$

2.1. DEFINITION. *Function* w *is a charge with respect to* P *if*

$$\mathbb{E}_P \sum_{n=0}^{\infty} |w(\underline{x}_n)| = \sum_{n=0}^{\infty} P^n|w| < \infty.$$

2.2. DEFINITION. *Function* f *is a potential w.r.t.* P *if there exists a charge* w *w.r.t.* P *such that*

$$f = \sum_{n=0}^{\infty} P^n w.$$

So function w is called a charge if the sum $\sum_{n=0}^{\infty} P^n w$ is well-defined. This sum is then a potential.

2.3. DEFINITION. *Function* f *is a*

$$\begin{array}{lcr} c - \textit{super} & & \geq \\ c - & \textit{harmonic function w.r.t. } P \textit{ if } f & = c + Pf. \\ c - \textit{sub} & & \leq \end{array}$$

2.4. DEFINITION. *Function* f *is a* c *- excessive function w.r.t.* P *if*

(2.4.1) c *is a charge w.r.t.* P

(2.4.2) $\sum_{n=0}^{\infty} P^n c \leq f$

(2.4.3) $c + Pf \leq f.$

So a c-superharmonic function with c a charge satisfying relation (2.4.2) is a c-excessive function. To see that c-excessive functions form an interesting class one should realize that when f is the value function of a stopping problem for a Markov chain with matrix of transition probabilities P and "cost" function c then relations (2.4.2) and (2.4.3) are fulfilled. This can be seen by noting that the left-hand side of (2.4.2) denotes the "return" in case we will never stop which is less than the value function. The left-hand side of (2.4.3) denotes the "return" if we wait one period and then continue in an optimal way. This may be a sub-optimal policy.

2.5. THEOREM. *Function* f *is a potential w.r.t.* P *iff* $w_P := f-Pf$ *is a charge w.r.t.* P *and* $\lim_{n\to\infty} P^n f = 0$.

PROOF. Suppose w is a charge such that $f = \sum_{n=0}^{\infty} P^n w$. Then by interchanging the order of summation (w is a charge) it follows that

$$f-Pf = \sum_{n=0}^{\infty} (P^n w - P^{n+1} w) = w.$$

Hence $w_P = w$ and consequently w_P is a charge. By iterating the equality

$$w_P + Pf = f$$

N times we find the equality

$$w_P + Pw_P + \ldots + P^N w_P + P^{N+1} f = f. \tag{2.5.1}$$

Since $f = \sum_{n=0}^{\infty} P^n w_P$, it then follows that $\lim_{n\to\infty} P^n f = 0$.*)

To show the converse, we note that $\sum_{n=0}^{\infty} P^n w_P$ is a potential since w_P is a charge. Moreover, it follows from (2.5.1) and $\lim_{n\to\infty} P^n f = 0$ that this potential equals f. □

It can be seen from the above proof that a potential uniquely determines its charge (if f is a potential then f-Pf is its charge).

*) For f_n, n=1,2,... a sequence of functions, we write $\lim_{n\to\infty} f_n = 0$ if $\lim_{n\to\infty} f_n(i) = 0$ for all $i \in E$.

2.6. THEOREM. *If to* $c \geq 0$ *there exists a nonnegative* c *- superharmonic function* v *w.r.t.* P *then* c *is a charge w.r.t.* P *and* $\sum_{n=0}^{\infty} P^n c \leq v$.

PROOF. The definition of a c-superharmonic function gives

$$c + Pv \leq v.$$

By iterating this inequality N times we find

$$c + Pc + \ldots + P^N c + P^{N+1} v \leq v.$$

Since $v \geq 0$ it follows then

$$\sum_{n=0}^{\infty} P^n c \leq v < \infty$$

and consequently c is a charge. □

As an illustration of theorem 2.6 we shall prove that relation (2.7.1) is sufficient for a Markov chain to be positive recurrent. In this way we recover the condition for positive recurrence as can be found in [Foster, theorem 2]. For a countable state space a condition similar to (2.7.1) can be found in [Kushner, theorem 8.6.5.7, p. 211]. There the condition is called a Liapunov function criterion.

2.7. FOSTER CRITERION - LIAPUNOV FUNCTION CRITERION

The Markov chain with transition matrix P *is positive recurrent if there exists a state* i_0 *and a nonnegative solution* y *of the inequalities*

$$e + \tilde{P}y \leq y, \tag{2.7.1}$$

where e *is defined by* $e(i) = 1$ *for all* i *and* $\tilde{P}$ *is the column-restriction of* P *to* $E\backslash\{i_0\}$ *i.e.*

$$\tilde{p}(i,j) := \begin{cases} 0 & \text{for } j = i_0 \\ p(i,j) & \text{for } j \neq i_0. \end{cases}$$

PROOF. Let $\underline{\tau}$ denote the reentry time of $\{i_0\}$, i.e. $\underline{\tau}$ is the least $n > 0$ if any with $\underline{x}_n = i_0$, and $\underline{\tau} = \infty$ if there is no such n. Then it is an easy check that

$$(2.7.2) \qquad \mathbb{P}_i[\underline{\tau} > n] = \tilde{P}^n e(i).$$

According to a well-known lemma

$$(2.7.3) \qquad \mathbb{E}_i[\underline{\tau}] = \sum_{n=0}^{\infty} \mathbb{P}_i[\underline{\tau} > n].$$

By (2.7.2) and (2.7.3) we have

$$(2.7.4) \qquad \mathbb{E}_i[\underline{\tau}] = \sum_{n=0}^{\infty} \tilde{P}^n e(i).$$

The Markov chain is a positive recurrent class ([Chung, p. 31]) if

$$(2.7.5) \qquad \mathbb{E}_i[\underline{\tau}] < \infty \text{ for all } i \in E.$$

To prove this it is by (2.7.4) sufficient to show that $\sum_{n=0}^{\infty} \tilde{P}^n e < \infty$ (i.e. all components are finite). Now theorem 2.6 says that relation (2.7.1) implies that e is a charge w.r.t. $\tilde{P}$. □

A Liapunov function criterion for the existence of an invariant probability measure in the case of a Markov process with a metric state space is given in [Hordijk and Van Goethem].

2.8. THEOREM. *If there exists a* c*-superharmonic function* f *w.r.t.* P*, for* c *a majorant of a charge then*

a. $h := \lim_{n\to\infty} P^n f$ *exists and* $-\infty \le h(i) < \infty$ *for all* $i \in E$

b. *if* $h(i) > -\infty$ *for all* $i \in E$ *then* c *is a charge w.r.t.* P

c. *if* $h \ge 0$ *) *then* f *is* c*-excessive w.r.t.* P.

*) We write $x \ge y$ if $x(i) \ge y(i)$ for all i and denote 0 for the vector with each component equal to 0.

PROOF. a. Let w be a charge such that $w \le c$. For $w_P := f-Pf$ it holds that $c_P := w_P-w \ge w_P-c \ge 0$ and

$$(2.8.1)\qquad w + c_P + Pf = f.$$

By iterating this equality N times we find

$$(2.8.2)\qquad \sum_{n=0}^{N} P^n(w+c_P) + P^{N+1}f = f,$$

w is a charge and $c_P \ge 0$ so $\lim_{N\to\infty} \sum_{n=0}^{N} P^n(w+c_P)(i)$ exists (and cannot be $-\infty$) and consequently also $\lim_{N\to\infty} P^N f(i)$ exists (and cannot be $+\infty$), for each $i \in E$.

b. If $\lim_{N\to\infty} P^N f(i)$ is finite for all $i \in E$ then $\sum_{n=0}^{\infty} P^n c_P < \infty$ and it follows that the nonnegative function c_P is a charge and so is w_P. Let

$$c^+ = \max(c,0) \text{ and } c^- = -\min(c,0).$$

Since w, w_P are charges and $w \le c \le w_P$, we have

$$(2.8.3)\qquad \sum_{n=0}^{\infty} P^n c^- \le \sum_{n=0}^{\infty} P^n w^- < \infty$$

$$(2.8.4)\qquad \sum_{n=0}^{\infty} P^n c^+ \le \sum_{n=0}^{\infty} P^n w_P^+ < \infty.$$

Relations (2.8.3) and (2.8.4) together imply that $\sum_{n=0}^{\infty} P^n|c| < \infty$ and hence c is a charge.

c. By iterating the inequality $c + Pf \le f$ we find

$$\sum_{n=0}^{N} P^n c + P^{N+1} f \le f.$$

If $\lim_{n\to\infty} P^n f \ge 0$ then we have that

$$\sum_{n=0}^{\infty} P^n c \le f.$$

Consequently c and f satisfy the relations (2.4.1), (2.4.2), (2.4.3) and f is a c - excessive function w.r.t. P. □

With $c \equiv 0$, the following theorem is similar to a theorem in classical potential theory due to M. Riesz (see [Helms, theorem 6.18]).

2.9. THEOREM. *A* c *- excessive function w.r.t.* P *is the sum of a potential w.r.t.* P *with charge not less than* c *and a nonnegative harmonic function w.r.t.* P.

PROOF. Let $w_P := f-Pf$ for f a c - excessive function w.r.t. P. Then f is a w_P-harmonic function and $w_P \geq c$. Relation (2.4.2) implies that

$$P^N f \geq P^N \sum_{n=0}^{\infty} P^n c = \sum_{n=N}^{\infty} P^n c.$$

This yields that (theorem 2.8 shows the existence of the limit) $\lim_{N\to\infty} P^N f \geq 0$ and we conclude by theorem 2.8 that w_P is a charge. From $w_P+Pf = f$ it follows by iteration $f = \sum_{n=0}^{\infty} P^n w_P + h$, with $h = \lim_{N\to\infty} P^N f$. Since

$$f - \sum_{n=0}^{\infty} P^n w_P^- \leq P^N f \leq f + \sum_{n=0}^{\infty} P^n w_P^- ,$$

it follows by the dominated convergence theorem that $Ph = h$ and consequently h is a harmonic function. □

We note that the above representation of a c - excessive function as the sum of a potential and a harmonic function is unique. Indeed, if $f = \sum_{n=0}^{\infty} P^n w + h$, with w a charge and h a harmonic function. Then $Pf = \sum_{n=1}^{\infty} P^n w + Ph = f-w$. Hence $w = f-Pf$ and the potential $\sum_{n=0}^{\infty} P^n w$ is uniquely determined by f. And so is $h = f - \sum_{n=0}^{\infty} P^n w$.

2.10. THEOREM. *If* c *is a charge and* f *is a* c *-superharmonic function w.r.t.* P *then the following assertions are equivalent*

a. $\lim_{n\to\infty} P^n f \geq 0$

b. $\lim_{n\to\infty} P^n f^- = 0$

c. f *is a* c *- excessive function.*

PROOF. According to theorem 2.8 we have that condition a implies condition c. If $\sum_{n=0}^{\infty} P^n c \leq f$ then $-f \leq \sum_{n=0}^{\infty} P^n(-c)$. Hence

$$(-f)^+ \le \left(\sum_{n=0}^{\infty} P^n(-c)\right)^+ \le \sum_{n=0}^{\infty} (P^n(-c))^+ \le \sum_{n=0}^{\infty} P^n(-c)^+.$$

Using that for arbitrary function g it holds that $(-g)^+ = g^-$ we have

$$0 \le f^- \le \sum_{n=0}^{\infty} P^n c^-.$$

Since c is a charge it follows then

$$\lim_{n\to\infty} P^n f^- \le \lim_{n\to\infty} P^n \sum_{k=0}^{\infty} P^k c^- = \lim_{n\to\infty} \sum_{k=n}^{\infty} P^k c^- = 0.$$

Hence c implies b.

To conclude we note that according to theorem 2.8 $\lim_{n\to\infty} P^n f$ exists and hence condition b implies $\lim_{n\to\infty} P^n f = \lim_{n\to\infty} P^n f^+ \ge 0$.

2.11. THEOREM. *If* f *is a* c*-superharmonic function with* c *a charge w.r.t.* P *then the following assertions are equivalent*

a. $\lim_{n\to\infty} P^n f = 0$

b. $\lim_{n\to\infty} P^n |f| = 0$

c. f *is a potential.*

PROOF. According to theorem 2.8 $\lim_{n\to\infty} P^n f$ does exist and $\lim_{n\to\infty} P^n f = 0$ implies that f is a c-excessive function. Theorem 2.10 then gives $\lim_{n\to\infty} P^n f^- = 0$. Together with $\lim_{n\to\infty} P^n f = \lim_{n\to\infty} P^n(f^+ - f^-) = 0$ this implies that $\lim_{n\to\infty} P^n f^+ = 0$. Consequently $\lim_{n\to\infty} P^n |f| = \lim_{n\to\infty} P^n(f^+ + f^-) = 0$ and so condition a implies condition b. Since f is also a (f-Pf)-superharmonic function and $c \le f-Pf$ it follows from theorem 2.8 that condition a implies that f-Pf is a charge. By theorem 2.5 it then follows that f is a potential. As b implies a, we now have that a implies c. Also from theorem 2.5 we have that condition c implies condition a. □

In the following sections we want to study Markov decision processes. Since each stationary policy corresponds to a Markov process we will extend the notions charge, potential and excessivity to collections of Markov processes.

2.12. DEFINITION. *When for each* P *element of a collection of Markov matrices* P *we have a function* c_P *we will speak of a cost structure* c_P.*) *The cost structure* c_P *is a charge structure if*

$$\mathbb{E}_R \sum_{n=0}^{\infty} |c(\underline{x}_n)| = \sum_{n=0}^{\infty} P_0 \cdots P_{n-1}|c_{P_n}| < \infty$$

for each $R = (P_0, P_1, \ldots)$.

2.13. DEFINITION. *For* c_P *a cost structure we call function* f *a*

$$\begin{array}{lcl} c_P - \textit{super} & & \geq \\ c_P - & \textit{harmonic function if } f & = c_P + Pf \textit{ for all } P \in P. \\ c_P - \textit{sub} & & \leq \end{array}$$

2.14. DEFINITION. *Function* f *is a* c_P*-excessive function if*

(2.14.1) c_P *is a charge structure*

(2.14.2) $\mathbb{E}_R \sum_{n=0}^{\infty} c(\underline{x}_n) \leq f$ *for all* R

(2.14.3) $c_P + Pf \leq f$ *for all* P.

2.15. DEFINITION. *Function* f *is a potential w.r.t.* P *if there exists a charge structure* c_P *such that*

$$f = \mathbb{E}_R \sum_{n=0}^{\infty} c(\underline{x}_n) \quad \textit{for all } R.$$

At first sight this definition looks very restrictive. In section 7 (theorem 7.3) it is shown that there are natural examples of potentials w.r.t. P.

2.16. THEOREM. *Function* f *is a potential w.r.t.* P *iff* $w_P := f - Pf$, $P \in P$, *defines a charge structure and* $\lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_n) = 0$ *for each* R.

PROOF. Suppose c_P is a charge structure such that $f = \mathbb{E}_R \sum_{n=0}^{\infty} c(\underline{x}_n)$ for all

*) In the following sections $c_P(i)$ will denote the cost when choosing the action or decision $p(i,.)$ in state i.

R. In particular for $R = (P,P,\ldots)$ it then follows that $f = \sum_{n=0}^{\infty} P^n c_P$ and consequently (cf. theorem 2.5) $c_P = f - Pf$. Hence $w_P = c_P$ for all $P \in \mathcal{P}$ and therefore $w_{\mathcal{P}}$ is a charge structure w.r.t. $\mathcal{P}$. By definition we have $w_P + Pf = f$ for all $P \in \mathcal{P}$. By iterating this equality we find

$$(2.16.1) \qquad \sum_{n=0}^{N} P_0 \ldots P_{n-1} w_{P_n} + P_0 \ldots P_N f = f.$$

For arbitrary policy $R = (P_0,P_1,\ldots)$ we conclude from (2.16.1) that

$$(2.16.2) \qquad f = \sum_{n=0}^{\infty} P_0 \ldots P_{n-1} w_{P_n} = \mathbb{E}_R \sum_{n=0}^{\infty} w(\underline{x}_n) \text{ iff}$$

$$\lim_{n\to\infty} P_0 \ldots P_n f = \lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_{n+1}) = 0. \square$$

2.17. THEOREM. *If* $c_{\mathcal{P}}$ *is a charge structure and* f *is a* $c_{\mathcal{P}}$*-superharmonic function then the following assertions are equivalent*

a. $\lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_n) \geq 0$ *for all* R

b. $\lim_{n\to\infty} \mathbb{E}_R f^-(\underline{x}_n) = 0$ *for all* R

c. f *is a* $c_{\mathcal{P}}$*-excessive function.*

PROOF. Let $w_P := f - Pf - c_P$ then $w_P \geq 0$ and

$$c_P + w_P + Pf = f \quad \text{for all P.}$$

By iterating this equality we find

$$(2.17.1) \qquad \sum_{n=0}^{N} P_0 \ldots P_{n-1}(c_{P_n} + w_{P_n}) + P_0 \ldots P_N f = f.$$

Since $c_{\mathcal{P}}$ is a charge structure and $w_P \geq 0$ for all P the first term in relation (2.17.1) has a limit. This implies that for policy $R = (P_0,P_1,\ldots)$ $\lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_n) = \lim_{n\to\infty} P_0 \ldots P_{n-1} f$ exists. Hence we conclude that $\lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_n)$ exists for all policies R. If moreover condition a is satisfied then we have

$$(2.17.1) \qquad \sum_{n=0}^{\infty} P_0 \ldots P_{n-1}\, c_P \leq \sum_{n=0}^{\infty} P_0 \ldots P_{n-1}(c_P + w_P) \leq f$$

and consequently $\mathbb{E}_R \sum_{n=0}^{\infty} c(\underline{x}_n) \le f$ for arbitrary $R = (P_0,P_1,\ldots)$. By definition it follows that f is a c_P - excessive function.

Assume c, then $\mathbb{E}_R \sum_{n=0}^{\infty} c(\underline{x}_n) \le f$ for all R. Rewriting this for $R = (P_N,P_{N+1},\ldots)$ we have

$$\sum_{n=N}^{\infty} P_N \cdots P_{n-1} c_{P_n} \le f.$$

Similar to the proof of theorem 2.10 we conclude from this

$$P_0 \cdots P_{N-1} f^- \le P_0 \cdots P_{N-1} \sum_{n=N}^{\infty} P_N \cdots P_{n-1} c^-_{P_n}.$$

Since c_P is a charge structure we have that the right-hand side of this inequality tends to zero as N tends to infinity. From this it follows that condition b is satisfied. It is obvious that condition b implies condition a. □

Theorems 2.16 and 2.17 are similar to theorems 2.5 and 2.10. Also a theorem similar to theorem 2.11 can be proved.

The remaining theorems of this section for the case of a cost structure identically zero and a collection P consisting of one Markov matrix (so in the case of a Markov process) are well-known in Markov potential theory (see [Blumenthal and Getoor], [Dynkin and Juschkewitsch], [Hunt]).

2.18. THEOREM. *If* f *is a potential w.r.t.* P *and* $\underline{\tau}$ *is a Markov time then for arbitrary* R

$$\mathbb{E}_R\, f(\underline{x}_{\underline{\tau}}) = \mathbb{E}_R \sum_{n=\underline{\tau}}^{\infty} w(\underline{x}_n) \text{ with } w_P := f-Pf, \qquad P \in P. \tag{2.18.1}$$

PROOF. For arbitrary policy $R = (P_0,P_1,\ldots)$ we write $R_n = (P_n,P_{n+1},\ldots)$. Since $\underline{\tau}$ is a Markov time we have that

$$\mathbb{E}_R\, [w(\underline{x}_{\underline{\tau}+k}) \mid \underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n] = \tag{2.18.2}$$

$$= \sum_{\ell} P_n P_{n+1} \cdots P_{n+k-1}(j,\ell)\, w_{P_{n+k}}(\ell) = \mathbb{E}_{j,R_n}[w(\underline{x}_k)].$$

Summing this for k = 0 to ∞ and using theorem 2.16, in particular relation (2.16.2), we find

$$(2.18.3) \qquad \sum_{k=0}^{\infty} \mathbb{E}_R\,[w(\underline{x}_{\underline{\tau}+k}) \mid \underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n] = f(j).$$

Now

$$\mathbb{E}_R\,[\sum_{n=\underline{\tau}}^{\infty} w(\underline{x}_n)] \overset{(1)}{=} \sum_{n=0}^{\infty}\sum_j \mathbb{P}_R\,[\underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n]\,\mathbb{E}_R\,[\sum_{k=0}^{\infty} w(\underline{x}_{\underline{\tau}+k}) \mid \underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n] =$$

$$\overset{(2)}{=} \sum_{n=0}^{\infty}\sum_j \mathbb{P}_R\,[\underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n] \sum_{k=0}^{\infty} \mathbb{E}_R\,[w(\underline{x}_{\underline{\tau}+k}) \mid \underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n] =$$

$$\overset{(3)}{=} \sum_{n=0}^{\infty}\sum_j \mathbb{P}_R\,[\underline{x}_{\underline{\tau}}=j,\ \underline{\tau}=n]\, f(j) =$$

$$= \mathbb{E}_R\, f(\underline{x}_{\underline{\tau}}),$$

where equality (1) comes from taking the expectation of the conditional expectation w.r.t. $(\underline{x}_{\underline{\tau}},\underline{\tau})$, equality (2) follows from Fubini's theorem on interchanging the order integration (or summation), equality (3) is direct from relation (2.18.3). □

From relation (2.18.1) it follows for f a potential w.r.t. P and $\underline{\tau} \le \underline{\tau}^*$ Markov times that for arbitrary R and charge structure c_P

$$(2.18.4) \qquad \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + f(\underline{x}_{\underline{\tau}})] - \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}^*-1} c(\underline{x}_n) + f(\underline{x}_{\underline{\tau}^*})] =$$

$$= \mathbb{E}_R\,[\sum_{n=\underline{\tau}}^{\underline{\tau}^*-1} (w(\underline{x}_n)-c(\underline{x}_n))].$$

For the case that f is c_P - superharmonic we have $w(\underline{x}_n)-c(\underline{x}_n) \ge 0$. Substituting this in (2.18.4) we find that the second term on the left-hand side of (2.18.4) is less than the first term. This important property will be proved for excessive functions in the next theorem.

2.19. LEMMA. *For each policy* R *and bounded Markov time* $\underline{\tau}$ *) *we have for an*

---

*) We call a Markov time $\underline{\tau}$ bounded when there exists an integer N such that $\underline{\tau} \le$ N.

arbitrary function r

$$r = \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}}) \right], \tag{2.19.1}$$

where $w_P = r - Pr$ *for all* $P \in \mathcal{P}$.

PROOF. The proof is given by induction on the upper bound of the Markov times. Suppose (2.19.1) is valid for all policies R and all Markov times $\underline{\tau}$ with $\underline{\tau} \le N$. Now let $\underline{\tau} \le N+1$ and $R = (P_0, P_1, \ldots)$ and $R_1 = (P_1, P_2, \ldots)$. We prove (2.19.1) for arbitrary state i. Since $\underline{\tau}$ is a Markov time we have on the event $\underline{x}_0 = i$ whether $\underline{\tau} = 0$ or $\underline{\tau} > 0$. When $\underline{\tau} = 0$ then relation (2.19.1) is obvious for starting state i. When $\underline{\tau} > 0$ on $\underline{x}_0 = i$, we define a new stochastic variable

$$\underline{\tau}^*(i_0, i_1, \ldots) := \underline{\tau}(i, i_0, i_1, \ldots) - 1. \tag{2.19.2}$$

It is easy to check that $\underline{\tau}^*$ is a Markov time and $\underline{\tau}^* \le N$. By the induction hypothesis we then have

$$r = \mathbb{E}_{R_1} \left[\sum_{n=0}^{\underline{\tau}^*-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}^*}) \right]. \tag{2.19.3}$$

Now

$$\mathbb{E}_{i,R} \left[\sum_{n=0}^{\underline{\tau}-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}}) \right] =$$

$$= w_{P_0}(i) + \sum_j p_0(i,j)\, \mathbb{E}_{j,R_1} \left[\sum_{n=0}^{\underline{\tau}^*-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}^*}) \right] =$$

$$= w_{P_0}(i) + \sum_j p_0(i,j)\, r(j) = r(i),$$

where the first equation follows from (2.19.2) with the Markov property, the second from (2.19.3) and the third from the definition of w_{P_0}. □

2.20. THEOREM. *If* r *is a* c_P*-excessive function and* $\underline{\tau}$, $\underline{\tau}^*$ *are Markov times with* $\underline{\tau} \le \underline{\tau}^*$ *then*

$$(2.20.1) \qquad \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})] \geq \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}^*-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}^*})],$$

for each policy R.

PROOF. For any integer N let $\underline{\tau}_N = \underline{\tau} \wedge N$ and $\underline{\tau}_N^* = \underline{\tau}^* \wedge N$. Then $\underline{\tau}_N$ and $\underline{\tau}_N^*$ are bounded Markov times. Lemma 2.19 yields

$$\mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_N})] = \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N^*-1} w(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_N^*})],$$

where $w_P = r - Pr$ for all $P \in \mathcal{P}$. Rearranging this equation, writing $r(\underline{x}_{\underline{\tau}_N})\,\chi(\underline{\tau} \leq N) + r(\underline{x}_N)\,\chi(\underline{\tau} > N)$ for $r(\underline{x}_{\underline{\tau}_N})$ and inserting sums like $\sum_{n=0}^{\underline{\tau}_N-1} c(\underline{x}_n)$ on both sides we find

$$(2.20.2) \qquad \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_N})\,\chi(\underline{\tau} \leq N) - \sum_{n=0}^{\underline{\tau}_N^*-1} c(\underline{x}_n) - r(\underline{x}_{\underline{\tau}_N^*})\,\chi(\underline{\tau}^* \leq N)] =$$

$$= \mathbb{E}_R\,[\sum_{n=\underline{\tau}_N}^{\underline{\tau}_N^*-1} (w(\underline{x}_n) - c(\underline{x}_n)) + r(\underline{x}_N)(\chi(\underline{\tau}^* > N) - \chi(\underline{\tau} > N))].$$

The limit as $N \to \infty$ of the first half of this equation is just the difference of the first and second term of (2.20.1). Hence we have to prove that this limit is nonnegative. Since r is $c_{\mathcal{P}}$ - superharmonic we have that $w_P - c_P \geq 0$ for all $P \in \mathcal{P}$ and this implies that the first term of the right-hand side of (2.20.2) has a nonnegative lim inf as $N \to \infty$. According to theorem 2.17 it follows with $\underline{\tau}^* \geq \underline{\tau}$ that

$$\liminf_{N\to\infty} \mathbb{E}_R\,[r(\underline{x}_N)(\chi(\underline{\tau}^* > N) - \chi(\underline{\tau} > N))] \geq -\lim_{N\to\infty} \mathbb{E}_R\, r^-(\underline{x}_N) = 0.$$

Consequently both terms on the right-hand side have a nonnegative lim inf and the proof is complete. □

We state a direct consequence of this theorem.

2.21. THEOREM. *If* r *is a* $c_{\mathcal{P}}$ *- excessive function then for each Markov time* $\underline{\tau}$

$$(2.21.1) \qquad r \geq \sup_R \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})]$$

PROOF. Substitute $\underline{\tau} \equiv 0$ in (2.20.1) then

$$r \geq \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}^*-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}^*})\right],$$

for each policy R and Markov time $\underline{\tau}^*$. Upon taking the supremum over all R the above inequality is relation (2.21.1). □

2.22. THEOREM. *If* r *is a* c_p *- excessive function then for arbitrary Markov time* $\underline{\tau}$

$$f := \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})\right]$$

is also a c_p *- excessive function.*

PROOF. To prove that f is a c_p - excessive function we have to check the relations (2.14.1), (2.14.2) and (2.14.3). The proof of (2.14.3), i.e. the proof that f is a c_p - superharmonic function, is postponed to the proof of theorem 3.1. There a slightly more general result has to be proved. By definition c_p is a charge structure and hence relation (2.14.1) is satisfied. To prove relation (2.14.2) substitute $\underline{\tau}^* \equiv \infty$ in (2.20.1) then

$$f \geq \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})\right] \geq \mathbb{E}_R \left[\sum_{n=0}^{\infty} c(\underline{x}_n)\right] \text{ for all R. } \square$$

2.23. THEOREM. *Let* $\underline{\tau}_A$ *be the entry time of set* A, *i.e.* $\underline{\tau}$ *is the least* $n \geq 0$ *if any with* $\underline{x}_n \in A$, *and* $\underline{\tau}_A = \infty$ *if there is no such* n. *If* r *is a* c_p *- excessive function then*

$$f := \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}_A-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_A})\right]$$

is the minimum of the c_p *- excessive functions that majorize* r *on* A.

PROOF. According to theorem 2.22 f is a c_p - excessive function. From the definition of $\underline{\tau}_A$ it follows immediately that f = r on A. Suppose g is a c_p - excessive function that majorizes **r on** A. Then for each policy R we have $\mathbb{E}_R\, g(\underline{x}_{\underline{\tau}_A}) \geq \mathbb{E}_R\, r(\underline{x}_{\underline{\tau}_A})$. Since g is c_p - excessive it follows from theorem 2.21 that

$$g \geq \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}_A - 1} c(\underline{x}_n) + g(\underline{x}_{\underline{\tau}_A}) \right] \geq$$

$$\geq \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}_A - 1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_A}) \right] = f.$$

Hence f is the minimum of the $c_{\mathcal{P}}$ - excessive functions that majorize r on A. □

3. ON THE VALUE FUNCTION OF AN OPTIMAL CONTROL PROBLEM

In the sections 3 and 4 we deal with the optimal control problem: given a cost structure c_P which is a charge structure and given a reward function r with $\mathbb{E}_R\,|r(\underline{x}_{\underline{\tau}})| < \infty$ for all R and $\underline{\tau}$, find a policy R and stopping time $\underline{\tau}$ ($\underline{\tau} = \infty$ with positive probability is admissible, with zero reward) such that

$$\mathbb{E}_R\,[\sum_{n=0}^{\tau-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})]$$

is maximized. In this section we investigate properties of the *value function*

$$v := \sup_{R,\underline{\tau}} \mathbb{E}_R\,[\sum_{n=0}^{\tau-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})]. \tag{3.0.1}$$

We assume that $-\infty < v(i) < +\infty$ and $P|v|(i) < +\infty$ for all $i \in E$ and all $P \in \mathcal{P}$. In section 13 we give some conditions implying these assumptions (cf. lemma 13.4).

As far as the author knows this general problem has not been studied previously. Related work can be found in [Bellman], [Blackwell (1967)] [Dubins and Savage], [Dynkin and Juschkewitsch], [Hinderer] and [Strauch]. The sections 3 and 4 extend the work of Dynkin and others on optimal stopping problems to allow for control of the transitions of the Markov process as well as its stopping time. They extend the work of Dubins and Savage and others on gambling models to allow for a cost structure along with a reward function.

3.1. THEOREM. *The function* v *is the minimum of the* c_P*-excessive functions that majorize* r.

PROOF. We first prove that v is a c_P-excessive function by verifying that the relations (2.14.1), (2.14.2) and (2.14.3) are satisfied. Relation (2.14.1) is true by definition. Relation (2.14.2) follows upon substituting $\underline{\tau} \equiv \infty$ from (3.0.1). To prove that v is a c_P-superharmonic function we choose an $\varepsilon > 0$. Then there exist policies R_i and stopping times $\underline{\tau}_i$, $i \in E$, such that

$$\mathbb{E}_{i,R_i}\,[\sum_{n=0}^{\underline{\tau}_i-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_i})] \geq v(i) - \varepsilon. \tag{3.1.1}$$

Define

$$\underline{\tau}(i_0,i_1,\dots) = 1 + \underline{\tau}_{i_1}(i_1,i_2,\dots),$$

then $\underline{\tau}$ is a Markov time. For P an arbitrary element of $\mathcal{P}$ let R be the policy that chooses decision rule P at time 0 and uses policy R_i from time 1 when the state at time 1 is i. For a more formal definition of R let $R_i = (P_{i0},P_{i1},\dots)$, $i \in E$, then the decision rule at time n+1 given the history $\underline{x}_0 = i_0$, $\underline{x}_1 = i_1,\dots,$ $\underline{x}_{n+1} = i_{n+1}$ is $P_{i_1 n}$. It is important to realize that R is not a memoryless policy and as such rather unique in this monograph. Now by the definition of v, the Markov property and relation (3.1.1) we have

$$\begin{aligned} v(i) &\geq \mathbb{E}_{i,R}\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})] = \\ &= c_P(i) + \sum_j p(i,j)\,\mathbb{E}_{j,R_j}\,[\sum_{n=0}^{\underline{\tau}_j-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_j})] \geq \\ &\geq c_P(i) + \sum_j p(i,j)\; v(j) - \varepsilon, \end{aligned}$$

since $\sum_j p(i,j) \leq 1$. Because ε and P were arbitrarily chosen, this means that v is a $c_{\mathcal{P}}$-superharmonic function.

Substituting $\underline{\tau} \equiv 0$ in (3.0.1) gives $v \geq r$ and hence v majorizes r. To prove that v is the minimum of the $c_{\mathcal{P}}$-excessive functions that majorize r we suppose that a certain function g is $c_{\mathcal{P}}$-excessive and majorizes r. Then according to theorem 2.21 and the fact that $g \geq r$

$$g \geq \sup_{R,\underline{\tau}} \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + g(\underline{x}_{\underline{\tau}})] \geq v. \;\square$$

We call a policy R together with a stopping time $\underline{\tau}$ a *strategy*. In many cases an *optimal strategy*, i.e. a strategy $(R,\underline{\tau})$ such that $v = \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})]$, can be determined when the value function v is known. So it is important to characterize the function v. We gave in the above theorem a characterization. Some more theorems which may be helpful in computing v will be given below.

3.2. DEFINITION. *Let* $T : x \to Tx$ *be the operator defined by*

$$Tx := r \vee \sup_P (c_P+Px).^{*)}$$

3.3. DEFINITION. *The optimal control problem is stable w.r.t.* x *if*

$$\lim_{N\to\infty} T^N x = v.$$

3.4. THEOREM. *Suppose the problem is stable w.r.t.* x. *If* $v \geq x$ *then* v *is the minimum of the* c_P*-superharmonic functions that majorize* $x \vee r$.

PROOF. Suppose $g \geq x \vee r$ and is c_P-superharmonic. Then $g \geq Tg \geq Tx$ which implies by iterating these inequalities that $g \geq T^N x$ for all N. Thus $g \geq v = \lim_{N\to\infty} T^N x$. Since v majorizes $x \vee r$ if $v \geq x$ this proves the theorem. □

3.5. THEOREM. *The value function* v *is a solution to Bellman's optimality equation*

$$v = r \vee \sup_P (c_P+Pv). \tag{3.5.1}$$

REMARK. The above assertion can also be stated as: v *is a fixed point of* T.

