CWI Tracts

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

CWI Tract 33

Stochastic games with finite state and action spaces

O.J. Vrieze

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 90D05, 90D15, 93E05, 93E20.

ISBN 90 6196 313 3

Printed in the Netherlands

ACKNOWLEDGEMENTS

This tract is a slightly revised version of my thesis which was written under the supervision of professor Henk Tijms and professor Stef Tijs. I am very grateful to both of them. With Henk Tijms I had very stimulating discussions during the realization of my thesis. Several chapters of this tract evolved from joint work with Stef Tijs. His never ceasing stream of inspiring fresh impulses always guarantees further investigations.

Also, I am indebted to professor dr. G. de Leve for having given me the splendid opportunity to start my research at the Mathematical Centre.

I thank the Centre for Mathematics and Computer Science for the opportunity to publish this tract in their series CWI Tracts and all those who have contributed to its technical realization.

O.J. Vrieze

CONTENTS

Part I. Introduction and description of the model.

1. Introduction.

In this monograph two-person zerosum stochastic games are considered. With the exception of sections 5.3 and 5.4 both the state space and the spaces of pure actions of the players are assumed to be finite sets.

This monograph consists of three parts, supplemented by an appendix.

In part I the model is described. Further, the different types of strategies and evaluation functions are introduced. Analysed subsequently is what happens when a player fixes his strategy in advance of the play.

In part II we study discounted stochastic games. The theory of stochastic games originated in 1953 with the fundamental paper of Shapley (1953). He considered stopping stochastic games, i.e. games for which, in each state and for each pair of actions of the players, the game stops with a positive probability. Discounted stochastic games can be regarded as special cases of stopping stochastic games. Shapley proved that a stopping game has a value and that both players possess optimal stationary strategies.

In the introductory section of part II we give an alternative proof of Shapley's result, using non-linear programming techniques. Further, in part II the emphasis lies on two subjects. Firstly we investigate structural properties of the class of discounted stochastic games, and secondly we consider algorithms.

Given the value of a discounted stochastic game, optimal stationary strategies can be constructed by taking optimal actions in certain matrix games. This fact enables us to extend the structural properties for matrix games and games in normal form to stochastic games. Particularly, the results of Bohnenblust, Karlin & Shapley (1950), Shapley & Snow (1950), Vilkas (1963) and Tijs (1976b, 1981) are enlarged to stochastic games. For computational reasons special attention is paid to the relations between the game parameters and the solution of the game, and to the influence of small perturbations of the game parameters on the solution of the game.

Algorithms for discounted stochastic games are mainly based on successive approximation methods (e.g. Van der Wal (1977)). Other iterative procedures that can be mentioned are the algorithms of Hoffman & Karp

(1966) and Pollatschek & Avi-Itzhak (1969). Parthasarathy & Raghavan (1981) studied one-player-control stochastic games. For the discounted case they gave a linear programming problem, the solution of which corresponds to the solution of the associated game.

In chapter 6, after a short review of existing solution methods for discounted stochastic games, we present two algorithms. One can be characterized by the term fictitious play for discounted stochastic games. This algorithm can be seen as the extension to stochastic games of the fictitious play scheme which Brown (1949, 1951) suggested as a solution concept for matrix games. The other algorithm of chapter 6 can be applied to the subclass of discounted switching control stochastic games, giving a finite procedure for deriving the solutions of these games. The class of switching control stochastic games was introduced by Filar (1981).

In part III we consider undiscounted stochastic games. This part is also built up of a chapter on structural properties and a chapter on algorithms.

The theory of undiscounted zerosum stochastic games began with Gillette (1957). For a long time it was an open question whether undiscounted stochastic games possess a value. This question was recently answered in the affirmative by Mertens & Neyman (1981). Independently a weaker version of their result was elaborated by Monash (1979).

In a valuable paper Blackwell & Ferguson (1968) studied an example of an undiscounted stochastic game (the big match) and showed that no optimal strategy exists for one of the players. Even if one wishes to play ε-optimal, $\varepsilon>0$, in general one has to use complicated history dependent strategies. Hence it is natural to examine subclasses of stochastic games for which ε-optimal or optimal stationary strategies exist and where these strategies can easily be calculated.

With respect to the structural properties, we characterize games having, for $\varepsilon\geq 0$, ε-optimal stationary strategies for one or both players. Further it appears that for each of the players there is at least one state which is easy for him. Here, easy for a player means that, starting in such a state, he can guarantee himself the value of the game with a stationary strategy. Also studied in detail is the subclass of stochastic games for which the value does not depend on the initial state. Next relations between the solution of the limit discount equation and the asymptotic behaviour of the value of the τ-step game are investigated.

In deriving our results we make use of the field of real Puiseux series. Bewley & Kohlberg (1976a, 1976b, 1978) have elegantly introduced this field of real Puiseux series in stochastic games. Their results enclose nearly all earlier results on stochastic games.

Concerning algorithms for undiscounted stochastic games, the aforementioned result of Blackwell & Ferguson (1968) showed that it is very hard to find ε-optimal strategies in general. The algorithms of Hoffman & Karp (1966), Federgruen (1978) and Van der Wal (1980) approximate the value and give ε-optimal stationary strategies for special subclasses of stochastic games. Also for undiscounted stochastic games, Parthasarathy & Raghavan introduced the one-player-control stochastic game and Filar (1981) introduced the switching control stochastic game.

In chapter 9 we examine the above mentioned two subclasses of stochastic games, which can be solved relatively easily. The one-player-control stochastic game can be solved by a linear programming problem. The switching control stochastic game can be solved by a finite sequence of linear programming problems. For both classes the orderfield property arises in a natural way from the algorithms.

In the appendix we give the necessary concepts and well-known facts for matrix games (section A.1) and Markov decision problems (section A.2) that will be used in this monograph. In section A.3 recent literature on structured stochastic games is outlined.

2. Stochastic games; the model.

2.1. THE MODEL OF THE TWO-PERSON ZEROSUM STOCHASTIC GAME.

In this monograph we study two-person zerosum stochastic games with finite state space and finite action spaces for both players. We begin by defining a stochastic game situation, which will serve as the framework of a particular stochastic game.

2.1.1. DEFINITION. *A finite two-person zerosum stochastic game situation is an ordered quintuple* $< S, \{A_s | s \in S\}, \{B_s | s \in S\}, r, p >$, *where* S, A_s *and* B_s *are finite non-empty sets,* r *is a real-valued function on the set* $H := \{(s,i,j) | s \in S, i \in A_s, j \in B_s\}$ *and where* p *is a map* $p: H \to \mathcal{P}(S)$ *with* $\mathcal{P}(S)$ *the family of probability distributions on the space* S.

The game parameters have the following meaning.

$S = \{1,2,\ldots,z\}$ is called the state space.

$A_s = \{1,2,\ldots,m_s\}$ is called the action set of player 1 in state s.

$B_s = \{1,2,\ldots,n_s\}$ is called the action set of player 2 in state s.

$r : H \to \mathbb{R}$ is called the payoff function; if in state s player 1 chooses action $i \in A_s$ and player 2 chooses $j \in B_s$, then player 2 pays player 1 the amount $r(s,i,j)$ (if $r(s,i,j)<0$, then player 2 receives $-r(s,i,j)$ from player 1).

$p : H \to \mathcal{P}(S)$ is called the transition map. $\mathcal{P}(S)$ can be identified with the set $\{w | w \in \mathbb{R}^z, w_s \geq 0$, each $s \in S$ and $\Sigma_{s=1}^{z} w_s = 1\}$. Therefore, for each $(s,i,j) \in H$, we identify $p(s,i,j)$ with the vector $(p(1|s,i,j), p(2|s,i,j), \ldots, p(z|s,i,j))$. Here $p(t|s,i,j)$ represents the probability that the system jumps to state t if in state s player 1 chooses action $i \in A_s$ and player 2 action $j \in B_s$. Hence $p(t|s,i,j) \geq 0$ and $\Sigma_{t=1}^{z} p(t|s,i,j) = 1$.

We usually omit the adjective finite for a stochastic game, since with the exception of the sections 5.3 and 5.4 we only consider finite stochastic games.

Such a stochastic game corresponds to a dynamic system which can be in different states and where at certain decision epochs the players can influence the course of the play. We consider the infinite horizon model and the set of decision epochs is assumed to be identical with the set $\mathbb{N}=\{0,1,2,\ldots\}$.

The game runs as follows. We assume that the initial state s_0 at decision epoch 0 is known to the players. The players select simultaneously and independently of one another (possibly by a chance experiment) an action $i_0 \in A_{s_0}$ and $j_0 \in B_{s_0}$ respectively. Now two things happen, both depending on the current state s_0 and the subsequently chosen actions i_0 and j_0.

(a) player 2 pays player 1 the amount $r(s_0,i_0,j_0)$.

(b) the system jumps to the next state s_1 according to the outcome of a chance experiment. The probability that the next state will be state t equals $p(t|s_0,i_0,j_0)$.

Subsequently, prior to the next decision epoch 1, both players are informed of the previous actions chosen by the players and of the new state s_1. At decision epoch 1, the above procedure repeats itself, etc.

We assume that the game is of perfect recall, i.e. at each decision epoch each player remembers all past actions chosen by all players and all past states that have occurred.

Note that for finite two-person zerosum stochastic games, we have for each state a similarity with matrix games, in the sense that r(s,i,j) denotes the (possibly negative) amount which player 2 pays player 1 if in state s the players select actions i and j respectively. However, contrary to the situation with matrix games, the game does not exist of a single play, but jumps according to the probability measure $p(.|s,i,j)$ to the next state and continues dynamically. So in choosing an action in a certain state a player not only takes into account the immediate reward, but also his possibilities in the future states.

Also like in matrix games, when selecting an action, the players are allowed to randomize their pure actions. At the different decision epochs this randomization may depend on the history of the game up to that epoch. In the next section we discuss the types of strategies that a player may use.

2.2. STRATEGIES IN STOCHASTIC GAMES.

2.2.1. DEFINITION. *The set of possible histories up to a decision epoch* τ *consists of all sequences* $h_\tau=(s_0,i_0,j_0,s_1,i_1,j_1,\ldots,s_{\tau-1},i_{\tau-1},j_{\tau-1})$ *that could have actually occurred up to time* τ, $\tau\geq 1$. *Here* s_k *represents the state and* i_k *and* j_k *the action of player* 1 *and player* 2 *respectively at time* k, $k=0,1,\ldots,\tau-1$.

Obviously the set of histories up to time τ equals H^τ, i.e. the τ-fold Cartesian product of H.