PROOF. Since v is a c_P-excessive function and v majorizes r (see theorem 3.1) we have by relation (2.14.3) that

$$v \geq r \vee \sup_P (c_P+Pv). \tag{3.5.2}$$

To prove the reverse inequality, note that given any $\varepsilon > 0$ and any state i there exists a strategy $(R,\underline{\tau})$ with $R = (P_0,P_1,\ldots)$ such that

$$\mathbb{E}_{i,R}\left[\sum_{n=0}^{\tau-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})\right] \geq v(i) - \varepsilon. \tag{3.5.3}$$

---

*) For vectors x and y the vector $x \vee y$ resp. $x \wedge y$ has i^{th} component $\max(x(i),y(i))$ resp. $\min(x(i), y(i))$.

Since $\underline{\tau}$ is a Markov time we have on the event $\underline{x}_0 = i$ whether $\underline{\tau} = 0$ or $\underline{\tau} > 0$. When $\underline{\tau} = 0$ then from (3.5.3) $r(i) \geq v(i)-\varepsilon$. When $\underline{\tau} > 0$ on $\underline{x}_0 = i$ we define a new stochastic variable

$$\underline{\tau}^*(i_0,i_1,\ldots) := \underline{\tau}(i,i_0,i_1,\ldots) - 1.$$

Then $\underline{\tau}^*$ is a Markov time and it follows from the Markov property and the definition of v when $R_1 := (P_1,P_2,\ldots)$ that

$$v(i)-\varepsilon \leq c_{P_0}(i) + \sum_j p_0(i,j)\, \mathbb{E}_{j,R_1} [\sum_{n=0}^{\underline{\tau}^*-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}^*})] \leq$$

$$\leq c_{P_0}(i) + \sum_j p_0(i,j)\, v(j).$$

Hence we conclude that

$$v(i) \leq r(i) \vee \sup_{P} (c_P+Pv)(i) + \varepsilon$$

Since ε and i were arbitrarily chosen it follows that

$$\text{(3.5.4)} \qquad v \leq r \vee \sup_{P} (c_P+Pv).$$

The relations (3.5.2) and (3.5.4) together prove the theorem. □

The next theorem gives conditions under which the supremum of c_P+Pv, $P \in \mathcal{P}$, is actually attained.

3.6. THEOREM. *Suppose $\mathcal{P}$ is compact and c_P is upper semicontinuous (i.e. $c_P(i)$ is an upper semicontinuous function of P for all $i \in E$). For v to be a solution of the functional equation*

$$\text{(3.6.1)} \qquad v = r \vee \max_{P} (c_P+Pv),$$

each of the following four conditions is sufficient

a. $c_P + Pv$ *is an upper semicontinuous function of* P

b. $\limsup_{P\to P_0} Pv^+ \leq P_0v^+$ *for all* $P_0 \in \mathcal{P}$

c. *Except for at most a finite number of states the function* v *is non-positive*

d. $\lim_{P\to P_0} Pe = P_0 e$ *for all* $P_0 \in \mathcal{P}$ *and* v *is bounded from above or* v^+ *is uniformly integrable w.r.t.* $\mathcal{P}(i)$, *where*

(3.6.2) $\qquad \mathcal{P}(i) := \{p(i,.) : P \in \mathcal{P}\},\ i \in E.$

PROOF. Let $w := \sup_{\mathcal{P}} (c_P + Pv)$. A well-known theorem says that an upper semicontinuous function attains its supremum over a compact set. Hence condition a implies the existence of a Q with $c_Q + Qv = w$. The proof proceeds now by proving that the other three conditions imply the upper semicontinuity of Pv and hence of $c_P + Pv$.

There is also a well-known theorem which says that the limit of a nonincreasing sequence of upper semicontinuous functions is again upper-semicontinuous. For any state j is $p(i,j)\ v(j)$ a continuous function of P. Hence $P(-v^-)$ is upper semicontinuous. By assumption b then also Pv^+ is upper semicontinuous and consequently so is Pv.

It is easily seen that condition c implies condition b. According to a theorem due to [Scheffé] (see also lemma 4.11)

$$\lim_{P\to P_0} \sum_j p(i,j) = \sum_j p_0(i,j)$$

implies that the convergence of $p(i,j)$ to $p_0(i,j)$ is uniformly in $j \in E$. Hence v^+ bounded or uniformly integrable w.r.t. $\mathcal{P}_i$, $i \in E$, is sufficient for condition b. □

3.7. DEFINITION. *Function* f *has the property anne (asymptotic nonnegative expectation) if*

(3.7.1) $\qquad \liminf_{n\to\infty} \mathbb{E}_R f(\underline{x}_n) \ge 0$ *for all* R.

We proved in theorem 2.17 that if f is a $c_{\mathcal{P}}$-superharmonic function then for all R

(3.7.2) $\qquad \lim_{n\to\infty} \mathbb{E}_R f(\underline{x}_n)$

exists. Moreover, relation (3.7.1) is equivalent to the c_p-excessivity of f. Thus we have the following theorem.

3.8. THEOREM. *Let* f *be a* c_p*-superharmonic function. The function* f *has the property anne if and only if* f *is a* c_p*-excessive function.*

3.9. THEOREM. *The value function* v *is the minimum of the* c_p*-superharmonic functions that majorize* r *and have the property anne. The value function* v *is the minimum of the solutions of Bellman's optimality equation that have the property anne.*

PROOF. Since v is according to theorem 3.1 the minimum of the c_p-excessive functions that majorize r, the first assertion follows from theorem 3.8. Since a solution of (3.5.1) is a c_p-superharmonic function that majorizes r, the class of solutions of the optimality equation is a subset of the c_p-superharmonic functions that majorize r. Hence the second assertion is a consequence of the first assertion and theorem 3.5. □

It may be difficult to check whether a solution of the optimality equation has property anne. In the case one knows that $v \geq 0$ it is perhaps easier to use the following consequence of theorem 3.9: v is the smallest nonnegative solution of the optimality equation.

3.10. THEOREM. *Suppose the problem is stable w.r.t.* x. *If* $v \leq x$ *then* v *is the unique solution of the optimality equation that minorizes* x *and has the property anne.*

PROOF. Suppose $g \leq x$ and $Tg = g$ and g has property anne. We will show that $g = v$. Indeed, according to theorem 3.9 we have $v \leq g$. To prove the reverse inequality we use the fact that T is a monotone operator, i.e. if $x \leq y$ then $Tx \leq Ty$. Hence $g = \lim_{N\to\infty} T^N g \leq \lim_{N\to\infty} T^N x = v$, the last equality is from the stability w.r.t. x. □

The discounted dynamic programming problem (see section 6) with bounded cost structure is stable w.r.t. x for each bounded function x. Moreover, each bounded function has the property anne. This means that according to theorem 3.10 the value function v is the only bounded solution of the optimality equation.

It follows from a result of [Schäl] that the negative dynamic programming problem (see section 6) is stable w.r.t. 0 when $\mathcal{P}$ is compact and c_P is continuous (i.e. $c_P(i)$ is a continuous function of P, for all $i \in E$). In view of theorem 3.10 we then have that v is the only nonpositive solution of the optimality equation with the property anne.

3.11. THEOREM. *Suppose the value function* v *is a bounded solution of* (3.6.1). *If each* $P \in \mathcal{P}$ *is absorbing* (*i.e.* $\lim_{n\to\infty} P^n e = 0$ *for each* $P \in \mathcal{P}$) *then* v *is the unique bounded solution to* (3.6.1).

PROOF. Suppose w is another bounded solution of (3.6.1), then v-w is bounded. Hence there exists a constant b with $|v-w| \le be$. Let $v = r \vee (c_{P_1} + P_1 v)$ and $w = r \vee (c_{P_2} + P_2 w)$. Since w is a solution of (3.6.1) we have $w \ge r \vee (c_{P_1} + P_1 w)$. Hence it follows

$$(3.11.1) \qquad v-w \le P_1|v-w|.$$

Similarly we have

$$(3.11.2) \qquad w-v \le P_2|v-w|.$$

From the fact that $\mathcal{P}$ has the product property it follows that there exists a matrix $Q \in \mathcal{P}$ such that

$$(3.11.3) \qquad Q|v-w| = P_1|v-w| \vee P_2|v-w|.$$

The relations (3.11.1), (3.11.2) and (3.11.3) together imply

$$|v-w| \le Q|v-w|.$$

Iterating this inequality and using $|v-w| \le be$ yields

$$|v-w| \le Q^N be \text{ for } N = 1,2,\dots .$$

We assumed that Q is absorbing and hence

$$|v-w| \le \lim_{N\to\infty} Q^N be = 0.$$

Consequently $v = w$ and the theorem follows. $\square$

4. EXISTENCE OF OPTIMAL STRATEGIES

In this section we investigate the existence of optimal strategies of the optimal control problem introduced in section 3. The notions "to conserve", "to equalize" and "thrifty" are adapted from [Dubins and Savage]. The relation with previous work is indicated in the introduction of section 3.

As in section 3 we assume that $c_{\mathcal{P}}$ is a charge structure and $\mathbb{E}_R |r(\underline{x}_{\underline{\tau}})| < \infty$ for all policies R and all Markov times $\underline{\tau}$. In this section we assume for the value function v that $\mathbb{E}_R |v(\underline{x}_{\underline{\tau}})| < \infty$ for each strategy $(R,\underline{\tau})$. In section 13 we give some conditions implying this assumption (cf. lemma 13.4).

We shall systematically use the notation

$$w_P := v - Pv, \quad P \in \mathcal{P},$$

where v is the value function.

To make certain that expectations and sums are well-defined when using $w_{\mathcal{P}}$ as cost structure, we show that $w_{\mathcal{P}}$ is a charge structure. According to the theorems 3.1 and 2.17 and relation 2.17.1 we have that

$$(4.0.1) \qquad \mathbb{E}_R \sum_{n=0}^{\infty} w(\underline{x}_n) \le v.$$

Since v is $c_{\mathcal{P}}$ - superharmonic, it follows that $w_P = v - Pv \ge c_P$, $P \in \mathcal{P}$. Hence $w_P^- \le c_P^-$ for all $P \in \mathcal{P}$, which implies

$$(4.0.2) \qquad \mathbb{E}_R \sum_{n=0}^{\infty} w^-(\underline{x}_n) \le \mathbb{E}_R \sum_{n=0}^{\infty} c^-(\underline{x}_n).$$

Because $v < \infty$ and $c_{\mathcal{P}}$ is a charge structure we obtain

$$\mathbb{E}_R \sum_{n=0}^{\infty} w^+(\underline{x}_n) < \infty \text{ and } \mathbb{E}_R \sum_{n=0}^{\infty} w^-(\underline{x}_n) < \infty.$$

According to definition 2.12 $w_{\mathcal{P}}$ is a charge structure.

4.1. THEOREM. *Suppose* $Q \in P$ *is such that*

$$(4.1.1) \qquad c_Q(i) + Qv(i) = v(i)$$

when $i \notin \Gamma := \{i : r(i) = v(i)\}$. *Let* Q^∞ *be the policy* $(Q,Q,\ldots)$ *and* $\underline{\tau}_\Gamma$ *the entry time of set* Γ. *Each of the following two conditions is sufficient for strategy* $(Q^\infty,\underline{\tau}_\Gamma)$ *to be optimal*

a. *value function* v *is a potential w.r.t.* Q

b. *there exists a constant* c *such that* $\underline{\tau}_\Gamma \le c$, $\mathbb{P}_Q$ *almost surely.*

PROOF. Let us first show that conditions a and b both imply

$$(4.1.2) \qquad v = \mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} w(\underline{x}_n) + v(\underline{x}_{\underline{\tau}_\Gamma}) \right].$$

As to condition b relation (4.1.2) is direct from lemma 2.19. If we assume a and take for collection P in theorem 2.18 the set $\{Q\}$ then relation (4.1.2) follows from relation (2.18.1).

By the definition of Γ we have

$$\mathbb{E}_Q [r(\underline{x}_{\underline{\tau}_\Gamma})] = \mathbb{E}_Q [v(\underline{x}_{\underline{\tau}_\Gamma})].$$

From relation (4.1.1) it follows that $w_Q(i) = c_Q(i)$ as $i \notin \Gamma$. Hence

$$\mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n) \right] = \mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} w(\underline{x}_n) \right].$$

Substituting the above equalities in (4.1.2) yields

$$v = \mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_\Gamma}) \right].$$

Thus strategy $(Q^\infty,\underline{\tau}_\Gamma)$ is optimal. □

In order to make a more thorough investigation of the existence of optimal strategies we introduce the following notions.

4.2. DEFINITION. P *conserves* v *if* $c_P = v - Pv$. *Strategy* $(R,\underline{\tau})$, *where* $R = (P_0,P_1,\ldots)$, *conserves* v *if* $i \in E_m$ *implies* $c_{P_m}(i) = v(i) - P_m v(i)$, *where* $E_m := \{j : \mathbb{P}_{\ell,R} [\underline{x}_m = j, \underline{\tau} > m] > 0$ *for some* $\ell \in E\}$.

When the policy maker (or gambler or manager) chooses decision rule P at time 0 and proceeds optimally thereafter then the expectation of his earnings is c_P+Pv. It is clear that this can not be larger than the maximum of the expected return, i.e. v (in mathematical terms v is c_P - superharmonic). When $c_P+Pv < v$ then the decision rule P cannot be a part of an optimal strategy. The decision maker made an irremediable mistake. Strategies not containing such mistakes are v conserving.

4.3. DEFINITION. *Strategy* $(R,\underline{\tau})$ *is thrifty if* $(R,\underline{\tau})$ *is* v *conserving and* $\mathbb{E}_R\, r(\underline{x}_\tau) = \mathbb{E}_R\, v(\underline{x}_\tau)$.

In a state where $r(i) < v(i)$ it is suboptimal to choose the stopping decision, because stopping gives $r(i)$ and one might expect to receive $v(i)$. So a strategy for which the policy R does not make irremediable decisions and for which the stopping time $\underline{\tau}$ does not give irremediable losses is called thrifty. Intuitively it is clear that an optimal strategy must have this property. As we shall show the following converse is true. If $(R,\underline{\tau})$ is thrifty and $\underline{\tau}$ is bounded then $(R,\underline{\tau})$ is optimal. In the case of an unbounded stopping time $\underline{\tau}$ we also need that the amount we actually receive in the time period up to time N has limit v as N tends to infinity. One might say that here the "actually received" and the "promised" earnings equalize. This property can be formalized in the following way.

4.4. DEFINITION. *Strategy* $(R,\underline{\tau})$ *is equalizing if* $\lim_{n\to\infty} \mathbb{E}_R\, [v(\underline{x}_n)\ \chi(\underline{\tau}>n)] = 0$.

4.5. THEOREM. *Strategy* $(R,\underline{\tau})$ *is thrifty if and only if*

$$(4.5.1)\qquad \mathbb{E}_R\, [\sum_{n=0}^{\tau-1} c(\underline{x}_n) + r(\underline{x}_\tau)] = \mathbb{E}_R\, [\sum_{n=0}^{\tau-1} w(\underline{x}_n) + v(\underline{x}_\tau)].$$

PROOF. The value function v is c_P - superharmonic and majorizes r. Hence $w_P = v-Pv \ge c_P$, $P \in \mathcal{P}$, and $v \ge r$. These inequalities imply that relation (4.5.1) is equivalent to the following relations (4.5.2) and (4.5.3) together

$$(4.5.2)\qquad \mathbb{E}_R\, [\sum_{n=0}^{\tau-1} (w(\underline{x}_n)-c(\underline{x}_n))] = 0$$

$$(4.5.3)\qquad \mathbb{E}_R\, [v(\underline{x}_\tau) - r(\underline{x}_\tau)] = 0.$$

Relation (4.5.2) is equivalent to the assertion that $(R,\underline{\tau})$ conserves v and the theorem is proved. □

4.6. THEOREM. *Strategy* $(R,\underline{\tau})$ *is optimal if and only if* $(R,\underline{\tau})$ *is thrifty and equalizing.*

PROOF. For N = 1,2,... let $\underline{\tau}_N$ denote $\underline{\tau} \wedge N$. Given any strategy R we have

$$(4.6.1)\qquad \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})] = \lim_{N\to\infty} \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})\ \chi(\underline{\tau}\le N)].$$

We rewrite the right-hand side of this equality. Using relation (2.19.1) with function v instead of r, i.e.

$$v = \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N-1} w(\underline{x}_n) + v(\underline{x}_{\underline{\tau}_N})],$$

and using the relation

$$\mathbb{E}_R\,[v(\underline{x}_{\underline{\tau}_N})] = \mathbb{E}_R\,[v(\underline{x}_{\underline{\tau}})\ \chi(\underline{\tau}\le N)] + \mathbb{E}_R\,[v(\underline{x}_N)\ \chi(\underline{\tau}> N)]$$

we obtain for the second part of equality (4.6.1)

$$\lim_{N\to\infty} \{v - \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}_N-1} (w(\underline{x}_n)-c(\underline{x}_n))] +$$

$$- \mathbb{E}_R\,[(v(\underline{x}_{\underline{\tau}})-r(\underline{x}_{\underline{\tau}}))\ \chi(\underline{\tau}\le N)] - \mathbb{E}_R\,[v(\underline{x}_N)\ \chi(\underline{\tau}> N)]\}.$$

This limit equals

$$(4.6.2)\qquad v - \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} (w(\underline{x}_n)-c(\underline{x}_n))] - \mathbb{E}_R\,[v(\underline{x}_{\underline{\tau}})-r(\underline{x}_{\underline{\tau}})] +$$

$$- \lim_{N\to\infty} \mathbb{E}_R\,[v(\underline{x}_N)\ \chi(\underline{\tau}> N)].$$

If $(R,\underline{\tau})$ is thrifty then the second and third term of expression (4.6.2) are zero. If $(R,\underline{\tau})$ is in addition equalizing then also the fourth term of (4.6.2) is zero and the expression equals v. Hence $(R,\underline{\tau})$ is optimal. To prove the converse we note that according to the theorems 3.1 and 2.17

$$\lim_{N\to\infty} \mathbb{E}_R\,[v(\underline{x}_N)\ \chi(\underline{\tau}> N)] \ge - \lim_{N\to\infty} \mathbb{E}_R\,[v^-(\underline{x}_N)] = 0.$$

This means that the fourth term in relation (4.6.2) is nonnegative. It is easy to verify that the second and third term are also nonnegative. If $(R,\underline{\tau})$ is optimal then the sum of the last three terms is zero and consequently they are all three zero. Hence $(R,\underline{\tau})$ is thrifty and equalizing. □

Using the above theorem it is rather easy to deduce sufficient conditions for $(Q^{\infty},\underline{\tau}_{\Gamma})$ as introduced in relation (4.1.1) to be an optimal strategy. These are given in the next two theorems.

4.7. THEOREM. *Strategy* $(Q^{\infty},\underline{\tau}_{\Gamma})$ *is optimal if and only if* $\lim_{N\to\infty} \tilde{Q}^N v = 0$, *with* $\tilde{Q}$ *the restriction of* Q *to the complement of* Γ, *i.e.*

$$(4.7.1) \qquad \tilde{q}(i,j) := \begin{cases} q(i,j) & \textit{if } i \notin \Gamma \textit{ and } j \notin \Gamma \\ 0 & \textit{otherwise.} \end{cases}$$

PROOF. From (4.1.1) we see that Q conserves v outside of Γ. Thus $(Q^{\infty},\underline{\tau}_{\Gamma})$ conserves v. Since v = r on Γ it follows then that $(Q^{\infty},\underline{\tau}_{\Gamma})$ is thrifty. Hence $(Q^{\infty},\underline{\tau}_{\Gamma})$ is optimal if and only if $(Q^{\infty},\underline{\tau}_{\Gamma})$ is equalizing.

From the definition of entry time $\underline{\tau}_{\Gamma}$ ($\underline{\tau}_{\Gamma}$ is the least $n \geq 0$ if any with $\underline{x}_n \in \Gamma$, and $\underline{\tau}_{\Gamma} = \infty$ if none) and relation (4.7.1) we have for $N = 1,2,\ldots$

$$\mathbb{E}_Q \, [v(\underline{x}_N) \, \chi(\underline{\tau}>N)] = \tilde{Q}^N v.$$

From this relation the theorem is obvious. □

4.8. THEOREM. *Each of the following two conditions ensures that* $(Q^{\infty},\underline{\tau}_{\Gamma})$ *is optimal*

a. *the value function* v *is bounded and* $\mathbb{P}_{i,Q}\,[\underline{\tau}_{\Gamma} < \infty] = 1$ *for all* $i \notin \Gamma$

b. *the value function* v *is bounded and* Q *is absorbing.*

PROOF. According to theorem 4.7 it is sufficient to show that $\lim_{N\to\infty} \tilde{Q}^N v = 0$. Since v is bounded it is sufficient to show that $\lim_{N\to\infty} \tilde{Q}^N e = 0$.

b. Because Q is absorbing we have

$$\lim_{N\to\infty} \tilde{Q}^N e \leq \lim_{N\to\infty} Q^N e = 0.$$

a. From

$$\mathbb{P}_Q\,[\underline{\tau}_\Gamma > N] = \tilde{Q}^N e,$$

the second part of condition a and $\mathbb{P}_{i,Q}\,[\underline{\tau}_\Gamma=0] = 1$ if $i \in \Gamma$, we find

$$\lim_{N\to\infty} \tilde{Q}^N e = \lim_{N\to\infty} \{e - \mathbb{P}_Q\,[\underline{\tau}_\Gamma \le N]\} = 0. \ \square$$

In most cases it is difficult to determine the value function. Sometimes one can make a guess at the optimal strategy and one is able to compute the expected reward for that strategy. In such a case one can use the following theorem with f the expected return. If the conditions of the theorem are satisfied then the theorem guarantees that the guess was correct and one knows the optimal strategy and the value function.

4.9. THEOREM. *Suppose* f *is a* c_P*-superharmonic function that majorizes* r *and has property anne. Suppose* $Q \in P$ *is such that*

$$c_Q(i) + Qf(i) = f(i) \text{ if } i \notin \Gamma := \{i : r(i) = f(i)\}$$

and $\lim_{N\to\infty} \tilde{Q}^N f = 0$ *with* $\tilde{Q}$ *the restriction of* Q *to the complement of* Γ (see 4.7.1). *Then* $f = v$ *and* $(Q^\infty, \underline{\tau}_\Gamma)$ *is an optimal strategy.*

PROOF. According to the theorems 3.8 and 3.1 we have that $f \ge v$. Similarly as in the proof of theorem 4.7 one can show that

$$f = \mathbb{E}_Q\,[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_\Gamma})]. \tag{4.9.1}$$

Hence $f \le v$, since v is not less than the expected return of $(Q^\infty, \underline{\tau}_\Gamma)$. We conclude that $f = v$ and then it follows with (4.9.1) that $(Q^\infty, \underline{\tau}_\Gamma)$ is optimal. $\square$

Throughout the sections 3 and 4 we assumed that c_P is a charge structure and $\mathbb{E}_R\,|r(\underline{x}_{\underline{\tau}})| < \infty$ for all R and all $\underline{\tau}$. These assumptions are superfluous in the next theorem because they follow from the assumptions of the theorem.

4.10. THEOREM. *Suppose* P *is compact and* c_P *is continuous. If there exists a function* $y \geq |r|$ *such that*

(4.10.1) $\qquad |c_P| + Py \leq y,$

(4.10.2) $\qquad \lim_{N\to\infty} P^N y = 0$ *for all* $P \in P$ *and*

(4.10.3) $\qquad \lim_{P\to P_0} Py = P_0 y$ *for all* $P_0 \in P$

then c_P *is a charge structure and there is a strategy* $(Q^\infty, \underline{\tau}_\Gamma)$ *as in* (4.1.1) *which is optimal.*

In the proof of this theorem we need the following result: if $0 \leq x \leq y$ then (4.10.3) implies $\lim_{P\to P_0} Px = P_0 x$. In order to prove this we first state three lemmas.

4.11. LEMMA. *If* $a_n(i) \geq 0$, $i=1,2,\ldots$ *and* $n=1,2,\ldots$, $\lim_{n\to\infty} a_n(i) = a_\infty(i)$, $i=1,2,\ldots$ *and* $\lim_{n\to\infty} \sum_{i=1}^\infty a_n(i) = \sum_{i=1}^\infty a_\infty(i) < \infty$ *then*

$$\lim_{n\to\infty} \sum_{i\in B} a_n(i) = \sum_{i\in B} a_\infty(i)$$

uniformly for each subset B *of the positive integers.*

PROOF. The assertion of this lemma is equivalent to

(4.11.1) $$\lim_{n\to\infty} \sum_{i=1}^\infty |a_n(i)-a_\infty(i)| = 0.$$

Suppose (4.11.1) is false. Then

(4.11.2) $$c := \limsup_{n\to\infty} \sum_{i=1}^\infty |a_n(i)-a_\infty(i)| > 0.$$

Take N such that

$$\sum_{i=N}^\infty a_\infty(i) < \frac{c}{3}.$$

Since

$$\lim_{n\to\infty} \sum_{i=N}^\infty a_n(i) = \lim_{n\to\infty} [\sum_{i=1}^\infty a_n(i) - \sum_{i=1}^{N-1} a_n(i)] = \sum_{i=N}^\infty a_\infty(i),$$

there exists an n_0 such that for $n \geq n_0$

$$\sum_{i=N}^{\infty} a_n(i) < \frac{c}{3}$$

and

$$\sum_{i=1}^{N-1} |a_n(i)-a_\infty(i)| < \frac{c}{3} .$$

Hence, for $n \geq n_0$,

$$\sum_{i=1}^{\infty} |a_n(i)-a_\infty(i)| \leq \sum_{i=1}^{N-1} |a_n(i)-a_\infty(i)| + \sum_{i=N}^{\infty} (a_n(i)+a_\infty(i)) < c.$$

This is in contradiction with (4.11.2). □

4.12. LEMMA. *If* $0 \leq b_n(i) \leq a_n(i)$, $i=1,2,\dots$ and $n=1,2,\dots$; $a_\infty(i) := \lim_{n\to\infty} a_n(i)$ *and* $b_\infty(i) := \lim_{n\to\infty} b_n(i)$;
and

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} a_n(i) = \sum_{i=1}^{\infty} a_\infty(i) < \infty$$

then

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} b_n(i) = \sum_{i=1}^{\infty} b_\infty(i).$$

PROOF. Given any $\varepsilon > 0$, let N be such that $\sum_{i=N}^{\infty} a_\infty(i) \leq \frac{1}{2}\varepsilon$. From lemma 4.11 it follows that there is an M such that $\sum_{i=N}^{\infty} a_n(i) \leq \varepsilon$ for $n \geq M$. Since $0 \leq b_n(i) \leq a_n(i)$ we have then

(4.12.1) $$\sum_{i=N}^{\infty} b_n(i) \leq \varepsilon \text{ for } n = M, M+1, \dots, \infty.$$

Since $\lim_{n\to\infty} \sum_{i=1}^{N} b_n(i) = \sum_{i=1}^{N} b_\infty(i)$ the relation (4.12.1) implies that the limitpoints of $\{\sum_{i=1}^{\infty} b_n(i)\}_{n=1}^{\infty}$ differ at most ε from $\sum_{i=1}^{\infty} b_\infty(i)$.
Hence

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} b_n(i) = \sum_{i=1}^{\infty} b_\infty(i). \; \square$$

The following lemma is for future reference stated slightly more general than we need here.

4.13. LEMMA. *If* $0 \le x_P \le y_P$, x_P, y_P *continuous in* P *and* $\lim_{P \to P_\infty} Py_P = P_\infty y_{P_\infty}$ *then* $\lim_{P \to P_\infty} Px_P = P_\infty x_{P_\infty}$.

PROOF. It is sufficient to prove that $\lim_{n\to\infty} \sum_j p_n(i,j)\, x_{P_n}(j) = \sum_j p_\infty(i,j)\, x_{P_\infty}(j)$ for an arbitrary state i and an arbitrary sequence $P_n \to P_\infty$. Now substitute $b_n(j) := p_n(i,j)\, x_{P_n}(j)$ and $a_n(j) := p_n(i,j)\, y_{P_n}(j)$, $j=1,2,\dots$ and $n=1,2,\dots,\infty$ in lemma 4.12. □

The above lemma is a discrete analogue of theorem 1 in [Pratt].

PROOF OF THEOREM 4.10. From relation (4.10.1) we have that the nonnegative function y is $|c_P|$ - superharmonic. Hence by theorem 3.8

$$\mathbb{E}_R \sum_{n=0}^{\infty} |c(\underline{x}_n)| \le y \text{ for all } R$$

and so c_P is a charge structure. Since also $y \ge |r|$, it follows by (2.21.1)

$$|v| \le \sup_{R,\underline{\tau}} \mathbb{E}_R \left[\sum_{n=0}^{\tau-1} |c(\underline{x}_n)| + |r(\underline{x}_{\underline{\tau}})|\right] \le y.$$

Now according to lemma 4.13 relation (4.10.3) implies $\lim_{P \to P_0} Pv^+ = P_0 v^+$. In view of theorem 3.6 we then have that

$$v = r \vee \max_P (c_P + Pv).$$

Consequently strategy $(Q^\infty, \underline{\tau}_\Gamma)$ as in (4.1.1) exists. Moreover from (4.10.2) $\lim_{N\to\infty} Q^N|v| \le \lim_{N\to\infty} Q^N y = 0$ and according to theorem 4.7 the strategy $(Q^\infty, \underline{\tau}_\Gamma)$ is optimal. □

In section 5 we need the following corollary of theorem 4.10.

4.14. COROLLARY. *Suppose P is compact and* c_P *is continuous. If there exists a function* $y \ge 0$ *such that relations* (4.10.1), (4.10.2) *and* (4.10.3) *hold, then there exists a stationary policy* Q^∞ *such that*

$$\mathbb{E}_Q \left[\sum_{n=0}^{\infty} c(\underline{x}_n)\right] = \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\infty} c(\underline{x}_n)\right]. \tag{4.14.1}$$

PROOF. In order to make it possible to apply theorem 4.10 we introduce a reward function r such that $r := \sup_R \mathbb{E}_R [\sum_{n=0}^{\infty} c(\underline{x}_n)] - e$. By theorem 4.10 the cost structure is a charge structure. Given any policy $(R,\underline{\tau})$ we have according to theorem 2.20 with v the value function of the optimal control problem

$$v \geq \mathbb{E}_R \left[\sum_{n=0}^{\tau-1} c(\underline{x}_n) + v(\underline{x}_{\tau})\right] \geq$$

$$\geq \mathbb{E}_R \left[\sum_{n=0}^{\tau-1} c(\underline{x}_n) + r(\underline{x}_{\tau})\right] + \mathbb{E}_R\, e(\underline{x}_{\tau}).$$

Hence, since Markov time $\underline{\tau} = \infty$ is allowed

$$v = \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\infty} c(\underline{x}_n)\right].$$

Moreover, since $\Gamma = \emptyset$ it follows that $\underline{\tau}_{\Gamma} = \infty$ and according to theorem 4.10 Q^{∞} as in (4.1.1) satisfies (4.14.1). □

5. SEMI-MARKOV DECISION PROCESSES WITH AVERAGE RETURN CRITERION

In this section we are concerned with sequential decision processes for which the times between transitions are random. Earlier (in section 1) if at time t the system had been observed in state i and action $p(i,.)$ had been chosen, the system transferred to a state j at time t+1 with probability $p(i,j)$. Now this transition takes place at random time $t+\underline{\tau}$, where the random time $\underline{\tau}$ only depends on i,j and P and not on the past history of the process. Let $F_P(.|i,j)$ where P is an element of $\mathcal{P}$ with i^{th} row $p(i,.)$, denote the distribution of the random time $\underline{\tau}$. At time $t+\underline{\tau}$ again an action $p(j,.) \in \mathcal{P}(j)$ has to be chosen, etc.

When using a stationary policy this decision process is a semi-Markov process.

Let $\underline{x}_n$, n=0,1,..., denote the state after the n^{th} transition. We write $c_P(i)$ for the expectation of the cost incurred between the n^{th} and the $(n+1)^{th}$ transition when $\underline{x}_n = i$ and the action taken after the n^{th} transition is the i^{th} row of P. We obtain for the expected duration of this transition interval

$$t_P(i) = \sum_j p(i,j) \int_0^\infty y \; dF_P(y|i,j).$$

It is assumed that for some positive constant a,

$$a \le t_P(i) < \infty, \text{ for all i and all P}$$

(cf. [Ross (1970), condition 1, p. 157]).

The optimality criterion we use in this section is the long-run average return per unit time. Actually we take

$$(5.0.1) \qquad \limsup_{N\to\infty} \frac{\mathbb{E}_R \sum_{n=0}^{N} c(\underline{x}_n)}{\mathbb{E}_R \sum_{n=0}^{N} t(\underline{x}_n)} .$$

This is the largest limit point as $N \to \infty$ of the expected cost over the first N + 1 transition intervals divided by the expected duration of the first N + 1 transition intervals (see [Ross (1970), p. 159] for a discussion of this criterion). The question we are mainly concerned with in this section is the question whether there exists an optimal policy. We give conditions that guarantee the existence of a stationary optimal policy. In our opinion these conditions are easy to verify. To illustrate this we solve a waiting line problem.

5.1. THEOREM. *Suppose* $\mathcal{P}$ *is compact,* c_P *is continuous and* $p(i,E) = 1$ *for all* i *and* P. *If there is some state* i_0 *and a function* $y \geq 0$ *such that*

$$(5.1.1) \qquad |c_P| + t_P + \tilde{P}y \leq y,$$

$$(5.1.2) \qquad \lim_{N\to\infty} \tilde{P}^N y = 0 \text{ for all } P \in \mathcal{P} \text{ and}$$

$$(5.1.3) \qquad \lim_{P\to P_0} \tilde{P}y = \tilde{P}_0 y \text{ for all } P_0 \in \mathcal{P}$$

where $\tilde{P}$ *denotes the column-restriction of* P *to* $E\backslash\{i_0\}$ (see 2.7). *Then there exists a stationary optimal policy.*

For the proof of this theorem we have to establish several results which are interesting on their own and will be given as lemmas. In 5.3 to 5.7 the conditions of 5.1 are assumed to hold.

5.2. LEMMA. *If for* $P \in \mathcal{P}$, $\sum_{n=0}^{\infty} \tilde{P}^n P\, |c_P|(i_0) < \infty$ *or* $\sum_{n=0}^{\infty} \tilde{P}^n P\, t_P(i_0) < \infty$ *then*

$$\lim_{N\to\infty} \frac{\mathbb{E}_{i_0,P} \sum_{n=1}^{N} |c(\underline{x}_n)|}{\mathbb{E}_{i_0,P} \sum_{n=1}^{N} t(\underline{x}_n)} \text{ exists and equals } \frac{\sum_{n=0}^{\infty} \tilde{P}^n P\, |c_P|(i_0)}{\sum_{n=0}^{\infty} \tilde{P}^n P\, t_P(i_0)}.$$

PROOF. Let $f \geq 0$. Since $\tilde{P}^k Pf(i_0)$ is the restricted expectation of $f(\underline{x}_{k+1})$ when visits to state i_0 at times $1,2,\ldots k$ are excluded; we find by applying the "last exit decomposition" of state i_0 (cf. [Chung. p. 46])

$$P^{n+1}f(i_0) = \sum_{k=0}^{n} p^k(i_0,i_0)\, \tilde{P}^{n-k}P\, f(i_0).$$

Summing over $n = 0$ to N and changing the order of summation gives

$$(5.2.1) \qquad \sum_{n=0}^{N} P^{n+1}\, f(i_0) = \sum_{k=0}^{N} p^k(i_0,i_0) \sum_{m=0}^{N-k} \tilde{P}^m P\, f(i_0).$$

Since $0 \leq p^k(i_0,i_0) \leq 1$ we have, whether $\sum_{k=0}^{\infty} p^k(i_0,i_0)$ converges or not, that $\lim_{N\to\infty} p^N(i_0,i_0)(\sum_{k=0}^N p^k(i_0,i_0))^{-1} = 0$. As an application of the regularity property of the Nörlund-means (see [Hardy, p. 64]) it follows from (5.2.1) that

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N} P^{n+1} f(i_0)}{\sum_{k=0}^{N} p^k(i_0,i_0)} = \sum_{m=0}^{\infty} \tilde{P}^m P f(i_0). \tag{5.2.2}$$

To complete the proof we write

$$\frac{\mathbb{E}_{i_0,P} \sum_{n=1}^{N} |c(\underline{x}_n)|}{\mathbb{E}_{i_0,P} \sum_{n=1}^{N} t(\underline{x}_n)} = \frac{\sum_{n=0}^{N-1} P^{n+1} |c_P|(i_0)}{\sum_{k=0}^{N-1} p^k(i_0,i_0)} \cdot \frac{\sum_{k=0}^{N-1} p^k(i_0,i_0)}{\sum_{n=0}^{N-1} P^{n+1} t_P(i_0)} .$$

Next we apply relation (5.2.2) once with $f = |c_P|$ and once with $f = t_P$. □

The above lemma is called a mean ergodic theorem. It says that the average expected absolute cost per unit time when starting in state i_0 equals the expected absolute cost divided by the expected length of the time until the first return to state zero. In most proofs of this lemma it is assumed that both expectations are finite.

5.3. LEMMA. *For each stationary policy the corresponding Markov chain is positive recurrent.*

PROOF. From $t_P \geq ae$ for some $a > 0$ and (5.1.1) it follows that

$$ae + \tilde{P}y \leq y.$$

Hence for $y^* := a^{-1}y$

$$e + \tilde{P}y^* \leq y^*.$$

So according to 2.7, the Markov chain with matrix of transition probabilities P is a positive recurrent chain. □

5.4. LEMMA. *For* $f \geq 0$

$$(5.4.1) \qquad \sum_{m=0}^{\infty} \tilde{P}^m P\, f(i_0) = \sum_{m=0}^{\infty} \tilde{P}^m f(i_0).$$

Moreover, c_P *and* t_P *are charge structures with respect to* $\tilde{\mathcal{P}} = \{\tilde{P} : P \in \mathcal{P}\}$.

PROOF. By the definition of $\tilde{P}$ we have

$$(5.4.2) \qquad \sum_{m=0}^{\infty} \tilde{P}^m P\, f = \sum_{m=0}^{\infty} \tilde{P}^{m+1} f + \sum_{m=0}^{\infty} \sum_{j} \tilde{p}^m(.,j)\, p(j,i_0)\, f(i_0).$$

As $p(i,E) = 1$ for all $i \in E$, we can write for the second term on the right-hand side $\sum_{m=0}^{\infty} \tilde{P}^m(e-\tilde{P}e)\, f(i_0)$. According to lemma 5.3 and 2.7 we have that $\sum_{m=0}^{\infty} \tilde{P}^m\, e < \infty$ and hence this term equals $f(i_0)$. Herewith relation (5.4.1) is proved.

Similar to theorem 4.10 the second assertion follows directly from relation (5.1.1). □

Define

$$(5.4.3) \qquad g_0 := \sup_{\mathcal{P}} \frac{\sum_{n=0}^{\infty} \tilde{P}^n\, c_P(i_0)}{\sum_{n=0}^{\infty} \tilde{P}^n\, t_P(i_0)}\,.$$

Then in view of the lemmas 5.2 (by writing $c_P = c_P^+ - c_P^-$) and 5.4 we have that g_0 is the supremum of the long-run average return per unit time over the *stationary* policies when the system starts in state i_0.

As in theorem 2.6 it follows from (5.1.1) that

$$(5.4.4) \qquad \sum_{n=0}^{\infty} \tilde{P}^n |c_P| \leq y \text{ and } \sum_{n=0}^{\infty} \tilde{P}^n\, t_P \leq y \text{ for all } P.$$

Since $t_P \geq ae$ we have that $\sum_{n=0}^{\infty} \tilde{P}^n\, t_P \geq ae$. Consequently $0 \leq g_0 \leq a^{-1} y(i_0) < \infty$. Define

$$(5.4.5) \qquad v := \sup_{\mathcal{P}} \sum_{n=0}^{\infty} \tilde{P}^n [c_P - g_0 t_P].$$

It is easy to verify that

$$(5.4.6) \qquad v(i_0) = 0.$$

From (5.4.4) it follows that

$$\sum_{n=0}^{\infty} \tilde{P}^n[|c_P|+|g_0|t_P] \le (|g_0|+1)y \text{ for all } P.$$

Hence, for $y^* := (|g_0|+1)y$,

$$(5.4.7) \qquad |c_P - g_0 t_P| + \tilde{P}y^* \le y^* \text{ for all } P.$$

It is rather straightforward to verify that relations (5.1.2), (5.1.3) and (5.4.7) imply that the conditions of corollary 4.14 are satisfied. Together with (5.4.6) the corollary 4.14 implies the existence of a policy Q^∞ with

$$(5.4.8) \qquad \sum_{n=0}^{\infty} \tilde{Q}^n[c_Q - g_0 t_Q](i_0) = 0.$$

In view of (5.4.3) we now have that Q^∞ is average-optimal in the class of stationary policies if we start in state i_0.

5.5. LEMMA. *There exists a stationary policy* Q^∞ *such that* Q^∞ *is optimal with respect to the average return criterion in the class of all stationary policies.*

PROOF. It follows directly from (5.4.8) and the definition of g_0 that Q^∞ is optimal in the class of all stationary policies when the system starts in state i_0. Since for each $P \in \mathcal{P}$ the state i_0 can be reached from each state we obtain that the associated Markov chain does not have disjoint closed sets. This implies, as is well-known, that the average expected return per unit time does not depend on the starting state from which the lemma follows. □

The rest of the proof of theorem 5.1 consists of proving that the policy Q^∞ is average-optimal in the class of all policies. The essential part is to show that the scalar g_0 in combination with the function v is a solution of the optimality equation for the average return criterion which satisfies an auxiliary condition.

5.6. LEMMA. *For* (g_0,v) *we have*

$$(5.6.1)\qquad v = \sup_P (c_P - g_0 t_P + Pv).$$

Thus (g_0,v) *is a solution of the optimality equation for the average return criterion. Moreover,* Q *satisfies*

$$(5.6.2)\qquad v = c_Q - g_0 t_Q + Qv.$$

PROOF. We found that v is the value function of the optimal control problem with cost function $c_P - g_0 t_P$. Since $r = v - e$ (cf. 4.14) it follows from (3.5.1) that

$$v = \sup_P (c_P - g_0 t_P + \tilde{P}v).$$

In view of relation (5.4.6) we may write Pv instead of $\tilde{P}v$ and hence relation (5.6.1) follows. For Q^∞, the optimal policy for the control problem, we have

$$v = \sum_{n=0}^{\infty} \tilde{Q}^n (c_Q - g_0 t_Q).$$

Hence

$$v = c_Q - g_0 t_Q + \tilde{Q}v = c_Q - g_0 t_Q + Qv. \quad \square$$

5.7. LEMMA. *Let* f *be such that* $f(i_0) = 1$ *and* $f(i) = 0$ *for* $i \neq i_0$ *then*

$$(5.7.1)\qquad \lim_{N\to\infty} \mathbb{E}_{i,R} |v(\underline{x}_N)| \,\{\mathbb{E}_{i,R} \sum_{n=0}^{N} f(\underline{x}_n)\}^{-1} = 0 \text{ for all } i \text{ and all } R.$$

PROOF. We first show that for each $i \in E$

$$(5.7.2)\qquad \lim_{N\to\infty} \tilde{P}_0 \tilde{P}_1 \cdots \tilde{P}_N |v|(i) = 0$$

uniformly in $R = (P_0, P_1, \ldots)$. Define

$$x_0 = y^* \text{ and } x_{n+1} = \sup_P \tilde{P}x_n \text{ for } n = 0,1,\ldots .$$

Since by relation (5.1.1) and the definition of y^*

$$\tilde{P}y^* \le y^* - (|c_P| + |g_0|t_P) \le y^* \text{ for all } P \in \mathcal{P},$$

it follows that $x_1 \le x_0$.

Now suppose $x_n \le x_{n-1}$ then $\tilde{P}x_n \le \tilde{P}x_{n-1}$ for all $P \in \mathcal{P}$ and hence $x_{n+1} = \sup_{\mathcal{P}} \tilde{P}x_n \le \sup_{\mathcal{P}} \tilde{P}x_{n-1} = x_n$. Thus by induction x_n, $n=0,1,\ldots$, is a decreasing sequence of functions. Consequently $x := \lim_{n\to\infty} x_n$ exists. It is easy to see that $0 \le x_n \le y^*$, $n=0,1,\ldots$. Using dominated convergence we find $x \ge \tilde{P}x$ for all $P \in \mathcal{P}$. Thus

$$x \ge \sup_{\mathcal{P}} \tilde{P}x.$$

Next we prove the reverse inequality. Let P_n be such that

$$x_{n+1} \le \tilde{P}_n x_n + n^{-1}e.$$

Now choose a converging subsequence of $\tilde{P}_n$, say $\tilde{P}_{n_k} \to \tilde{P}_0$ as $k \to \infty$ ($\mathcal{P}$ is compact by the assumptions of theorem 5.1). Since $\lim_{k\to\infty} \tilde{P}_{n_k} y^* = \tilde{P}_0 y^*$ and $0 \le x_n \le y^*$, according to lemma 4.13, we have

$$x \le \tilde{P}_0 x.$$

Consequently

$$x = \max_{\mathcal{P}} \tilde{P}x = \tilde{P}_0 x. \tag{5.7.3}$$

Relation (5.7.3) implies

$$x = \lim_{N\to\infty} \tilde{P}_0^N x \le \lim_{N\to\infty} \tilde{P}_0^N y^* = 0,$$

by relation (5.1.2). By induction it is straightforward to establish that (use $|v| \le y^*$ and the inequality at the top of this page)

$$\tilde{P}_0 \tilde{P}_1 \ldots \tilde{P}_N |v| \le x_{N+1} \tag{5.7.4}$$

for all N and all $R = (P_0, P_1, \ldots)$. But then assertion (5.7.2) follows.

Using again the "last exist decomposition" of state i_0 (see lemma 5.2) and recalling that $v(i_0) = 0$, we find with (5.7.4)

$$P_0 \ldots P_N |v|(i) = \sum_{n=0}^{N} [P_0 \ldots P_{n-1}](i, i_0)\, [\tilde{P}_n \ldots \tilde{P}_N |v|](i_0) \leq$$

$$\leq \sum_{n=0}^{N} P_0 \ldots P_{n-1}\, (i, i_0)\, x_{N-n}(i_0).$$

As a second application of the regularity property of the Nörlund-means it follows then

$$\lim_{N\to\infty} \frac{P_0 \ldots P_N |v|(i)}{\sum_{n=0}^{N} P_0 \ldots P_{n-1}(i, i_0)} = 0,$$

which is relation (5.7.1) in a different notation. □

PROOF OF THEOREM 5.1. From relation (5.6.1) we have

$$c_P - g_0 t_P + Pv \leq v \text{ for all } P.$$

Iterating this inequality we obtain

$$\sum_{n=0}^{N} P_0 \ldots P_{n-1}\, (c_{P_n} - g_0 t_{P_n}) + P_0 \ldots P_N v \leq v.$$

By rewriting this we find

$$\frac{\sum_{n=0}^{N} P_0 \ldots P_{n-1}\, c_{P_n}(i)}{\sum_{n=0}^{N} P_0 \ldots P_{n-1}\, t_{P_n}(i)} \leq$$

$$\leq g_0 + \left\{ \frac{v(i)}{\sum_{n=0}^{N} P_0 \ldots P_{n-1}\, t_{P_n}(i)} - \frac{P_0 \ldots P_N v(i)}{\sum_{n=0}^{N} P_0 \ldots P_{n-1}\, t_{P_n}(i)} \right\}$$

for all $i \in E$.

In order to prove that the largest limit point as $N \to \infty$ of the left-hand side does not increase g_0 we show that the term between brackets has limit zero. Indeed, since $\sum_{n=0}^{N} P_0 \cdots P_{n-1}\, t_{P_n}(i) \to \infty$ as $N \to \infty$ the first term tends to zero. By lemma 5.7 the second term converges to zero. Consequently $g_0 e$ is an upper bound of the average expected return per unit time. Moreover, since $c_Q - g_0 t_Q + Qv = v$ it can be proved in a similar way that the average expected return corresponding to policy Q^∞ actually equals $g_0 e$. The latter was already shown in lemma 5.5. □

5.8. WAITING LINE MODEL WITH CONTROLLABLE INPUT

The idea of "reduction of queues through the use of price" comes from [Leeman]. Here we shall restrict ourselves to show the applicability of our conditions (5.1.1), (5.1.2) and (5.1.3). A more detailed study of this type of control problems can be found in [Low].

Assume that the arrival process is a Poisson process with expected number of arrivals per unit time λ_p where p denotes the service price. Thus the input process can be controlled by the service price. It seems reasonable to assume that λ_p decreases as p increases. Let us assume further that the price p lies between the bounds a and b, i.e. $a \le p \le b$. Let F be the distribution of the service time $\underline{s}$. The times at which a decision on the price has to be taken are the times a person completes service. The state at that time is the number of people the departing customer leaves behind. We assume that the service time is independent of p.

The transition probabilities corresponding to price p equal

$$(5.8.1) \qquad p(i,j) = \begin{cases} 0 & \text{for } j < i-1, \\ k_{j-i+1}(p) & \text{for } j \ge i-1, \end{cases}$$

where $k_r(p)$ denotes the probability of r people arriving during a service period, i.e.

$$(5.8.2) \qquad k_r(p) = \int_0^\infty e^{-\lambda_p s} (\lambda_p s)^r (r!)^{-1}\, dF(s).$$

For future reference we state that (5.8.2) implies

$$(5.8.3) \qquad \sum_{r=k}^{\infty} r(r-1)\cdots(r-k+1)\, k_r(p) = \lambda_p^k\, \mathbb{E}\, \underline{s}^k,$$

where it is assumed that $\mathbb{E}\,\underline{s}^k$ exists. Since $k_r(p)$, $r=0,1,\ldots$, is a continuous function of λ_p it follows directly that P is compact if λ_p is a continuous function of p.

The following assumptions are made:

$$(5.8.4) \qquad \rho^{-1} := 1 - \lambda_a \,\mathbb{E}\,\underline{s} > 0,$$

$$(5.8.5) \qquad \lambda_p \text{ is a continuous function of p for } a \le p \le b,$$

$$(5.8.6) \qquad c_P(i) \text{ is a continuous function of P for all } i \in E.$$

5.9. BOUNDED COSTS

Suppose constant d is such that $|c_P| \le de$ for all $P \in P$. In view of condition (5.1.1) we need a nonnegative function y such that

$$(5.9.1) \qquad |c_P| + t_P + \tilde{P}y \le y$$

with $\tilde{P}$ the column-restriction to $E\backslash\{i_0\}$. Since $t_P(i) = \mathbb{E}\,\underline{s} < \infty$ for all $i \in E$ and $|c_P| \le de$ it is sufficient to find a y with

$$(5.9.2) \qquad e + \tilde{P}y \le y,$$

because in that case $y^* := (d + \mathbb{E}\,\underline{s})y$ will satisfy (5.9.1).

A funtion y satisfying (5.9.2) with state 0 for i_0 is an upper bound of the expected number of transitions to the state zero (cf. 2.6 and 2.7). Hence y(i) is equal to some constant times the number of steps to the point zero (i.e. equal to i) seems a good candidate. We try y(i) = i, then for service price p and i ≥ 1

$$(5.9.3) \qquad \sum_{j\neq 0} p(i,j)j = \sum_{j=i-1}^{\infty} k_{j-i+1}(p)\, j =$$

$$= \sum_{r=0}^{\infty} k_r(p)(r+i-1) =$$

$$= i - (1 - \lambda_p \,\mathbb{E}\,\underline{s}).$$

From assumption (5.8.4) it follows that $\rho(1 - \lambda_p \mathbb{E}\, \underline{s}) \geq 1$ for all p. Hence

$$(5.9.4) \qquad 1 + \sum_{j \neq 0} p(i,j)\rho j \leq \rho i \text{ for } i \geq 1$$

and $y(0) := \sum_{j\neq 0} p(0,j)\rho j$, $y(i) := \rho i$ for $i \geq 1$ satisfies (5.9.2).

In order to verify the condition (5.1.2) we note that for $\underline{\tau}$ the busy period, i.e. the return time to $\{0\}$, the inequality $\mathbb{E}_{i,P}[\underline{\tau}] \geq i\,\mathbb{E}\,\underline{s}$ holds. Moreover, in view of (2.7.4) and Wald's equation

$$\mathbb{E}_P[\underline{\tau}] = \mathbb{E}\,\underline{s} \sum_{n=0}^{\infty} \tilde{P}^n e.$$

Hence

$$\tilde{P}^n y \leq \rho \tilde{P}^n \sum_{k=0}^{\infty} \tilde{P}^k e.$$

Since the right-hand side tends to zero as $n \to \infty$ for each $P \in \mathcal{P}$ we find that also condition (5.1.2) is true.

To check that $\tilde{P}y$ depends continuously on price p it is in view of (5.9.3) sufficient to verify that $\lambda_p \mathbb{E}\,\underline{s}$ is a continuous function of p. This is a direct consequence of assumption (5.8.5). We conclude that theorem 5.1 can be applied.

Before we treat the case of unbounded costs we state a lemma which does not depend on any previous assumption made in this section.

5.10. LEMMA. *If* $c \geq 0$ *and* $x \geq 0$ *are such that*

$$(5.10.1) \qquad c + {}_H Q\, x \leq x,$$

with ${}_H Q$ *the column-restriction to a certain subset* H, *i.e.*

$$(5.10.2) \qquad {}_H q(i,j) = \begin{cases} q(i,j) & \text{for } j \in H, \\ 0 & \text{for } j \notin H, \end{cases}$$

then

$$(5.10.3) \qquad \sum_{n=0}^{\infty} Q^n c(i) \leq x(i) + \sum_{k=1}^{\infty} \sum_{j \in H^c} q^k(i,j)\, x(j) \text{ for all } i \in E.$$

PROOF. From (5.10.1) it follows that

$$\sum_{n=0}^{N} Q^n c \le \sum_{n=0}^{N} Q^n(x - {}_HQx) \le$$

$$\le x + \sum_{n=1}^{N} Q^{n-1}(Q - {}_HQ)x - Q^N {}_HQx.$$

But $Qx - {}_HQx = Qx^*$ where $x^*(j) = x(j)$ if $j \in H^c$ and $x^*(j) = 0$ otherwise. Hence

$$\sum_{n=0}^{N} Q^n c \le x + \sum_{n=1}^{N} Q^n x^* \le x + \sum_{n=1}^{\infty} Q^n x^* \text{ for all } N.$$

This completes the proof of the lemma. □

5.11. COSTS BOUNDED BY A LINEAR FUNCTION

Suppose for some constant d we have that $|c_P(i)| \le di$ for all $i \in \{1,2,\ldots\}$ and all $P \in \mathcal{P}$. It is now sufficient to find a function $y \ge 0$ such that

$$(5.11.1) \qquad i + \sum_{j\neq 0} p(i,j)\, y(j) \le y(i) \text{ for all } i.$$

We try $y(i) = i(i+1)$, then for service price p and $i \ge 1$

$$(5.11.2) \qquad \sum_{j\neq 0} p(i,j)\, j(j+1) = \sum_{r=0}^{\infty} k_r(p)(r+i-1)(r+i).$$

Since $(r+i-1)(r+i) = r(r-1) + 2ir + i^2-i$ we find, when using (5.8.3), that the right-hand side equals $\lambda_p^2\, \mathbb{E}\, \underline{s}^2 + 2i\, \lambda_p\, \mathbb{E}\, \underline{s} - 2i + i + i^2$. Rewriting this we find

$$(5.11.3) \qquad i(i+1) - i\{2(1 - \lambda_p\, \mathbb{E}\, \underline{s}) - i^{-1}\, \lambda_p^2\, \mathbb{E}\, \underline{s}^2\}.$$

According to assumption (5.8.4) we can find an integer i_0 such that

$$(5.11.4) \qquad 2(1 - \lambda_p\, \mathbb{E}\, \underline{s}) - i^{-1}\, \lambda_p^2\, \mathbb{E}\, \underline{s}^2 \ge \rho^{-1} \text{ for } i \ge i_0.$$

Then

$$(5.11.5)\qquad i + \sum_{j\neq 0} p(i,j)\ \rho\ j(j+1) \le \rho\ i(i+1) \text{ for } i \ge i_0.$$

In order to find a function which satisfies (5.11.1) for all $i \ge 1$ we apply lemma 5.10 with

$$H^c := \{0,1,\dots,i_0-1\},\ c(i) := i,\ Q := \tilde{P},$$

i.e. the column-restriction of P to $E\backslash\{0\}$ for P an arbitrary element of $\mathcal{P}$,

$$x(i) := \begin{cases} \rho\ i(i+1) & \text{for } i \ge i_0, \\ m & \text{for } i=0,1,\dots,i_0-1, \end{cases}$$

where

$$m := \max\ \{i + \sum_{j\in H} q(i,j)\ \rho\ j(j+1)\ :\ i=0,1,\dots,i_0-1\}.$$

It can be verified that

$$c + {}_H Qx \le x \text{ on } E\backslash\{0\}.$$

Hence according to (5.10.3)

$$(5.11.6)\qquad \sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j)\ j \le x(i) + \sum_{n=1}^{\infty} \sum_{j\in H^c} \tilde{p}^n(i,j)\ x(j).$$

By relation (5.9.4) (cf. 2.6) we have $\sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j) \le \rho i$. Using this inequality it follows from (5.11.6) that there is some constant ρ^* such that

$$(5.11.7)\qquad \sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j)\ j \le \rho^*\ i(i+1) \text{ for } i \ge 1.$$

Define

$$(5.11.8)\qquad y(i) := \sup_{\mathcal{P}} \sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j)\ j.$$

It follows from theorem 13.6 that the supremum in (5.11.8) equals the supremum over all policies. According to theorem 2.22 with $\underline{\tau} = \infty$ we then have that y is superharmonic, i.e.

$$i + \sum_j \tilde{p}(i,j)\, y(j) \le y(i) \text{ for all } i \text{ and all } P. \tag{5.11.9}$$

Thus y is a function that satisfies (5.11.1). Moreover, since (5.11.7) was deduced for an arbitrary P we have by (5.11.7)

$$y(i) \le \rho^* \, i(i+1) \text{ for } i \ge 1. \tag{5.11.10}$$

To check condition (5.1.2) we note that

$$\sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j)\, j \ge \tfrac{1}{2} i(i+1),$$

since the system must pass the states i,i-1,...,1 to reach state 0 from state i and $\sum_{k=1}^{i} k = \frac{1}{2}i(i+1)$. Hence for $i \ge 1$ and some constant ρ^{**}

$$y(i) \le \rho^{**} \sum_{n=0}^{\infty} \sum_j \tilde{p}^n(i,j)\, j.$$

Since the series on the right-hand side converges we have

$$\sum_j \tilde{p}^n(i,j)\, y(j) \le \rho^{**} \sum_{k=n}^{\infty} \sum_j \tilde{p}^n(i,j)\, j$$

and this tends to zero as $n \to \infty$. Finally by (5.11.2), (5.11.3) and condition (5.8.5)

$$\sum_j \tilde{p}(i,j)\, j(j+1)$$

is a continuous function of the price p. By lemma 4.13 also

$$\sum_j \tilde{p}(i,j)\, y(j)$$

depends in a continuous way on p. Herewith the conditions (5.1.1), (5.1.2) and (5.1.3) are verified and theorem 5.1 can be applied.

5.12. REMARKS

As in section 5.11 it can be proved that for a quadratic cost function we can apply theorem 5.1 if the third moment of the service time exists.

Thus $\mathbb{E}|\underline{s}|^3 < \infty$ implies the existence of a stationary optimal policy with respect to the average return per unit time. In general it seems that in addition to the assumptions already made in this section the finiteness of the $(k+1)^{th}$ absolute moment of $\underline{s}$ implies the existence of an optimal stationary policy when the cost function is bounded by a polynomial of degree k.

Condition (5.1.1) for bounded costs and $t_P = e$, $P \in \mathcal{P}$, is equivalent to: *There is some state* i_0 *and a function* $0 \le y < \infty$ *such that for all* $P \in \mathcal{P}$

$$(5.12.1) \qquad e + \tilde{P}y \le y,$$

with $\tilde{P}$ *the column-restriction to* $E\backslash\{i_0\}$.

In the case that $\mathcal{P}$ consists of one element, say P, this condition reduces to the Foster or Liapunov function criterion of section 2. It turns out that in waiting time models when the embedded Markov chain approach is used, the Foster criterion is very useful in proving ergodicity. In many cases however one needs a weaker condition than given by [Foster]. Such conditions can be found in several places in the literature: [Moustafa, theorem 2.I], [Crabill, theorem 1], [Pakes, theorem 1 and theorem 2], [Cohen, ii, p. 25], [Kushner, theorem 7, p. 211]. The weakest form can already be found in [Moustafa], in our notation:

If for some $\varepsilon > 0$ *there exist a function* $y \ge 0$ *and a state* i_0 *such that*

$$\sum_{j=0}^{\infty} p(i,j)\, y(j) \le y(i) - \varepsilon \textit{ for } i > i_0$$

and

$$\sum_{j=0}^{\infty} p(i,j)\, y(j) < \infty \textit{ for } i \le i_0,$$

then the irreducible Markov chain is positive recurrent.

With the use of [Chung, theorem 3, p. 47] and our lemma 5.10 the above condition can be slightly weakened to:

If for some $\varepsilon > 0$ *and some finite set* H *there exists a function* $y \ge 0$ *such that*

$$\sum_{j \notin H} p(i,j)\, y(j) \le y(i) - \varepsilon \textit{ for all } i,$$

then the irreducible Markov chain is positive recurrent.

Indeed by lemma 5.10 we have for $\tilde{P}$ the column-restriction to $E\backslash\{i_0\}$ (cf. 2.7)

$$\varepsilon \sum_{n=0}^{\infty} \sum_{j} \tilde{p}^{n}(i_0,j) \le \sum_{n=0}^{\infty} \sum_{j \in H} \tilde{p}^{n}(i,j)\, y(j).$$

Since H is finite and $\sum_{n=0}^{\infty} \tilde{p}^{n}(i,j) < \infty$ for each j (cf. [Chung, p. 47]) we have that the right-hand side of this inequality is finite. Hence the expectation of the return time to $\{i_0\}$ is finite (cf. 2.7) and the chain is positive recurrent.

In lemma 5.3 we proved that condition (5.12.1) implies the Markov chain is positive recurrent for each $P \in \mathcal{P}$. Since state i_0 can be reached under each P we have (when we forget about transient states) that for each $P \in \mathcal{P}$ the Markov chain consists of one positive recurrent class. The question then is whether the converse of the above assertion is also true, i.e. if $\mathcal{P}$ is compact and each $P \in \mathcal{P}$ consists of one positive recurrent class then there exists a function y satisfying (5.12.1). [Fisher] showed by an ingenious proof that the answer is "yes" when in each state there is only a finite number of possible decisions. In general the answer is "no" which is shown by the following counterexample.

COUNTEREXAMPLE.

$$\mathcal{P} = \{P_1, P_2, \ldots, P_\infty\};$$

$$E = \{0,1,2,\ldots\};$$

$$p_k(n+1,n) = 1 \text{ for } k \in \{1,2,\ldots,\infty\} \text{ and } n \in \{0,1,\ldots\};$$

for $k \in \{1,2,\ldots\}$,

$$p_k(0,n) = \begin{cases} 2^{-n} & \text{for } 1 \le n \le k, \\ 2^{-k}\,(4^k - k)^{-1} & \text{for } k+1 \le n \le 4^k; \end{cases}$$

$$p_\infty(0,n) = 2^{-n} \text{ for all } n \in \{1,2,\ldots\}.$$

The expectation of the return time to 0 under P_k (notation $\mu_k(0,0)$) equals

$$\mu_k(0,0) = 1 + \sum_{n=1}^{\infty} p_k(0,n)\, n =$$
$$= 1 + \sum_{n=1}^{k} 2^{-n}\, n + 2^{-k}(4^k - k)^{-1} \sum_{n=k+1}^{4^k} n.$$

Since the third term on the right-hand side goes to infinity as $k \to \infty$ we have $\lim_{k\to\infty} \mu_k(0,0) \neq \mu_\infty(0,0)$ which is the sum of the first two terms after the equality sign. Finally it is easily checked that $\lim_{k\to\infty} P_k = P_\infty$ and thus P is compact. Since condition (5.12.1) should imply

$$\mu_k(0,0) \leq y \text{ for all } k \in \{0,1,\dots,\infty\}$$

we find that such a function y does not exist.

In verifying the conditions (5.1.1), (5.1.2) and (5.1.3) for the waiting line model it turned out that conditions (5.1.2) and (5.1.3) were relatively easily checked. Condition (5.1.1) seems to be the most important one. May the other two conditions be omitted in theorem 5.1? A counter-example of [Fisher and Ross] and the result of [Fisher] show that the answer is negative.

Ergodic theorems have been known for a long time in probability theory. The use of an ergodic theorem to convert a Markov decision problem with average return criterion into one with total return criterion, the author learned from [Breiman]. In [Lippman] the same technique is used. The results in this section are related to those of [Lippman]. There the state space is a Borel subset of a metric space. For the case of a countable state space our conditions are more general.

6. DISCOUNTED AND NON-DISCOUNTED DYNAMIC PROGRAMMING

In this section we return to the optimal control model of sections 3 and 4. Again it is assumed that c_P is a charge structure. In this section we focus on strategies with stopping time $\underline{\tau}$ equal to infinity. So the decision-maker is not allowed to stop the system. Hence the value function becomes

$$(6.0.1) \qquad v := \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\infty} c(\underline{x}_n) \right].$$

In order to make it possible to use results from the sections 3 and 4 we introduce a reward function r with $r := v-e$. Then v equals the value function of the optimal control problem with cost structure c_P and reward function r (cf. the proof of 4.14). Moreover, the interesting stopping times will automatically be equal to infinity. The results of this section are direct consequences of theorems in the sections 3 and 4.

As for the cases of discounted dynamic programming, positive dynamic programming and negative dynamic programming they are known (see [Blackwell (1965)], [Blackwell (1967)], [Hinderer] and [Strauch]). The other results seem to be new.

As a consequence of 3.1 and 2.17 we have that for all policies $\lim_{N\to\infty} \mathbb{E}_R\, v(\underline{x}_N)$ exists and, moreover, this limit is nonnegative.

6.1. THEOREM. *The value function* v *is a solution to Bellman's optimality equation*

$$(6.1.1) \qquad v = \sup_{P} (c_P + Pv).$$

Moreover, if P *is compact,* c_P *is upper semicontinuous and*

$$\limsup_{P\to P_0} Pv^+ \le P_0 v^+ \textit{ for all } P_0 \in P,$$

then v *satisfies*

$$(6.1.2) \qquad v = \max_{P} (c_P + Pv).$$

PROOF. Since $v > r$ relation (6.1.1) follows immediately from (3.5.1). Relation (6.1.2) is an implication of theorem 3.6. □

Next we introduce some useful terminology. The above model will be called:

discounted dynamic programming (d.d.p.) with discount factor $0 < \alpha < 1$ *if*

$$p(i,E) = \alpha \text{ for all } i \text{ and all } P;$$

positive dynamic programming (p.d.p.) if $c_P(i) \geq 0$ *for all* i *and* P;

negative dynamic programming (n.d.p.) if $c_P(i) \leq 0$ *for all* i *and* P.

6.2. THEOREM. *In d.d.p. with bounded cost structure, i.e. for some constant* b

$$|c_P(i)| \leq b \text{ for all } i \text{ and } P, \tag{6.2.1}$$

if $\mathcal{P}$ *is compact and* c_P *is upper semicontinuous then the value function* v *is the unique bounded solution of* (6.1.2).

PROOF. If α is the discountfactor then from (6.0.1) and (6.2.1) we have $|v| \leq be(1-\alpha)^{-1}$, thus v is bounded. Since $P^n e = \alpha^n e$ we have, moreover, that each $P \in \mathcal{P}$ is absorbing and the assertion is an implication of theorem 3.11 and theorem 6.1. □

We note that using a well-known result on contraction mappings the following generalization of theorem 6.2 can be proved.
In d.d.p. with bounded cost structure the value function is the unique bounded solution of (6.1.1) (see [Denardo]).

If for some $Q \in \mathcal{P}$ it holds that $c_Q + Qv = v$ then we say that Q *satisfies the optimality equation*. We remark that in view of theorem 3.6 for n.d.p. with $\mathcal{P}$ compact and c_P upper semicontinuous such a Q always exists. By the following theorem then Q^∞ is optimal.

6.3. THEOREM. *If* Q *satisfies the optimality equation then each of the following assumptions imply that policy* Q^∞ *is optimal:*

a. $\lim_{N\to\infty} Q^N v \leq 0$

b. *d.d.p. with bounded cost structure*

c. *value function* v *is nonpositive*

d. *n.d.p.*

PROOF. At the beginning of this section we noted that $\lim_{n\to\infty} \mathbb{E}_R\, v(\underline{x}_n)$ always exists and, moreover, is nonnegative. Since Q satisfies the optimality equation it follows with $v > r$ that Q in combination with stopping time $\underline{\tau}_\infty :\equiv \infty$ satisfies relation (4.1.1). According to theorem 4.7 we have that policy Q^∞ is optimal if and only if $\lim_{N\to\infty} Q^N v = 0$. Since $\lim_{N\to\infty} Q^N v \geq 0$ it follows that

(6.3.1) $$\lim_{N\to\infty} Q^N v \leq 0$$

is a criterion for the optimality of Q^∞. It is straightforward that assumptions a, c and d (in n.d.p. we have $v \leq 0$) imply relation (6.3.1).