First we shall describe the different types of strategies that a player may use and next give a formal definition. A behaviour strategy μ of player 1 specifies for each decision epoch τ, each state s_τ on time τ and each history h_τ a probability distribution $\mu_\tau(h_\tau,s_\tau)$ on the action space A_{s_τ} of player 1 in state s_τ. Then $\mu_\tau(i|h_\tau,s_\tau)$ is the probability with which player 1 chooses action $i\in A_{s_\tau}$ at time τ if state s_τ and history h_τ have occurred.

A semi-Markov strategy for player 1 is a behaviour strategy for which $\mu_\tau(h_\tau,s_\tau)$ only depends on h_τ through s_0; so $\mu_\tau(h_\tau,s_\tau)$ is of the form $\mu_\tau(s_0,s_\tau)$.

A Markov strategy for player 1 is a semi-Markov strategy for which $\mu_\tau(s_0,s_\tau)$ does not depend on s_0; so $\mu_\tau(s_0,s_\tau)$ is of the form $\mu_\tau(s_\tau)$.

A stationary strategy for player 1 is a Markov strategy for which $\mu_\tau(s_\tau)$ does not depend on τ; so $\mu_\tau(s_\tau)$ is of the form $\mu(s_\tau)$. In the sequel a stationary strategy for player 1 shall be denoted by the symbol ρ. Then $\rho=(\rho_1,\ldots,\rho_z)$, where ρ_s is a probability measure on the action space A_s for each $s\in S$. So $\rho_s\in P(A_s)$. If player 1 decides to play the stationary strategy ρ, then every time that the system is in state s, player 1 selects his pure action according to ρ_s. A stationary strategy ρ is called pure if ρ_s is pure for each $s\in S$, i.e. $\rho_s(i_s)=1$ for some $i_s\in A_s$.

Strategies for player 2 are defined analogously. For player 2 a behaviour strategy is denoted by ν and a stationary strategy by σ. Formally the above concepts lead to the following definition.

2.2.2. DEFINITION. *A behaviour strategy* μ *for player* 1 *is a sequence* $\mu_0, \mu_1, \mu_2, \ldots$, *where* $\mu_0 \in X_{s=1}^{z} P(A_s)$ *and* $\mu_\tau : H^\tau \to X_{s=1}^{z} P(A_s)$ *for* $\tau \geq 1$.
A semi-Markov strategy μ^{SM} *for player* 1 *is a sequence* $\mu_0^{SM}, \mu_1^{SM}, \mu_2^{SM}, \ldots$, *where* $\mu_0^{SM} \in X_{s=1}^{z} P(A_s)$ *and* $\mu_\tau^{SM} : S \to X_{s=1}^{z} P(A_s)$ *for* $\tau \geq 1$.
A Markov strategy μ^M *for player* 1 *is a sequence* $\mu_0^M, \mu_1^M, \mu_2^M, \ldots$, *where* $\mu_\tau^M \in X_{s=1}^{z} P(A_s)$ *for* $\tau \geq 0$.
A stationary strategy ρ *for player* 1 *is an element of* $X_{s=1}^{z} P(A_s)$.
A pure stationary strategy ρ^p *for player* 1 *is an element of* $X_{s=1}^{z} A_s$.
Strategies for player 2 *are defined analogously.*

For player ℓ, $\ell = 1,2$ we denote by ST_ℓ, $SMST_\ell$, MST_ℓ, SST_ℓ and $PSST_\ell$ respectively the classes of behaviour strategies, semi-Markov strategies, Markov strategies, stationary strategies and pure stationary strategies.

2.2.3. REMARK. It should be noted, that the set ST_ℓ is not the most general class of strategies. If we were to represent a stochastic game in the so-called extensive form (see e.g. Von Neuman & Morgenstern (1944)), then this would yield a tree of infinite depth for the infinite horizon game. On this tree pure and mixed strategies could be defined in the sense of Kuhn (1953) and Aumann (1964). This procedure would lead to a class of strategies for which the set ST_ℓ is a proper subset. However, Aumann (1964) has proved for a certain class of games with perfect recall, including stochastic games with finite state space and finite action spaces as a special case, that every mixed strategy defined for the game in extensive form has an equivalent behaviour strategy. Here two strategies of a player are called equivalent if, for all strategies of the other players and for all starting states, both strategies yield on each decision epoch the same expected payoff for that player.

2.3. CRITERIA FUNCTIONS FOR STOCHASTIC GAMES.

A pair of strategies (μ,ν) induces for fixed starting state s and each time epoch τ a probability measure $\mathbb{P}_{s\mu\nu}(\tau)$ on the finite product space H^τ. By the Kolmogorov extension theorem (Kolmogorov (1933)) the sequence $\mathbb{P}_{s\mu\nu}(0), \mathbb{P}_{s\mu\nu}(1), \ldots$ can be extended in the classical way to a unique

probability measure $\mathbb{P}_{s\mu\nu}$ on the inifinite product space H^{∞}.

Given that player 1 and 2 choose strategy μ and strategy ν respectively, we define the following stochastic variables:

$X^{\tau}_{\mu\nu}$, representing the action of player 1 at epoch τ.

$Y^{\tau}_{\mu\nu}$, representing the action of player 2 at epoch τ.

$Z^{\tau}_{\mu\nu}$, representing the state at epoch τ.

Obviously the marginal distributions of $X^{\tau}_{\mu\nu}$, $Y^{\tau}_{\mu\nu}$ and $Z^{\tau}_{\mu\nu}$, for each $\tau\in\mathbb{N}$, are determined by $\mathbb{P}_{s\mu\nu}$. For initial state s the expected payoff at decision epoch τ is given by

$$(2.3.1)\qquad V^{\tau}_{s\mu\nu} := \mathbb{E}_s\{r(Z^{\tau}_{\mu\nu},X^{\tau}_{\mu\nu},Y^{\tau}_{\mu\nu})\}$$

$$= \sum_{(t,i,j)\in H} r(t,i,j).\mathbb{P}_{s\mu\nu}\{Z^{\tau}_{\mu\nu}=t;\ X^{\tau}_{\mu\nu}=i;\ Y^{\tau}_{\mu\nu}=j\}.$$

The way in which the stream of payoffs is evaluated specifies a particular game.

2.3.1. DEFINITION. *A discounted two-person zerosum stochastic game with interest rate* $\alpha\in(0,\infty)$ *is a two-person zerosum stochastic game situation for which the stream of expected payoffs is evaluated by*

$$V_{s\mu\nu} := \sum_{\tau=0}^{\infty} \left(\frac{1}{1+\alpha}\right)^{\tau}.V^{\tau}_{s\mu\nu}.$$

Note in definition 2.3.1 that $V_{s\mu\nu}$ equals the total discounted expected payoff when the discount factor equals $(1+\alpha)^{-1}$, the starting state is s and the players choose μ and ν respectively as their strategies. Since the state and action spaces are assumed to be finite, $V_{s\mu\nu}$ exists for all (μ,ν).

2.3.2. DEFINITION. *An average reward two-person zerosum stochastic game is a stochastic game situation for which the stream of payoffs is evaluated by*

$$W_{s\mu\nu} := \liminf_{k\to\infty} \frac{1}{k+1} \sum_{\tau=0}^{k} V^{\tau}_{s\mu\nu}.$$

Note in definition 2.3.2 that $W_{s\mu\nu}$ equals the average expected payoff per unit time when the starting state is s and the players choose μ and ν respectively as their strategies. Obviously $W_{s\mu\nu}$ exists for all (μ,ν). In chapter III of this monograph we mention some other possibilities of averaging the stream of immediate payoffs.

The two types of games defined in definitions 2.3.1 and 2.3.2 will be studied in the chapters II and III respectively.

Other evaluation functions are possible. For example Groenewegen (1981) analysed the total expected payoff without discounting, where the expected immediate payoffs are simply added up (i.e. interest rate $\alpha=0$).

Whatever the evaluation function, player 1 clearly wishes to maximize this function and player 2 wishes to minimize it. The following definition applies to both discounted and average reward games. Compare the similarity of definition 2.3.3 with definition A.1.3. Note that, given the evaluation function and the initial state, the stochastic game can be identified with a game in normal form, namely $< ST_1, ST_2, G_{s\mu\nu} >$, where G represents the evaluation function.

2.3.3. DEFINITION. *Let* G *represent the evaluation function for a two-person zerosum stochastic game. The game is said to have a value if for each initial state* $s\in S$:

$$\sup_{\mu\in ST_1} \inf_{\nu\in ST_2} G_{s\mu\nu} = \inf_{\nu\in ST_2} \sup_{\mu\in ST_1} G_{s\mu\nu}.$$

For games which have a value, say $G^* \in \mathbb{R}^Z$, *for given* $\varepsilon \geq 0$, *a strategy* μ_ε *of player* 1 *and a strategy* ν_ε *of player* 2 *are called* ε*-optimal respectively if for each* $s \in S$ *respectively*:

$$\inf_{\nu\in ST_2} G_{s\mu_\varepsilon\nu} \geq G^*_s - \varepsilon \qquad \sup_{\mu\in ST_1} G_{s\mu\nu_\varepsilon} \leq G^* + \varepsilon.$$

Zero-optimal strategies are called optimal.

The next theorem follows at once from theorem A.1.4.

2.3.4. THEOREM. *Let* G *be the evaluation function. If there exist a vector* $v\in\mathbb{R}^Z$ *and strategies* $\hat{\mu}$ *and* $\hat{\nu}$ *such that* $G_{s\mu\hat{\nu}} \leq v_s \leq G_{s\hat{\mu}\nu}$ *for all* μ, ν *and* s, *then* v *equals the value of the game and* $\hat{\mu}$ *and* $\hat{\nu}$ *are optimal strategies for player* 1 *and player* 2 *respectively.*

Also theorem A.1.5 can be extended to stochastic games (see Hordijk, Vrieze & Wanrooij (1976, 1983) for a proof).

2.3.5. THEOREM. *Let* G *be the evaluation function. If, for each* $\varepsilon>0$, *there exists* μ_ε *and* ν_ε *such that for each* μ, ν *and* s:

$$G_{s\mu\nu_\varepsilon}-\varepsilon \leq G_{s\mu_\varepsilon\nu_\varepsilon} \leq G_{s\mu_\varepsilon\nu}+\varepsilon,$$

then the value of the game exists and for the specific game with starting state $s\in S$ *this value equals* $\lim_{\varepsilon\downarrow 0} G_{s\mu_\varepsilon\nu_\varepsilon}$.

3. Playing against a fixed strategy

In this section, we consider how player 1 can profit from the announcement by player 2 of the strategy he intends to play.

When in a finite two-person zerosum stochastic game player 2 plays a behaviour strategy, then Monash (1979; theorem 1, page 6) has proved that player 1 can suffice with non-randomized strategies. Actually, Monash proved this fact for the average reward case, but in a similar way it can be shown for the discounted case. Below we examine what happens when player 2 fixes a semi-Markov strategy or a stationary strategy.