In the proof of theorem 6.2 we showed that in d.d.p. with bounded cost structure the function v is bounded and each P is absorbing. Hence $\lim_{N\to\infty} Q^N v = 0$ which is stronger than relation (6.3.1). □

As a consequence of the above theorem we have

6.4. THEOREM. *In d.d.p. with bounded cost structure there exists an optimal stationary policy if* $\mathcal{P}$ *is compact and* c_P *is upper semicontinuous.*

PROOF. By the theorems 6.2 and 6.1 the value function v satisfies

$$v = \max_{\mathcal{P}} (c_P + Pv).$$

Since $\mathcal{P}$ has the product property it follows now that there is a Q such that $v = c_Q + Qv$. According to theorem 6.3 policy Q^∞ is optimal. □

In [Hordijk and Tijms (1972)] it is shown by means of a counterexample that the boundedness condition in the above theorem cannot be omitted.

The following notation is introduced

$$v_P := \sum_{n=0}^{\infty} P^n c_P.$$

6.5. THEOREM. *Policy* Q^∞ *is optimal if and only if* v_Q *has the property anne* (definition 3.7) *and, moreover, satisfies*

(6.5.1) $$v_Q = \max_{\mathcal{P}} (c_P + Pv_Q).$$

PROOF. Suppose Q^∞ is optimal. Then $v = v_Q = c_Q+Qv_Q$ and hence with (6.1.1) it follows that v_Q satisfies relation (6.5.1). Moreover, by the theorems 3.1 and 3.8 the function $v_Q = v$ has the property anne. To prove the converse we note that (6.5.1) implies that v_Q is a c_P - superharmonic function. If v_Q has in addition the property anne then in view of theorem 3.8 v_Q is c_P - excessive. By the definition of the value function we have $v_Q \le v$. Hence according to theorem 3.1 we conclude that $v_Q = v$ and consequently Q^∞ is optimal. □

6.6. THEOREM. *In d.d.p. with bounded cost structure and in p.d.p. the stationary policy* Q^∞ *is optimal if and only if* v_Q *satisfies relation* (6.5.1).

PROOF. This assertion follows immediately from theorem 6.5 if we show that v_Q has the property anne. Now in p.d.p. this is obvious. In order to show it for d.d.p. with bounded cost structure, let $|c_P(i)| \le b$ for all i and all P and some constant b, then $|v_Q| \le (1-\alpha)^{-1}$ be and hence $\mathbb{E}_R \, |v_Q(\underline{x}_n)| \le \alpha^n(1-\alpha)^{-1}$ be when α is the discountfactor. Thus $\lim_{n\to\infty} \mathbb{E}_R \, v_Q(\underline{x}_n) = 0$ and v_Q has the property anne. □

6.7. THEOREM. *If for some function* f *with the property anne and some policy* Q^∞ *we have that*

$$(6.7.1) \qquad \sup_P \; c_P+Pf \le f = c_Q+Qf$$

and if in addition $\lim_{n\to\infty} Q^n f \le 0$, *then* $f = v = v_Q$ *and* Q^∞ *is optimal.*

PROOF. Relation (6.7.1) implies that f is c_P - superharmonic. Since f has the property anne it follows by theorem 3.8 that f is c_P - excessive. Consequently, according to theorem 3.1 we have $v \le f$. Iterating the equality $c_Q+Qf = f$ we obtain

$$\sum_{n=0}^{N} Q^n c_Q + Q^{N+1} f = f.$$

With $\lim_{n\to\infty} Q^n f \le 0$ we find

$$v \geq v_Q = \sum_{n=0}^{\infty} Q^n c_Q \geq f \geq v.$$

Hence $f = v = v_Q$ and Q^∞ is optimal. $\square$

7. ON POTENTIALS, ABSORBING POLICIES AND CHARGE STRUCTURES

In section 2 we defined a potential w.r.t. $\mathcal{P}$. By introducing this analogue of a well-known notion a natural question is raised. If f is a potential w.r.t. P for each $P \in \mathcal{P}$ is it then true that f is a potential w.r.t. $\mathcal{P}$? Only for a particular case we are able to answer this question (theorem 7.1). Similar results for absorbing and transient policies are obtained in the theorems 7.3 and 7.4. Together with the two corollaries 7.5 and 7.6 they generalize results of [Veinott (1969)].

In this section we assume that $\mathcal{P}$ is compact.

7.1. THEOREM. *If* f *is a potential with nonnegative charge w.r.t.* P *for each* $P \in \mathcal{P}$ *then* f *is a potential w.r.t.* $\mathcal{P}$.

PROOF. Define $w_P := f - Pf$, $P \in \mathcal{P}$, then since w_P is the charge of f w.r.t. P we have that $w_P \geq 0$ for all $P \in \mathcal{P}$ and $f \geq 0$. Iterating the equality $w_P + Pf = f$ we find that $\sum_{n=0}^{N} P_0 \cdots P_{n-1} w_{P_n} + P_0 \cdots P_N f = f$ for all N, $P_0, \ldots, P_N$ and hence $\sup_R \mathbb{E}_R \sum_{n=0}^{\infty} w(\underline{x}_n) \leq f$. Consequently $w_{\mathcal{P}}$ is a charge structure and so is $-w_{\mathcal{P}}$. Let us study the n.d.p. problem with cost structure $-w_{\mathcal{P}}$, then the value function is defined as

$$v := \sup_R \mathbb{E}_R \sum_{n=0}^{\infty} -w(\underline{x}_n).$$

Since $f \geq 0$ it holds that Pf is a lower semicontinuous function and hence w_P is upper semicontinuous. According to section 6 there is a Q^∞ which is optimal.

Now since f is a potential with charge w_Q w.r.t. Q we conclude that $v = -f$. As a consequence of theorem 2.17 we obtain $\lim_{N\to\infty} \mathbb{E}_R\, v^-(\underline{x}_N) = \lim_{N\to\infty} \mathbb{E}_R\, f(\underline{x}_N) = 0$ and by theorem 2.16 f is a potential w.r.t. $\mathcal{P}$. $\square$

7.2. DEFINITION. *Policy* $R = (P_0, P_1, \ldots)$ *is absorbing if* $\lim_{N\to\infty} P_0 \cdots P_N e = 0$; *it is transient if* $\sum_{n=0}^{\infty} P_0 \cdots P_n(i,j) < \infty$ *for all* i,j.

7.3. THEOREM. *If each stationary policy is absorbing then each policy is absorbing and* e *is a potential w.r.t.* $\mathcal{P}$.

PROOF. For each $P \in \mathcal{P}$ we have that the function e is an excessive function w.r.t. P. Since in addition $\lim_{n\to\infty} P^n e = 0$ it follows that e is a potential w.r.t. P. Thus we can apply theorem 7.1 and find that e is a potential w.r.t. $\mathcal{P}$. By theorem 2.16 we obtain $\lim_{n\to\infty} \mathbb{E}_R\, e(\underline{x}_n) = 0$ for each policy R. □

It is well-known that if a stationary Markov chain is absorbing then it is transient. We do not know whether this is also true for non-stationary policies.

7.4. THEOREM. *If each stationary policy is absorbing and if* $\lim_{P\to P_0} Pe = P_0 e$ *for all* $P_0 \in \mathcal{P}$ *then each policy is transient.*

PROOF. As in the first part of the proof of lemma 5.7 it can be shown that for each $i \in E$

$$(7.4.1) \qquad \lim_{N\to\infty} \mathbb{E}_{i,R}\, e(\underline{x}_N) = 0,$$

uniformly in R. Hence for arbitrary state j there is an integer m such that

$$(7.4.2) \qquad \mathbb{E}_{j,R}\, e(\underline{x}_m) \le a,$$

for some $a < 1$ and all policies R.

Let $(P_0,P_1,\ldots)$ be an arbitrary policy, then for $w_n := e-P_n e$, $n=0,1,\ldots$, we have

$$(7.4.3) \qquad \sum_{k=1}^{m} P_{n+1}\cdots P_{n+k-1}\, w_{n+k} = e - P_n\cdots P_{n+m}\, e \text{ for all n.}$$

From (7.4.2) and (7.4.3) we find

$$(7.4.4) \qquad \sum_{k=1}^{m} P_{n+1}\cdots P_{n+k-1}\, w_{n+k}(j) \ge 1-a > 0 \text{ for all n.}$$

The probability that the system "breaks down" before time t+1 when at time t decision $p(i,.)$ is taken in state i equals $1 - \sum_j p(i,j)$. The probability that the system is in state j at time n and "breaks down" between times n+1 and m is not larger than the probability that the system "breaks down" between times n+1 and m. Hence

$$(7.4.5) \qquad \sum_{k=1}^{m} [P_0 \ldots P_n](i,j)\, [P_{n+1} \ldots P_{n+k-1}\, w_{n+k}](j) \le$$

$$\le \sum_{k=1}^{m} P_0 \ldots P_{n+k-1}\, w_{n+k}(i).$$

From (7.4.4) and (7.4.5) we obtain

$$P_0 \ldots P_n(i,j) \le (1-a)^{-1} \sum_{k=1}^{m} P_0 \ldots P_{n+k-1}\, w_{n+k}(i).$$

Consequently

$$(7.4.6) \qquad \sum_{n=0}^{\infty} P_0 \ldots P_n(i,j) \le (1-a)^{-1} \sum_{n=0}^{\infty} \sum_{k=1}^{m} P_0 \ldots P_{n+k-1}\, w_{n+k}(i).$$

Since $\sum_{n=0}^{\infty} P_0 \ldots P_{n+k-1}\, w_{n+k}(i)$ denotes the probability that the system "breaks down" after time k, we find from (7.4.6)

$$\sum_{n=0}^{\infty} P_0 \ldots P_n(i,j) \le m(1-a)^{-1} < \infty.$$

This proves the assertion. □

7.5. THEOREM. *If* E *has a finite number of states and if each stationary policy is transient then each policy is transient.*

PROOF. We use a well-known argument. If E is finite and P is transient then $\sum_{n=0}^{\infty} P^n(i,j) < \infty$ for all j and hence $\sum_{n=0}^{\infty} P^n e < \infty$. It follows that $\lim_{N\to\infty} P^N e = 0$ and thus P is absorbing. The rest of the proof follows from theorem 7.4. □

As a direct consequence we state the following theorem.

7.6. THEOREM. *If* E *has a finite number of states and if each stationary policy is transient then each bounded cost structure is a charge structure. Moreover, for each upper semicontinuous cost structure there is a stationary optimal policy (strategy).*

PROOF. According to theorem 7.5 we have that $\sum_{n=0}^{\infty} P_0 \ldots P_n(i,j) < \infty$ for all i,j. Hence $\sum_{n=0}^{\infty} P_0 \ldots P_n |c_{P_{n+1}}| < \infty$ from which the first assertion follows.

To prove the second statement we note that automatically the value function v is bounded and each P is absorbing. Thus $\lim_{N\to\infty} P^N v = 0$ and according to the theorems 6.1 and 6.3 there exists a stationary optimal policy (strategy). □

8. RECURRENCE FOR A DECISION PROCESS

In this section we generalize the notion of recurrence for one Markov chain to a collection of Markov chains. It seems to us that the extension of well-known theorems for one Markov chain to a collection of Markov chains has important implications in the theory of Markov decision processes.

The term communicating system stems from [Bather]. His paper makes it clear that similar to the minimal closed sets in a Markov chain the notion of communicating system plays a basic role in Markov decision processes especially when the average return criterion is used.

In [Hordijk (1972)] an earlier version of several theorems of this section can be found. There the less striking name C-minimal closed set instead of communicating system was used. Theorem 8.6 for finite E was obtained independently in [Bather].

8.1. DEFINITION. *For* $A \subset E$ *let* $f_P(i,A)$ *denote the probability of reaching subset* A *from state* i *for the Markov chain with matrix of transition probabilities* P. *We take for all* $P \in \mathcal{P}$, $f_P(i,A) = 1$ *if* $i \in A$ *and write* $f_P(i,i)$ *for* $f_P(i,\{i\})$.

Subset $A \subset E$ *is called a communicating class w.r.t.* $\mathcal{P}$ *if*

$$f_P(i,A^c) = 0 \text{ for all } i \in A, \text{ all } P \in \mathcal{P}$$

and if

for each pair of states $i,j \in A$ *there is a matrix* $P \in \mathcal{P}$ *and a nonnegative integer* n *such that* $p^n(i,j) > 0$.

If state space E *is a communicating class w.r.t.* $\mathcal{P}$ *then we speak of the communicating system* $(E,\mathcal{P})$.

State j *is recurrent w.r.t.* $\mathcal{P}$ *if for each* $i \in E$ *with* $f_P(j,i) > 0$ *for some* $P \in \mathcal{P}$, *it holds that* $\sup_{\mathcal{P}} f_P(i,j) = 1$.

If A *is a communicating class w.r.t.* $\mathcal{P}$ *and if each element of* A *is a* *recurrent state w.r.t.* P *then we call* A *a recurrent class w.r.t.* P.

If state space E *is a recurrent class w.r.t.* P *then we speak of the recurrent system* (E,P).

The following two theorems are generalizations to collections of Markov chains of the theorems I.8.5 and I.8.6 in [Chung]. Note that an excessive function is a c_P - excessive function with $c_P \equiv 0$.

8.2. THEOREM. a) *If* u *is an excessive function w.r.t.* P *and if* $u(j) > 0$ *then*

$$(8.2.1) \qquad u(i)/u(j) \geq \sup_{P} f_P(i,j).$$

b) *If*

$$(8.2.2) \qquad w(i) := \sup_{P} f_P(i,j) \text{ for all } i \in E.$$

then

$$(8.2.3) \qquad w(i) = \sup_{P} Pw(i) \text{ for } i \neq j$$

and

$$(8.2.4) \qquad w(j) \geq \sup_{P} Pw(j).$$

Hence w *is an excessive function w.r.t.* P.

PROOF. a) Define

$$u^*(i) := u(i)/u(j) \text{ for } i \in E;$$

then clearly u^* is also an excessive function w.r.t. P, moreover $u^*(j) = 1$. Now let us focus on the optimal control problem as introduced in section 3, with cost structure $c_P \equiv 0$ and $r(i) = \delta(i,j)$.*)

*) The Kronecker delta function is defined by

$$\delta(i,j) := \begin{cases} 0 \text{ for } i \neq j, \\ 1 \text{ for } i = j. \end{cases}$$

Then the value function of this problem equals

$$v(i) = \sup_R f_R(i,j) \text{ for all } i,$$

where $f_R(i,j)$ denotes the probability that state j is ever reached from state i when policy R is used ($f_R(i,i) = 1$ for all i and all R). According to theorem 3.1 we have that v is the least excessive function with $v(j) \geq 1$ and hence $v \leq u^*$. From this relation (8.2.1) follows.

b) We again consider the above introduced optimal control problem. By theorem 13.6 it follows that $\sup_{\mathcal{P}} f_P(i,j) = \sup_R f_R(i,j)$. Hence function w of (8.2.2) equals the value function v. Since $r(i) = 0$ for $i \neq j$ it follows by theorem 3.5 that relation (8.2.3) holds. Relation (8.2.4) follows from the fact that v is an excessive function. □

In the following two theorems it is assumed that $(E,\mathcal{P})$ is a communicating system.

8.3. THEOREM. a) *If* E *is a recurrent system w.r.t.* $\mathcal{P}$*, then every excessive funtion w.r.t.* $\mathcal{P}$ *is a constant function.*
b) *If* E *is a nonrecurrent system w.r.t.* $\mathcal{P}$*, and contains more than one state then there exists a nonnegative, nonconstant, bounded function* w *satisfying the relations* (8.2.3) *and* (8.2.4).

PROOF. a) Suppose u is an excessive function, then $u \geq 0$. If $u \not\equiv 0$ then there is some state j with $u(j) > 0$. By (8.2.1) and the definition of recurrence we obtain

$$u(i)/u(j) \geq \sup_{\mathcal{P}} f_P(i,j) = 1$$

and hence $u(i) \geq u(j)$ for all i. Consequently $u(i) > 0$ for all i, and by interchanging i and j we get $u(i) = u(j)$ for all i. Hence u is a constant function.

b) If E is a nonrecurrent system then by definition there is a pair of states (i,j) such that

$$(8.3.1) \qquad \sup_{\mathcal{P}} f_P(i,j) < 1.$$

Now we consider again the optimal control problem with $c_P \equiv 0$ and $r(i) = \delta(i,j)$. As shown in the proof of theorem 8.2.b we then have that the value function v equals the function w defined in (8.2.2). In virtue of (8.3.1) then $v(i) < v(j) = 1$ and thus v is a nonconstant, bounded excessive function. Moreover, as in the proof of theorem 8.2 the function v satisfies (8.2.3) and (8.2.4). □

As a consequence of theorem 8.3 we state the following theorem, which provides a criterion for recurrence w.r.t. $\mathcal{P}$. It generalizes theorem 6 of [Foster]. We note that the adjective bounded may be inserted in the criterion.

8.4. THEOREM. E *is a nonrecurrent system w.r.t.* $\mathcal{P}$ *if and only if there exists a nonconstant (bounded) excessive function w.r.t.* $\mathcal{P}$.

The next theorem is an application of theorem 8.3 to optimal control problems.

8.5. THEOREM. *If* E *is a recurrent system then for the optimal control problem with* $c_P \equiv 0$ *and* $r \geq 0$ *the value function* v *is a constant function with*

$$v(i) = \sup_j r(j) \text{ for all } i.$$

PROOF. The value function is by theorem 3.1 the least excessive function that majorizes r. According to theorem 8.3 we conclude that v is a constant function. Consequently v is the least constant function that majorizes r and hence $v(i) = \sup_j r(j)$ for all i. □

The above theorem remains valid for c_P nonnegative. However, when $c_P(i) > 0$ for some i and P, it follows that $v \equiv \infty$.

In Markov decision problems with average return criterion it is often desirable that the "maximal" average return does not depend on the starting state, i.e. the function

$$(8.5.1) \qquad g(i) := \sup_R \liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \mathbb{E}_{i,R}\, c(\underline{x}_n)$$

is a constant function. The next theorem provides a condition guaranteeing this. Although it is not uncommon to define the "maximal" average expected return as in (8.5.1) one might prefer to take the largest limit points, i.e. limes superior instead of limes inferior (cf. 5.0.1). Actually we are forced to take the lim inf in the next proof.

8.6. THEOREM. *If* E *is a recurrent system and if* c_P *is bounded from below then* g *as defined in* (8.5.1) *is a constant function.*

PROOF. Since c_P is bounded from below there is some constant c such that $g^* := g+ce$ is nonnegative. The proof proceeds now by showing that g is superharmonic. Then g^* is a superharmonic function and hence g^* is excessive. According to theorem 8.3a then g^* is a constant function and so is g.

Given any $\varepsilon > 0$ there is for each $i \in E$ a policy R_i such that

$$\liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \mathbb{E}_{i,R_i}\, c(\underline{x}_n) \geq g(i)-\varepsilon.$$

For P an arbitrary element of $\mathcal{P}$ let R be the policy that chooses decision rule P at time 0 and uses policy R_i from time 1 when the state at time 1 is i (as in theorem 3.1 we use here non-memoryless policies to show that g is superharmonic).

We have for R and arbitrary $i \in E$

$$g(i) \geq \liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \mathbb{E}_{i,R}\, c(\underline{x}_n) =$$

$$= \liminf_{N\to\infty} \left\{\frac{c_P(i)}{N+1} + \sum_j p(i,j) \left[\frac{1}{N+1} \sum_{m=0}^{N-1} \mathbb{E}_{j,R_j}\, c(\underline{x}_m)\right]\right\}.$$

From Fatou's lemma

$$g(i) \geq \sum_j p(i,j) \left[\liminf_{N\to\infty} \frac{1}{N+1} \sum_{m=0}^{N-1} \mathbb{E}_{j,R_j}\, c(\underline{x}_m)\right] \geq$$

$$\geq \sum_j p(i,j)\,(g(j)-\varepsilon).$$

Since ε and P were arbitrarily chosen we conclude that g is superharmonic.

□

As a generalization of the notation introduced in definition 8.1 let $f_R(i,A)$ denote the probability that subset A is ever reached from state i when policy R is used ($f_R(i,A) = 1$ for $i \in A$ and all R; we write $f_P(i,A)$ for $f_{P^\infty}(i,A)$).

8.7. THEOREM. *If* $p(i,E) = 1$ *for all* i *and* P *and if for some subset* A

$$\inf_i \sup_R f_R(i,A) > 0.$$

then there exists a $Q \in \mathcal{P}$ *with*

$$f_Q(i,A) = 1 \text{ for all } i.$$

PROOF. We consider the optimal control problem with $c_P \equiv 0$ and $r = \chi(A)$, i.e. $r(i) = 1$ if $i \in A$, $r(i) = 0$ otherwise. For the value function v we have $v(i) = \sup_R f_R(i,A) > 0$ for all i. According to theorem 13.7 there is a policy Q^∞ and an entry time in some subset $B \subset E$, say $\underline{\tau}_B$, such that

$$v_Q := \mathbb{E}_Q\,[r(\underline{x}_{\underline{\tau}_B})] \geq (1-\varepsilon)v, \tag{8.7.1}$$

for $0 < \varepsilon < 1$. Since $r = 0$ outside A and $v(i) > 0$ for all i it follows from (8.7.1) that $B \subset A$. Indeed, if $i \in B\backslash A$ then $v_Q(i) = \mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})] = \mathbb{E}_{i,Q}\,[r(\underline{x}_0)] = r(i) = 0 < (1-\varepsilon)\,v(i)$.

Suppose $\inf_i \sup_R f_R(i,A) = a > 0$ then $v(i) \geq a$ for all i. Let $\tilde{Q}$ be the column-restriction of Q to the complement of B (cf. 5.10); then for arbitrary $i \in B^c$

$$\begin{aligned} \mathbb{P}_{i,Q}\,[\underline{\tau}_B > n] &= \tilde{Q}^n\, e(i) \leq a^{-1}\, \tilde{Q}^n\, v(i) \leq \\ &\leq a^{-1}(1-\varepsilon)^{-1}\, \tilde{Q}^n\, v_Q(i) = \\ &= a^{-1}(1-\varepsilon)^{-1}\, \mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})\, \chi(\underline{\tau}_B > n)]. \end{aligned} \tag{8.7.2}$$

Since

$$\mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})] = \mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})\, \chi(\underline{\tau}_B \leq n)] + \mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})\, \chi(\underline{\tau}_B > n)]$$

and because the first term on the right-hand side tends to $\mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_B})]$ as n tends to infinity, we find that the second term tends to zero as n tends to infinity. In virtue of (8.7.2) we obtain

$$\lim_{n\to\infty} \mathbb{P}_{i,Q}\,[\underline{\tau}_B > n] = 0.$$

Since $q(j,E) = 1$ for all $j \in E$ we have that

$$\mathbb{P}_{i,Q}\,[\underline{\tau}_B \le n] = 1 - \mathbb{P}_{i,Q}\,[\underline{\tau}_B > n]$$

tends to one as n tends to infinity. Hence $f_Q(i,B) = 1$ and a fortiori $f_Q(i,A) = 1$. Since by definition $f_Q(i,B) = 1$ for $i \in B$, this completes the proof. □

The above theorem can be seen as a generalization of theorem 1 in [Chung and Derman].

In the sequel of this section it is assumed that $p(i,E) = 1$ for all $i \in E$ and all $P \in \mathcal{P}$.

The next theorem shows that recurrence w.r.t. $\mathcal{P}$ is a class property.

8.8. THEOREM. *Let* E^* *be a communicating class w.r.t.* $\mathcal{P}$. *If for some* $j \in E^*$

$$\inf_{i\in E^*} \sup_R f_R(i,j) > 0$$

then E^* *is a recurrent class w.r.t.* $\mathcal{P}$.

PROOF. Let i_0 be an arbitrary element of E^*. Since E^* is a communicating class there exists a matrix P and a subset $B = \{j,i_1,i_2,\ldots,i_n,i_0\}$ such that $p(j,i_1)\,p(i_1,i_2)\,\ldots\,p(i_n,i_0) = \alpha$ for some positive constant α. Since E^* is a communicating class we can apply theorem 8.7 with E^* for E and $\{j\}$ for A and we find a matrix Q with $f_Q(i,j) = 1$ for all $i \in E^*$.

Define matrix Q^* as follows

$$q^*(i,j) = \begin{cases} q(i,j) & \text{for } i \notin B \\ p(i,j) & \text{for } i \in B. \end{cases}$$

Then

$$f_{Q^*}(i,i_0) \geq f_{Q^*}(i,B) \underset{\ell \in B, \ell \neq 0}{\text{minimum}} f_{Q^*}(\ell,i_0)$$

The first factor on the right-hand side equals 1 and the second factor is not less than α. Hence

$$\inf_i f_{Q^*}(i,i_0) \geq \alpha.$$

Now we apply theorem 8.7 with E^* for E, $\{i_0\}$ for A and with $\{Q^*\}$ for the collection of Markov matrices P and find that $f_{Q^*}(i,i_0) = 1$ for all $i \in E^*$. Hence the theorem follows. □

The remaining theorems of this section are corollaries of the foregoing results. They assert the existence of optimal strategies under various conditions.

8.9. THEOREM. *If* E *is a recurrent system then there exists a stationary optimal strategy for the optimal control problem with* $c_P \equiv 0$ *and* r *such that* $r(i) \leq r(i_0)$ *for some state* $i_0 \in E$ *and all* $i \in E$.

PROOF. By the definition of recurrence and theorem 8.7 there exists a matrix Q with $f_Q(i,i_0) = 1$ for all $i \in E$. The stationary policy Q^∞ in combination with the entry time of $\{i_0\}$ provides a stationary optimal strategy. □

8.10. THEOREM. *If* E *has a finite number of states and is a communicating system then every optimal control problem with* $c_P \equiv 0$ *has a stationary optimal strategy.*

PROOF. When E is finite and a communicating system then

$$\min_{i,j} \sup_R f_R(i,j) > 0.$$

As a consequence of theorem 8.8 we have that E is a recurrent system. Now we can apply theorem 8.9 and the assertion is proved. □

The following theorem can be found in [Dubins and Savage, theorem 3.8.5, p. 56].

8.11. THEOREM. *If for some optimal control problem we have that* $c_P \equiv 0$ *and* $r = \chi(A)$ *for a subset* A *and* $\inf_i v(i) > 0$ *then there exists a stationary optimal strategy.*

PROOF. The value function of the optimal control problem with $c_P \equiv 0$ and $r = \chi(A)$ equals

$$v(i) = \sup_R f_R(i,A) \text{ for all } i \in E.$$

Since $\inf_i v(i) > 0$ it follows from theorem 8.7 that there exists a policy Q^∞ such that $f_Q(i,A) = 1$ for all $i \in E$. Hence policy Q^∞ in combination with the entry time of A is optimal. □

The last theorem provides a sufficient condition in the case that the cost structure is not identically zero.

8.12. THEOREM. *For the optimal control problem with charge structure* c_P, *reward function* r *and bounded value function* v *let*

$$\Gamma := \{i : r(i)=v(i)\}$$

and

$$P^* := \{P : c_P(i) + \sum_j p(i,j)\, v(j) = v(i) \text{ for } i \notin \Gamma \text{ and } P \in \mathcal{P}\}.$$

If

$$\inf_i \sup_{P^*} f_P(i,\Gamma) > 0$$

then there exists a stationary optimal policy.

We note that the strategies $(R,\underline{\tau}_\Gamma)$ with $R = (P_0,P_1,\ldots)$, $P_n \in P^*$ for $n \geq 0$ and $\underline{\tau}_\Gamma$ the entry time of Γ are *thrifty* strategies.

PROOF. It is easy to verify that P^* has the product property. By applying theorem 8.7 with P^* we find a $Q \in P^*$ with $f_Q(i,\Gamma) = 1$ for all $i \in E$. In view of theorem 4.8.a we conclude that $(Q^\infty, \underline{\tau}_\Gamma)$ is optimal. □

9. EXPONENTIALLY BOUNDED STOPPING TIMES

A property that holds for most of the sequential decision problems is that the infimum of the sampling costs over the various experiments is positive. This property, in combination with a boundedness condition on the loss function (in our terminology the reward function) implies that the optimal stopping time $\underline{\tau}$ (or the random number of observations) is exponentially bounded, i.e. there are positive constants a and b such that for stopping time $\underline{\tau}$

$$(9.0.1)\qquad \mathbb{P}\,[\underline{\tau} > n] \le a \exp(-bn) \text{ for all } n \in \{0,1,\ldots\}.$$

The "sequential probability ratio test" as introduced by [Wald] can be identified with the optimal strategy in an optimal control problem (cf. [Lehmann, p. 104]). Thus the well-known property that the number of observations in Wald's test is exponentially bounded follows also from the results in this section. In fact there is a wide class of problems for which the assumptions of this section are satisfied. They all have optimal stopping times with the nice property (9.0.1). A result related to theorem 9.5 can be found in [Ross (1971), theorem 6.13, p. 136].

In this section we make the assumption

$$(9.0.2)\qquad c_0 := \underset{i\in E, P\in \mathcal{P}}{\text{infimum}} - c_P(i) > 0.$$

If $p(i,E) = 1$ for all $i \in E$, then the above assumption implies that $\sum_{n=0}^{\infty} P^n c_P = -\infty$ and c_P is not a charge w.r.t. P. So in this section we do not assume that $c_{\mathcal{P}}$ is a charge structure.

9.1. THEOREM. *If a stationary strategy* $(Q^\infty, \underline{\tau}_A)$ *with* $\underline{\tau}_A$ *the entry time of* A, *is such that*

$$(9.1.1)\qquad \mathbb{E}_{i,Q}\,[\sum_{n=0}^{\underline{\tau}_A - 1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_A})] \ge r(i) \textit{ for all } i \in A^c$$

and if in addition r *is bounded from below and* $\mathbb{E}_Q\,[r(\underline{x}_{\underline{\tau}_A})]$ *is bounded from above on* A^c, *then* $\underline{\tau}_A$ *is exponentially bounded under policy* Q^∞.

PROOF. Assumption (9.0.2) and relation (9.1.1) together imply

$$(9.1.2)\qquad \mathbb{E}_{i,Q} \sum_{n=0}^{\underline{\tau}_A-1} c_0 e \le \mathbb{E}_{i,Q}\,[r(\underline{x}_{\underline{\tau}_A})] - r(i) \text{ for all } i \in A^c.$$

The right-hand side is bounded from above on A^c; let constant c_1 be an upper bound. The left-hand side equals constant c_0 multiplied by

$$(9.1.3)\qquad \mathbb{E}_{i,Q}\,[\underline{\tau}_A] = \sum_{n=0}^{\infty} \mathbb{P}_{i,Q}\,[\underline{\tau}_A > n].$$

By the Markov inequality or alternatively directly from (9.1.3), since $\mathbb{P}_{i,Q}\,[\underline{\tau}_A > n]$ is monotone nonincreasing in n, we have that

$$(9.1.4)\qquad \mathbb{P}_{i,Q}\,[\underline{\tau}_A > n] \le \frac{1}{n}\,\mathbb{E}_{i,Q}\,[\underline{\tau}_A].$$

Let N be such that

$$(9.1.5)\qquad \alpha := N^{-1} c_0^{-1} c_1 < 1;$$

then we obtain from the relations (9.1.2), (9.1.4) and (9.1.5)

$$(9.1.6)\qquad \mathbb{P}_{i,Q}\,[\underline{\tau}_A > N] \le \alpha \text{ for all } i \in A^c.$$

Let $\tilde{Q}$ denote the restriction of Q to A^c; then by rewriting the left-hand side of (9.1.6) we get

$$(9.1.7)\qquad \tilde{Q}^N e \le \alpha e.$$

Because

$$\tilde{Q}^{kN} e = \tilde{Q}^{(k-1)N}\,\tilde{Q}^N e \le \alpha \tilde{Q}^{(k-1)N} e$$

it is immediate from (9.1.7) that (recall that $\underline{\tau}_A=0$ when $\underline{x}_0 \in A$)

$$(9.1.8)\qquad \mathbb{P}_Q\,[\underline{\tau} > n] = \tilde{Q}^n e \le \alpha^{[nN^{-1}]},$$

where [x] denotes the largest integer not exceeding x.

Let $a := \alpha^{-1/N}$ and $b := N^{-1} \log \alpha^{-1}$; then from (9.1.8) we have that

$$(9.1.9) \qquad \mathbb{P}_Q [\underline{\tau} > n] \le a \exp(-bn). \ \square$$

The reward for stopping immediately in state i equals the right-hand side of relation (9.1.1), whereas the left-hand side denotes the value of strategy $(Q^\infty, \underline{\tau}_A)$. Thus each strategy which does at least as well as stopping immediately satisfies relation (9.1.1).

Section 4 suggests, that $(Q^\infty, \underline{\tau}_\Gamma)$ is a good candidate for an optimal strategy, if $\underline{\tau}_\Gamma$ is the entry time of the stopping set

$$(9.1.10) \qquad \Gamma = \{i : r(i)=v(i)\}$$

and Q satisfies

$$(9.1.11) \qquad c_Q(i) + Qv(i) = v(i) \text{ for all } i \in \Gamma^c,$$

where v stands for the value function of the optimal control problem. In particular this strategy $(Q^\infty, \underline{\tau}_\Gamma)$ will satisfy relation (9.1.1) if it is optimal.

9.2. THEOREM. *If the value function* v *and* $\mathbb{E}_Q [r(\underline{x}_{\underline{\tau}_\Gamma})]$ *are bounded from above and* r *is bounded from below on* Γ^c *then strategy* $(Q^\infty, \underline{\tau}_\Gamma)$ *is optimal and, moreover, the optimal stopping time* $\underline{\tau}_\Gamma$ *is exponentially bounded under the stationary optimal policy* Q^∞.

PROOF. As we argued above the only thing we have to prove in order that we can apply theorem 9.1 is that $(Q^\infty, \underline{\tau}_\Gamma)$ is optimal.

Denoting $\underline{\tau}_N := \underline{\tau}_\Gamma \wedge N$, we have according to lemma 2.19 and the relations (9.1.10) and (9.1.11)

$$(9.2.1) \qquad v = \mathbb{E}_Q \Big[\sum_{n=0}^{\underline{\tau}_N - 1} c(\underline{x}_n) + v(\underline{x}_{\underline{\tau}_N}) \Big].$$

Hence

$$(9.2.2)\qquad \mathbb{E}_Q \sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n) = \lim_{N\to\infty} \mathbb{E}_Q \sum_{n=0}^{\underline{\tau}_N - 1} c(\underline{x}_n) =$$

$$= v - \lim_{N\to\infty} \mathbb{E}_Q\,[v(\underline{x}_{\underline{\tau}_\Gamma})\ \chi(\underline{\tau}_\Gamma \le N)] - \lim_{N\to\infty} \mathbb{E}_Q\,[v(\underline{x}_N)\ \chi(\underline{\tau}_\Gamma > N)].$$

Since $v = r$ on Γ the second term on the right-hand side equals $\mathbb{E}_Q\,[r(\underline{x}_{\underline{\tau}_\Gamma})]$. It is assumed that this expectation is finite. Further it is assumed that v is bounded from above on Γ^c and hence the third term has a lim sup which is less than plus infinity. Consequently, because of $c_Q \le 0$

$$\sum_{n=0}^{\infty} \tilde{Q}^n c_Q = \mathbb{E}_Q \sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n),$$

with $\tilde{Q}$ as in 9.1, is finite. With assumption (9.0.2) we then obtain

$$\lim_{n\to\infty} \tilde{Q}^n e = 0.$$

Since v is bounded on Γ^c then also

$$(9.2.3)\qquad \lim_{n\to\infty} \tilde{Q}^n v = \lim_{n\to\infty} \mathbb{E}_Q\,[v(\underline{x}_n)\ \chi(\underline{\tau}_\Gamma > n)] = 0.$$

In virtue of the relations (9.2.2) and (9.2.3) we then have

$$\mathbb{E}_Q\,[\sum_{n=0}^{\underline{\tau}_\Gamma - 1} c(\underline{x}_n) + v(\underline{x}_{\underline{\tau}_\Gamma})] = v$$

Since $v = r$ on Γ we find that $(Q^\infty, \underline{\tau}_\Gamma)$ is optimal. □

9.3. DEFINITION. *Let* v_N *denote the supremum over the expected values of the strategies* $(R, \underline{\tau})$ *with* $\underline{\tau} \le N$, *i.e.*

$$v_N = \underset{R,\underline{\tau}\le N}{\text{supremum}}\ \mathbb{E}_R\,[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})].$$

An important tool in computing the value function is the approximation of v by v_N for N sufficiently large. This can only work if $\lim_{N\to\infty} v_N = v$.

9.4. DEFINITION. *An optimal control problem is stable if*

$$\lim_{N\to\infty} v_N = v.$$

In section 3 we defined stability w.r.t. x.

It can be proved that (cf. [Ross (1971), p. 136]; for the definition of T see 3.2)

$$Tv_N = v_{N+1}.$$

From this relation it follows easily that a stable problem is at least stable w.r.t. all x such that

$$\sup_{P} (c_P+Px) \le x \text{ and } x \le r.$$

Verbally, the problem is stable for all x that are c_P - superharmonic and minorize r.

In [Starr] the rate of convergence of v_N to v for a special problem is numerically analyzed. It is noticed that the convergence is quite rapid. The following theorem asserts that it is exponentially fast.

9.5. THEOREM. *Under the assumptions of theorem 9.2 the optimal control problem is stable. Moreover,* v_N *tends exponentially fast to* v *as* $N \to \infty$.

PROOF. For $(Q^\infty, \underline{\tau}_\Gamma)$ as in theorem 9.2 we have with $\underline{\tau}_N = \underline{\tau}_\Gamma \wedge N$ by definition of v_N that

$$v_N \ge \mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_N-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_N}) \right].$$

Using the fact that $c_P \le 0$ for all $P \in \mathcal{P}$ and $\underline{\tau}_\Gamma \ge \underline{\tau}_N$ we find by rewriting the right-hand side of this inequality

$$(9.5.1) \qquad v_N \ge \mathbb{E}_Q \left[\sum_{n=0}^{\underline{\tau}_\Gamma-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}_\Gamma}) \right] +$$

$$- \mathbb{E}_Q \left[r(\underline{x}_{\underline{\tau}_\Gamma}) \, \chi(\underline{\tau}_\Gamma > N) \right] + \mathbb{E}_Q \left[r(\underline{x}_N) \, \chi(\underline{\tau}_\Gamma > N) \right].$$

Since the first term on the right-hand side is by theorem 9.2 equal to v and since by definition $v \geq v_N$ it is sufficient to prove that the second term on the right-hand side has a nonpositive lim sup and the third term has a nonnegative lim inf.

Indeed in view of the Markov property the second term equals

$$(9.5.2) \qquad \tilde{Q}^N \, \mathbb{E}_Q \, [r(\underline{x}_{\underline{\tau}_\Gamma})],$$

with the same notation as in theorem 9.2. The expectation $\mathbb{E}_Q \, [r(\underline{x}_{\underline{\tau}_\Gamma})]$ is bounded from above and by the relations (9.1.8) and (9.1.9) $\tilde{Q}^N e$ tends exponentially fast to zero. Hence the positive part of (9.5.2) tends exponentially fast to zero. In a similar way it can be proved that the negative part of the third term on the right-hand side of inequality (9.5.1) tends exponentially fast to zero. Consequently v_N tends to v as $N \to \infty$ and the rate of convergence is at least exponential. □

We note that in the case of a bounded reward function all boundedness conditions in the foregoing theorems are satisfied. This section is concluded with a theorem about the uniqueness of the value function v as solution of the optimality equation.

9.6. THEOREM. *If the nonnegative function* w *satisfies*

$$(9.6.1) \qquad w = \max\,(r, c_{Q_0} + Q_0 w) \; \textit{for some } Q_0 \in \mathcal{P}$$

and

$$(9.6.2) \qquad w \geq c_P + Pw \; \textit{for all } P \in \mathcal{P}$$

and if in addition the functions w *and* w-v *are bounded from above on the complement of* $\Gamma_0 := \{i : r(i) = w(i)\}$, *then* w *is equal to* v *and* $(Q_0^\infty, \underline{\tau}_{\Gamma_0})$ *is optimal.*

PROOF. If c_P is a charge structure then it follows from the theorems 3.1 and 3.8 that $w \geq v$. Since $w \geq 0$ this inequality is true in general. Indeed, proceeding in a similar way as in lemma 2.19 one can prove that for each policy R and bounded Markov time $\underline{\tau}$

$$\mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}-1} c(\underline{x}_n) + w(\underline{x}_{\underline{\tau}}) \right] \le w.$$

Hence we have for arbitrary Markov time $\underline{\tau}$ with $\underline{\tau}_N := \underline{\tau} \wedge N$ in view of $w \ge r$ and $w \ge 0$, that

$$\mathbb{E}_R \left[\sum_{n=0}^{\underline{\tau}_N-1} c(\underline{x}_n) + r(\underline{x}_{\underline{\tau}})\, \chi(\underline{\tau} \le N) \right] \le w\,.$$

Letting N tend to infinity we find that w majorizes the expected value of an arbitrary strategy. Hence by definition of v we have $v \le w$ and consequently $\Gamma_0 \subset \Gamma = \{i : r(i)=v(i)\}$. Let $\tilde{Q}_0$ be the restriction of Q to Γ_0^c; then as in the proof of theorem 9.2 it can be shown that

$$\text{(9.6.3)} \qquad \lim_{n\to\infty} \tilde{Q}_0^n\, e = 0.$$

From $v \le w$, $\Gamma_0 \subset \Gamma$, $c_{Q_0} + Q_0 v \le v$ and the relations (9.6.1) and (9.6.2) it follows that

$$\text{(9.6.4)} \qquad 0 \le w-v \le \tilde{Q}_0(w-v).$$

By assumption w-v is bounded on Γ_0^c and therefore by (9.6.3) and (9.6.4) v equals w. Since $\lim_{n\to\infty} \tilde{Q}_0^n\, v = 0$ we have as in theorem 9.2 that $(Q_0^\infty, \underline{\tau}_{\Gamma_0})$ is optimal. □

10. SUFFICIENT CONDITIONS FOR THE EXISTENCE OF AN OPTIMAL POLICY WITH RESPECT TO THE AVERAGE RETURN CRITERION

In this section we investigate the existence of optimal policies with respect to the average return criterion. A policy will be called optimal if it maximizes

$$(10.0.1)\qquad \liminf_{N\to\infty} \frac{1}{N+1}\,\mathbb{E}_R \left[\sum_{n=0}^{N} c(\underline{x}_n)\right].$$

The limes inferior rather than the limes superior is chosen, in order to be able to prove relation (10.6.1). In section 12 we will show that under rather general conditions the two criteria lead to the same supremum.

This section uses results from [Hordijk (1971)] and [Hordijk (1972)].

It is assumed in this section that c_P is continuous and bounded, i.e.

$$(10.0.2)\qquad |c_P(i)| \le b \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

Furthermore it is assumed that $\mathcal{P}$ is compact and $p(i,E) = 1$ for all $i \in E$ and $P \in \mathcal{P}$. In this section a probability measure $p(.)$ on E is always a proper probability measure, i.e. $p(E) = 1$.

Let g denote the supremum over all policies of the average expected return

$$(10.0.3)\qquad g(i) := \sup_R \liminf_{N\to\infty} \frac{1}{N+1}\,\mathbb{E}_{i,R} \left[\sum_{n=0}^{N} c(\underline{x}_n)\right] \text{ for } i \in E.$$

In the following subsection we state several assumptions we need in the sequel. These assumptions will be discussed afterwards.

10.1. ASSUMPTIONS

A. $\pi_P(i,j)$, *defined by*

$$\pi_P(i,j) := \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} p^n(i,j)$$

is a (proper) probability measure, i.e. $\pi_P(i,E) = 1$ *for all* $i \in E$ *and all* $P \in \mathcal{P}$.

B. *The Cesaro-limit* π_P *depends continuously on* P, *i.e.*

$$\lim_{P \to P_0} \pi_P = \pi_{P_0} \quad \textit{for all } P_0 \in \mathcal{P}$$

(*i.e.* $\lim_{P \to P_0} \pi_P(i,j) = \pi_{P_0}(i,j)$ *for all* $i,j \in E$ *and all* $P_0 \in \mathcal{P}$).

C. *For each* $i \in E$

$$\{\pi_P(i,.) : P \in \mathcal{P}\}$$

is a tight collection of probability measures.

D. *The system* $(E,\mathcal{P})$ *is recurrent.*

E. *For each* $P \in \mathcal{P}$ *the associated Markov chain does not have disjoint closed sets.*

F. *The collection of probability measures*

$$\{p(i,.) : i \in E,\ P \in \mathcal{P}\}$$

is tight.

It is well-known that the Cesaro-limit in 10.1.A. always exists. However, it may be that $\pi_P(i,.)$ is not a probability measure. A Markov chain for which $\pi_P(i,E) = 1$ for all $i \in E$ is called *non-dissipative* (cf. [Chung]). So assumption A can be stated in the following form: *for each stationary policy the Markov chain is non-dissipative.*

It is not difficult to construct counterexamples for which π_P is not a continuous function of P. The counterexample in 5.12 provides one.

A collection of probability measures $\mathcal{A}$ on a metric space is called *tight* if for each positive ε there exists a compact set K such that $P(K) \geq 1-\varepsilon$ for all P in $\mathcal{A}$ (cf. [Billingsley]). It is obvious that the state space E can be seen as a discrete topological space. Then each compact set has a finite number of elements. A theorem of Prohorov says that in a separable and complete metric space collection $\mathcal{A}$ is tight if and only if $\mathcal{A}$ is relatively compact, i.e. every sequence of elements of $\mathcal{A}$ contains a weakly convergent subsequence. In the case 10.1.C this implies that if $P_n \in \mathcal{P}$ for all $n \in \{1,2,\ldots\}$ then for each $i \in E$ there exists a probability measure $\pi(i,.)$ such that for some subsequence n_k, $k=1,2,\ldots$, of the positive integers

$$\lim_{k\to\infty} \pi_{P_{n_k}}(i,j) = \pi(i,j) \text{ for all } j \in E.$$

Although it follows from the general theorem of Prohorov, this is easily verified in our discrete state space.

It is also easy to check that assumptions A and B imply assumption C. Alternatively this follows from the well-known fact that a continuous image of a compact set is also a compact set. Hence for each $i \in E$ the collection in assumption C is compact, so a fortiori relatively compact and thus by Prohorov's theorem tight, if assumptions A and B are satisfied.

Under an additional assumption the converse is also true. By definition we have that C implies A and, moreover, as a corollary of the following lemma we obtain that the assumptions C and E together imply assumption B.

10.2. LEMMA. *If* $\lim_{n\to\infty} P_n = P_\infty$ *and* P_∞ *has no disjoint closed sets then under assumption* C *we have* $\lim_{n\to\infty} \pi_{P_n} = \pi_{P_\infty}$.

PROOF. By assumption C there is for a fixed $i \in E$ a probability measure $\pi(i,.)$ such that for some sequence of the positive integers n_k, $k=1,2,\ldots$,

$$\lim_{k\to\infty} \pi_k(i,j) = \pi(i,j) \text{ for all } j \in E, \tag{10.2.1}$$

where $\pi_k(i,j)$ is just another notation for $\pi_{P_{n_k}}(i,j)$. It is well-known that

$$\sum_{\ell} \pi_k(i,\ell)\, P_{n_k}(\ell,j) = \pi_k(i,j) \text{ for all } j \in E.$$

Letting k tend to infinity we find by lemma 4.13 that

$$\sum_{\ell} \pi(i,\ell)\, p_\infty(\ell,j) = \pi(i,j) \text{ for all } j \in E.$$

Hence $\pi(i,.)$ is an invariant probability measure with respect to P_∞. Since P_∞ has no disjoint closed sets the probability measure $\pi_{P_\infty}(i,.)$ is the only invariant probability measure for P_∞ and consequently

$$\pi(i,j) = \pi_{P_\infty}(i,j) \text{ for all } j \in E.$$

The assertion follows now by relation (10.2.1). □

The above lemma can be strengthened in the following sense. Let n_k denote the number of minimal closed sets of P_k. If $n_k < \infty$ for all $k \in \{1,2,\dots\}$ and $\infty > \liminf_{k\to\infty} n_k \geq n_\infty$ then $\lim_{k\to\infty} n_k = n_\infty$ and the lemma remains true. So what one has to prevent is the creation of an extra minimal closed set as $k \to \infty$. Related results for finite Markov chains can be found in [Schweitzer].

Assumption F was introduced because the assumptions B and C are awkward to check. We have the following connection between C and F.

10.3. LEMMA. *Assumption* F *implies assumption* C *and hence also assumption* **A.**

PROOF. Choose any $\varepsilon > 0$. Let K be a finite set such that

$$p(i,K) \geq 1-\varepsilon \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

Hence

$$p^2(i,K) = \sum_{j\in E} p(i,j)\, p(j,K) \geq (1-\varepsilon)\, p(i,E) \geq 1-\varepsilon$$

for all $i \in E$ and all $P \in \mathcal{P}$. Clearly we then have

$$p^n(i,K) \geq 1-\varepsilon \text{ for all } i \in E, \text{ all } P \in \mathcal{P} \text{ and all } n \in \{1,2,\dots\}.$$

Consequently

$$\sum_{j\in K} \frac{1}{N} \sum_{n=1}^{N} p^n(i,j) \geq 1-\varepsilon \text{ for all } N$$

and since K is finite also the limit

$$\sum_{j\in K} \pi_P(i,j) \geq 1-\varepsilon \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

This proves the assertion. □

We next give the main result of this section. Since the proof is rather long we divide it into subsections.

10.4. THEOREM. *Assumptions* A *and* B *or assumptions* C *and* D *imply the existence of a stationary optimal policy.*

Condition A will be assumed in all subsections. As to the other assumptions we will indicate where we need them.

10.5. LEMMA. *Under the assumption* C, *there exist* α_n *and* P_n *with* $0 < \alpha_n < 1$ *and* $P_n \in \mathcal{P}$ *for* $n = 1,2,\dots$ *such that*

$$\lim_{n\to\infty} \alpha_n = 1, \tag{10.5.1}$$

P_n *is* α_n *- discounted optimal, i.e.*

$$v_n := \sum_{k=0}^{\infty} \alpha_n^k P_n^k c_{P_n} \geq \sum_{k=0}^{\infty} \alpha_n^k \mathbb{E}_R [c(\underline{x}_k)] \text{ for each policy } R \tag{10.5.2}$$

$$\lim_{n\to\infty} P_n = P_\infty \text{ for some } P_\infty \in \mathcal{P}, \tag{10.5.3}$$

$$\lim_{n\to\infty} \Pi_n = \Pi \text{ for some stochastic matrix } \Pi \tag{10.5.4}$$

with $\pi(i,E) = 1$ *for all* $i \in E$ *and* $\Pi_n := \Pi_{P_n}$,

$$\lim_{n\to\infty} (1-\alpha_n)\, v_n = x \text{ for some vector } \mathbf{x}. \tag{10.5.5}$$

PROOF. The proof proceeds by showing that each of the above relations can be obtained by choosing an appropriate subsequence. Suppose we have a sequence α_n with $\lim_{n\to\infty} \alpha_n = 1$. According to theorem 6.4 there are matrices P_n, $n=1,2,\dots$, such that P_n^∞ is an α_n-discounted optimal policy. Since $\mathcal{P}$ is compact there is a subsequence of P_n, $n=1,2,\dots$, which converges to an element of $\mathcal{P}$. Now suppose α_n, P_n, $n=1,2,\dots$, satisfy the relations (10.5.1), (10.5.2) and (10.5.3). We assumed that assumption C holds. Hence by the relative compactness there is a subsequence satisfying relation (10.5.4). As to relation (10.5.5) we note that by relation (10.0.2) we have

$$(1-\alpha) \sum_{k=0}^{\infty} \mathbb{E}_R [\alpha^k c(\underline{x}_k)] \leq (1-\alpha) \sum_{k=0}^{\infty} \alpha^k be \leq be \text{ for all } R. \tag{10.5.6}$$

Hence the sequence $(1-\alpha_n)\, v_n$ is bounded and by the diagonal procedure we can choose a subsequence satisfying the relation (10.5.5). □

10.6. LEMMA. *The supremum over the expected average return does not exceed the vector* x, *i.e.* $g \le x$.

PROOF. For $R = (P_0, P_1, P_2, \ldots)$ an arbitrary policy it follows from an Abelian theorem or alternatively a Tauberian theorem that (for a proof see [Hordijk (1971)])

$$\liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} P_0 \ldots P_{n-1}\, c_{P_n}(i) \le \tag{10.6.1}$$

$$\le \liminf_{\alpha\uparrow 1} (1-\alpha) \sum_{k=0}^{\infty} \alpha^k\, P_0 \ldots P_{k-1}\, c_{P_k}(i) \text{ for all } i \in E.$$

By definition the right-hand side of this inequality does not exceed $(1-\alpha_n)\, v_n(i)$ when $\alpha = \alpha_n$. Hence the supremum over all policies of the left-hand side term is not larger than $x(i)$. □

10.7. LEMMA. *For* $\Pi_\infty := \Pi_{P_\infty}$ *we have* $\Pi_\infty x = x$.

PROOF. From

$$v_n = \sum_{k=0}^{\infty} \alpha_n^k\, P_n^k\, c_{P_n}$$

it is readily seen that

$$(1-\alpha_n)\, v_n = (1-\alpha_n)\, c_{P_n} + (1-\alpha_n)\, P_n\, \alpha_n\, v_n. \tag{10.7.1}$$

Letting n tend to infinity we find that the first term of (10.7.1) tends to x, the second tends to zero and the third term, since $(1-\alpha_n)\, v_n$ is bounded, tends to $P_\infty x$. Hence $P_\infty x = x$ and by iterating this equality we obtain $P_\infty^k x = x$ for all $k \in \{1,2,\ldots\}$. Consequently also $\frac{1}{N}\sum_{k=1}^{N} P_\infty^k x = x$ for all $N \in \{1,2,\ldots\}$ and hence $\Pi_\infty x = x$. □

10.8. LEMMA. *For* $c_\infty := c_{P_\infty}$ *we have* $\Pi c_\infty = \Pi x$.

PROOF. Using the well-known relation $\Pi_P P^k = \Pi_P$ for all $k \in \{1,2,\ldots\}$ and by interchanging the order of summation (this is allowed since all series are absolute convergent) we obtain

$$\Pi_P (1-\alpha) \sum_{k=0}^{\infty} \alpha^k P^k c_P = (1-\alpha) \sum_{k=0}^{\infty} \alpha^k \Pi_P P^k c_P = \Pi_P c_P.$$

By substituting α_n for α we find

$$\Pi_n (1-\alpha_n) v_n = \Pi_n c_n, \text{ where } c_n := c_{P_n}.$$

From the boundedness of $(1-\alpha_n)v_n$ and c_n and the fact that $\lim_{n\to\infty} \pi_n(i,E) = \pi(i,E) = 1$ for all $i \in E$ we find in view of lemma 4.13 by letting n tend to infinity $\Pi c_\infty = \Pi x$. □

10.9. LEMMA. *Under the assumptions* A *and* B *the policy* $(P_\infty, P_\infty, \ldots)$ *is optimal.*

PROOF. The assumptions A and B together imply C. Hence 10.4 to 10.8 are valid. Since $\lim_{n\to\infty} P_n = P_\infty$, we have from the assumption B that $\lim_{n\to\infty} \Pi_n = \Pi_\infty$ and hence by (10.5.4) we obtain $\Pi_\infty = \Pi$. By using this equality we obtain from 10.7 and 10.8 that $\Pi_\infty c_\infty = x$. Since

$$\Pi_\infty c_\infty = \lim_{N\to\infty} \frac{1}{N+1} \mathbb{E}_{P_\infty} \sum_{n=0}^{N} c(\underline{x}_n) \le g,$$

we have $x = \Pi_\infty c_\infty \le g \le x$, where the last inequality is from 10.6. Hence $\Pi_\infty c_\infty = g$ and by the definition of g we have that P_∞^∞ is a stationary optimal policy.

10.10. LEMMA. *Under* A *and* C *we have* $\Pi\Pi_\infty = \Pi$.

PROOF. From $\Pi_n P_n = \Pi_n$ by letting n tend to infinity we find $\Pi P_\infty = \Pi$. By iterating this equality we find $\Pi P_\infty^k = \Pi$ for all $k \in \{1,2,\ldots\}$ and hence $\frac{1}{N+1} \sum_{n=0}^{N} \Pi P_\infty^n = \Pi$ for all $N \in \{1,2,\ldots\}$. By letting N tend to infinity we obtain $\Pi\Pi_\infty = \Pi$. □

10.11. LEMMA. *For* $i \in E$ *let* D_i *be the support of* $\pi(i,.)$, *i.e.*

$D_i = \{j : \pi(i,j) > 0\}$, *then* D_i *is a closed set w.r.t.* P_∞ *and, moreover,* $\Pi_\infty c_\infty$ *equals* g *on* $D := \bigcup_i D_i$.

PROOF. The first assertion is immediate from the fact that $\pi(i,.)$ is for each $i \in E$ an invariant probability measure w.r.t. P_∞ (cf. 10.10). From 10.8 and 10.10 it follows that $\Pi(x-\Pi_\infty c_\infty) = 0$. From $\Pi_\infty c_\infty \leq g$ and $g \leq x$ (by 10.6) we have $x - \Pi_\infty c_\infty \geq 0$. Hence x equals $\Pi_\infty c_\infty$ on each D_i and consequently we have that g equals $\Pi_\infty c_\infty$ on D. □

It is clear from 10.9 *that theorem* 10.4 *is true if the assumptions* A *and* B *hold. We now prove the existence of a stationary optimal policy under the assumptions* C *and* D.

10.12. PROOF OF THE THEOREM. When D = E there is nothing left to prove. By assumption D we can apply theorem 8.7 to assert that there is a policy Q^∞ with $f_Q(i,D) = 1$ for all $i \in E$. Define matrix $\tilde{P}_\infty$ by

$$\tilde{p}_\infty(i,j) = \begin{cases} q(i,j) & \text{for } i \notin D, \\ p_\infty(i,j) & \text{for } i \in D. \end{cases}$$

Then the states of D^c are all transient states for the Markov chain with matrix of transition probabilities $\tilde{P}_\infty$. According to theorem 8.6 it follows from assumption D that g is a constant function. From these facts together with 10.11 we can see that $\Pi_{\tilde{P}_\infty} c_{\tilde{P}_\infty} = g$. Hence the policy $(\tilde{P}_\infty,\tilde{P}_\infty,\ldots)$ is optimal. □

10.13. REMARKS ON THEOREM 10.4. Under the assumptions A and B or, alternatively, under the assumptions C and D, if in addition the set D contains all positive recurrent states w.r.t. P_∞, the policy $(P_\infty,P_\infty,\ldots)$ is optimal. Hence we obtained an optimal policy as limit of discounted optimal policies. This can have nice consequences. For example if there are discounted optimal policies of (s,S) type in an inventory model with a certain cost structure, then there exists an (s,S) policy which is optimal with respect to the average return criterion if our theorem applies.

The argument of 10.9 also shows that: *If the assumptions* A *and* B *(or the assumptions* C *and* E *implying the assumptions* A *and* B*) are satisfied then each limit policy obtained from discounted optimal policies with dis-*

countfactor tending to one, is an optimal policy with respect to the average return criterion.

For arbitrary $i \in E$ the set D_i contains at least one positive recurrent class w.r.t. P_∞ because $\pi(i,.)$ is an invariant probability measure. Now let $\tilde{D}$ be such a positive recurrent class. Then from 10.11 we have $\Pi_\infty c_\infty = g$ on $\tilde{D}$. If one proceeds as in subsection 10.12 for $\tilde{D}$ instead of D one finds a stationary optimal policy which has no disjoint closed sets. Consequently: *If the assumptions* C *and* D *are satisfied then there exists a stationary optimal policy for which the corresponding Markov chain has no disjoint closed sets.*

10.14. COROLLARY. *Each of the following three combinations of assumptions is also sufficient for the existence of a stationary optimal policy*: (C,E), (D,F), (E,F).

PROOF. In view of the comments on the assumptions and lemmas 10.2 and 10.3 one easily can show that each of the above combinations implies the assumptions (A,B) and/or (C,D). Hence by theorem 10.4 the assertion follows.□

10.15. AN INFINITE PERIOD STATIONARY INVENTORY MODEL WITH BACKLOGGING

We conclude this section by showing that in this model our theorem can be applied.

Let $\underline{y}_t$ denote the level of inventory at time t and let $\underline{\Delta}_t$ be the amount ordered after observing $\underline{y}_t$. Assume that delivery of the ordered units is instantaneous. Thus after the moment of ordering, the inventory level is $\underline{y}_t+\underline{\Delta}_t$. Suppose the sequence of demands $\underline{d}_t$, $t=1,2,\ldots$, for the product during each of the periods is a sequence of independent and identically distributed random variables with

$$\mathbb{P}\,[\underline{d}_t=j] = p_j \text{ for } j = 0,1,\ldots \text{ with } \sum_{j=0}^{\infty} p_j = 1.$$

We allow negative inventory, i.e. backlogging of demand, and consequently have a denumerable state space.

The decision which has to be made at times $t = 0,1,\ldots$ is the amount to be ordered. Now let $p_k(i,j)$ denote the transition probability to inventory level j when i units are available and k units are ordered. Then

$$p_k(i,j) = \mathbb{P}\,[\text{demand equals } i+k-j] = \begin{cases} p_{i+k-j} & \text{for } i+k \geq j \\ 0 & \text{otherwise.} \end{cases}$$

In all practical cases there will be a finite storage capacity. Also an infinitely large backlogging will not be convenient and so it seems that the following condition is natural. The set $K(i)$ of available ordering decisions in state i satisfies

(10.15.1) $\quad K(i) = \{k : a \leq i+k \leq b\}$ for all $i \in E$ for some integers a,b.

This relation implies that the collection of probability measures

(10.15.2) $\quad \{p_k(i,.) : i \in E,\ k \in K(i)\}$

is tight, and hence assumption F is satisfied. Indeed, given any $\varepsilon > 0$, let n be such that

$$\sum_{j=0}^{a+n} p_j \geq 1-\varepsilon,$$

then

$$\sum_{j=-n}^{b} p_k(i,j) = \sum_{j=0}^{i+k+n} p_j \geq 1-\varepsilon \text{ for all } i \in E \text{ and all } k \in K(i).$$

If $p_j > 0$, $j=0,1,\ldots$, then each stationary policy has no disjoint closed sets and assumption E is satisfied. It follows from this argument that corollary 10.14 applies.

A stationary rule which prescribes no ordering in state i when $i \geq s$ and prescribes an order of $S-i$ units when $i < s$ is called an (s,S) policy. It is easily seen that under an (s,S) policy the state space does not have disjoint closed sets.

Under certain conditions on the cost function it can be proved that there exist optimal (s,S) policies with respect to the expected discounted return (see for instance [Johnson],[Tijms] and [Veinott (1966)]). According to theorem 10.4, under those conditions there also exists an optimal (s,S) policy with respect to the average return criterion.

11. SIMULTANEOUS DOEBLINCONDITION

In this section we introduce a condition which can be seen as an extension of Doeblin's condition (cf. [Doob, p. 192]) to a collection of Markov chains. We call it the simultaneous Doeblincondition (sim D).

It will be shown that the condition sim D implies assumption C of section 10 and, moreover, sim D in combination with the condition that $(E,\mathcal{P})$ is a communicating system is sufficient for the existence of an optimal policy with respect to the average return criterion.

In the next section it will be pointed out that the condition sim D gives the connection between sufficient conditions for existence of average-optimal policies, which can be found in the literature and the conditions in section 10.

Although we restrict ourselves also in this section to a countable state space E, we shall introduce the Doeblincondition and an equivalent one for a general measurable space $(E,\mathcal{F})$.

CONDITION D (introduced in [Doeblin]). *There exist a finite measure* ϕ, *a positive integer* n *and a positive real number* ε *such that, for each* $A \in \mathcal{F}$, $\phi(A) \le \varepsilon$ *implies* $p^n(x,A) \le 1-\varepsilon$ *for all* x.

Actually Doeblin introduced this condition with ϕ Lebesgue measure on a Borelset with finite measure in a finite dimensional Euclidean space. In [Doob] this is generalized to a finite measure on a measurable space.

For P a transition probability function, the formula $Pf(x) = \int P(x,dy)\, f(y)$ defines a positive endomorphism on the Banach space B of bounded measurable functions on $(E,\mathcal{F})$ with $\|f\| = \sup_E |f(x)|$ (cf. [Neveu, p. 179]).

CONDITION K-B (introduced in [Kryloff and Bogolioubofff]). *There exist a compact endomorphism* Q *on the Banach space* B *and a positive integer* n *such that* $\|P^n-Q\| < 1$.

If this condition is satisfied then P is called *quasi-compact*.

In [Yosida and Kakutani] it is proved that the Doeblincondition with ϕ the Lebesgue measure implies the condition K-B. Moreover, they showed that for a quasi-compact transition probability function P the strong ergodic theorem holds. In [Neveu, p. 185] it is pointed out that the conditions D and K-B are equivalent.

It is rather easy to verify that condition D can be given in the following formulation if $p(x,E) = 1$ for all $x \in E$ (cf. [Neveu, p. 185]): *There exist a probability measure* μ *on* $(E,\mathcal{F})$, *a positive integer* n, *and two real numbers* $0 < \theta < 1$ *and* $\eta > 0$ *such that, for* $F \in \mathcal{F}$, $\mu(F) \geq \theta$ *implies* $P^n(x,F) \geq \eta$ *for all* $x \in E$.

When E is a countable set with $\mathcal{F}$ the σ-algebra of all subsets then this can be simplified to: *There exist a finite set* K, *a positive integer* n, *and a positive real number* c *such that* $P^n(i,K) \geq c$ *for all* $i \in E$.

Now we return to our collection of Markov matrices $\mathcal{P}$ and introduce the following condition.

11.1. SIMULTANEOUS DOEBLINCONDITION (sim D). *There exist a finite set* K, *a positive integer* n, *and a positive real number* c *such that* $p^n(i,K) \geq c$ *for all* $i \in E$ *and all* $P \in \mathcal{P}$.

It is easy to see that assumption F (section 10) implies the condition sim D. For our discrete state space E it is possible to give a more precise assertion.

11.2. LEMMA. *If* $p(i,E) = 1$ *for all* $i \in E$ *for a Markov matrix* P *then the collection* $\{p(i,.) : i \in E\}$ *is tight if and only if* P *is (strongly) compact.*

PROOF. Assume that $\{p(i,.) : i \in E\}$ is tight. To prove that P is compact we have to show (by definition cf. [Neveu, p. 179]) that the unit ball of the Banach space B has a relatively compact image under the endomorphism P. Since in a metric space each sequentially compact set is compact, it is sufficient to show that given any sequence of functions with $\|f_n\| \leq 1$ for $n=1,2,\ldots$, there exists a subsequence f_{n_k} such that Pf_{n_k} converges in norm as $k \to \infty$. Indeed by the diagonal procedure there is a subsequence f_{n_k} such that $\lim_{k\to\infty} f_{n_k}(i) = f(i)$ for some function f and all $i \in E$. Given any $\varepsilon > 0$ there are a finite set K and an integer N such that

$$p(i,K) > 1-\varepsilon \text{ for all } i \in E \text{ and } \max_{j\in K} |f_{n_k}(j)-f(j)| < \varepsilon \text{ for } k > N. \tag{11.2.2}$$

Hence for $k > N$

$$|\sum_j p(i,j) \{f_{n_k}(j)-f(j)\}| \le$$

$$\le |\sum_j p(i,j) f_{n_k}(j) - \sum_{j \in K} p(i,j) f_{n_k}(j)| +$$

$$+ |\sum_{j \in K} p(i,j) \{f_{n_k}(j)-f(j)\}| + |\sum_{j \in K} p(i,j) f(j) - \sum_j p(i,j) f(j)|,$$

where each of the three terms on the right-hand side of this inequality is less than ε for all $i \in E$ by relation (11.2.2) and the fact that $||f|| \le 1$ and $|| f_{n_k}|| \le 1$ for $k=1,2,\ldots$. This proves that Pf_n converges in norm to Pf.

To prove the converse let the sequence of finite subsets K_n, $n=1,2,\ldots$, be such that

$$K_n \subset K_{n+1}, \; n=1,2,\ldots, \text{ and } \bigcup_{n=1}^{\infty} K_n = E.$$

For

$$f_n(i) = \begin{cases} 1 \text{ for } i \in K_n \\ 0 \text{ for } i \notin K_n \end{cases}$$

we have that

$$\lim_{n\to\infty} Pf_n(i) = \lim_{n\to\infty} p(i,K_n) = 1 \text{ for all } i \in E.$$

Since P is compact it follows that this convergence is uniform in $i \in E$. Consequently, given any $\varepsilon > 0$ there exists a finite subset K_n with $p(i,K_n) \ge 1-\varepsilon$ for all $i \in E$ and hence the collection $\{p(i,.) : i \in E\}$ is tight. □

The infinite period stationary inventory model with backlogging, as treated in section 10, satisfies assumption F. Consequently there are non-trivial Markov decision processes for which sim D holds. The condition sim D with the triple (K,n,c) implies that the finite set K can be reached in n steps with probability at least c.