The theorems of this section have been extracted from Hordijk, Vrieze & Wanrooij (1976, 1983). The statements of that paper are extensions to stochastic game situations of results of Derman & Strauch (1966) for Markov decision situations. A version of this extension, which is similar to theorem 3.1, is also indicated in Groenewegen & Wessels (1976). As shown in Hordijk, Vrieze & Wanrooij (1976), the results of this section hold for N-person games with countable state space and countable action spaces. But since this thesis is mainly concerned with finite two-person zerosum stochastic games, we have projected these results on this last model.

Theorem 3.3 states that, if one player plays a stationary strategy, then the other player can restrict himself to solving a Markov decision problem associated to that stationary strategy.

3.1. THEOREM. *For a two-person zerosum stochastic game situation, let* ν *be a semi-Markov strategy for player* 2. *Then for each behaviour strategy* μ *of player* 1, *there exists a semi-Markov strategy* μ^{SM} *such that*

$$V^{\tau}_{s\mu\nu} = V^{\tau}_{s\mu^{SM}\nu} \qquad \textit{for each } s\in S \textit{ and } \tau=0,1,\ldots .$$

PROOF. Fix a starting state $\hat{s}\in S$ and fix a strategy μ of player 1. We will abbreviate $\mathbb{P}_{\hat{s}\mu\nu}$, $X^{\tau}_{\mu\nu}$, $Y^{\tau}_{\mu\nu}$ and $Z^{\tau}_{\mu\nu}$ to $\mathbb{P}_{\hat{s}}$, X^{τ}, Y^{τ} and Z^{τ} respectively. (cf. section 2.3 for the meaning of these variables).
For all $\tau=0,1,2,\ldots$ and each s, i and j:

$$\mathbb{P}_{\hat{s}}(Z^{\tau}=s;\ X^{\tau}=i;\ Y^{\tau}=j) = \tag{3.1}$$

$$\mathbb{P}_{\hat{s}}(X^{\tau}=i \,|\, Z^{\tau}=s;\ Y^{\tau}=j).\mathbb{P}_{\hat{s}}(Z^{\tau}=s;\ Y^{\tau}=j).$$

Since ν is a semi-Markov strategy, the random variables X^τ and Y^τ, given $\hat{s}$ and s, are independent. Then

$$\mathbb{P}_{\hat{s}}(X^\tau=i \,|\, Z^\tau=s;\ Y^\tau=j) = \mathbb{P}_{\hat{s}}(X^\tau=i \,|\, Z^\tau=s).$$

So (3.1) becomes

$$\mathbb{P}_{\hat{s}}(Z^\tau=s;\ X^\tau=i;\ Y^\tau=j) = \mathbb{P}_{\hat{s}}(X^\tau=i \,|\, Z^\tau=s)\,\mathbb{P}_{\hat{s}}(Z^\tau=s;\ Y^\tau=j). \tag{3.2}$$

Now, define μ^{SM} as follows. If the initial state is $\hat{s}$ and the state at time τ is s, then choose action i with probability $\mathbb{P}_{\hat{s}}(X^\tau=i \,|\, Z^\tau=s)$.
We will abbreviate $\mathbb{P}_{\hat{s}\mu^{SM}\nu}$ to $\mathbb{P}^*_{\hat{s}}$.
By induction with respect to τ, we first show that

$$\mathbb{P}^*_{\hat{s}}(Z^\tau=s;\ X^\tau=i;\ Y^\tau=j) = \mathbb{P}_{\hat{s}}(Z^\tau=s;\ X^\tau=i;\ Y^\tau=j). \tag{3.3}$$

This equality is easily reached for $\tau=0$; suppose it holds for some τ, then

$$\mathbb{P}_{\hat{s}}(Z^{\tau+1}=t) = \tag{3.4}$$

$$\sum_{s,i,j} \mathbb{P}_{\hat{s}}(Z^\tau=s;\ X^\tau=i;\ Y^\tau=j).p(t \,|\, s,i,j) =$$

$$\sum_{s,i,j} \mathbb{P}^*_{\hat{s}}(Z^\tau=s;\ X^\tau=i;\ Y^\tau=j).p(t \,|\, s,i,j) =$$

$$\mathbb{P}^*_{\hat{s}}(Z^{\tau+1}=t).$$

Since ν is a semi-Markov strategy, (3.4) leads to

$$\mathbb{P}_{\hat{s}}(Z^{\tau+1}=t;\ Y^{\tau+1}=j) = \mathbb{P}^*_{\hat{s}}(Z^{\tau+1}=t;\ Y^{\tau+1}=j). \tag{3.5}$$

Now the definition of μ^{SM}, (3.2) for $\tau+1$ and (3.5) imply equality (3.3) for $\tau+1$. But then the theorem follows from the definitions of $V^\tau_{\hat{s}\mu\nu}$ and $V^\tau_{\hat{s}\mu^{SM}\nu}$ (see (2.3.1) and (3.3)).

□

3.2. THEOREM. *Consider a two-person zerosum stochastic game with either the total discounted payoff or the average reward as criterion function. Suppose that for the game where the both players are restricted to playing semi-Markov strategies, the value exists. Then for the unrestricted game the value also exists and equals the value of the restricted game. Moreover an* ε*-optimal strategy, for given* $\varepsilon \geq 0$*, for a player in the restricted game is also* ε*-optimal in the original game.*

PROOF. Let G represent the evaluation function and let G^* be the value of the restricted game. Let ν_ε be an ε-optimal semi-Markov strategy for player 2 in the restricted game, given $\varepsilon>0$. Such a strategy exists since the value exists. By theorem 3.1 there exists for each behaviour strategy μ of player 1 a semi-Markov strategy μ^{SM} such that for each $s \in S$:

$$G_{s\mu\nu_\varepsilon} = G_{s\mu^{SM}\nu_\varepsilon}. \tag{3.6}$$

On the other hand for each $s \in S$:

$$G_{s\mu^{SM}\nu_\varepsilon} \leq G_s^* + \varepsilon, \tag{3.7}$$

so

$$G_{s\mu\nu_\varepsilon} \leq G_s^* + \varepsilon \quad \text{for each } \mu \text{ and } s \in S. \tag{3.8}$$

Similarly

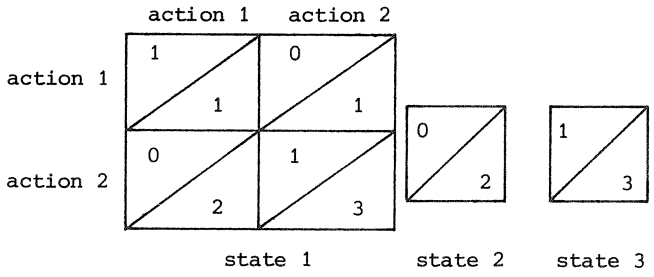
$$G_s^* - \varepsilon \leq G_{s\mu_\varepsilon\nu} \quad \text{for each } \nu \text{ and } s \in S, \tag{3.9}$$

where μ_ε is ε-optimal for player 1 in the restricted game. Since $\varepsilon>0$ is arbitrary, we may apply theorem 2.3.5 to the combination of (3.8) and (3.9). Hence the value of the unrestricted game exists and equals $\lim_{\varepsilon \downarrow 0} G_{s\mu_\varepsilon\nu_\varepsilon} = G_s^*$, $s \in S$. Furthermore by (3.8) and (3.9) (which also hold in the case $\varepsilon=0$) it follows that μ_ε and ν_ε are also ε-optimal in the original game.

□

The converse of this theorem is not true, i.e. the existence of the value of a game needs not imply the existence of the value of the restricted game. An example of such a game is the "big match" of Blackwell & Ferguson (1968).

3.3. EXAMPLE.



In this example there are three states, state 1, 2 and 3; in state 1 both players have two actions while in the states 2 and 3 both players have 1 action. In examples always player 1-th actions correspond with the rows and player 2-th actions with the columns of the matrices. An entry $\begin{array}{|c/c|}\hline \gamma & \\ \hline & \delta \\ \hline \end{array}$ of a matrix means immediate payoff γ and the next state will be state δ with probability 1. So for this example

$$S = \{1,2,3\},\ A_1 = B_1 = \{1,2\},\ A_2 = A_3 = B_2 = B_3 = \{1\}$$

and

$$p(1|1,1,1) = p(1|1,1,2) = p(2|1,2,1) = p(3|1,2,2) =$$

$$p(2|2,1,1) = p(3|3,1,1) = 1,$$

while the other transition probabilities are zero.

The average reward criterion is considered. Obviously, by the results of Mertens & Neyman (1980) this game has a value. Here $W^* = (\frac{1}{2},0,1)$. But if the players are restricted to semi-Markov strategies, then the value of the game does not exist. This can be shown as follows (cf. Hordijk, Vrieze & Wanrooij (1976)).

In the big match the set of semi-Markov strategies for a player is the same as the set of Markov strategies, since the states 2 and 3 are absorbing. Let $\mu^M = (\mu_0^M, \mu_1^M, \ldots)$ be a Markov strategy for player 1. Let $p_\tau(\mu^M)$ be the probability that in state 1 player 1 chooses action 2 for the first time at epoch τ. Let $p(\mu^M) = \Sigma_{\tau=0}^{\infty} p_\tau(\mu^M)$ and abbreviate $p_\tau(\mu^M)$ and $p(\mu^M)$ to p_τ and p.

For each $\delta>0$ there exists a $\hat{\tau}$ such that $\Sigma_{\tau=0}^{\hat{\tau}} p_\tau \geq p-\delta$. We construct a Markov strategy ν_δ for player 2 as follows:

choose in state 1 action 1 at the epochs $0,1,\ldots,\hat{\tau}$ and action 2 thereafter.

If player 1 plays μ^M and player 2 plays ν_δ, the game reduces to a stochastic process that realizes exactly one of the following events:

(i) player 1 uses action 2 before time $\hat{\tau}+1$.

(ii) player 1 uses action 2 for the first time at $\hat{\tau}+1$ or thereafter.

(iii) player 1 never uses action 2.

The probability that (i) occurs is at least $p-\delta$ and the average reward in that case is 0. Event (ii) has average return 1 but probability at most δ. Event (iii) has probability $1-p$ and average return 0. Hence the overall average payoff is at most δ, so $\inf_{\nu\in MST_2} W_{1\mu^M\nu} \leq \delta$ which by the arbitrariness of $\delta>0$ results in $\inf_{\nu\in MST_2} W_{1\mu^M\nu} \leq 0$. Hence $\sup_{\mu\in MST_1} \inf_{\nu\in MST_2} W_{1\mu\nu} \leq 0$.