For the waiting line model, as introduced in section 5, it is clear that to reach state n from state m for $n < m$ takes as least m-n steps. Hence for this problem the condition sim D is not satisfied. However it is our opinion that for each honest Markovian decision problem there exist a subset A and a stationary policy R such that when using policy R outside the set A the embedded Markovian decision problem on A satisfies the condition sim D. And moreover, when using policy R outside the set A then each (nearly-)optimal policy remains (nearly-)optimal.

In the remainder of this section we will investigate properties of condition 11.1. As in section 10 we assume that $\mathcal{P}$ is compact and $p(i,E) = 1$ for all $i \in E$ and all $P \in \mathcal{P}$.

The following results are from [Hordijk (1972)] in which also an elementary proof of the strong ergodic theorem for discrete spaces can be found.

It is useful to have available the following notations and relations:

(11.2.3) $\quad {}_Ap^n(i,B) := \mathbb{P}_P\,[\underline{x}_n \in B,\ \underline{x}_m \notin A,\ 1 \le m < n \mid \underline{x}_0 = i]$

(11.2.4) $\quad m_P(i,A) := \sum_{n=1}^{\infty} {}_Ap^n(i,E)$

(11.2.5) $\quad {}_Ap^n(i,A^c) = {}_Ap^{n+1}(i,E).$

For $\underline{\tau}$ the reentry time of subset A (i.e. $\underline{\tau}$ is the least $n > 0$, if any, with $\underline{x}_n \in A$, and $\underline{\tau} = \infty$ if there is no such n) we find with $\mathbb{P}_{i,P}\,[\underline{\tau}>n] = {}_Ap^n(i,A^c)$ and the relations (11.2.5) and (2.7.3) that $\mathbb{E}_{i,P}\,[\underline{\tau}] = m_P(i,A)$.

As ${}_Ap^n(i,A)$ denotes the probability of reaching the set A for the first time at the n^{th} step, we have the relation

(11.2.6) $\quad \sum_{k=1}^{n} {}_Ap^k(i,A) + {}_Ap^{n+1}(i,E) = 1.$

11.3. THEOREM. *The following four conditions are equivalent*

a. *simultaneous Doeblincondition*;

b. *there exist a finite set* K, *an integer* N *and a positive real number* c *such that*

$$\sum_{n=1}^{N} {}_Kp^n(i,K) \ge c \text{ for all } i \in E \text{ and all } P \in \mathcal{P};$$

c. *there exist a finite set* K *and a real number* b *such that*

$$m_P(i,K) \le b \text{ for all } i \in E \text{ and all } P \in \mathcal{P};$$

d. *given any* $\varepsilon > 0$ *there exist a finite set* $K(\varepsilon)$ *and an integer* $N(\varepsilon)$ *such that*

$$P^{N(\varepsilon)}(i,K(\varepsilon)) \ge 1-\varepsilon \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

PROOF. Assume condition a is true for the triple (K,N,c). Then

$$(11.3.1) \qquad \sum_{n=1}^{N} {}_K p(i,K) = \mathbb{P}_{i,P}\,[\bigcup_{n=1}^{N} \{\underline{x}_n \in K\}] \ge \mathbb{P}_{i,P}\,[\underline{x}_N \in K] \ge c.$$

Hence the triple (K,N,c) satisfies condition b.

Next assume condition b is true for the triple (K,N,c). From relation (11.2.6) it then follows:

$${}_K P^{N+1}(i,E) \le 1-c \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

Since

$${}_K P^{n+m}(i,E) = \sum_{j\in K^c} {}_K P^n(i,j)\; {}_K P^m(j,E)$$

we obtain (cf. relation (9.1.8))

$$(11.3.2) \qquad {}_K P^n(i,E) \le (1-c)^{[n(N+1)^{-1}]}$$

for all $i \in E$, all $P \in \mathcal{P}$ and all $n \in \{1,2,\ldots\}$. Hence

$$(11.3.3) \qquad m_P(i,K) = \sum_{n=1}^{\infty} {}_K P^n(i,E) \le (N+1)c^{-1} \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

Consequently condition c is satisfied for $(K,(N+1)c^{-1})$.

Next assume condition c is true for (K,b). Given any $\delta > 0$ there is an integer M such that (recall that ${}_K P^n(i,E)$ is nonincreasing in n)

$${}_K P^{M+1}(i,E) \le \delta \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

In view of relation (11.2.6) then

$$(11.3.4) \qquad \sum_{n=1}^{M} {}_Kp^n(i,K) \geq 1-\delta \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

In the beginning of section 10 we pointed out that as a consequence of the compactness of $\mathcal{P}$ the collection $\{p(i,.) : P \in \mathcal{P}\}$ is tight for each $i \in E$. Moreover, since for any integer k and state i the probability measure $p^k(i,.)$ is a continuous function of P we have that $\{p^k(i,.) : P \in \mathcal{P}\}$ is tight for all $i \in E$ and all $k \in \{1,2,\ldots\}$. Because the union of a finite number of tight collections is again a tight collection, it then follows that $\{p^n(i,.) : i \in K, 1 \leq n \leq M \text{ and } P \in \mathcal{P}\}$ is tight. Consequently there exists a finite set A such that

$$(11.3.5) \qquad p^n(i,A) \geq 1-\delta \text{ for all } i \in K, \text{ all } n \text{ with } 1 \leq n \leq M \text{ and all } P \in \mathcal{P}.$$

Using the "first entrance decomposition" of the set K we find

$$p^{M+1}(i,A) = {}_Kp^{M+1}(i,A) + \sum_{n=1}^{M} \sum_{j \in K} {}_Kp^n(i,j)\, p^{M+1-n}(j,A).$$

According to the relations (11.3.4) and (11.3.5) the last term of this relation is at least $(1-\delta)^2$. Hence given any $\varepsilon > 0$, we choose $\delta > 0$ such that $(1-\delta)^2 \geq 1-\varepsilon$, then the condition d is satisfied with $N(\varepsilon) = M+1$ and $K(\varepsilon) = A$.

It is evident that the condition d implies the condition a. □

With the above theorem we can prove that assumption C of section 10 is implied by the condition sim D. We can even prove the following stronger result.

11.4. THEOREM. *The condition* sim D *implies that the collection* $\{\pi_P(i,.) : i \in E, P \in \mathcal{P}\}$ *is tight.*

PROOF. Given any $\varepsilon > 0$, there exist by theorem 11.3d a finite set K and an integer N such that

$$p^N(i,K) \geq 1-\varepsilon \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

Hence (cf. the proof of lemma 10.3) for all $n > N$ the same relation holds.

Consequently

$$\pi_P(i,K) = \lim_{k\to\infty} \frac{1}{k} \sum_{n=1}^{k} p^{N+n}(i,K) \geq 1-\varepsilon$$

for all $i \in E$ and all $P \in \mathcal{P}$. □

The following lemma is related to proposition 6.1 of [Orey, p. 29].

11.5. LEMMA. *If for some subset* A, *some positive integer* N *and a positive number* c *it holds that*

(11.5.1) $$\sum_{n=1}^{N} {}_Ap^n(i,A) \geq c \text{ for all } i \in E \text{ and all } P \in \mathcal{P},$$

then

(11.5.2) $$\sum_{n=1}^{m} {}_Ap^n(i,A) \to 1$$

uniformly in i *and* P *as* $m \to \infty$, *and*

(11.5.3) $$\sum_{n=1}^{\infty} {}_Ap^n(i,E)$$

is uniformly bounded in i *and* P.

PROOF. The proof proceeds similar to the proof of "b implies c" in theorem 11.3. Indeed, similar to (11.3.2) we have that

(11.5.4) $${}_Ap^n(i,E) \leq (1-c)^{[n(N+1)^{-1}]} \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

In view of (11.2.6) we then have

$$\sum_{n=1}^{m-1} {}_Ap^n(i,A) \geq 1 - (1-c)^{[m(N+1)^{-1}]} \text{ for all } i \in E \text{ and all } P \in \mathcal{P}.$$

From this the first assertion follows.

Similar to (11.3.3) we obtain from (11.5.4) that

$$\sum_{n=1}^{\infty} {}_Ap^n(i,E) \leq (N+1)c^{-1} \text{ for all } i \in E \text{ and all } P \in \mathcal{P},$$

which proves the second assertion. □

As a consequence of the next lemma we have that if for some Markov chain with matrix of transition probabilities P, some subset A can be reached from each state i then the Doeblincondition implies relation (11.5.1) with $\mathcal{P} = \{P\}$. Hence the Markov chain is uniformly ϕ-recurrent (in the terminology of [Orey]) with ϕ the counting measure if all the states are communicating. [Orey, proposition 6.1 (p. 26) and theorem 7.1 (p. 30)] provides then another proof of the strong ergodic theorem in this case.

11.6. LEMMA. *If some set* A *can be reached from each state* i *under each stationary policy then the condition* sim D *implies the relation* (11.5.1).

PROOF. Assume that the condition sim D holds with triple (K,n,d). So

(11.6.1) $\quad p^n(i,K) \geq d$ for all $i \in E$ and all $P \in \mathcal{P}$.

For $i \in E$ and $P \in \mathcal{P}$ we define

$$n(i,P) := \min\{n \geq 0 : p^n(i,A) > 0\}.$$

It can be seen that for each $i \in E$, $n(i,P)$ is an upper semicontinuous function of P and hence attains its supremum over the compact set $\mathcal{P}$. For $i \in E$ let

$$n(i) := \max_{\mathcal{P}} n(i,P).$$

Since the set A can be reached from each state i under each stationary policy we find $n(i) < \infty$ for all $i \in E$. Because the set K is finite we have

$$m := \max_{i \in K} n(i) < \infty.$$

The sum $\sum_{k=1}^{m} {}_A p^k(i,A)$ is a continuous function of P. Moreover, from the definition of m it follows that this sum is positive for all $i \in K$ and all $P \in \mathcal{P}$. For $i \in K$ define

$$\varepsilon(i) := \min_{\mathcal{P}} \sum_{k=1}^{m} {}_A p^k(i,A).$$

Then $\varepsilon(i)$ is positive for all $i \in K$. Hence, so is

$$\varepsilon := \min_{i \in K} \varepsilon(i).$$

Using relation (11.6.1) we find for arbitrary state $i \in E$ and arbitrary matrix $P \in \mathcal{P}$

$$\sum_{k=1}^{n+m} {}_A p^k(i,A) \geq \sum_{j \in K} p^n(i,j) \sum_{k=1}^{m} {}_A p^k(j,A) \geq d\varepsilon.$$

Hence relation (11.5.1) is satisfied with $N = n+m$ and $c = d\varepsilon$. □

Using the foregoing lemmas we shall prove in the next theorem that the condition sim D guarantees any communicating system to be a recurrent system (cf. section 8).

11.7. THEOREM. *If the condition* sim D *is satisfied, then the system is recurrent if and only if it is communicating.*

PROOF. Since in general the property recurrent is stronger than communicating we have only to prove that the latter implies the former. Assume the system is communicating. Let i_0 be an arbitrary state. Using only the assumption that the system is communicating, we shall prove that for some stationary policy Q^∞ state i_0 can be reached from each state $i \in E$. The proof proceeds by induction. Let $E = \{i_0, i_1, i_2, \ldots\}$ and assume that i_0 can be reached from states $i_1, \ldots, i_m$ under policy P^∞. Hence there are states (called j-states)

(11.7.1) $\quad j_{kn}$ for $k = 1, \ldots, m$ and $1 \leq n \leq n_k$

such that

(11.7.2) $\quad p(i_k, j_{k1})\, p(j_{k1}, j_{k2}) \ldots p(j_{k(n_k-1)}, j_{kn_k})\, p(j_{kn_k}, i_0) > 0.$

Since the system is communicating the state i_0 can be reached from state i_{m+1} under some policy P_*. Now there are two possibilities:

a. Going from state i_{m+1} to state i_0 we reach state i_0 without passing any of the j-states of (11.7.1). In this case we can take matrix $P_{**} \in \mathcal{P}$ (recall that $\mathcal{P}$ has the product property (cf. p. 1)) such that P_{**} is equal to P in the j-states and is equal to P_* in the other states. Then i_0 can be reached from states $i_1, \ldots, i_{m+1}$ under policy P_{**}.

b. Going from state i_{m+1} to state i_0 we pass a j-state. Let j^* be the j-state which is passed first when going from i_{m+1} to i_0. Because j^* can be reached from i_{m+1} under policy P_*^∞ and i_0 can be reached from j^* under policy P^∞ we have that state i_0 can be reached from state i_{m+1} under the policy P_{**}^∞ as introduced under a.

This completes the proof that for some policy Q^∞ state i_0 can be reached from each state $i \in E$. Applying the lemmas 11.5 and 11.6 for $P = \{Q\}$ (so P is a collection consisting of one element) we obtain with $A = \{i_0\}$

$$f_Q(i,i_0) = \sum_{n=1}^{\infty} {}_{i_0}q^n(i,i_0) = 1 \text{ for all } i \in E \text{ with } i \neq i_0.$$

Hence the state i_0 is recurrent. □

We conclude this section by collecting some other combinations of conditions which imply the assumptions of theorem 10.4.

11.8. THEOREM. *Each of the following two conditions implies the existence of a stationary optimal policy with respect to the average expected return*

the system is communicating and the condition sim D *holds*

the assumption E *of section* 10 *and the condition* sim D *hold.*

PROOF. The first assertion follows from the theorems 10.4, 11.4 and 11.7. The second assertion is a consequence of lemma 10.2 and the theorems 10.4 and 11.4. □

It is evident that for a finite state space the condition sim D is always satisfied. As a consequence of the first part of the above theorem we obtain that for a finite state space it is sufficient for the existence of a stationary optimal policy that the system is communicating. This result was obtained independently in [Bather].

12. CONNECTION WITH THE WORK OF DERMAN, ROSS, TAYLOR AND VEINOTT

In this section we point out some of the relations between conditions introduced in [Derman (1966)], [Derman and Veinott], [Taylor] and [Ross (1968)] and the assumptions made in section 10. In our opinion the condition sim D plays a basic role here.

In the second part of the section another assumption which implies the existence of an average-optimal policy is given. The section concludes by answering a question raised in section 10.

12.1. In the above given references it is assumed that in each state there is only a finite number of possible decisions. In our notation we then have

$$\mathcal{P}(i) := \{p(i,.) : P \in \mathcal{P}\}$$

is a finite set of probability measures for all $i \in E$. It is now easily deduced that $\mathcal{P}$ is compact and $c_{\mathcal{P}}$ is continuous.

Since all results from the literature to be cited in this section can be generalized to infinite sets $\mathcal{P}(i)$ such that $\mathcal{P}$ is compact and $c_{\mathcal{P}}$ is continuous we use these assumptions from the outset.

12.2. In [Derman (1966)] it was proved that the following four conditions together imply the existence of a constant g and a bounded function v such that

$$v = \max_{\mathcal{P}} (c_P - ge + Pv).$$

In the sequel of this section we shall call the pair (g,v) a (bounded) solution of the optimality equation.

I. $c_{\mathcal{P}}$ *is bounded*;

II. E *is a positive recurrent class for each* $P \in \mathcal{P}$;

III. *for each* $P \in \mathcal{P}$ *there exist a constant* g_P *and bounded function* v_P *such that* $v_P = c_P - g_P e + Pv_P$;

IV. *there exist constants* b_1, b_2 *such that* $|g_P| \le b_1$ *and* $|v_P(i)| \le b_2$ *for all* $i \in E$ *and all* $P \in \mathcal{P}$.

It can also be found in [Derman (1966)] that a bounded solution of the optimality equation implies the existence of a stationary optimal policy. Indeed, iterating the inequality

$$c_P - ge + Pv \le v, \qquad P \in \mathcal{P}$$

we obtain

$$\sum_{n=0}^{N} P_0 \cdots P_{n-1} (c_{P_n} - ge) + P_0 \cdots P_N v \le v \text{ for all } P_0, P_1, \ldots, P_N \in \mathcal{P}.$$

Hence

$$(12.2.1) \qquad \frac{1}{N+1} \sum_{n=0}^{N} P_0 \cdots P_{n-1} c_{P_n} \le ge + \frac{1}{N+1} (v - P_0 \cdots P_n v).$$

Consequently ge is an upper bound of the set of limitpoints of the first term. Moreover, for

$$\Pi_P := \lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} P^n$$

we have by the condition II

$$\Pi_P P = P \Pi_P = \Pi_P \Pi_P = \Pi_P.$$

And hence if Q satisfies the optimality equation, i.e.

$$c_Q - ge + Qv = v,$$

then by multiplying with Π_Q we obtain

$$\Pi_Q c_Q = ge.$$

From which it follows that Q^∞ is optimal.

12.3. From theorem 1 in [Derman and Veinott] it follows that conditions I and II together with the condition

V. *there exists some state* i_0 *such that the expected number* $m_P(i, i_0)$ *of steps from state* i *to state* i_0 *under policy* P^∞ *is uniformly bounded* *in* i *and* P,

imply the conditions III and IV. If in addition to condition II the return-time from state i_0 to state i_0 has a finite second moment (in [Kemeny, Snell and Knapp, p. 274] this is called strong ergodicity) then conversely the conditions III and IV for every bounded cost structure imply condition V. This can be shown by using theorem 2 of [Derman and Veinott].

12.4. In [Taylor] the following condition is introduced (cf. [Taylor, lemma 3.2, p. 1684]).

VI. $v_\alpha(i) - v_\alpha(j)$ *with* $v_\alpha(i) := \sup_P \sum_{n=0}^{\infty} \alpha^n P^n c_P$ *is uniformly bounded in* $i,j \in E$ *and* $\alpha \in (0,1)$.

As follows from arguments in [Taylor] the condition VI implies the existence of a bounded solution of the optimality equation (see also [Ross (1968)]). Indeed, since (cf. theorem 6.1)

$$v_\alpha(i) = \max_P [c_P(i) + \alpha \sum_j p(i,j)\, v_\alpha(j)] \text{ for all } i \in E,$$

we obtain by subtracting $v_\alpha(0)$ from both sides

$$v_\alpha(i) - v_\alpha(0) =$$

$$= \max_P [c_P(i) - (1-\alpha)\, v_\alpha(0) + \alpha \sum_j p(i,j)\, (v_\alpha(j) - v_\alpha(0))].$$

From I and VI we have that $(1-\alpha)\, v_\alpha(0)$ and $v_\alpha(i) - v_\alpha(0)$ are uniformly bounded in α and i. The diagonal procedure then provides a sequence $\{\alpha_n\}$ with $0 < \alpha_n < 1$, $\alpha_n \to 1$ as $n \to \infty$ and a constant g together with a bounded function v such that

$$\lim_{n\to\infty} (1-\alpha_n)\, v_{\alpha_n}(0) = g \text{ and } \lim_{n\to\infty} v_{\alpha_n}(i) - v_{\alpha_n}(0) = v(i).$$

Hence

$$v(i) = \max_P [c_P(i) - g + \sum_j p(i,j)\, v(j)] \text{ for all } i \in E.$$

12.5. In [Ross (1968)] it is proved that the conditions I and V together imply the condition VI.

12.6. According to theorem 11.3 we have that condition V implies the condition sim D. Moreover, under the assumption that state i_0 can be reached from any state i under any policy P^∞ we have in view of the lemmas 11.6 and 11.5 and relation (11.2.4) that the condition V is equivalent to the condition sim D. From lemma 10.2 and theorem 11.4 it follows also that the condition V implies the continuity of Π_P as a function of P. Thus theorem 10.4 applies in this case also. However, it follows from the results of the sections 10 and 11 that the existence of a state which is always accessible (i.e. from each state under each stationary policy) is an unnecessarily strong assumption. It seems to us that in cases where some state i_0 is always accessible, the approach of section 5 is better. Theorem 5.1 allows unbounded cost structures as well and for bounded cost structures one does not need to be sure beforehand that state i_0 is positive recurrent. On the contrary, relation (5.1.1) can serve as a criterion for uniform positive recurrency (see subsection 5.13).

If sim D holds and state i_0 is always accessible then the assumptions of theorem 5.1 are satisfied for each bounded cost structure. Indeed, according to the lemmas 11.6 and 11.5 and relation (11.2.4) we have for some constant b_1 and all $P \in \mathcal{P}$

$$\sum_{n=0}^{\infty} \tilde{P}^n e \leq b_1 e,$$

where P denotes the column-restriction of P to $E\backslash\{i_0\}$ (cf. subsection 2.7). Hence

$$y := \sup_{\mathcal{P}} \sum_{n=0}^{\infty} \tilde{P}^n e$$

is bounded and satisfies (cf. theorems 6.1 and 13.6)

$$(12.6.1) \qquad y = \sup_{\mathcal{P}} (e+\tilde{P}y).$$

Consequently for $|c_P(i)| \leq b_2$ for all $i \in E$ and all $P \in \mathcal{P}$ we have with $y^* = (b_2+1)y$ and $t_P = e$, $P \in \mathcal{P}$,

$$(12.6.2) \qquad |c_P| + t_P + \tilde{P}y^* \leq y^* \text{ for all } P \in \mathcal{P}.$$

Since y^* is bounded it is obvious that the relations (5.1.2) and (5.1.3) are also satisfied.

The condition sim D implies (12.6.2) for some bounded function y^*. When c_P is bounded away from zero (i.e. for some constant a $|c_P(i)| \geq a > 0$ for all $i \in E$ and all $P \in \mathcal{P}$) then also the converse is true. In this case (12.6.2) for bounded y^* implies the condition sim D. Indeed, then the function y defined in (12.6.1) is bounded and hence $m_P(i,i_0)$ is uniformly bounded in i and P, and according to theorem 11.3 the condition sim D is valid.

Using the following lemma we obtain in theorem 12.8 another condition implying the existence of a stationary optimal policy w.r.t. the average return criterion.

12.7. LEMMA. *If for some constant* b $|c_P(i)| \leq b$ *for all* $i \in E$ *and all* $P \in \mathcal{P}$ *then*

(12.7.1) $$v_\alpha(i) - v_\alpha(j) \geq -2b \inf_R m_R(i,j) \text{ for all } \alpha \in (0,1),$$

where $m_R(i,j)$ *denotes the expected number of steps from state* i *to state* j *under policy* R.

PROOF. This proof is related to the proof in [Ross (1968), theorem 1.4] (cf. [Ross (1970), theorem 6.19, p. 148]). According to definition 2.14 the function v_α is c_P-excessive w.r.t. $\mathcal{P}^* := \{\alpha P : P \in \mathcal{P}\}$. It follows then from theorem 2.21 that for any Markov time $\underline{\tau}$ and policy R we have

$$v_\alpha(i) \geq \mathbb{E}_{i,R}\left[\sum_{n=0}^{\tau-1} \alpha^n c(\underline{x}_n) + \alpha^{\underline{\tau}} v_\alpha(\underline{x}_{\underline{\tau}})\right].$$

For $\underline{\tau}$ the entry time of $\{j\}$ we can weaken this inequality to

(12.7.2) $$v_\alpha(i) - v_\alpha(j) \geq -b\, \mathbb{E}_R\, \underline{\tau} - (1 - \mathbb{E}_R\, \alpha^{\underline{\tau}})\, |v_\alpha(j)|.$$

By Jensen's inequality $\mathbb{E}_R\, \alpha^{\underline{\tau}} \geq \alpha^{\mathbb{E}_R \underline{\tau}}$ and also $1 - \alpha^x \leq (1-\alpha)x$ for $x \geq 1$ and $0 < \alpha < 1$, thus

(12.7.3) $$(1 - \mathbb{E}_R\, \alpha^{\underline{\tau}}) \leq (1-\alpha)\, \mathbb{E}_R\, \underline{\tau}.$$

Because $(1-\alpha)\ |v_\alpha(j)| \le b$ we then find by substituting (12.7.3) in (12.7.2) the relation (12.7.1). □

This lemma is used in theorem 12.8 to provide another condition implying the condition VI.

12.8. THEOREM. *If* c_P *is bounded and for some constant* a

$$(12.8.1) \qquad \inf_R m_R(i,j) \le a \text{ for all } i,j \in E,$$

then there exists a bounded solution of the optimality equation and hence a stationary optimal policy.

PROOF. It is evident from lemma 12.7 that the assumptions of the theorem imply the condition VI. The rest of the proof proceeds as in 12.4 and 12.2. □

12.9. REMARK. If E is a finite set and for some policy R state j can be reached from each state $i \in E$ then $m_R(i,j) < \infty$ for all $i \in E$. Hence for a communicating system and finite set E we have (cf. the proof of theorem 11.7)

$$(12.9.1) \qquad \max_{i,j\in E} \inf_R m_R(i,j) < \infty.$$

Consequently in this case theorem 12.8 applies.

12.10. REMARK. It can be seen from the subsections 12.2 and 12.4 that under the conditions I, II and VI we have for each sequence of discountfactors tending to one, a subsequence $\{\alpha_n\}$ such that

$$\lim_{n\to\infty} (1-\alpha_n)\ v_{\alpha_n}(i) = \sup_R \limsup_{N\to\infty} \frac{1}{N+1} \mathbb{E}_{i,R} \left[\sum_{n=0}^{N} c(\underline{x}_n)\right] = \Pi_Q c_Q(i)$$

for some $Q \in P$ and all $i \in E$. Hence $\lim_{\alpha\uparrow 1} (1-\alpha)\ v_\alpha(i)$ exists for all $i \in E$ and does not depend on i.

In the rest of this section we shall prove

$$(12.10.1)\qquad \sup_R \liminf_{N\to\infty} \frac{1}{N+1}\mathbb{E}_{i,R}[\sum_{n=0}^{N} c(\underline{x}_n)] = \lim_{\alpha\uparrow 1}(1-\alpha)\, v_\alpha(i) = $$

$$= \sup_R \limsup_{N\to\infty} \frac{1}{N+1}\mathbb{E}_{i,R}[\sum_{n=0}^{N} c(\underline{x}_n)]$$

for all $i \in E$ (and $\lim_{\alpha\uparrow 1}(1-\alpha)\, v_\alpha(i)$ does not depend on i), under weaker conditions. We assume in the sequel of this section that c_P is bounded.

12.11. LEMMA. *If*

$$(12.11.1)\qquad \sup_{i\in E} \inf_R m_R(i,j) < \infty \text{ for all } j \in E$$

then

$$(12.11.2)\qquad \sup_R \limsup_{N\to\infty} \frac{1}{N+1}\mathbb{E}_R\ [\sum_{n=0}^{N} c(\underline{x}_n)] \le \liminf_{\alpha\uparrow 1}(1-\alpha)\, v_\alpha.$$

PROOF. Given an arbitrary state j we choose the sequence $\{\alpha_n\}$ of discount-factors such that

$$(12.11.3)\qquad \lim_{n\to\infty}(1-\alpha_n)\, v_{\alpha_n}(j) = \liminf_{\alpha\uparrow 1}(1-\alpha)\, v_\alpha(j).$$

According to lemma 12.7 the difference $v_\alpha(i) - v_\alpha(j)$ is bounded uniformly in $\alpha \in (0,1)$. Consequently, there is a subsequence $\{\alpha_m^*\}$ of $\{\alpha_n\}$ such that for some constant g and some function v

$$(12.11.4)\qquad \lim_{m\to\infty}(1-\alpha_m^*)\, v_m(j) = g$$

and

$$(12.11.5)\qquad \lim_{m\to\infty}\{v_m(i) - v_m(j)\} = v(i) \text{ for all } i \in E,$$

where $v_m(i) = v_{\alpha_m^*}(i)$.

In view of (12.11.1) and lemma 12.7 we find that v is bounded from below. Hence for each $P \in \mathcal{P}$ we have $Pv^- < \infty$ and Pv can be defined as $Pv^+ - Pv^-$ and is possibly $+\infty$. As in 12.4 we obtain that

$$(12.11.6)\qquad c_P - ge + Pv \le v \text{ for all } P \in \mathcal{P}$$

(hence $Pv < \infty$). This relation implies (12.2.1) and hence (recall that v is bounded from below) for an arbitrary policy R it holds that

$$(12.11.7) \qquad \limsup_{N\to\infty} \frac{1}{N+1} \mathbb{E}_R \left[\sum_{n=0}^{N} c(\underline{x}_n) \right] \le g.$$

The relations (12.11.3), (12.11.4) and (12.11.7) together imply (12.11.2) since state j and policy R were arbitrarily chosen. □

12.12. COROLLARY. *If* $(E,\mathcal{P})$ *is a communicating system and each* $P \in \mathcal{P}$ *satisfies the Doeblincondition then the relation* (12.11.2) *is true.*

PROOF. We shall show that relation (12.11.1) is satisfied. Given any state j there is a stationary policy Q^∞ such that j can be reached from each state $i \in E$ under policy Q^∞ (see the proof of theorem 11.7). Since Q satisfies the Doeblincondition it follows from the lemmas 11.6 and 11.5 and the relation (11.2.4) with $\mathcal{P} = \{Q\}$ and $A = \{j\}$ that $\sup_i m_Q(i,j) < \infty$. □

12.13. THEOREM. *If* $(E,\mathcal{P})$ *is a communicating system and if the condition* sim D *holds then the relation* (12.10.1) *is valid.*

PROOF. It is straightforward from lemma 12.7 that

$$\lim_{\alpha\uparrow 1} (1-\alpha) [v_\alpha(i)-v_\alpha(j)] = 0 \text{ for all } i,j \in E.$$

Consequently the function x as introduced in theorem 10.4 is a constant function. It follows from the proof of theorem 10.4, in particular the subsections 10.10 and 10.11, that (cf. (10.0.3))

$$(12.13.1) \qquad \sup_R \liminf_{N\to\infty} \frac{1}{N+1} \mathbb{E}_R \left[\sum_{n=0}^{N} c(\underline{x}_n) \right] = x.$$

Because we can start in subsection 10.5 with an arbitrary sequence of discountfactors tending to one, it follows from (12.13.1) that $\lim_{\alpha\uparrow 1} (1-\alpha)\, v_\alpha$ exists and equals the left-hand side of (12.13.1). In view of the relation (12.11.2) we obtain that the relation (12.10.1) is valid. □

12.14. REMARK. Define for $N = 0,1,\ldots$

$$w_{N+1} := \sup_R \frac{1}{N+1} \mathbb{E}_R \left[\sum_{n=0}^{N} c(\underline{x}_n) \right].$$

In [Hordijk (1973)] it is proved that if for constants c and α_0

$$|(1-\alpha_1)\, v_{\alpha_1}(i) - (1-\alpha_2)\, v_{\alpha_2}(i)| \leq |\alpha_1 - \alpha_2| c$$

for all $i \in E$ and all α_1, α_2 with $\alpha_0 < \alpha_1 < \alpha_2 < 1$, which is certainly satisfied if $(1-\alpha)\, v_\alpha$ has a bounded derivative with respect to α, then $\lim_{n\to\infty} w_n$ exists and, moreover,

$$\lim_{n\to\infty} w_n = \lim_{\alpha\uparrow 1} (1-\alpha)\, v_\alpha.$$

13. RANDOMIZATION AND NEARLY OPTIMAL POLICIES

In this section we will collect several results on various topics which were needed in the foregoing sections.

We write E for the set of all stochastic matrices on state space E. Let the function d on pairs $(P_1,P_2) \in E$ be defined by

$$d(P_1,P_2) = \sum_{i,j} 2^{-(i+j)} |p_1(i,j)-p_2(i,j)|,$$

where for convenience we have identified the state space E with the set of positive integers. It can be seen that this function d defines a metric on P and, moreover, that P with this metric is a separable metric space. The weak convergence defined in section 1 is convergence with respect to this metric.

As usual, we will call an element of the smallest σ-algebra containing all open subsets of E, a Borel set. We assume that P is a Borel set.

For notational convenience we introduce a set A of actions such that there is a one-to-one correspondence between A and P. We use the set A to index P, i.e. P_a, $a \in A$, is that P which is in correspondence with a. Then (A,F), with F the Borel subsets of A, is a measurable space and P_a is a measurable mapping.

We write $M(A)$ for the set of all probability measures on (A,F). Define

$$\hat{P} = \{\hat{P} : \hat{p}(i,.) = \int_A p_a(i,.)\, d\mu_i(a),$$

$$\mu_i(.) \in M(A) \text{ for all } i \in E\}.$$

Verbally $\hat{P}(i) := \{\hat{p}(i,.) : \hat{P} \in \hat{P}\}$ can be described as the set of all randomizations of the decisions in state i.

If $P(i)$ is a compact set then using the metric induced by d on $P(i)$, it can be seen that $P(i)$ is a compact, separable metric space. $\hat{P}(i)$ is a quotient space (cf. [Kelley, p. 97]) of the space of all probability measures on $P(i)$, say $M(P(i))$. It follows from a theorem of Prohorov (cf. [Billingsley, p. 37]) that $M(P(i))$ is relatively compact. By definition, $M(P(i))$ is closed and thus compact. Hence $\hat{P}(i)$ is compact. Consequently, if P is compact then $P(i)$ is compact for all $i \in E$, hence $\hat{P}(i)$ is compact for all $i \in E$ and according to a theorem of Tychonoff (cf. [Kelley, p. 143])

$\hat{\mathcal{P}}$ is compact with respect to the metric d.

In the introduction we have identified the decision to be taken with the probability measure on E that is induced by it. In practical problems there may be several decisions with the same probability measure but different costs. In order to fit our model we then have to choose an appropriate cost and to assign this cost to the probability measure. Since costs are maximized in our model, it is obvious that the supremum over the different costs is appropriate here.

We proceed in a similar way when allowing randomizations. For each $i \in E$ let $\tilde{c}_P(i)$ be the minimum of the concave functions on $\hat{\mathcal{P}}$ that majorize $c_P(i)$ on $\mathcal{P}$. Then under reasonable regularity conditions for $P_0 \in \hat{\mathcal{P}}$ (we write $c_a(i)$ resp. $\tilde{c}_a(i)$ for $c_{P_a}(i)$ resp. $\tilde{c}_{P_a}(i)$)

$$\text{(13.0.1)} \qquad \tilde{c}_{P_0}(i) = \sup \{\int_A c_a(i)\, d\mu(a) : \mu \in \mathcal{M}(A) \text{ and } p_0(i,.) = \int_A p_a(i,.)\, d\mu(a)\}$$

and

$$\text{(13.0.2)} \qquad \tilde{c}_{P_0}(i) \geq \int_A \tilde{c}_a(i)\, d\mu(a)$$

for all $\mu \in \mathcal{M}(A)$ with $p_0(i,.) = \int_A p_a(i,.)\, d\mu(a)$.