The value of the restricted game, if it exists, equals the value of the original game by theorem 3.2. Since for the original game $W_1^* = \frac{1}{2}$, we see that the restricted game has no value.

Now we come to analyse what happens when a player fixes a stationary strategy. With a fixed stationary strategy $\sigma\in SST_2$, we can associate a Markov decision situation MDS(σ) in the following way:
Let $< S, \{A_s \mid s\in S\}, \{B_s \mid s\in S\}, r, p >$ be the original stochastic game situation. Then MDS(σ) $:= <\bar{S}, \{\bar{A}_s \mid s\in S\}, \bar{r}, \bar{p}>$ is defined as $\bar{S} := S$; $\bar{A}_s := A_s$ for each $s\in S$; $\bar{r}(s,i) := \Sigma_{j\in B_s} r(s,i,j).\sigma_s(j)$ and $\bar{p}(t|s,i) = \Sigma_{j\in B_s} p(t|s,i,j).\sigma_s(j)$.

3.4. THEOREM. *Suppose that in a two-person zerosum stochastic game player* 2 *fixes a stationary strategy* σ. *Then for as well the discounted as the average reward criterion player* 1 *cannot do better than solving the Markov decision problem corresponding to* MDS(σ).

PROOF. We denote a strategy with respect to MDS(σ) by $\bar{\mu}$. Let G be the evaluation function for the game and $\bar{G}$ for the Markov decision problem corresponding to MDS(σ). From theorem 3.1 we derive, that for each strategy μ there exists a semi-Markov strategy μ^{SM} such that for each starting state $s\in S$:

(3.10) $\quad G_{s\mu\sigma} = G_{s\mu^{SM}\sigma}.$

Now observe that there is a one-to-one correspondence between the set of semi-Markov strategies in MDS(σ) and the set of semi-Markov strategies for player 1 in the original game. Simply by adding to a $\bar{\mu}^{SM}=(\bar{\mu}_0^{SM},\bar{\mu}_1^{SM},\ldots)\in\overline{SMST}$ that $\mu^{SM}=(\mu_0^{SM},\mu_1^{SM},\ldots)\in SMST_1$ for which $\mu_\tau^{SM}=\bar{\mu}_\tau^{SM}$ for each $\tau=0,1,\ldots$.
Note that such a one-to-one correspondence does not exist for the sets $\overline{ST}$ and ST_1.
So if we can prove for corresponding $\mu^{SM}\in SMST_1$ and $\bar{\mu}^{SM}\in\overline{SMST}$ that for each $s\in S$:

(3.11) $\quad G_{s\mu^{SM}\sigma} = \bar{G}_{s\bar{\mu}^{SM}}$,

then (3.11) combined with (3.10) will yield the theorem.

Fix an initial state $\hat{s}\in S$. We abbreviate $\mathbb{P}_{\hat{s}\mu^{SM}\sigma}$ and $\bar{\mathbb{P}}_{\hat{s}\bar{\mu}^{SM}}$ to $\mathbb{P}$ and $\bar{\mathbb{P}}$ respectively. We first show by induction with respect to τ that

(3.12) $\quad \mathbb{P}(Z^\tau=s) = \bar{\mathbb{P}}(\bar{Z}^\tau=s)$

for corresponding μ^{SM} and $\bar{\mu}^{SM}$ and each $s\in S$.
Equality (3.12) is clearly true for $\tau=0$; suppose it holds for a certain τ, then:

$$\mathbb{P}(Z^{\tau+1}=t) = \sum_s \mathbb{P}(Z^\tau=s).\mathbb{P}(Z^{\tau+1}=t\,|\,Z^\tau=s) =$$

$$\sum_s \bar{\mathbb{P}}(\bar{Z}^\tau=s).\sum_{i,j} p(t\,|\,s,i,j).\mathbb{P}(X^\tau=i\,|\,Z^\tau=s).\sigma_s(j) =$$

$$\sum_s \bar{\mathbb{P}}(\bar{Z}^\tau=s).\sum_i\Big(\sum_j p(t\,|\,s,i,j).\sigma_s(j)\Big).\bar{\mathbb{P}}(\bar{X}^\tau=i\,|\,\bar{Z}^\tau=s) =$$

$$\sum_s \bar{\mathbb{P}}(\bar{Z}^\tau=s).\sum_i \bar{p}(t\,|\,s,i).\bar{\mathbb{P}}(\bar{X}^\tau=i\,|\,\bar{Z}^\tau=s) =$$

$$\bar{\mathbb{P}}(\bar{Z}^{\tau+1}=t).$$

Now (3.11) follows from the fact, that for each τ:

$$V^{\tau}_{s\mu SM_\sigma} = \sum_{s,i,j} r(s,i,j).\mathbb{P}(Z^\tau=s;\ X^\tau=i,\ Y^\tau=j) =$$

$$\sum_{s,i}\Big(\sum_{j} r(s,i,j).\sigma_s(j)\Big).\mathbb{P}(Z^\tau=s).\mathbb{P}(X^\tau=i \,|\, Z^\tau=s) =$$

$$\sum_{s,i} \bar{r}(s,i).\overline{\mathbb{P}}(\overline{Z}^\tau=s).\overline{\mathbb{P}}(\overline{X}^\tau=i \,|\, \overline{Z}^\tau=s) = \overline{V}^{\tau}_{s\bar{\mu}SM}.$$

□

The well-known results for Markov decision problems (see the theorems A.2.5 and A.2.6) lead to the following corollary of theorem 3.4.

3.5. COROLLARY. *For a two-person zerosum stochastic game with evaluation function* G, *which represents discounting or averaging, and a fixed stationary strategy* σ *of player* 2 *we have for each* $s \in S$:

$$\sup_{\mu \in ST_1} G_{s\mu\sigma} = \max_{\rho \in PSST_1} G_{s\rho\sigma}.$$

Part II. *Discounted stochastic games.*

4. Review of discounted stochastic games.

4.1. INTRODUCTION.

Stochastic games were introduced by Shapley (1953). He considered both finite and infinite horizon two-person zerosum stochastic games with finite state set and finite action sets. Shapley proved that such games have a value and that both players possess optimal stationary strategies with respect to the discounted payoff criterion.

Extensions of Shapley's model with respect to the conditions on state space and action spaces are exposed by Kushner & Chamberlain (1969) (finite state space, compact action spaces), Maitra & Parthasarathy (1970) (compact metric state space and compact action spaces plus continuity conditions on the payoff function and the transition map), Wessels (1977) (countable state space and finite action spaces) and Groenewegen & Wessels (1976) (countable state space and countable action spaces).

Rogers (1969) has extended Shapley's model to two-person non-zerosum games with finite state space and finite action spaces. He has proved the existence of equilibrium points of stationary strategies. Parthasarathy (1971) considered Roger's model with a countable state space.

Sobel (1971) introduced N-person stochastic games. For the model with finite state space and finite action spaces he showed the existence of an equilibrium point of stationary strategies (cf. also Federgruen (1978)). Vrieze (1976) considered countable person games with a countable state space and compact action spaces. Under appropriate continuity assumptions he showed the existence of an equilibrium point of stationary strategies. Rieder (1979) also considered countable person games. Using a measurable selection theorem he showed the existence of a stationary equilibrium point for the model where the state and action spaces are Borelsets and the transition law is given by a bounded transition measure. Tijs (1980) treated N-person games with a finite state space and with metric action spaces. Under certain continuity assumptions he proved the existence of a stationary ε-equilibrium point for the model where in each state at most one of the action spaces is topologically big, while the other action spaces are topologically small (finite, compact, precompact).

Finally we mention the work of Whitt (1977). He approximated games with uncountable state space and uncountable action spaces for the players by games with countable state space and countable action spaces. Subsequently he showed that under a number of conditions equilibrium points of the approximating game are ε-equilibrium points of the original game.

In the sequel $SG(S,\alpha)$ denotes the class of two-person zerosum stochastic games with finite state space S, finite action spaces for the both players and where the stream of payoffs are discounted according to an interest rate α.

In section 4.2 some well-known facts are given concerning games belonging to the class $SG(S,\alpha)$. In section 4.3 some extensions of this model are considered. In section 4.4 an alternative proof of the existence of the value and of optimal stationary strategies is worked out, using the Kuhn-Tucker conditions for non-linear programs.

In chapter 5 the emphasis lies on structural properties of the solutions of discounted stochastic games. In section 5.1 relations between the game parameters and the solutions are investigated, which smoothes the way for an analysis of the construction of games with given solution. In section 5.2 we give an axiomatic characterization of the value function on the class $SG(S,\alpha)$. In section 5.3 we treat to what extent small perturbations of the game parameters influence the solution of the game. In section 5.4 games with a unique pair of optimal stationary strategies stand central.

Sections 5.3 and 5.4 are the only sections of this monograph where we consider another model than the two-person zerosum stochastic game with finite state and finite actions spaces. In those sections the model has the following dimensions: countable state space, compact metric action spaces and measurable payoff and transition functions.

4.2. EXISTENCE OF VALUE AND OPTIMAL STRATIONARY STRATEGIES.

In discounted stochastic games as treated by Shapley (1953), a discount factor $\beta\in[0,1)$ is specified. Then a reward r earned on decision epoch τ is discounted by the factor β^{τ}, $\tau=0,1,\ldots$. The idea is that, when

looking ahead at time zero, a payoff r on decision epoch τ is worth at decision epoch 0 only $\beta^{\tau}.r$, this because of e.g. inflation. One can also use the concept of interest rate. Say $\alpha\in(0,\infty)$ is the interest rate per unit time. Then an amount r on decision epoch 0 has grown to an amount $(1+\alpha)^{\tau}.r$ at decision epoch α. This is the same as saying that in order to have an amount r on epoch τ, one should start with an amount $(1+\alpha)^{-\tau}.r$ at time 0. So an interest rate α corresponds to a discount factor $(1+\alpha)^{-1}$. This equivalence will be used in the sequel. We consider only infinite horizon games.

For the rest of this section we fix a game $\Gamma\in SG(S,\alpha)$. Let $M:=\max_{s,i,j}|r(s,i,j)|$. Then it follows that for each pair of strategies μ and ν the total discounted payoff is bounded by $(1+\alpha)\alpha^{-1}M$.