We shall investigate whether the value function of an optimal control problem remains the same when allowing randomizations of decision rules.

13.1. THEOREM. *If*

$$\text{(13.1.1)} \qquad w := \sup_R \mathbb{E}_R \sum_{n=0}^{\infty} c^-(\underline{x}_n) < \infty$$

and f *is a* $c_{\mathcal{P}}$*-excessive function, then* f *is also a* $\tilde{c}_{\hat{\mathcal{P}}}$*-excessive function.*

PROOF. According to definition 2.14 we have to verify that

$$\text{(13.1.2)} \qquad \tilde{c}_P,\ P \in \hat{\mathcal{P}}, \text{ is a charge structure w.r.t. } \hat{\mathcal{P}};$$

$$\text{(13.1.3)} \qquad \sum_{n=0}^{\infty} P_0 \cdots P_{n-1}\, \tilde{c}_{P_n} \leq f \text{ for all } P_0, P_1, \ldots \in \hat{\mathcal{P}};$$

(13.1.4) $\quad \tilde{c}_P + Pf \le f$ for all $P \in \hat{\mathcal{P}}$.

For an arbitrary function g (with $P|g| < \infty$ for all $P \in \mathcal{P}$) and $\mu \in \mathcal{M}(A)$ it holds that

(13.1.5) $$\int_A c_a(i)\, d\mu(a) + \int_A \sum_j p_a(i,j)\, g(j)\, d\mu(a) \le \sup_{\mathcal{P}} (c_P(i)+Pg(i))$$

for all $i \in E$. Hence with (13.0.1) we obtain

(13.1.6) $$\tilde{c}_P + Pg \le \sup_{\mathcal{P}} (c_P+Pg) \text{ for all } P \in \hat{\mathcal{P}}.$$

Relation (13.1.4) is a direct consequence of (13.1.6) and the fact that f is $c_{\mathcal{P}}$ - superharmonic.

By (13.1.1) and theorem 2.22 (with $\underline{\tau}\equiv\infty$) we have that w is $c_{\mathcal{P}}^-$ - superharmonic, i.e.

(13.1.7) $\quad c_P^- + Pw \le w$ for all $P \in \mathcal{P}$.

Since $c_P \le \tilde{c}_P$ for all $P \in \mathcal{P}$,

$$\tilde{c}_P^- + Pw \le w \text{ for all } P \in \mathcal{P}.$$

Hence from (13.1.5) with w instead of g we find

$$\tilde{c}_P^- + Pw \le w \text{ for all } P \in \hat{\mathcal{P}}.$$

Iterating this inequality we find for each positive integer N

$$\sum_{n=0}^{N} P_0 \dots P_{n-1}\, \tilde{c}_{P_n}^- + P_0 \dots P_N\, w \le w \text{ for all } P_0,\dots,P_N \in \hat{\mathcal{P}}.$$

Consequently for each $R = (P_0,P_1,\dots)$ with $P_n \in \hat{\mathcal{P}}$ for all n

(13.1.8) $$\sum_{n=0}^{\infty} P_0 \dots P_{n-1}\, \tilde{c}_{P_n}^- \le w.$$

Now assume that relation (13.1.2) does not hold. Then for some $(P_0,P_1,\dots)$ with $P_n \in \hat{\mathcal{P}}$ for all n and some state i_0 we have in view of (13.1.8)

$$\sum_{n=0}^{\infty} P_0 \cdots P_{n-1}\, \tilde{c}^{+}_{P_n}(i_0) = \infty = \sum_{n=0}^{\infty} P_0 \cdots P_{n-1}\, \tilde{c}_{P_n}(i_0).$$

Choose N_0 such that

$$\sum_{n=0}^{N_0} P_0 \cdots P_{n-1}\, \tilde{c}_{P_n}(i_0) > f(i_0) + w(i_0).$$

Define

$$x_N = \sup \{ \sum_{n=0}^{N} P_0 \cdots P_{n-1}\, c_{P_n} : P_0, \ldots, P_N \in \mathcal{P}\}$$

and

$$\hat{x}_N = \sup \{ \sum_{n=0}^{N} P_0 \cdots P_n\, \tilde{c}_{P_n} : P_0, \ldots, P_N \in \hat{\mathcal{P}}\}.$$

It can be shown by induction on N that $x_{N+1} = \sup_{\mathcal{P}} (c_P + Px_N)$ and $\hat{x}_{N+1} = \sup_{\hat{\mathcal{P}}} (\tilde{c}_P + Px_N)$ for all $N \in \{1,2,\ldots\}$. Hence with $x_0 = \hat{x}_0 = 0$ and using (13.1.6) it follows by induction on N that $x_N = \hat{x}_N$ for all $N \in \{0,1,\ldots\}$. In particular $x_{N_0} = \hat{x}_{N_0}$ and consequently there are matrices $Q_0, \ldots, Q_N \in \mathcal{P}$ such that

$$\sum_{n=0}^{N_0} Q_0 \cdots Q_{n-1}\, c_{Q_n} > f(i_0) + w(i_0).$$

Given any sequence $Q_{N_0+1}, Q_{N_0+2}, \ldots \in \mathcal{P}$ we have by (13.1.1)

$$\sum_{n=0}^{\infty} Q_0 \cdots Q_{n-1}\, c_{Q_n} > f(i_0).$$

This is in contradiction with the fact that f is a $c_{\mathcal{P}}$ - excessive function. Hence relation (13.1.2) is true.

Define

$$v = \sup_R \mathbb{E}_R \left[\sum_{n=0}^{\infty} c(\underline{x}_n) \right]$$

and

$$v^* = \sup_R \mathbb{E}_R \sum_{n=0}^{\infty} |c(\underline{x}_n)|.$$

Note that the assumptions of this lemma imply $-\infty < v(i) < +\infty$ and

$-\infty < v^*(i) < +\infty$. According to the theorems 2.22 (with $\underline{\tau}\equiv\infty$) and 2.21 and the nonnegativity of $|c_P|$ we find

$$(13.1.9) \qquad P_{a_0}\ldots P_{a_n} v^* \le v^* \text{ for all } n \in \{0,1,2,\ldots\}$$

and all $a_0,a_1,\ldots,a_n \in A$. In order to prove the relation (13.1.3) it is according to theorem 2.17 sufficient to show that for each sequence $P_0,P_1,\ldots \in \hat{P}$

$$(13.1.10) \qquad \lim_{n\to\infty} P_0P_1\ldots P_n f^- = 0.$$

Since f is c_P - excessive we have by (2.14.2) $v \le f$ and hence $f^- \le v^-$. Instead of (13.1.10) we shall prove the stronger relation

$$(13.1.11) \qquad \lim_{n\to\infty} P_0P_1\ldots P_n v^- = 0 \text{ for all } P_0,P_1,\ldots \in \hat{P}.$$

Choose an arbitrary sequence $\hat{P}_0,\hat{P}_1,\ldots \in \hat{P}$. In the rest of this proof $\hat{R} := (\hat{P}_0,\hat{P}_1,\ldots)$ is a fixed policy. For $n = 0,1,2,\ldots$ let $\hat{P}_n$ be obtained from μ_{ni}, $i \in E$, i.e.

$$(13.1.12) \qquad \hat{p}_n(i,.) = \int_A p_a(i,.)\, d\mu_{ni}(a) \text{ for all } i \in E.$$

We introduce the probability product space (cf. [Neveu, proposition V.1.1, p. 162])

$$(13.1.13) \qquad (\prod_{t=0}^{\infty} A_t, \bigotimes_{t=0}^{\infty} F_t, \mu),$$

where (A_t,F_t), $t = 0,1,\ldots$, are copies of (A,F) and the restriction of μ to $\Pi_{t=0}^{N} A_t$ is determined by the probabilities on rectangles $\Pi_{t=0}^{N} F_t$. These probabilities are given by

$$\int_{F_0}\cdots\int_{F_N} \sum_{i_1,\ldots,i_N} d\mu_{0i_0}(a_0)\, p_{a_0}(i_0,i_1)\, d\mu_{1i_1}(a_1)\ldots$$

$$\ldots p_{a_{N-1}}(i_{N-1},i_N)\, d\mu_{Ni_N}(a_N),$$

with i_0 some fixed state in E. Next we define a sequence of measurable functions on this product space by

$$g_n(a_0,a_1,\dots,a_n,\dots) := P_{a_0}P_{a_1}\dots P_{a_n}\, v^-(i_0).$$

According to the theorems 2.22 (with $\underline{\tau}\equiv\infty$) and 2.17 ($c_{\mathcal{P}}$ is a charge structure because $-\infty < v^* < +\infty$) we have $\lim_{n\to\infty} g_n = 0$ for all elements of $\Pi_{t=0}^{\infty} A_t$. Using a bounded convergence theorem on the product space we find with $v^- \le v^*$ and (13.1.9)

$$(13.1.14)\quad \lim_{n\to\infty} \int_{\Pi_{t=0}^{\infty} A_t} g_n(\omega)\, d\mu(\omega) = 0.$$

The relation (13.1.14) in the usual notation is

$$\lim_{n\to\infty} \hat{P}_0\dots\hat{P}_n\, v^-(i_0) = 0.$$

This completes the proof. □

In section 3 we proved that the supremum of the expected return over all policies including the non-memoryless is a $c_{\mathcal{P}}$ - excessive function (theorem 3.1). According to theorem 13.1 the function v is also a $\tilde{c}_{\hat{\mathcal{P}}}$ - excessive function when relation (13.1.1) is true. Consequently, including all policies defined on $\hat{\mathcal{P}}$, i.e. all randomized policies, does not increase the value function when (13.1.1) is satisfied.*)

The following theorem, which is adapted from [Derman and Strauch] and [Derman (1970)], makes it evident why we focussed attention on memoryless policies.

13.2. THEOREM. *Assume that* $\mathcal{P}$ *contains all randomized decision rules (i.e.* $\mathcal{P}=\hat{\mathcal{P}}$*). Given any sequence of policies* $R_1,R_2,\dots$ *and any sequence of non-negative real numbers* $a_1,a_2,\dots$ *with* $\sum_{i=1}^{\infty} a_i = 1$ *there exists for each state* $i_0 \in E$ *a memoryless policy* R_0 *such that*

$$(13.2.1)\quad \mathbb{P}_{R_0}\,[\underline{x}_n=i,\underline{y}_n\in F\,|\,\underline{x}_0=i_0] = \sum_{k=1}^{\infty} a_k\, \mathbb{P}_{R_k}\,[\underline{x}_n=i,\underline{y}_n\in F\,|\,\underline{x}_0=i_0]$$

for all $i \in E$, *all* $F \in \mathcal{F}$ *and all* $n \in \{0,1,2,\dots\}$; $\underline{y}_n$ *denotes the decision at time* n.

*) Randomization becomes important when constraints are introduced. Cf. Neyman-Pearson lemma [Lehmann, p. 63] and [Derman (1970), chapter 7].

PROOF. For any nonnegative integer n and each state $i \in E$ we define a randomized decision by

$$(13.2.2)\qquad \mu_{ni}(F) := \{\sum_{k=1}^{\infty} a_k \mathbb{P}_{R_k} [\underline{x}_n=i,\underline{y}_n\in F|\underline{x}_0=i_0]\}\cdot$$

$$\cdot\{\sum_{k=1}^{\infty} a_k \mathbb{P}_{R_k} [\underline{x}_n=i|\underline{x}_0=i_0]\}^{-1},$$

when the denominator is positive; otherwise let $\mu_{ni}(.)$ be an arbitrary probability measure on (A,F). For $n \in \{0,1,\ldots\}$ let $\hat{P}_n$ be the associated decision rule, i.e.

$$(13.2.3)\qquad \hat{p}_n(i,.) = \int_A p_a(i,.)\, d\mu_{ni}(a) \text{ for all } i.$$

Define R_0 as $(\hat{P}_0,\hat{P}_1,\ldots)$.

The proof of relation (13.2.1) proceeds by induction on n. For $n = 0$ and $i \neq i_0$ both sides of equality (13.2.1) are equal to zero. If $i = i_0$ then

$$\mathbb{P}_{R_0} [\underline{x}_0=i_0,\underline{y}_0\in F|\underline{x}_0=i_0] = \mathbb{P}_{R_0} [\underline{y}_0\in F|\underline{x}_0=i_0] =$$

$$= \mu_0(F|i_0) = \sum_{k=1}^{\infty} a_k \mathbb{P}_{R_k} [\underline{x}_0=i,\underline{y}_0\in F|\underline{x}_0=i_0].$$

Assume that relation (13.2.1) holds for $n = m$, i.e.

$$(13.2.4)\qquad \sum_{k=1}^{\infty} a_k \mathbb{P}_{R_k} [\underline{x}_m=i,\underline{y}_m\in F|\underline{x}_0=i_0] = \mu_{mi}(F)\, \hat{P}_0\ldots\hat{P}_{m-1}(i_0,i).$$

We first prove that

$$(13.2.5)\qquad \sum_{k=1}^{\infty} a_k \mathbb{P}_{R_k} [\underline{x}_{m+1}=j|\underline{x}_0=i_0] = \hat{P}_0\ldots\hat{P}_m\,(i_0,j).$$

Since

$$\mathbb{P}_{R_k} [\underline{x}_{m+1}=j|\underline{x}_0=i_0,\underline{x}_m=i,\underline{y}_m=a] = p_a(i,j)$$

for all $k \in \{1,2,\ldots\}$ we find (by conditioning on $\underline{x}_m$, $\underline{y}_m$ and (13.2.4)) that the left-hand side of (13.2.5) equals

$$\sum_i \hat{P}_0 \cdots \hat{P}_{m-1}\ (i_0,i) \int_A p_a(i,j)\ d\mu_{mi}(a).$$

Hence with (13.2.3) the relation (13.2.5) follows.

According to relation (13.2.2) we have

$$\sum_{k=1}^{\infty} a_k\, \mathbb{P}_{R_k}\ [\underline{x}_{m+1}=j, \underline{y}_{m+1} \in F \mid \underline{x}_0 = i_0] =$$

$$= \mu_{(m+1)j}(F) \sum_{k=1}^{\infty} a_k\, \mathbb{P}_{R_k}\ [\underline{x}_{m+1}=j \mid \underline{x}_0 = i_0].$$

In view of (13.2.5) the second part of this equality can be written as

$$\hat{P}_0 \cdots \hat{P}_m\ (i_0,j)\ \mu_{(m+1)j}(F).$$

This equals

$$\mathbb{P}_{R_0}\ [\underline{x}_{m+1}=j, \underline{y}_{m+1} \in F \mid \underline{x}_0 = i_0]. \ \square$$

We call a cost structure $c_{\mathcal{P}}$ concave if for each $i \in E$ and $\mu \in M(A)$ it holds that

$$(13.2.6) \qquad c_{P_0}(i) \geq \int_A c_a(i)\ d\mu(a) \text{ with } p_0(i,.) = \int_A p_a(i,.)\ d\mu(a).$$

13.3. COROLLARY. *If $c_{\mathcal{P}}$ is a concave charge structure and $\mathcal{P}$ contains all randomized decision rules, then*

$$(13.3.1) \qquad \sup_R \mathbb{E}_R\ [\sum_{n=0}^{\infty} |c(\underline{x}_n)|] < \infty.$$

PROOF. Assume that the relation (13.3.1) does not hold. Then there is a state i_0 and a sequence of policies R_k^* such that

$$(13.3.2) \qquad \mathbb{E}_{i_0,R_k^*}\ [\sum_{n=0}^{\infty} |c(\underline{x}_n)|] > 2^k.$$

Next we apply theorem 13.2 with $a_k = 2^{-k}$ and $R_k = R_k^*$ for $k = 1,2,\ldots$ and we obtain a policy R_0 satisfying (13.2.1). In view of (13.2.6) it follows that

$$\mathbb{E}_{i_0,R_0}\ |c(\underline{x}_n)| \geq \sum_{k=1}^{\infty} 2^{-k}\, \mathbb{E}_{i_0,R_k^*}\ |c(\underline{x}_n)| \text{ for all } n \in \{0,1,\ldots\}.$$

Hence with (13.3.2) we have

$$\sum_{n=0}^{\infty} \mathbb{E}_{i,R_0} \; |c(\underline{x}_n)| = \infty.$$

This is in contradiction with the assumption that c_P is a charge structure. □

A similar reasoning as in corollary 13.3 shows that the relation (13.1.1) is a necessary condition for $\tilde{c}_{\hat{P}}$ to be a charge structure w.r.t. $\hat{P}$. Hence the relation (13.1.1) is a necessary and sufficient condition for the c_P - excessive function f to be $\tilde{c}_{\hat{P}}$ - excessive.

The results of this section are also true for the optimal control problem. In section 6 we treated the total return model by introducing an auxiliary function r. Here we show that the converse is also true. Each optimal control problem can be converted into a total return model by introducing an auxiliary state s and defining $p_a(s,s) = 1$ and $c_a(s) = 0$, $a \in A$. So s is an absorbing state. Further we introduce a new action or decision τ which we identify with the stopping decision, i.e. $p_\tau(i,s) = 1$ and $c_\tau(i) = r(i)$, $i \in E$. Then a stationary strategy for the optimal control model becomes a stationary policy for the total return model.

13.4. LEMMA. *If* P *contains all randomized decision rules,* c_P *is a concave charge structure and* $\sup_{\underline{\tau}} \mathbb{E}_R \; |r(\underline{x}_{\underline{\tau}})| < \infty$ *for all policies* R, *then*

$$(13.4.1) \qquad v^* := \sup_{R,\underline{\tau}} \mathbb{E}_R \; [\sum_{n=0}^{\underline{\tau}-1} |c(\underline{x}_n)| + |r(\underline{x}_{\underline{\tau}})|] < \infty.$$

Moreover, for each policy R *and each Markov time* $\underline{\tau}$ *it holds that*

$$(13.4.2) \qquad \mathbb{E}_R \; v^*(\underline{x}_{\underline{\tau}}) < \infty.$$

REMARK. There is an asymmetry in the assumptions of this lemma. As to the policies we assume that the expectations of the absolute costs are finite for all policies, as to the Markov times we assume that the *supremum* of the absolute reward over all Markov times is finite. To get rid of this asymmetry one can use randomized Markov times. A randomized Markov time (stopping time) is obtained if at each time t one performs an auxiliary random experiment depending on $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_t$ in order to decide whether to stop or not. If $\mathbb{E}_R \; |r(\underline{x}_{\underline{\sigma}})|$ is finite for all randomized Markov times $\underline{\sigma}$ then the supremum of $\mathbb{E}_R\ |r(\underline{x}_{\underline{\tau}})|$ over all Markov times is finite and conversely.

PROOF. Converting the optimal control problem into a total return model, it is straightforward from corollary 13.3 and the above remark that the relation (13.4.1) is true.

Now suppose $\sum_j p(i,j)\ v^*(j) = \infty$ for some state i and matrix P. Then the policy R, as in the proof of theorem 3.1, would have an infinite absolute return, contradicting the relation (13.4.1). Hence

$$(13.4.3)\qquad Pv^* < \infty \text{ and } w_P := v^* - Pv^* \geq 0 \text{ for all } P \in \mathcal{P}.$$

With (13.4.3) it can be proved that for each bounded Markov time $\underline{\tau}$ (use induction on the upper bound of the Markov times and proceed as in lemma 2.19)

$$(13.4.4)\qquad v^* = \mathbb{E}_R\ [\sum_{n=0}^{\tau-1} w(\underline{x}_n) + v^*(\underline{x}_{\underline{\tau}})] \text{ for all policies R.}$$

For arbitrary policy R and Markov time $\underline{\tau}$ we have from (13.4.4) and the second part of (13.4.3)

$$\mathbb{E}_R\ v^*(\underline{x}_{\underline{\tau}}) = \lim_{n\to\infty} \mathbb{E}_R\ [v^*(\underline{x}_{\underline{\tau}})\ \chi(\underline{\tau}\leq n)] \leq v^* < \infty. \ \square$$

This section is concluded with an investigation of nearly optimal policies. The results collected here are adapted from [Blackwell (1967)], [Blackwell (1970)] and [Ornstein]. They are stated for the total return model. Using conversion of models it is obvious that analogue results hold for the optimal control problem. In the rest of this section we assume

$$\sup_R \mathbb{E}_R\ [\sum_{n=0}^{\infty} |c(\underline{x}_n)|] < \infty$$

(consequently $c_{\mathcal{P}}$ is a charge structure). For notational convenience we write

$$v_{R,\underline{\tau}} := \mathbb{E}_R\ [\sum_{n=0}^{\tau-1} c(\underline{x}_n)],\ v_R := \mathbb{E}_R\ [\sum_{n=0}^{\infty} c(\underline{x}_n)],$$

$$v_P := v_{P^\infty} \text{ and } v := \sup_R v_R.$$

13.5. DEFINITION. *Policy* R *is* ε*-optimal in state* i *if* $v_R(i) \geq v(i) - \varepsilon$. *If for any* $\varepsilon > 0$ *and any state* i *there is a stationary policy* Q^∞ *such that* $v_Q(i) \geq v(i) - \varepsilon$, *then we say that there exist stationary weak nearly optimal policies.*

Policy R *is* ε*-optimal if* $v_R(i) \geq v(i) - \varepsilon$ *for all* $i \in E$. *If for any* $\varepsilon > 0$ *there is a stationary policy* Q^∞ *which is* ε*-optimal, then we say that there exist stationary strong nearly optimal policies.*

13.6. THEOREM. *Each of the following three conditions is sufficient for the existence of stationary weak nearly optimal policies*

a. $\sup_{\mathcal{P}} \sum_{n=0}^{\infty} P^n e < \infty$ *and* $\limsup_{n\to\infty} P^n v \leq 0$ *for all* $P \in \mathcal{P}$;

b. $\sup_{\mathcal{P}} \sum_{n=0}^{\infty} nP^n c_P^- < \infty$ *and* c_P *is bounded*;

c. *the cost structure is nonnegative.*

PROOF. Assume condition a is valid. According to theorem 6.1 v satisfies Bellman's optimality equation

$$v = \sup_{\mathcal{P}} (c_P + Pv).$$

Now given any $\varepsilon > 0$ and any initial state i_0 choose Q such that

$$c_Q + Qv \geq v - \delta e, \tag{13.6.1}$$

with

$$\delta := \varepsilon\Big(\sup_{\mathcal{P}} \sum_{n=0}^{\infty} P^n e(i_0)\Big)^{-1}. \tag{13.6.2}$$

By iterating the inequality (13.6.1) we obtain

$$\sum_{n=0}^{N} Q^n c_Q + Q^{N+1} v \geq v - \delta \sum_{n=0}^{N} Q^n e \text{ for all } N \in \{1,2,\ldots\}. \tag{13.6.3}$$

Because of $\limsup_{n\to\infty} Q^n v \leq 0$, (13.6.2) and (13.6.3) imply

$$\sum_{n=0}^{\infty} Q^n c_Q(i_0) \geq v(i_0) - \varepsilon.$$

Hence Q is ε-optimal in state i_0.

Assume condition b is true. Given any $\varepsilon > 0$ and any initial state i_0 let policy R be such that

$$v_R(i_0) > v(i_0) - \frac{\varepsilon}{4} .$$

Let $0 < \alpha_0 < 1$ be such that

$$\mathbb{E}_{i_0,R} [\sum_{n=0}^{\infty} \alpha^n c(\underline{x}_n)] > v(i_0) - \frac{\varepsilon}{4}$$

for all α with $\alpha_0 \leq \alpha \leq 1$. Let α_1 with $0 < \alpha_1 < 1$ be such that

$$(13.6.4) \qquad (1-\alpha_1) \sup_P \sum_{n=0}^{\infty} nP^n c_P^-(i_0) < \frac{\varepsilon}{2} .$$

Choose an α with $\max(\alpha_0,\alpha_1) \leq \alpha < 1$. We apply the first part of the theorem for the discounted dynamic programming problem with discountfactor α. Hence there exists a Q with

$$\mathbb{E}_{i_0,Q} [\sum_{n=0}^{\infty} \alpha^n c(\underline{x}_n)] > \mathbb{E}_{i_0,R} [\sum_{n=0}^{\infty} \alpha^n c(\underline{x}_n)] - \frac{\varepsilon}{4} > v(i_0) - \frac{\varepsilon}{2} .$$

Because of $(1-\alpha^n) \leq (1-\alpha)n$ for $0 < \alpha < 1$ and $n = 1,2,\ldots$, we have with (13.6.4)

$$\sum_{n=0}^{\infty} (Q^n c_Q^- - \alpha^n Q^n c_Q^-) \leq (1-\alpha) \sum_{n=0}^{\infty} nQ^n c_Q^- < \frac{\varepsilon}{2} .$$

Consequently

$$\mathbb{E}_{i_0,Q} [\sum_{n=0}^{\infty} c(\underline{x}_n)] \geq v(i_0) - \varepsilon$$

and Q^∞ is ε-optimal in state i_0.

Assume condition c is satisfied. Given any $\varepsilon > 0$ and any initial state i_0 let R be such that

$$v_R(i_0) > v(i_0) - \frac{\varepsilon}{2} .$$

Let E_k, $k = 1,2,\ldots$, be finite subsets of E with $E_k \subset E_{k+1}$, $k = 1,2,\ldots$, and $\cup_{k=1}^{\infty} E_k = E$. For $\underline{\tau}_k$ the exit time of E_k, $k = 1,2,\ldots$, we have that $\lim_{k\to\infty} \underline{\tau}_k = \infty$. Hence

$$\lim_{k\to\infty} v_{R,\underline{\tau}_k} = v_R$$

and consequently for some k_0 we have $v_{R,\underline{\tau}_{k_0}}(i_0) > v(i_0) - \frac{\varepsilon}{2}$. Let us consider now the total return model with *finite* state space E_{k_0}. For this problem the cost structure is bounded. Since the cost function is nonnegative condition b is satisfied. Hence there exists a Q such that

$$v_{Q,\underline{\tau}_{k_0}}(i_0) > v_{R,\underline{\tau}_{k_0}}(i_0) - \frac{\varepsilon}{2}.$$

Hence

$$v_Q(i_0) \geq v_{Q,\underline{\tau}_{k_0}}(i_0) \geq v(i_0) - \varepsilon$$

and Q^∞ is ε-optimal in state i_0. □

As noted in the beginning of section 6 $\lim_{n\to\infty} P^n v$ with $P \in \mathcal{P}$ always exists and this limit is nonnegative. Hence the condition $\limsup_{n\to\infty} P^n v \leq 0$ for all $P \in \mathcal{P}$ is not weaker than assuming that $\lim_{n\to\infty} Pv = 0$ for all $P \in \mathcal{P}$.

13.7. THEOREM. *If* $c_P \geq 0$ *for all* $P \in \mathcal{P}$ *then given any* $\varepsilon > 0$ *there exists a stationary policy* Q^∞ *such that*

(13.7.1) $\quad v_Q(i) \geq (1-\varepsilon)\, v(i)$ *for all* $i \in E$.

If v *is bounded then there exist stationary strong nearly optimal policies.*

PROOF. The second assertion is an immediate consequence of the first one.

Choose an ε with $0 < \varepsilon < 1$. Let the elements of E be indexed by the positive integers, i.e. $E = \{i_1, i_2, \ldots\}$. To prove the first assertion we show the existence of sets E_k with $k = 1,2,\ldots$, and matrices $P_k \in \mathcal{P}$ with $k = 1,2,\ldots$ such that $i_k \in E_k$ for $k = 1,2,\ldots$,

(13.7.2) $\quad p_n(i,.) = p_m(i,.)$ for all $i \in E_n \cap E_m$

and for $\underline{\tau}_n$, the exit time of E_n,

$$(13.7.3) \qquad v_{P_n,\underline{\tau}_n}(i_n) \geq (1-\varepsilon_n)\, v(i_n)$$

with

$$(13.7.4) \qquad \varepsilon_n = \varepsilon(\tfrac{1}{2} + \ldots + \tfrac{1}{2^n}).$$

Let us assume for the moment that the relations (13.7.2) and (13.7.3) are proved. Define Q as follows

$$q(i_n,.) := p_n(i_n,.) \text{ for } n = 1,2,\ldots .$$

Then

$$v_Q(i_m) \geq v_{Q,\underline{\tau}_m}(i_m) = v_{P_m,\underline{\tau}_m}(i_m) \geq (1-\varepsilon)\, v(i_m)$$

for all $m \in \{1,2,\ldots\}$ and consequently, Q^∞ satisfies relation (13.7.1). The proof of (13.7.2) and (13.7.3) proceeds by induction on n. Assume $E_1,\ldots,E_n$ and $P_1,\ldots,P_n$ are known and satisfy (13.7.2.) and (13.7.3). Define

$$\mathcal{P}_n = \{P : p(i,.) = p_k(i,.) \text{ if } i \in E_k \text{ for } k = 1,2,\ldots,n\},$$

let C_n be the set of policies with decision rules in $\mathcal{P}_n$ and take

$$v_n = \sup_{R \in C_n} v_R.$$

Under the assumption that

$$(13.7.5) \qquad v_n \geq (1-\varepsilon_n)\, v,$$

we shall show that relations similar to (13.7.2), (13.7.3) and (13.7.5) can be established for n+1.

According to theorem 13.6 there is a $P_{n+1} \in \mathcal{P}_n$ such that

$$(13.7.6) \qquad v_{P_{n+1}}(i_{n+1}) \geq (1-\delta^2)\, v_n(i_{n+1})$$

with

(13.7.7) $\quad \delta = 2^{-(n+2)}\ \varepsilon.$

Define

(13.7.8) $\quad B = \{i : v_{P_{n+1}}(i) < (1-\delta)\ v_n(i)\}$

and

(13.7.9) $\quad E_{n+1} = E \backslash B.$

Then E_{n+1} and P_{n+1} satisfy (13.7.3) for n+1 as will be proved. Indeed, by (13.7.6) we have that the expected return when starting in i_{n+1} and using policy P_{n+1} until entering B plus the expected return thereafter together exceed $(1-\delta^2)\ v_n(i_{n+1})$. Hence with $\underline{\tau}_{n+1}$ the exit time of E_{n+1} we have

(13.7.10)
$$v_{P_{n+1},\underline{\tau}_{n+1}}(i_{n+1}) + \sum_{j\in B} \mathbb{P}_{i_{n+1},P_{n+1}} [\underline{x}_{\underline{\tau}_{n+1}}=j]\ v_{P_{n+1}}(j) \geq$$
$$\geq (1-\delta^2)\ v_n(i_{n+1}).$$

By the definition of E_{n+1} we have

(13.7.11) $\quad v_{P_{n+1}}(j) \geq (1-\delta)\ v_n(j)$ for all $j \in E_{n+1}$.

Since v_n is the value function corresponding to C_n it follows from the theorems 3.1 and 2.21 (note that c_P is a charge structure since $\sup_R \mathbb{E}_R [\sum_{n=0}^{\infty} c(\underline{x}_n)] < \infty$) that for any policy $R \in C_n$ and any Markov time $\underline{\tau}$

(13.7.12) $\quad v_n \geq v_{R,\underline{\tau}} + \mathbb{E}_R\ v_n(\underline{x}_{\underline{\tau}}).$

Substituting P_{n+1} and $\underline{\tau}_{n+1}$ in (13.7.12) gives

(13.7.13)
$$v_n(i_{n+1}) \geq v_{P_{n+1},\underline{\tau}_{n+1}}(i_{n+1}) + \sum_{j\in B} \mathbb{P}_{i_{n+1},P_{n+1}} [\underline{x}_{\underline{\tau}_{n+1}}=j]\ v_n(j).$$

Substituting $(1-\delta)\ v_n(j)$ for $v_{P_{n+1}}(j)$ in the second term of (13.7.10), we find with (13.7.8) and (13.7.13)

$$(13.7.14)\quad \sum_{j\in B} \mathbb{P}_{i_{n+1},P_{n+1}}[\underline{x}_{\underline{\tau}_{n+1}}=j]\, v_n(j) \le \delta v_n(i_{n+1}).$$

Since $v_{P_{n+1}} \le v_n$ a similar relation with $v_{P_{n+1}}$ instead of v_n in the left-hand side holds. Together with (13.7.10) this yields

$$(13.7.15)\quad v_{P_{n+1},\underline{\tau}_{n+1}}(i_{n+1}) \ge (1-2\delta)\, v_n(i_{n+1}).$$

Since

$$(13.7.16)\quad (1-2\delta)(1-\varepsilon_n) \ge (1-\varepsilon_{n+1})$$

it follows with (13.7.5) that relation (13.7.3) is satisfied for n+1.

In the remainder of the proof we establish relation (13.7.5) for n+1. Define

$$\mathcal{P}_{n+1} = \{P : P \in \mathcal{P}_n \text{ and } p(i,.) = p_{n+1}(i,.) \text{ for } i \in E_{n+1}\},$$

let C_{n+1} be the policies with decision rules in $\mathcal{P}_{n+1}$ and take

$$v_{n+1} = \sup_{R\in C_{n+1}} v_R.$$

By relation (13.7.11) we have that

$$v_{n+1}(i) \ge (1-\delta)\, v_n(i) \text{ for all } i \in E_{n+1}.$$

To prove that a similar inequality also holds outside E_{n+1} we proceed as follows. Given any state i there exists in view of theorem 13.6 a policy $P \in \mathcal{P}_n$ such that

$$(13.7.17)\quad v_P(i) \ge (1-\delta)\, v_n(i).$$

Let R be the policy that chooses decisions according to P until the entry of E_{n+1} and uses decision rule P_{n+1} thereafter. Then with $\underline{\sigma}$ the entry time of E_{n+1} we have

$$(13.7.18)\quad v_R(i) \ge v_{P,\underline{\sigma}}(i) + \sum_{j\in E_{n+1}} \mathbb{P}_{i,P}[\underline{x}_{\underline{\sigma}}=j]\, v_{P_{n+1}}(j).$$

Using the relations (13.7.11) and (13.7.17) we derive from (13.7.18)

$$v_R(i) \geq v_{P,\underline{\sigma}}(i) + \sum_j \mathbb{P}_{i,P}\,[\underline{x}_{\underline{\sigma}}=j]\; v_P(j) - \delta v_n(i) =$$

$$= v_P(i) - \delta v_n(i).$$

Finally with (13.7.16) and (13.7.17) we obtain

$$v_{n+1}(i) \geq v_R(i) \geq (1-\varepsilon_{n+1})\; v(i). \;\square$$

We conclude this section by proving that in the positive dynamic programming case the existence of an optimal policy implies that some stationary policy is optimal. For the negative dynamic programming problem this is almost an immediate consequence of theorem 4.6. Indeed, when policy R is optimal then the decision rule for time 0, i.e. P_0, conserves v. Hence P_0^∞ is thrifty; since $v \leq 0$ we have that each policy is equalizing. Consequently P_0^∞ is optimal.