We now introduce a number of notations. For a pair of stationary strategies ρ and σ we denote by $P(\rho,\sigma)$ a stochastic $z\times z$-matrix whose (s,t)-th entry equals

$$(4.2.1)\qquad p(t|s,\rho_s,\sigma_s) := \sum_{i\in A_s}\sum_{j\in B_s} p(t|s,i,j).\rho_s(i).\sigma_s(j),$$

i.e. if in state s the players play ρ_s and σ_s respectively, then the probability that at the next decision epoch the system is in state t equals $p(t|s,\rho_s,\sigma_s)$.
Further $r(\rho,\sigma)$ will denote the z-vector with s-th coordinate:

$$(4.2.2)\qquad r(s,\rho_s,\sigma_s) := \sum_{i\in A_s}\sum_{j\in B_s} r(s,i,j).\rho_s(i).\sigma_s(j).$$

We now define two maps, which play an essential role in the proof of the existence of the value and of optimal stationary strategies.
$L_{\alpha\rho\sigma}:\mathbb{R}^z\to\mathbb{R}^z$ is the map such that for each $v\in\mathbb{R}^z$:

$$(4.2.3)\qquad L_{\alpha\rho\sigma}(v) := r(\rho,\sigma)+(1+\alpha)^{-1}.P(\rho,\sigma).v$$

$U_{\alpha}:\mathbb{R}^z\to\mathbb{R}^z$ is the map such that for each $v\in\mathbb{R}^z$ and $s\in S$:

$$(4.2.4)\qquad U_{\alpha s}(v) := \max_{\rho_s}\min_{\sigma_s}\{r(s,\rho_s,\sigma_s)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\rho_s,\sigma_s).v_t\}$$
$$= \max_{\rho_s}\min_{\sigma_s}(L_{\alpha\rho\sigma}(v))_s.$$

Observe that $U_{\alpha s}(v)$ is the value of the $m_s \times n_s$-matrix game whose (i,j)-th entry equals $r(s,i,j)+(1+\alpha)^{-1}.\Sigma_{t=1}^{z} p(t|s,i,j).v_t$. This matrix game is denoted by $[G_{s\alpha}(v)]$.

Note that a stochastic game Γ in fact is a collection of games, namely $\{\Gamma_s | s \in S\}$, where Γ_s refers to the specific stochastic game with initial state s. This collection of games is solved simultaneously.
Let $V(\mu,\nu)$ be the z-vector whose s-th component equals $V_{s\mu\nu}$ (cf. definition 2.3.1).

4.2.1. THEOREM. *The maps* $L_{\alpha\rho\sigma}$ *and* U_α *are monotone contraction operators on* $\mathbb{R}^z$. *Hence both maps have a unique fixed point. The unique fixed point of* $L_{\alpha\rho\sigma}$ *equals the discounted payoff vector* $V(\rho,\sigma)$.

PROOF. If $v \leq w$, then

$$L_{\alpha\rho\sigma}(v) = r(\rho,\sigma)+(1+\alpha)^{-1}.P(\rho,\sigma).v \leq r(\rho,\sigma)+(1+\alpha)^{-1}.P(\rho,\sigma).w$$

$$= L_{\alpha\rho\sigma}(w).$$

By lemma A.1.7(c) for each $s \in S$:

$$U_{\alpha s}(v) = Val(G_{s\alpha}(v)) \leq Val(G_{s\alpha}(w)) = U_{\alpha s}(w).$$

So $L_{\alpha\rho\sigma}$ and U_α are monotone.
For $v,w \in \mathbb{R}^z$ we have

$$d(L_{\alpha\rho\sigma}(v),L_{\alpha\rho\sigma}(w)) = \|L_{\alpha\rho\sigma}(v)-L_{\alpha\rho\sigma}(w)\| = (1+\alpha)^{-1}.\|P(\rho,\sigma)(v-w)\| \leq$$

$$(1+\alpha)^{-1}\|v-w\| = (1+\alpha)^{-1}d(v,w).$$

Using lemma A.1.8:

$$d(U_\alpha(v),U_\alpha(w)) = \|U_\alpha(v)-U_\alpha(w)\| = \max_s |Val(G_{s\alpha}(v))-Val(G_{s\alpha}(w))|$$

$$\leq (1+\alpha)^{-1} \max_{s,i,j} \sum_{t=1}^{z} p(t|s,i,j).|(v_t-w_t)| \leq (1+\alpha)^{-1}\|v-w\|$$

$$= (1+\alpha)^{-1}d(v,w).$$

So both $L_{\alpha\rho\sigma}$ and U_α are contraction operators. Then, by the Banach-Picard fixed point theorem it follows that there exist unique vectors $v^*_{\rho\sigma}$ and V^* such that

$$L_{\alpha\rho\sigma}(v^*_{\rho\sigma}) = v^*_{\rho\sigma} \quad \text{and} \quad U_\alpha(V^*) = V^*.$$

To prove that $v^*_{\rho\sigma}$ equals $V(\rho,\sigma)$ we first note, that the (s,t)-th element of the matrix $P^\tau(\rho,\sigma)$ gives the probability that at time τ the system is in state t when the starting state is state s and the players choose the strategies ρ and σ respectively.
Then

$$(4.2.5) \qquad V(\rho,\sigma) = \sum_{\tau=0}^{\infty} (1+\alpha)^{-\tau} P^\tau(\rho,\sigma).r(\rho,\sigma)$$

$$= r(\rho,\sigma) + (1+\alpha)^{-1}.P(\rho,\sigma).\sum_{\tau=0}^{\infty} (1+\alpha)^{-\tau} P^\tau(\rho,\sigma).r(\rho,\sigma)$$

$$= r(\rho,\sigma) + (1+\alpha)^{-1}.P(\rho,\sigma).V(\rho,\sigma).$$

Hence $V(\rho,\sigma)$ equals the unique fixed point of $L_{\alpha\rho\sigma}$.

□

4.2.2. REMARK. By the contraction property of $L_{\alpha\rho\sigma}$ and U_α it follows that $V(\rho,\sigma)=\lim_{\tau\to\infty} L^\tau_{\alpha\rho\sigma}(x)$ and $V^*=\lim_{\tau\to\infty} U^\tau_\alpha(x)$ for each $x\in\mathbb{R}^Z$, where $L^\tau_{\alpha\rho\sigma}$ and U^τ_α are the τ-th iterates of $L_{\alpha\rho\sigma}$ and U_α respectively.

4.2.3. LEMMA. *If* $L_{\alpha\rho\sigma}(v)\leq v$ *then* $V(\rho,\sigma)\leq v$. *If* $L_{\alpha\rho\sigma}(v)<v$ *then* $V_{\rho\sigma}<v$. *The assertions remain true when the inequality signs are reversed.*

PROOF. By the monotonicity property of $L_{\alpha\rho\sigma}$ it follows by induction that $L^\tau_{\alpha\rho\sigma}(v)\leq v$ for each $\tau\geq 1$, when $L_{\alpha\rho\sigma}(v)\leq v$ $(<v)$. Hence $V(\rho,\sigma)=\lim_{\tau\to\infty} L^\tau_{\alpha\rho\sigma}(v)\leq v$ $(<v)$ in view of remark 4.2.2.

□

Now we have enough tools to prove Shapley's theorem.

4.2.4. THEOREM. *The unique fixed point* V^* *of the map* U_α *equals the value of the game. Stationary strategies* ρ^* *and* σ^* *are optimal for player* 1 *and player* 2 *respectively if, for all* ρ *and* σ:

(4.2.6) $\quad L_{\alpha\rho\sigma^*}(v^*) \leq v^* \leq L_{\alpha\rho^*\sigma}(v^*)$

i.e. if ρ_s^* *and* σ_s^* *are optimal actions for player* 1 *and player* 2 *respectively in the matrix game* $[G_{s\alpha}(v^*)]$, *for each* $s \in S$.

PROOF. Let v^* be the unique fixed point of U_α. Let ρ_s^* and σ_s^* be optimal actions for player 1 and player 2 respectively in the matrix game $[G_{s\alpha}(v^*)]$, so that (4.2.6) holds for ρ^* and σ^*. Then by lemma 4.2.3 we have for each ρ and σ:

$$v(\rho,\sigma^*) \leq v^* \leq v(\rho^*,\sigma).$$

But then by corollary 3.5 for each $s \in S$:

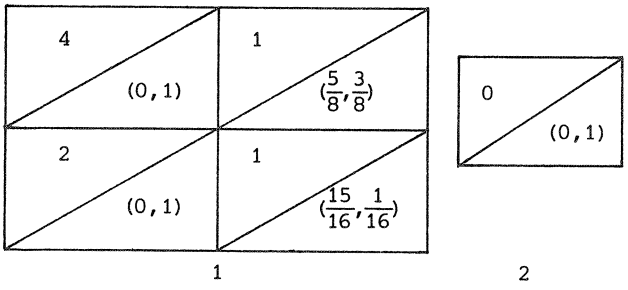
$$\sup_{\mu \in ST_1} v_{s\mu\sigma^*} \leq v_s^* \leq \inf_{\nu \in ST_2} v_{s\rho^*\nu}$$

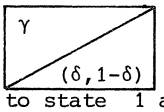
Hence by theorem 2.3.4 we see that v^* is the value of the game and that ρ^* and σ^* are optimal stationary strategies for player 1 and player 2 respectively.

□

The equation $v=U_\alpha(v)$ is often referred to as the optimality equation for discounted stochastic games.

4.2.5. REMARK. A game belonging to $SG(S,\alpha)$ with rational parameters need not have a rational value, as can be seen from the following game:



Here an entry 

means an immediate reward γ and a jump with probability δ to state 1 and a jump with probability $1-\delta$ to state 2. The interest rate equals $\alpha=\frac{1}{4}$, so the discount factor is $\beta = \frac{1}{1+\frac{1}{4}} = 4/5$. Obviously $V_2^*=0$ and then V_1^* is the solution of the equation

$$v = \mathrm{Val}\ \left(\begin{bmatrix} 4 & 1+\frac{1}{2}v \\ 2 & 1+\frac{3}{4}v \end{bmatrix}\right),$$

which results in $V_1^*=\sqrt{8}$.

4.3. GENERALIZATIONS.

In this section we discuss some generalizations of the model treated in section 4.2.

(i) Shapley (1953) himself considered the so-called stopping game, i.e. a finite two-person zerosum stochastic game situation where p has the property $\Sigma_{t=1}^{z} p(t|s,i,j)=p_{sij}<1$ for each s, i and j while $1-p_{sij}$ is the stopprobability. Subsequently the streng of immediate rewards is simply added up.

Such a game is equivalent with a discounted game in our sense, where in addition the interest rate may depend on the state and the actions. Namely if in state s the players choose actions i and j respectively, then the temporary interest rate equals α_{sij} defined by the equation $(1+\alpha_{sij})^{-1}=p_{sij}$. The transition probabilities are given by $p_{sij}^{-1}.p(t|s,i,j)$ when $p_{sij}>0$ and can be arbitrarily chosen when $p_{sij}=0$.

(ii) Van der Wal (1981) considered contracting games. He examined two-person zerosum stochastic games with countable state space and finite action spaces with the additional assumption that there exists a non-negative vector $\xi\in\mathbb{R}^z$, such that:

(a) $|r(\rho,\sigma)|\leq M\xi$ for some constant $M\geq 0$ and all ρ and σ.

(b) $P(\rho,\sigma).\xi\leq\beta\xi$ for some $\beta\in[0,1)$ and all ρ and σ.

Again, in Van der Wal's model the rewards are added up. Indeed, concerning the finite state case, this model is a real generalization of Shapley's model.

Consider one of the most simple games imaginable, namely the game $\boxed{0/1}$. This game fits in Van der Wal's frame work (ξ=0, M and β arbitrary) but not in Shapley's. However the slightly more difficult game $\boxed{0/1}\boxed{1/\frac{1}{2}}$ cannot be embedded in Van der Wal's model, since this game fails to have the uniform tail condition on page 65 in Van der Wal (1981). (Here the entry $\boxed{1/\frac{1}{2}}$ means immediate reward 1, probability $\frac{1}{2}$ to stay in the only state of the game and probability $\frac{1}{2}$ that the game stops). A next generalization on this subject should cover this last example.

(iii) Another generalization is concerned with stochastic renewal games (see Denardo (1971), Sobel (1973) and Federgruen (1978)). In renewal games the time until the next decision epoch is a random variable whose probability distribution function only depends on the current state and the subsequently chosen actions of the players. The immediate payoffs are discounted.

For the above mentioned three models the existence proof of the value and of optimal stationary strategies runs analogously to the proof of theorem 4.2.4. The only difference lies in finding a suitable Banach space and showing that $L_{\alpha\rho\sigma}$ and U_α are monotone contraction operators on that space.

4.4. AN ALTERNATIVE PROOF OF THE EXISTENCE OF THE VALUE.

In this section we assume that each immediate payoff is positive. Observe that this restriction can be made without loss of generality, since adding a constant c to each immediate payoff changes, for each pair of strategies of the players, the total discounted payoff by $\alpha^{-1}(1+\alpha)c1_z$. Hence the value changes by $\alpha^{-1}(1+\alpha)c1_z$, while all intrinsic game properties are remained.

Consider the following non-linear programming problem associated with a stochastic game belonging to SG(S,α).

4.4.1. NLP

variables $y = \{y(s,j) \mid s\in S,\ j\in B_s\}$ and $v=(v_1,\ldots,v_z)$.

min $\Sigma_{s=1}^{z} v_s$, subject to

$$\text{(i)} \quad \sum_{j=1}^{n_s} y(s,j)\Big[r(s,i,j)+(1+\alpha)^{-1}\sum_{t=1}^{z} p(t\mid s,i,j)v_t\Big]-v_s \le 0, \quad \text{for all } s\in S \text{ and } i\in A_s$$

$$\text{(ii)} \quad \sum_{j=1}^{n_s} y(s,j)-1 \le 0, \quad \text{for all } s\in S$$

$$\text{(iii)} \quad -\sum_{j=1}^{n_s} y(s,j)+1 \le 0, \quad \text{for all } s\in S$$

$$\text{(iv)} \quad -y(s,j) \le 0, \quad \text{for all } s\in S \text{ and } j\in B_s$$

This NLP is suggested by Rothblum (1979) as a solution method for stochastic games, whereby the existence of such a solution is pre-assumed. In this section we give a proof of the existence of the value and of optimal stationary strategies for both players with the aid of Kuhn-Tucker conditions with respect to optimal solutions of NLP 4.4.1.

First observe that there is a one-to-one correspondence between the set $F:=\{y=\{y(s,j)\mid s\in S,\ j\in B_s\}\mid y \text{ satisfies (ii), (iii) and (iv) of NLP 4.4.1}\}$ and the set $X_{s=1}^{z} P(B_s)$. This last set in turn corresponds to the set of stationary strategies of player 2 (cf. definition 2.2.2). For an element $y\in F$ we denote by σ^y the corresponding stationary strategy of player 2. For a stationary strategy σ we denote by y^σ the corresponding element of F. Obviously $\sigma=\sigma^{(y^\sigma)}$ and $y=y^{(\sigma^y)}$ for each $\sigma\in X_{s=1}^{z} P(B_s)$ and each $y\in F$.
In the following let $M:= \max_{s,i,j} r(s,i,j)$.

4.4.2. LEMMA. NLP 4.4.1 *is feasible.*

PROOF. Take $y\in F$. Let $v^M:=(v_1^M,\ldots,v_z^M)$ with $v_s^M:=\alpha^{-1}(1+\alpha)M$. The pair (y,v^M) satisfies (i)-(iv).

□

4.4.3. LEMMA. *Let* $(\bar{y},\bar{v})$ *be a local minimum, if it exists, of* NLP 4.4.1. *Then* $\Sigma_{s=1}^{z} \bar{v}_s \le z\alpha^{-1}(1+\alpha)M$ and $\bar{v}_s>0$ *for all* $s \in S$.

PROOF. For fixed $y=\bar{y}$ NLP 4.4.1 turns over into a linear programming problem, which we call LP($\bar{y}$). If $(\bar{y},\bar{v})$ is some local minimum of NLP 4.4.1, then $\bar{v}$ must be an optimal solution of LP($\bar{y}$). Let v^M be as in the proof of lemma 4.4.2. Then v^M is a feasible solution of LP($\bar{y}$), which shows half of the lemma, namely

$$\sum_{s=1}^{z} \bar{v}_s \leq \sum_{s=1}^{z} v_s^M = z\alpha^{-1}(1+\alpha)M.$$

On the other hand, let $\tilde{v}$ be a feasible solution of LP($\bar{y}$). Let state $\tilde{s}$ be such that $\tilde{v}_{\tilde{s}}=\min_{s\in S} \tilde{v}_s$. Then by condition (i):

$$\tilde{v}_{\tilde{s}} \geq \sum_{j=1}^{n_{\tilde{s}}} \bar{y}(\tilde{s},j)[r(\tilde{s},i,j)+(1+\alpha)^{-1} \sum_{t=1}^{z} p(t|\tilde{s},i,j)\tilde{v}_t]$$

$$\geq \sum_{j=1}^{n_{\tilde{s}}} \bar{y}(\tilde{s},j)[\min_{s,i,j} r(s,i,j)+(1+\alpha)^{-1}\tilde{v}_{\tilde{s}}] = \min_{s,i,j} r(s,i,j)+(1+\alpha)^{-1}\tilde{v}_{\tilde{s}}.$$

Hence for all $s \in S$: $\tilde{v}_s \geq \tilde{v}_{\tilde{s}} \geq \alpha^{-1}(1+\alpha) \min_{s,i,j} r(s,i,j) > 0$ (all rewards are assumed to be positive). Since $\tilde{v}$ is assumed to be an arbitrary feasible solution of LP($\bar{y}$) we can conclude $\bar{v}_s > 0$ for all $s \in S$.

□

4.4.4. LEMMA. NLP 4.4.1 *has a bounded global minimum, which is attained for some pair* (y^*,v^*).

PROOF. Add to NLP 4.4.1 the following constraint:

(v) $\qquad v_s - z\alpha^{-1}(1+\alpha)M \leq 0, \qquad$ for all $s\in S$.

The resulting non-linear programming problem is called NLP$^+$. From lemma 4.4.3 it can be seen that NLP$^+$ and NLP 4.4.1 have the same sets of local and global minima, if they exist.
Now for NLP$^+$ the feasible region is a non-empty compact set. Further the objective function is continuous on this feasible region, hence the minimum is attained for NLP$^+$ by some pair (y^*,v^*) satisfying (i)-(v). Then also for NLP 4.4.1 the minimum is attained for this pair (y^*,v^*).

□

Thus lemma 4.4.4 shows that NLP 4.4.1 has at least one local (or global) minimum point.

Observe that if (y,v) satisfies the constraints (i)-(iv), then, for each pure stationary strategy ρ^p of player 1, constraint (i) leads to

$$(4.4.1) \qquad v \geq r(\rho^p,\sigma^y)+(1+\alpha)^{-1}P(\rho^p,\sigma^y).v.$$

4.4.5. LEMMA. *If* (y,v) *satisfies the constraints* (i)-(iv) *then* $v \geq V(\rho^p,\sigma^y)$ *for each pure stationary strategy* ρ^p *of player* 1. *Further* $v \geq \sup_{\mu \in ST_1} V(\mu,\sigma^y)$.

PROOF. The first assertion follows from relation (4.4.1). If for v in (4.4.1) we repeatedly substitute the right hand side of (4.4.1), then we obtain for all $T \geq 1$:

$$(4.4.2) \qquad v \geq \sum_{\tau=0}^{T} (1+\alpha)^{-\tau} P^{\tau}(\rho^p,\sigma^y).r(\rho^p,\sigma^y)+(1+\alpha)^{-T-1}P^{T+1}(\rho^p,\sigma^y).v.$$

Letting $T \to \infty$ and using the boundedness of v and r and the fact that $\lim_{T\to\infty}(1+\alpha)^{-T-1}=0$ we obtain $v \geq V(\rho^p,\sigma^y)$.
The second assertion of the lemma follows from corollary 3.5. □

4.4.6. LEMMA. *For each local or global minimum point of* NLP 4.4.1 *the Kuhn-Tucker conditions hold.*

PROOF. We will show that the constraints (i)-(iv) satisfy the Arrow-Hurwicz-Uzawa constraint qualification (cf. Mangasarian (1969), page 102). Let $d=z+\Sigma_{s=1}^{z} n_s$. Let $(\bar{y},\bar{v})$ be feasible for NLP 4.4.1. Then the constraints satisfy the Arrow-Hurwicz-Uzawa constraint qualification at the point $(\bar{y},\bar{v})$ if the system

$$(4.4.3) \qquad \left\{ \begin{array}{l} \nabla g_K(\bar{y},\bar{v}).c > 0 \\ \\ \nabla g_L(\bar{y},\bar{v}).c \geq 0 \end{array} \right\} \text{ has a solution } c \in \mathbb{R}^d.$$

Here ∇ represents the gradient symbol,
$K:=\{(s,i) \mid g_{si}(\bar{y},\bar{v})=0$ and g_{si} is not concave at $(\bar{y},\bar{v})\}$,
$L:=\{(s,i) \mid g_{si}(\bar{y},\bar{v})=0$ and g_{si} is concave at $(\bar{y},\bar{v})\}$,

$$g_{si}(y,v) := \sum_{j=1}^{n_s} y(s,j)[r(s,i,j)+(1+\alpha)^{-1} \sum_{t=1}^{z} p(t|s,i,j)v_t]-v_s,$$

$g_K := \{g_{si} \mid (s,i)\in K\}$ and $g_L := \{g_{si} \mid (s,i)\in L\} \cup \{\text{linear constraints (ii)-(iv)}\}$.