13.8. THEOREM. *If* $c_P \geq 0$ *for all* $P \in \mathcal{P}$ *and there exists an optimal policy then there exists a stationary optimal policy.*

PROOF. According to theorem 4.6 (in fact the analogue of theorem 4.6 for the total return model) there is also an optimal policy R such that each decision rule of R is v conserving. Hence without loss of generality we can assume that $\mathcal{P}$ consists of v conserving matrices. According to theorem 13.7 there exists a Q such that for some $a > 0$

$$v_Q \geq a\, v.$$

Hence

$$\lim_{n\to\infty} Q^n v \leq \frac{1}{a} \lim_{n\to\infty} Q^n \sum_{k=0}^{\infty} Q^k c_Q = 0.$$

Thus Q^∞ is also equalizing and in view of theorem 4.6 we have that Q^∞ is optimal. $\square$

BIBLIOGRAPHY

BATHER, J.A. (1973). *Optimal decision procedures for finite Markov chains.*
Part I : *Examples.* Advances in Appl. Probability 5, 328-339.
Part II : *Communicating systems.*
Part III : *General Convex Systems.* To appear in Advances in Appl. Probability.

BELLMAN, R. (1957). *Dynamic programming.* Princeton University Press.

BILLINGSLEY, P. (1968). *Convergence of probability measures.* Wiley, New York.

BLACKWELL, D. (1961). *On the functional equation of dynamic programming.* J. Math. Anal. Appl. 2, 273-276.

BLACKWELL, D. (1962). *Discrete dynamic programming.* Ann. Math. Statist. 35, 863-865.

BLACKWELL, D. (1965). *Discounted dynamic programming.* Ann. Math. Statist. 36, 226-235.

BLACKWELL, D. (1967). *Positive dynamic programming.* Proc. Fifth Berkeley Sympos. Math. Stat. and Prob., Vol. 1, 415-418.

BLACKWELL, D. (1970). *On stationary policies.* J. Roy. Statist. Soc. Ser. A. 133, 33-38.

BLUMENTHAL, R.M. and R.K. GETOOR (1968). *Markov processes and potential theory.* Academic Press, New York.

BREIMAN, L. (1964). *"Stopping-rule problems" in Applied Combinatorial Mathematics.* Wiley, New York.

CHOW, Y.S., H. ROBBINS and D. SIEGMUND (1971). *Great expectations: The theory of optimal stopping.* Houghton Mifflin Company, Boston.

CHUNG, K.L. (1960). *Markov chains with stationary transition probabilities,* second edition. Springer, Berlin.

CHUNG, K.L. and C. DERMAN (1956). *Non-recurrent random walks.* Pacific. J. Math. 6, 441-447.

COHEN, J.W. (1969). *The single server queue.* North-Holland Publ. Co., Amsterdam.

CRABILL, T.B. (1968). *Sufficient conditions for positive recurrence and recurrence of specially structured Markov chains.* Operations Res. 16, 858-867.

DE GROOT, M.H. (1970). *Optimal statistical decisions.* McGraw-Hill, New York.

DE LEVE, G. (1964). *Generalized Markovian decision processes.*
Part I : *Model and Method*,
Part II: *Probabilistic Background*, Mathematical Centre Tracts no. 3 and 4, Amsterdam.

DE LEVE, G., H.C. TIJMS and P.J. WEEDA (1970). *Generalized Markovian decision processes, applications.* Mathematical Centre Tracts no. 5, Amsterdam.

DENARDO, E.V. (1967). *Contraction mappings in the theory underlying dynamic programming.* SIAM Rev. 9, 165-177.

DERMAN, C. (1962). *On sequential decisions and Markov chains.* Management Sci. 9, 16-24.

DERMAN, C. (1963). *Stable sequential control rules and Markov chains*, J. Math. Anal. Appl. 6, 257-265.

DERMAN, C. (1964). *On sequential control processes.* Ann. Math. Statist. 35, 341-349.

DERMAN, C. (1965). *Markovian sequential control processes-denumerable state space.* J. Math. Anal. Appl. 10, 295-302.

DERMAN, C. (1966). *Denumerable state Markovian decison processes-average cost criterion.* Ann. Math. Statist. 37, 1545-1554.

DERMAN, C. (1968). *Markovian decision processes-average cost criterion.* In: *Mathematics of the decision sciences*, G.B. DANTZIG and A.F. VEINOTT, Jr., editors. Am. Math. Soc., Providence, Rhode Island.

DERMAN, C. (1970). *Finite state Markovian decision processes.* Academic Press, New York.

DERMAN, C. and R. STRAUCH (1966). *A note on memoryless rules for controlling sequential control processes.* Ann. Math. Statist. 37, 276-278.

DERMAN, C. and A.F. VEINOTT, Jr. (1967). *A solution to a countable system of equations arising in Markovian decision processes.* Ann. Math. Statist. 38, 582-584.

DOEBLIN, W. (1937/38). *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples.* Bull. Soc. Math. Roumaine, 39-1, 57-115; 39-2, 3-61.

DOOB, J.L. (1953). *Stochastic Processes.* Wiley, New York.

DUBINS, L.E. and L.J. SAVAGE. (1965). *How to gamble if you must: inequalities for stochastic processes.* McGraw-Hill, New York.

DYNKIN, E.B. (1963). *The optimum choice of the instant for stopping a Markov process.* Transl. of Dokl. Acad. Sci. USSR. 4, 627-629.

DYNKIN, E.B. and A.A. JUSCHKEWITSCH. (1969). *Sätze und Aufgaben über Markoffsche Prozesse.* Springer-Verlag, Berlin.

FELLER, W.F. (1950, 1966). *An introduction to probability theory and its applications.* Vol. I, third edition. Vol. II, second edition. Wiley, New York.

FISHER, L. (1968). *On recurrent denumerable decision processes.* Ann. Math. Statist. 39, 424-434.

FISHER, L. and S.M. ROSS (1968). *An example in denumerable decision processes.* Ann. Math. Statist. 39, 674-675.

FOSTER, F.G. (1953). *On stochastic matrices associated with certain queuing processes.* Ann. Math. Statist. 24, 355-360.

HARDY, G.H. (1949). *Divergent Series.* Oxford.

HELMS, L.L. (1969). *Introduction to potential theory.* Wiley, New York.

HINDERER, K. (1970). *Foundations of non-stationary dynamic programming with discrete time parameter.* Springer-Verlag, Berlin.

HORDIJK, A. (1971). *A sufficient condition for the existence of an optimal policy with respect to the average cost criterion in Markovian decision processes.* Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 263-274, Academia, Praag.

HORDIJK, A. (1972). *On Doeblin's condition and its application in Markov decision processes.* Mathematical Centre Report BW 15/72, Amsterdam. (in Dutch).

HORDIJK, A. (1973). *On the convergence of the average expected return in dynamic programming.* To appear in J. Math. Anal. Appl.

HORDIJK, A., R. POTHARST and J.TH. RUNNENBURG (1973). *Optimal stopping of Markov chains.* Mathematical Centre Syllabus 19, Amsterdam. (in Dutch)

HORDIJK, A. and H.C. TIJMS (1972). *A counterexample in discounted dynamic programming.* J. Math. Anal. Appl. 39, 455-457.

HORDIJK, A. and H.C. TIJMS (1973). *A modified form of the iterative method of dynamic programming.* To appear in Ann. Statist.

HORDIJK, A. and P. VAN GOETHEM (1973). *A criterion for the existence of invariant probability measures in Markov processes.* Mathematical Centre Report SW 22/73. Submitted for publication.

HOWARD, R.A. (1960). *Dynamic programming and Markov processes.* Technology Press, Cambridge, Massachusetts.

HUNT, G.A. (1957/58). *Markov processes and potentials I, II, III.* Illinois J. Math. 1, 44-93, 316-369; 2, 151-213.

JOHNSON, E.L. (1968). *On (s,S) policies.* Management Sci. 15, 80-101.

KELLEY, J.L. (1955). *General topology.* Van Nostrand, Princeton, New Jersey.

KEMENY, J.G., J.L. SNELL and A.W. KNAPP (1966). *Denumerable Markov chains.* Van Nostrand, Princeton, New Jersey.

KRYLOFF, N. and N. BOGOLIOUBOFF (1937a). *Sur les propriétés en chaîne* C.R. Acad. Sci., Paris, 204, 1386-1388.

KRYLOFF, N. and N. BOGOLIOUBOFF (1937b). *Les propriétés ergodiques des suites des probabilités en chaîne.* C.R. Acad. Sci., Paris, 204, 1454-1456.

KUSHNER, H. (1971). *Introduction to stochastic control.* Holt, Rinehart and Winston, New York.

LEEMAN, W.A. (1964). *The reduction of queues through the use of price.* Operations Res. 12, 783-785.

LEHMANN, E.L. (1959). *Testing statistical hypotheses.* Wiley, New York.

LIPPMAN, S.A. (1971). *Maximal average-reward policies for semi-Markov decision processes with arbitrary state and action space.* Ann. Math. Statist. 42, 1717-1726.

LOW, D.W. (1972). *Optimal dynamic policies for an M/M/S queue with variable arrival rate*. I.B.M. report.

MAITRA, A. (1965). *Dynamic programming for countable state systems.* Sankhya 27A, 241-248.

MAITRA, A. (1968). *Discounted dynamic programming on compact metric spaces.* Sankhya 30A, 211-216.

MOUSTAFA, M.D. (1957). *Input-output Markov processes*, Proc. Kon. Nederl. Akad. Wet. Ser. A60, Indag. Math. 19, 112-118.

NEVEU, J. (1965). *Mathematical foundations of the calculus of probability.* Holden-Day, San Francisco.

OREY, S. (1971). *Limit theorems for Markov chain transition probabilities.* Van Nostrand Reinhold, London.

ORNSTEIN, D. (1969). *On the existence of stationary optimal strategies.* Proc. Amer. Math. Soc. 20, 563-569.

PAKES, A.G. (1968). *Some conditions for ergodicity and recurrence of Markov chains*. Operations Res. 17, 1058-1061.

PRATT, J.W. (1960). *On interchanging limits and integrals*. Ann. Math. Statist. 31, 74-77.

ROSS, S.M. (1968). *Non-discounted denumerable Markovian decision models.* Ann. Math. Statist. 39, 412-423.

ROSS, S.M. (1968). *Arbitrary state Markovian decision processes*. Ann. Math. Statist. 39, 2118-2122.

ROSS, S.M. (1970). *Applied probability models with optimization applications* Holden-Day, San Francisco.

ROSS, S.M. (1970a). *Average cost semi-Markov decision processes*. J. Appl. Probability 7, 649-656.

RUNNENBURG, J.TH. (1960). *On the use of Markov processes in one-server waiting-time problems and renewal theory*. Poortpers N.V., Amsterdam.

SCHÄL, M. (1973). *Dynamic programming under continuity and compactness assumptions*. Advances in Appl. Probability 5, 24-25.

SCHEFFÉ, H. (1947). *A useful convergence theorem for probability distributions*. Ann. Math. Statist. 18, 434-438.

SCHWEITZER, P.J. (1968). *Perturbation theory and finite Markov chains.* J. Appl. Probability 5, 401-413.

STRAUCH, R. (1966). *Negative dynamic programming.* Ann. Math. Statist. 37, 871-889.

STARR, N. (1972). *How to win a war if you must: optimal stopping based on success runs.* Ann. Math. Statist. 43, 1884-1893.

TAYLOR, H.M. (1965). *Markovian sequential replacement processes.* Ann. Math. Statist. 36, 1677-1694.

TIJMS, H.C. (1972). *Analysis of* (s,S) *inventory models.* Mathematical Centre Tracts no. 40, Amsterdam.

VEINOTT, A.F., Jr. (1966). *On the optimality of* (s,S) *inventory policies: new conditions and a new proof.* SIAM J. Appl. Math. 14, 1067-1083.

VEINOTT, A.F., Jr. (1969). *Discrete dynamic-programming with sensitive discount optimality criteria.* Ann. Math. Statist. 40, 1635-1660.

WALD, A. (1947). *Sequential analysis.* Wiley, New York.

YOSIDA, K. and S. KAKUTANI (1941). *Operator-theoretical treatment of Markoff's process and mean ergodic theorem.* Ann. of Math. 42, 188-228.

LIST OF NOTATIONS

x, y, f, g etc.	real-valued functions (also called vectors) on the state space E
$x(i)$	i^{th} component of vector x
e	vector with all components equal to 1
0	the real number zero and the vector with all components equal to zero
$x \vee y$	vector with i^{th} component $\max(x(i),y(i))$
$x \wedge y$	vector with i^{th} component $\min(x(i),y(i))$
x^+	$x \vee 0$
x^-	$-(x \wedge 0)$
$x \leq y$	$x(i) \leq y(i)$ for all $i \in E$
$x = y$	$x \leq y$ and $y \leq x$
$x < \infty$	$x(i) < \infty$ for all $i \in E$
$\mathcal{P}, \mathcal{P}(i)$	see page 1
$p(i,j)$	$(i,j)^{th}$ entry of stochastic matrix P
$p(i,.)$	i^{th} row-vector of P
$p(i,A)$	$\sum_{j \in A} p(i,j)$
$\sum_j$	summation over all $j \in E$
Px	vector with i^{th} component $\sum_j p(i,j)\, x(j)$
$P_0P_1 \ldots P_n$	matrix with $(i,j)^{th}$ entry $\sum_{\ell_1,\ldots,\ell_n} p_0(i,\ell_1)\, p_1(\ell_1,\ell_2) \ldots p_n(\ell_n,j)$
$\sup_{\mathcal{P}} (c_P+Px)$	vector with i^{th} component $\sup_{\mathcal{P}} [c_P(i) + \sum_j p(i,j)\, x(j)]$
$\limsup_{n\to\infty} x_n$	vector with i^{th} component $\limsup_{n\to\infty} x_n(i)$
$P \to P_0$	$p(i,j) \to p_0(i,j)$ for all $i,j \in E$; see page 1
P^∞	stationary policy $(P,P,\ldots)$

policy	see page 1
strategy	see page 21
entry time	see page 18
reentry time	see page 8
$c_{\mathcal{P}}$-excessive, $c_{\mathcal{P}}$-superharmonic, etc.	see the definitions in section 2. If $c_P = 0$ for all $P \in \mathcal{P}$ we write excessive, superharmonic, etc.
$c_{\mathcal{P}}$ is continuous	if $\lim\limits_{P\to P_0} c_P(i) = c_{P_0}(i)$ for all $i \in E$ and all $P_0 \in \mathcal{P}$
$c_{\mathcal{P}}$ is upper semicontinuous	if $\limsup\limits_{P\to P_0} c_P(i) \le c_{P_0}(i)$ for all $i \in E$ and all $P_0 \in \mathcal{P}$
A^c	the complement of subset $A \subset E$
$f_R(i,A)$	see page 69
$f_P(i,A)$	see page 64
$f_P(i,j)$	see page 64
$\mathbb{E}_R\, c(\underline{x}_n)$	for $R = (P_0,P_1,\ldots)$ equal to the vector $P_0\ldots P_{n-1}c_{P_n}$
$\mathbb{E}_{i,R}\, c(\underline{x}_n)$	the i^{th} component of the vector $\mathbb{E}_R\, c(\underline{x}_n)$, for $R = (P_0,P_1,\ldots)$ equal to $\sum\limits_{\ell_1,\ldots,\ell_n} p_0(i,\ell_1)\, p_1(\ell_1,\ell_2)\, \ldots\, p_{n-1}(\ell_{n-1},\ell_n)\, c_{P_n}(\ell_n)$
$\mathbb{E}_P\, [\ldots]$	abbreviation for $\mathbb{E}_{P^\infty}\, [\ldots]$
$\sum\limits_{n=0}^{N} P_0\ldots P_{n-1}c_{P_n}$	abbreviation for $[c_{P_0} + P_0 c_{P_1} + P_0 P_1 c_{P_2} + \ldots + P_0\ldots P_{N-1}c_{P_N}]$
$\underline{\tau}$	Markov time or stopping time, $\underline{\tau}$ equal to infinity is admissible
$\chi(\ldots)$	is equal to one on the event $(\ldots)$ and equal to zero otherwise
$f(\underline{x}_{\underline{\tau}})$	is equal to $f(\underline{x}_n)$ for $\underline{\tau} = n$, $n \in \{0,1,2,\ldots\}$ and equal to zero for $\underline{\tau} = \infty$, equivalently $f(\underline{x}_{\underline{\tau}})$ is equal to $f(\underline{x}_{\underline{\tau}})\, \chi(\underline{\tau}<\infty)$
$\mathbb{E}_{i,R}\, [\sum\limits_{k=0}^{\underline{\tau}-1} c(\underline{x}_k)]$	denotes for $R = (P_0,P_1,\ldots)$ the conditional expectation given $\underline{x}_0 = i$ of $\sum\limits_{k=0}^{\underline{\tau}-1} c_{P_k}(\underline{x}_k)$ under policy R

SUMMARY

This text contains the material presented in 1973 in the Colloquium on Probability Theory organized jointly by the Mathematical Centre and the Institute for Applications of Mathematics of the University of Amsterdam.

The central theme is the investigation of the existence of optimal policies or optimal strategies in various discrete time dynamic programming problems.

In section 2 some well-known theorems in Markov potential theory are generalized to collections of Markov chains. Most of the definitions and results in this section also play an important role in the sequel.

In sections 3 and 4 a discrete time optimal control problem is investigated. It is proved that the value function is the minimum of the c_p-excessive functions that majorize the reward function. Further it is shown that a strategy is optimal if and only if it is thrifty and equalizing.

Section 5 deals with a semi-Markov decision process having at least one state for which the expected cost until the system enters this state is uniformly bounded over all policies. Using results from the foregoing sections, we obtain a rather general condition guaranteeing the existence of optimal policies with respect to the average return criterion.

In section 6 some theorems on dynamic programming problems with total return criterion are collected.

Using results from section 6, we answer in section 7 some questions raised in connection with the notions introduced in section 2. The section is concluded with a theorem on the existence of optimal strategies for problems with a finite state space.

In section 8 the notions communicating and recurrent system are introduced. Similar to the notions communicating and recurrent class for one Markov chain, they play a basic role in Markov decision processes.

It is proved in section 9 for a wide class of sequential decision problems that the optimal stopping time is exponentially bounded under the optimal policy.

In section 10 we investigate again the discrete time dynamic programming problem with the supremum of the expected return per unit time as optimality criterion. If the invariant probability measures depend continuously on the decision rule or if they form a tight collection and the system is recurrent then there exists a stationary optimal policy.

A simultaneous Doeblincondition is investigated in section 11.

In section 12 it is pointed out that this notion provides the connection between conditions given in the literature and those of the sections 10 and 11.

In section 13 we collect several results announced in the foregoing sections. It is proved there that randomization does not increase the value function. Finally some theorems on the existence of weak and strong nearly optimal policies are given.

Stellingen bij het proefschrift *Dynamic programming and Markov potential theory* van A. Hordijk

I

Een topologisch halflichaam is een commutative topologische ring E, met een deelverzameling K die voldoet aan de volgende axioma's:

1. $K+\bar{K} \subset K$; $K.K \subset K$ ($\bar{K}$ is de afsluiting van K).
2. $K-K = E$.
3. $M \subset \bar{K}$ en $N = \cap\{\bar{K}+x : x \in M\} \Rightarrow N = \emptyset$ of $N = \bar{K}+y$ voor een $y \in \bar{K}$.
4. $a,b \in K \Rightarrow ax = b$ voor een $x \in K$.
5. $\bar{K} \cap (-\bar{K}) = \{0\}$.
6. Voor $a \in E$ zij $F_a = \{x \in E : ax \in \bar{K}\}$; de verzamelingen $b+F_a$, $a,b \in E$, vormen een subbasis voor de gesloten verzamelingen in E.

We noemen de axioma's 1-6 het axiomastelsel A. Axiomastelsel B wordt uit A verkregen door axioma 1 te vervangen door het zwakkere axioma

1^*. $K+K \subset K$; $K.K \subset K$

en toe te voegen het axioma

7. $x \in \bar{K}$ en x is inverteerbaar $\Rightarrow x \in K$.

Axiomastelsel C bestaat uit de axioma's 1^*, 2, 4, 5, 6 en de volgende verscherping van axioma 3

3^*. $M \subset \bar{K}$ en $N = \cap\{\bar{K}+x : x \in M\} \Rightarrow N = \emptyset$ of $N = \bar{K}+y$ voor een $y \in \bar{K}$; indien $M \subset K$ dan $y \in K$.

De axiomastelsels A, B en C zijn equivalent.

Uit: *Axiom systems for topological semifields*, A. Hordijk. Nieuw Archief voor Wiskunde (3), XV, 140-145.

II

Zij I de verzameling van alle idempotente elementen van het topologisch halflichaam E. Voor $e_1, e_2 \in I$ zij

$$V_{e_1,e_2} = \{e \in I : e-e_1 \in \bar{K} \text{ en } e_2-e \in \bar{K}\}.$$

Deze verzamelingen vormen een subbasis voor de gesloten verzamelingen van de relatieve topologie in I.

Uit: *Enige eigenschappen van topologische halflichamen*, A. Hordijk. Rapport nr.1, Wiskundig Seminarium der Vrije Universiteit.

III

Als de funktie f: $[0,1] \to \mathbb{R}$ analytisch voortzetbaar is op een voldoend grote omgeving van het interval $[0,1]$ in het complexe vlak, dan geldt voor ieder driehoekig schema van punten uit $[0,1]$ dat de bijbehorende rij van interpolatiepolynomen uniform op $[0,1]$ naar f convergeert (zie: *Interpolation and approximation*, P.J. Davis).
De volgende omkering geldt: indien de interpolatiepolynomen voor ieder driehoekig schema van punten uit $[0,1]$ uniform naar f convergeren op $[0,1]$ dan is f analytisch op $(0,1)$.

IV

Als $\underline{x}_1, \underline{x}_2, \ldots$ een rij van onderling onafhankelijk stochastische variabelen is met een tweemaal differentieerbare verdelingsfunktie F en als ϕ een monotoon naar oneindig stijgende funktie is met

$$\lim_{x\to\infty} \left[1 + \frac{F''(x)(1-F(x))}{(F'(x))^2}\right] \phi^2\left(\frac{1}{1-F(x)}\right) = 0,$$

dan geldt voor iedere rij $\{x_n\}$ met $x_n = O(\phi(n))$

$$\lim_{n\to\infty} \frac{\mathbb{P}[\{\max(\underline{x}_1,\ldots \underline{x}_n)-b_n\}nF'(b_n) > x_n]}{\mathbb{P}[\underline{z} > x_n]} = 1.$$

Hier is $\underline{z}$ een stochastische variabele met kansverdeling $\exp[-\exp(-x)]$ en is b_n gedefinieerd door $F(b_n) = 1-1/n$.

Indien F de normale verdeling is dan geldt bovenstaande bewering voor iedere rij met $x_n = o((\log n)^{\frac{1}{2}})$.

Uit: *The rate of growth of sample maxima*,
L. de Haan en A. Hordijk. Ann. Math. Statist. 43, 1185-1196.

V

Voor $H_1 \in \Lambda$ en $H_2 \in \Lambda$ waarbij Λ het aantrekkingsgebied van de dubbelexponentiële verdeling is, zij

$$A_i(x) = \int_x^\infty (1-H_i(t))dt/x(1-H_i(x)), \qquad i = 1,2;$$

$$c(x) = A_1(x)/A_2(x) \text{ en } d(x) = (1-H_1(x))/(1-H_2(x)).$$

Elk van de condities $\lim_{x\to\infty} c(x) = 1$, $\lim_{x\to\infty} d(x) = 0$ en $\lim_{x\to\infty} d(x) = \infty$ is een voldoende voorwaarde voor $H_1H_2 \in \Lambda$. Omgekeerd is het mogelijk dat $H_1 \in \Lambda$, $H_2 \in \Lambda$, $H_1H_2 \in \Lambda$ terwijl noch $c(x)$ noch $d(x)$ een limiet heeft als $x \to \infty$.

Uit: *Products of distribution functions attracted to extreme value laws*,
S.J. Resnick. J. Appl. Prob. 8, 781-793.

VI

Een stationair Markov proces kan op verschillende manieren gedefinieerd worden. Het schijnt dat een stationair Markov proces op $\mathbb{R}^n$ met continue paden een sterk Markov proces is indien de Markov eigenschap geldt voor alle tijden van eerste binnenkomst in Borelverzamelingen.

Zie: *Generalized Markovian Decision Processes, Part II*,
G. de Leve. Mathematical Centre Tracts 4, 2nd edition, errata and addenda.

Een deeltje beweegt zich volgens een rechtscontinu, sterk Markov proces in $\mathbb{R}^n$. Voor $U \subset \mathbb{R}^n$ en gesloten zij q een overgangswaarschijnlijkheidsfunktie van $(U, \mathcal{B}^n \cap U)$ naar $(U^c, \mathcal{B}^n \cap U^c)$. We veronderstellen dat U een "reflecterende" verzameling is. Iedere keer dat het deeltje in U binnenkomt, wordt het op stochastische wijze volgens $q(u,.)$ teruggekaatst als u het trefpunt is met U. Ter bestudering van de bewegingswetten van het "gestoorde" proces voeren we enkele begrippen in. Zij $\{P_0(t,x,B) : 0 \le t < \infty,\ x \in \mathbb{R}^n,\ B \in \mathcal{B}^n\}$ de bewegingswet behorende bij het proces dat verkregen wordt door het oorspronkelijke te "stoppen" op de eerste binnenkomsttijd van U. Veronderstel dat $P_0(t,x,U)$ voor $x \in U^c$ als funktie van t hoogstens eindig veel discontinuïteiten in een eindig interval heeft. Neem aan dat $\lim_{h \downarrow 0} h^{-1} P_0(h,x,B)$ bestaat (notatie $p_0(0,x,B)$); zij, voor $t > 0$,

$$p_0(t,x,B) = \int_{U^c} P_0(t,x,dy)\ p_0(0,y,B) \text{ en}$$

$$p_d(t,x,B) = P_0(t,x,B) - \lim_{h \downarrow 0} P_0(t-h,x,B)$$

voor alle $x \in U^c$ en alle $B \in \mathcal{B}^n \cap U$.

De kans dat het deeltje in het gestoorde proces zich op tijdstip s+t in de verzameling $B \in \mathcal{B}^n \cap U^c$ bevindt indien het op tijdstip s zich in toestand $x \in U^c$ bevindt, wordt gegeven door

$$P(t,x,B) = \sum_{k=0}^{\infty} P_k(t,x,B)$$

met

$$P_k(t,x,B) = \int_0^t d\tau \Big(\int_U p_0(\tau,x,du) \Big[\int_{U^c} q(u,dy)\ P_{k-1}(t-\tau,y,B) \Big] \Big) +$$

$$+ \sum_{0<\tau\le t} \int_U p_d(\tau,x,du) \Big[\int_{U^c} q(u,dy)\ P_{k-1}(t-\tau,y,B) \Big].$$

De overgangswaarschijnlijkheidsfunkties $\{P(t,x,B) : 0 \le t < \infty,\ x \in U^c,\ B \in \mathcal{B}^n \cap U^c\}$ voldoen aan de Chapman-Kolmogorov vergelijkingen.

VIII

Beschouw een aperiodieke, irreducibele Markov keten $\underline{x}_n$, n=0,1,..., met aftelbaar veel ergodische toestanden. Laat p de matrix van overgangswaarschijnlijkheden en π de invariante kansmaat zijn. Als f een niet-negatieve funktie op de toestandsverzameling van de Markov keten is en $\underline{\tau}_i$ het tijdstip van eerste binnenkomst in $\{i\}$, dan geldt:

$$\lim_{n\to\infty} \mathbb{E}[f(\underline{x}_n)\ \chi_{\{\underline{\tau}_i > n\}}] = 0$$

is een voldoende voorwaarde voor

$$\lim_{n\to\infty} \mathbb{E}f(\underline{x}_n) = \sum_j \pi(j)\ f(j).$$

Indien $\sum_j \pi(j)\ f(j) < \infty$ dan is bovenstaande voorwaarde ook noodzakelijk.

Uit: *On the convergence of moments in stationary Markov chains*, P.J. Holewijn en A. Hordijk. Verschijnt in: Stoch. Proc. Appl.

IX

Een mogelijke veralgemening van het eerste gedeelte van stelling luidt: Zij $\underline{x}(t)$ een meetbaar regeneratief proces op $\mathbb{R}^1$ met $\underline{t}$ de "vernieuwingstijd". Als $\underline{t}$ een verdeling F heeft met verwachting μ die geen rooster-verdeling is en als voor de niet-negatieve funktie f geldt dat $\mathbb{E}[f(\underline{x}(t))\ \chi_{\{\underline{t}>t\}}]$ als funktie van t direct Riemann-integreerbaar is dan bestaat $\lim_{t\to\infty} \mathbb{E}f(\underline{x}(t))$ en is gelijk aan $\frac{1}{\mu}\mathbb{E}[\int_0^{\underline{t}} f(\underline{x}(t))dt]$.

X

Het eerste gedeelte van het bewijs van stelling I.15.3 (pag. 93 e.v.) uit Chung's boek *Markov chains with stationary transition probabilities* (2^{nd} edition), is een ingewikkelde afleiding van de bekende gelijkheid van Wald voor een speciaal geval.
Relatie (9) uit deze stelling kan door gebruik te maken van taboo-kansen, eenvoudiger bewezen worden.

XI

Het 12 jaar geleden door Iglehart geuite vermoeden dat voor het klassieke voorraadmodel van Arrow, Harris en Marschak de minimale kosten over de eerste n perioden minus n maal de minimale gemiddelde verwachte kosten per tijdseenheid convergeren als n naar oneindig gaat, kan voor een discrete vraagverdeling in bevestigende zin beantwoord worden.

Uit: *Convergence results and approximations for optimal* (s,S) *policies*,
A. Hordijk en H.C. Tijms. Verschijnt in: Management Sci.

XII

Als $\mathcal{P}$ een compacte verzameling van Markov matrices is, zoals geïntroduceerd in dit proefschrift, met de toestandsverzameling een aperiodieke positief-terugkerende klasse voor iedere $P \in \mathcal{P}$ en als het infimum over alle Markov matrices van de verwachting van het aantal transities nodig om van toestand i naar toestand j te komen, een begrensde funktie van i en j is, dan geldt voor zekere begrensde funkties g(j) en v(i,j) dat

$$\lim_{n\to\infty} \{\underset{P_0,\ldots,P_n \in \mathcal{P}}{\text{supremum}} \sum_{\ell_1,\ldots,\ell_n} p_0(i,\ell_1)\, p_1(\ell_1,\ell_2)\ldots p_n(\ell_n,j) - ng(j)\} = v(i,j)$$

voor alle i en j.

XIII

Voor iedere α met $0 < \alpha < 1$ zij T_α een contractie-afbeelding van $\mathbb{R}^k$ naar $\mathbb{R}^k$ met contractiemodulus α en dekpunt $m(\alpha)$. Als voor zekere constante b geldt $\|m(\alpha)-m(\beta)\| \leq |\alpha-\beta| b$ voor α en β in een voldoend kleine linkeromgeving van 1 en als α_n een monotoon naar 1 stijgende rij is met

$$\prod_{n=1}^{\infty} \alpha_n = 0 \text{ en } \lim_{n\to\infty} \sum_{j=2}^{n} (\alpha_j \ldots \alpha_n)(\alpha_j - \alpha_{j-1}) = 0$$

dan geldt

$$\lim_{n\to\infty} T_{\alpha_n} \circ T_{\alpha_{n-1}} \circ \ldots \circ T_{\alpha_1} \circ x = \lim_{\alpha\uparrow 1} m(\alpha) \text{ voor alle } x \in \mathbb{R}^k.$$

XIV

Voor een Markov proces op een separabele metrische ruimte E met overgangswaarschijnlijkheidsfunktie p, een compacte verzameling A en een niet-negatieve reëelwaardige funktie ϕ op E veronderstellen we

$$1 + \int_{A^c} p(x,dy)\ \phi(y) \le \phi(x) \text{ voor alle } x \in A^c$$

en

$$\int_{A^c} p(x,dy)\ \phi(y) \text{ is begrensd op A.}$$

Als bovendien voor iedere begrensde continue en reëelwaardige funktie f op E ook de funkties $\int p(x,dy)\ f(y)$ en $\int_A p(x,dy)\ f(y)$ continu zijn dan heeft het Markov proces een invariante kansverdeling.

Uit: *A criterion for the existence of invariant probability measures in Markov processes*, A. Hordijk en P. Van Goethem. Mathematisch Centrum Rapport SW 22.

XV

Derman gebruikt in zijn boek *Finite state Markovian decision processes* het volgende, intuïtief aannemelijke argument. Indien zekere strategie minimale verwachte kosten per tijdseenheid heeft dan blijft dit zo indien de verzameling van beslissingen uitgebreid wordt, mits de aan de toegevoegde beslissingen verbonden kosten voldoende groot genomen worden. Dit argument is onjuist.

XVI

Beschouw de rij van a posteriori-kansen in het sekwente toetsingsprobleem met enkelvoudige nulhypothese tegen een enkelvoudige alternatieve hypothese. De klasse van alle continue harmonische funkties voor dit stochastische proces is precies de verzameling van alle lineaire funkties van $[0,1]$ naar $\mathbb{R}^1$.

Uit: *A Bayes process*, K.M. van Hee en A. Hordijk. Mathematisch Centrum Rapport SW 23.

XVII

Beschouw een optimaal stopprobleem op $\mathbb{N}$ met niet-negatieve opbrengstfunktie r. Als de waarde van het spel **v** een eindigwaardige funktie is dan geldt voor iedere ε met $0 < \varepsilon < 1$ en $\underline{\tau}_\varepsilon$, de binnenkomsttijd in $\{i \in \mathbb{N} : r(i) \geq (1-\varepsilon)v(i)\}$, dat

$$\mathbb{E}_i \, [r(\underline{x}_{\underline{\tau}_\varepsilon})] \geq (1-\varepsilon) \; v(i) \quad \text{voor alle } i \in \mathbb{N}.$$

Als r bovendien begrensd is dan geldt $\mathbb{P}_i[\underline{\tau}_\varepsilon < \infty] = 1$ voor iedere $i \in \mathbb{N}$.

XVIII

De door Dunford en Schwartz gegeven adstructie van de toepasbaarheid van Markov processen zegt meer over Dunford en Schwartz dan over Markov processen.

Zie: *Linear Operators*, N. Dunford en J.T. Schwartz.
Part I (1958) pag. 730.

XIX

Bij de wetenschapsbeoefening is de oefening vaak belangrijker dan de resultaten.

XX

In oude kulturen placht men tijdens het gehele leven een gedeelte van de dag te besteden aan het vertellen en aanhoren van verhalen. De "éducation permanente" is een terugkeer tot deze goede gewoonte.

XXI

De metrobouw in Amsterdam is een stadsondermijnende aktiviteit.