In our case we have that in general g_{si} is not concave (nor convex). For $c\in\mathbb{R}^d$ denote the successive components by $c_1,\ldots,c_z$, $c_{11},\ldots,c_{1n_1}$, $c_{21},\ldots,c_{z-1n_{z-1}}$, $c_{z1},\ldots,c_{zn_z}$.
From constraint (i) it can be deduced that for a feasible point $(\bar{y},\bar{v})$ we have

$$(4.4.4) \qquad \nabla g_{si}(\bar{y},\bar{v}).c = (1+\alpha)^{-1} \sum_{j=1}^{n_s} \bar{y}(s,j) \sum_{t=1}^{z} p(t|s,i,j)c_t + \sum_{j=1}^{n_s} [r(s,i,j)+(1+\alpha)^{-1} \sum_{t=1}^{z} p(t|s,i,j)\bar{v}_t]c_{sj}-c_s.$$

Now let $(\bar{y},\bar{v})$ be a local minimum of NLP 4.4.1. Put $\bar{c}_s=-\bar{v}_s$, $s\in S$ and $\bar{c}_{sj}=0$, $s\in S$ and $j\in B_s$. Then we obtain from (4.4.4) that for all $(s,i)\in K$ we have

$$(4.4.5) \qquad \nabla g_{si}(\bar{y},\bar{v}).\bar{c} = \bar{v}_s-(1+\alpha)^{-1} \sum_{j=1}^{n_s} \bar{y}(s,j) \sum_{t=1}^{z} p(t|s,i,j)\bar{v}_t > 0,$$

in view of constraint (i) and the fact that $\sum_{j=1}^{n_s} \bar{y}(s,j)r(s,i,j) > 0$. Clearly $\nabla g_L(\bar{y},\bar{v}).\bar{c}=0$ for the linear constraints (ii)-(iv). Hence the Arrow-Hurwicz-Uzawa constraint qualification holds in the point $(\bar{y},\bar{v})$. Since further the objective function and each constraint function of NLP 4.4.1 is differentiable at $(\bar{y},\bar{v})$ we can apply theorem 7.3.7 of Mangasarian (1969) in order to conclude that the Kuhn-Tucker conditions hold for $(\bar{y},\bar{v})$.

Fix for a moment a local minimum point $(\bar{y},\bar{v})$ of NLP 4.4.1. Such a point exists by lemma 4.4.4. By lemma 4.4.5 the Kuhn-Tucker conditions hold for such a point, i.e. there exist, for each $s\in S$, $i\in A_s$ and $j\in B_s$, numbers λ_{si}, ϕ_{s1}, ϕ_{s2}, $\Phi_{sj} \geq 0$, such that

$$\text{(a)} \qquad (\sum_{j=1}^{n_s} \bar{y}(s,j)[r(s,i,j)+(1+\alpha)^{-1} \sum_{t=1}^{z} p(t|s,i,j)\bar{v}_t]-\bar{v}_s)\lambda_{si} = 0,$$

for all $s\in S$ and $i\in A_s$

(b) $\left(\sum_{j=1}^{n_s} \bar{y}(s,j)-1\right)\phi_{s1} = 0, \qquad$ for all $s\in S$

(c) $\left(-\sum_{j=1}^{n_s} \bar{y}(s,j)+1\right)\phi_{s2} = 0, \qquad$ for all $s\in S$

(d) $-\bar{y}(s,j)\phi_{sj} = 0, \qquad$ for all $s\in S$ and all $j\in B_s$

and

(e) $$1+\sum_{s=1}^{z}\sum_{i=1}^{m_s}\sum_{j=1}^{n_s}(1+\alpha)^{-1}\lambda_{si}\bar{y}(s,j)p(t|s,i,j)-\sum_{i=1}^{m_t}\lambda_{ti} = 0,$$

for all $t\in S$

(f) $$\sum_{i=1}^{m_s}\lambda_{si}[r(s,i,j)+(1+\alpha)^{-1}\sum_{t=1}^{z}p(t|s,i,j)\bar{v}_t]+\phi_{s1}-\phi_{s2}-\phi_{sj} = 0,$$

for all $s\in S$ and all $j\in B_s$.

Multiplying (f) by $\bar{y}(s,j)$, summing over $j\in B_s$, using (d), comparing that expression with the summation of (a) over $i\in A_s$ leads to

(4.4.6) $$\bar{v}_s\sum_{i=1}^{m_s}\lambda_{si} = \phi_{s2}-\phi_{s1}, \qquad \text{for all } s\in S.$$

From (e) we obtain $\sum_{i=1}^{m_s}\lambda_{si}\geq 1$, all $s\in S$, and then from (f) and (4.4.6) we derive (remembering $\phi_{sj}\geq 0$):

(4.4.7) $$\sum_{i=1}^{m_s}(\lambda_{si}/\Sigma_{k=1}^{m_s}\lambda_{sk})[r(s,i,j)+(1+\alpha)^{-1}\sum_{t=1}^{z}p(t|s,i,j)\bar{v}_t] \geq \bar{v}_s$$

for all $s\in S$ and all $j\in B_s$.

Define the stationary strategy $\bar{\rho}$ for player 1 as

(4.4.8) $$\bar{\rho}_s(i) := \lambda_{si}/\sum_{k=1}^{m_s}\lambda_{sk}.$$

Then (4.4.7) is equivalent to

(4.4.9) $$r(\bar{\rho},\sigma^p)+(1+\alpha)^{-1}P(\bar{\rho},\sigma^p)\bar{v} \geq \bar{v}$$

for each pure stationary strategy σ^p of player 2.

Inequality (4.4.9) is the "player 2 version" of inequality (4.4.1). Then analogous to the proof of lemma 4.4.5 inequality (4.4.9) leads to

$$(4.4.10) \qquad \inf_{\nu \in ST_2} V(\bar{\rho},\nu) \geq \bar{v}.$$

The above derivations have paved the way to the following existence theorem.

4.4.7. THEOREM. *Each local minimum of* NLP 4.4.1 *is also global. For each global minimum* $(\bar{y},\bar{v})$ *the* $\bar{v}$*-part is unique and say equals* v^*. *The value of the corresponding stochastic game exists and equals* v^*. *Both players possess optimal stationary strategies.*

PROOF. Let $(\bar{y},\bar{v})$ be a local minimum of NLP 4.4.1, which exists by lemma 4.4.4. Let $\bar{\rho}$ be as defined in (4.4.8).
Combining lemma 4.4.5 and inequality (4.4.10) yields

$$\sup_{\mu} V(\mu,\sigma^{\bar{y}}) \leq \bar{v} \leq \inf_{\nu} V(\bar{\rho},\nu)$$

Application of theorem 2.3.4 gives that $\bar{v}$ equals the value of the stochastic game and that $\bar{\rho}$ and $\sigma^{\bar{y}}$ are optimal stationary strategies for player 1 and player 2 respectively. Since a game has a unique value the uniqueness of the $\bar{v}$-part of a local minimum $(\bar{y},\bar{v})$ follows.

□

Summarizing, we have seen in this section that the existence of the value and of optimal stationary strategies for a discounted stochastic game can be shown by analysing the associated non-linear programming problem NLP 4.4.1. Moreover a solution method which gives a solution of NLP 4.4.1 and the corresponding Kuhn-Tucker coefficients, provides the value and a pair of optimal stationary strategies for the players of the stochastic game.

5. Structural properties of discounted stochastic games.

5.1. RELATIONS BETWEEN THE GAME PARAMETERS AND THE SOLUTION OF THE GAME.

We derive relations between, on the one hand, the game parameters and, on the other hand, the value and the set of optimal stationary strategies of that game.

The results of this section, based mainly on Vrieze & Tijs (1980), can be seen as extensions of the work of Bohnenblust, Karlin & Shapley (1950) and Shapley & Snow (1950).

For the moment we fix a game $\Gamma \in SG(S,\alpha)$. Let $\mathcal{O}_\ell$ denote the set of optimal stationary strategies for player ℓ, $\ell=1,2$.
The last assertion of the next lemma shows the dimension relations between the solution sets of discounted stochastic games (cf. definition A.1.10).

5.1.1. THEOREM. *For $\ell=1,2$ the set $\mathcal{O}_\ell$ can be identified with the Cartesian product $X_{s=1}^{z}\mathcal{O}_\ell(s)$, where $\mathcal{O}_\ell(s)$ is the set of optimal actions for player ℓ in the matrix game $[G_{s\alpha}(v^*)]$, where v^* is the value of the game. The pair $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ has the (m_s,n_s)-BKS property for all $s\in S$.*

PROOF. In section 4.2, theorem 4.2.4 we have already proved that an element of $X_{s=1}^{z}\mathcal{O}_\ell(s)$ is optimal for player ℓ. Now let $\rho\notin X_{s=1}^{z}\mathcal{O}_1(s)$ and particularly let $\rho_{\hat{s}}\notin\mathcal{O}_1(\hat{s})$. Then there exists a $\sigma\in SST_2$ such that $L_{\alpha\rho\sigma}(v^*)<v^*$ (strict inequality at least in component $\hat{s}$). But then, by lemma 4.2.3, $V(\rho,\sigma)<v^*$. So ρ cannot be optimal, which proves $\mathcal{O}_1=X_{s=1}^{z}\mathcal{O}_1(s)$. Analogously one can show $\mathcal{O}_2=X_{s=1}^{z}\mathcal{O}_2(s)$. That the pair $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ has the (m_s,n_s)-BKS property follows at once from theorem A.1.11.

□

Now we wish to present an extension of the results of Shapley & Snow (1950), concerning the characterization of extreme optimal mixed actions for matrix games (see theorem A.1.9). From theorem 5.1.1 it is evident that $\rho\in\mathcal{O}_1$ is an extreme point of $\mathcal{O}_1$ if and only if ρ_s is an extreme point of $\mathcal{O}_1(s)$ for all $s\in S$. The subsequent theorem follows at once from theorem A.1.9.

5.1.2. THEOREM. *Let* $\hat{\rho}$ *be an extreme point of* $\mathcal{O}_1$ *and* $\hat{\sigma}$ *be an extreme point of* $\mathcal{O}_2$. *Then there exists a stochastic subgame situation* $\langle S, \{\hat{A}_s | s\in S\}, \{\hat{B}_s | s\in S\}, \hat{f}, \hat{p}\rangle$, *where*

(a) *for all* $s\in S$ *the sets* $\hat{A}_s$ *and* $\hat{B}_s$ *are subsets of* A_s *and* B_s *respectively with* $|\hat{A}_s| = |\hat{B}_s|$ *and*

(b) $\hat{f}$ *and* $\hat{p}$ *are restrictions of the maps* r *and* p *to the set* $\{(s,i,j) | s\in S,\ i\in\hat{A}_s,\ j\in\hat{B}_s\}$,

such that the value of this subgame equals the value v^* *of the original game and such that* $\hat{\rho}$ *and* $\hat{\sigma}$ *can be calculated in the Shapley-Snow manner from the square matrix game* $[\hat{G}_{s\alpha}(v^*)]$ (*cf. theorem* A.1.9). *Here* $[\hat{G}_{s\alpha}(v^*)]$ *is the restriction of* $[G_{s\alpha}(v^*)]$ *to the rows and columns corresponding to* $\hat{A}_s$ *and* $\hat{B}_s$ *respectively*.

This theorem also suggests a method of finding the finite number of extreme optimal stationary strategies of $\mathcal{O}_1$ and $\mathcal{O}_2$ when the value of the game is known. This can be done by looking at the finite number of stochastic subgames in which both players have an equal number of pure actions at each state.

Now we analyse two problems concerning the construction of games with prescribed solution.

5.1.3. PROBLEM. *Let* S, (A_s,B_s), $s\in S$, p, α, V *and* $\mathcal{O}_\ell(s)$, $s\in S$, $\ell\in\{1,2\}$ *be given, where* $\mathcal{O}_1(s)\subset\mathcal{P}(A_s)$ *and* $\mathcal{O}_2(s)\subset\mathcal{P}(B_s)$ *are convex polyhedra. The question is whether it is possible to construct a function* r *such that*

(P_1) *the value of the corresponding game equals* V.

(P_2) $X_{s=1}^{z}\mathcal{O}_\ell(s)$ *is the set of optimal stationary strategies for player* ℓ, $\ell=1,2$.

5.1.4. PROBLEM. *Let* S, (A_s,B_s), $s\in S$, r, α, V *and* $\mathcal{O}_\ell(s)$, $s\in S$, $\ell\in\{1,2\}$ *be given, where* $\mathcal{O}_1(s)\subset\mathcal{P}(A_s)$ *and* $\mathcal{O}_2(s)\subset\mathcal{P}(B_s)$ *are convex polyhedra. The question is whether it is possible to construct a map p such that* (P_1) *and* (P_2) *hold.*

It will result that problem 5.1.4 is more difficult than problem 5.1.3.

5.1.5. THEOREM. *Problem* 5.1.3 *can be solved if and only if for each* $s \in S$ *the pair* $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ *has the* (m_s,n_s)-BKS *property.*

PROOF. If there exists a function r with the desired properties then by theorem 5.1.1 $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ must have the (m_s,n_s)-BKS property, $s \in S$. Now let $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ have the (m_s,n_s)-BKS property for all $s \in S$. Then by theorem A.1.11, there exists a m_s,n_s-matrix game $[K_s]=[k_s(i,j)]$ such that $\mathrm{Val}(K_s)=V_s$ and such that $\mathcal{O}_1(s)$ and $\mathcal{O}_2(s)$ are exactly the optimal action sets for player 1 respectively player 2. Now put

$$r(s,i,j) = k_s(i,j) - (1+\alpha)^{-1} \sum_{t=1}^{z} p(t|s,i,j).V_t \quad , s \in S,\ i \in A_s,\ j \in B_s$$

Then theorem 4.2.4 and theorem 5.1.1 show that this function r has the desired properties.

□

With regard to problem 5.1.4 it is clear that this problem can only have a solution when the pair $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ has the (m_s,n_s)-BKS property for each $s \in S$. However r, α and V cannot be chosen independently, as the following theorem shows.

5.1.6. THEOREM. *Concerning problem* 5.1.4, *necessary and sufficient conditions for the existence of a map* p *such that property* (P_1) *holds, are given by the following system of inequalities*:

(5.1.1) $\quad \underline{m} \le (1+\alpha)(V_s-w_s) \le \bar{m}$, all $s \in S$,

where

$$\underline{m} := \min_s V_s,\ \bar{m} := \max_s V_s \text{ and } w_s := \mathrm{Val}\,(r(s,.,.)).$$

PROOF. First suppose that there exists a map p such that property (P_1) holds. Let $\hat{\rho}$ be optimal for player 1 in the corresponding stochastic game and let $\tilde{\sigma}_s$ be an optimal action for player 2 in the matrix game $[r(s,.,.)]$. Then

$$V_s \le r(s,\hat{\rho}_s,\tilde{\sigma}_s)+(1+\alpha)^{-1}. \sum_{t=1}^{z} p(t|s,\hat{\rho}_s,\tilde{\sigma}_s).V_t$$

$$\le r(s,\hat{\rho}_s,\tilde{\sigma}_s)+(1+\alpha)^{-1}.\bar{m} \le w_s+(1+\alpha)^{-1}.\bar{m}.$$

Analogously one can prove $V_s \geq w_s+(1+\alpha)^{-1}.\underline{m}$, so (5.1.1) holds.

Now suppose that (5.1.1) holds. Take $\underline{s}\in S$ such that $V_{\underline{s}}=\underline{m}$ and $\bar{s}\in S$ such that $V_{\bar{s}}=\bar{m}$. In view of (5.1.1), for each $s\in S$ there exists an $\varepsilon_s\in[0,1]$ such that

$$(1+\alpha)(V_s-w_s) = \varepsilon_s\underline{m}+(1-\varepsilon_s)\bar{m} = \varepsilon_s V_{\underline{s}}+(1-\varepsilon_s)V_{\bar{s}}.$$

Then choose the map p in such a way, that for all $(i,j)\in A_s\times B_s$:

$$\begin{aligned} p(\underline{s}|s,i,j) &= \varepsilon_s \\ p(\bar{s}|s,i,j) &= 1-\varepsilon_s \\ p(t|s,i,j) &= 0, \quad \text{if } t\neq\underline{s},\bar{s}. \end{aligned}$$

Now

$$\mathrm{Val}(G_{s\alpha}(V)) = \mathrm{Val}(r(s,.,.)+V_s-w_s) = V_s$$

for each $s\in S$. So using theorem 4.2.4 we see that property (P_1) holds. □

5.1.7. THEOREM. *Concerning problem* 5.1.4, *a necessary condition for the existence of a map* p *such that* (P_1) *and* (P_2) *hold, is that for each* $s\in S$:

$$\begin{aligned} (5.1.2)\qquad \underline{m} &\leq (1+\alpha)(V_s - \max_{\sigma_s\in\mathcal{O}_2(s)} \max_{\rho_s\in\mathcal{P}(A_s)} r(s,\rho_s,\sigma_s)) \\ &\leq (1+\alpha)(V_s - \min_{\rho_s\in\mathcal{O}_1(s)} \min_{\sigma_s\in\mathcal{P}(B_s)} r(s,\rho_s,\sigma_s)) \leq \bar{m}. \end{aligned}$$

PROOF. Let a stochastic game have value V^* and let $\delta\in\mathcal{O}_2$. Then for each $s\in S$ and each $\rho_s\in\mathcal{P}(A_s)$:

$$r(s,\rho_s,\delta_s)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\rho_s,\delta_s).V_t^* \leq V_s^*$$

leads to

$$r(s,\rho_s,\delta_s)+(1+\alpha)^{-1}.\underline{m} \leq V_s^*.$$

So

$$\max_{\sigma_s\in\mathcal{O}_2(s)} \max_{\rho_s\in\mathcal{P}(A_s)} r(s,\rho_s,\sigma_s)+(1+\alpha)^{-1}\cdot\underline{m} \leq V_s^*,$$

which proves the first inequality of (5.1.2). Analogously the last inequality of (5.1.2) can be proved.

However

$$\max_{\rho_s \in \mathcal{P}(A_s)} \max_{\sigma_s \in \mathcal{O}_2(s)} r(s,\rho_s,\sigma_s) \geq \max_{\rho_s \in \mathcal{O}_1(s)} \max_{\sigma_s \in \mathcal{O}_2(s)} r(s,\rho_s,\sigma_s)$$

$$\geq \min_{\rho_s \in \mathcal{O}_1(s)} \min_{\sigma_s \in \mathcal{O}_2(s)} r(s,\rho_s,\sigma_s) \geq \min_{\rho_s \in \mathcal{O}_1(s)} \min_{\sigma_s \in \mathcal{P}(B_s)} r(s,\rho_s,\sigma_s).$$

This completes the theorem.

□

Not only for actions belonging to $\mathcal{O}_1(s)$ and $\mathcal{O}_2(s)$ there is a restrictive relation, but also for actions belonging to $\mathcal{P}(A_s) \setminus \mathcal{O}_1(s)$ and $\mathcal{P}(B_s) \setminus \mathcal{O}_2(s)$ there are conditions.

5.1.8. THEOREM. *Concerning problem* 5.1.4, *a necessary condition for the existence of a map* p *such that* (P_1) *and* (P_2) *hold, is that for each* $\hat{\rho}_s \in \mathcal{P}(A_s) \setminus \mathcal{O}_1(s)$,

$$\underline{m} < (1+\alpha)(V_s - \min_{\sigma_s \in \mathcal{P}(B_s)} r(s,\hat{\rho}_s,\sigma_s)),$$

and for each $\hat{\sigma}_s \in \mathcal{P}(B_s) \setminus \mathcal{O}_2(s)$,

$$(1+\alpha)(V_s - \max_{\rho_s \in \mathcal{P}(A_s)} r(s,\rho_s,\hat{\sigma}_s)) < \bar{m}.$$

PROOF. Suppose $\hat{\rho}_s \notin \mathcal{O}_1(s)$ for a stochastic game with value V^*. Then by theorem 5.1.1:

$$\min_{\sigma_s \in \mathcal{P}(B_s)} \{r(s,\hat{\rho}_s,\sigma_s) + (1+\alpha)^{-1} \cdot \sum_{t=1}^{z} p(t|s,\hat{\rho}_s,\sigma_s).V_t^*\} < V_s^*.$$

Or

$$\min_{\sigma_s \in \mathcal{P}(B_s)} \{r(s,\hat{\rho}_s,\sigma_s) + (1+\alpha)^{-1}.\underline{m}\} < V_s^*.$$

Analogously the second inequality can be shown.

□

It is not clear whether the conditions of theorem 5.1.7 and theorem 5.1.8 are also sufficient for problem 5.1.4 to have a solution. Perhaps some condition which relates $\mathcal{O}_\ell(s)$, $\ell=1,2$ and the sets of optimal actions

for the players in the matrix games $[r(s,.,.)]$, should be formulated. The fact that such a relationship exists is shown in a convincing way by the following theorem.

5.1.9. THEOREM. *If for problem* 5.1.4 *the vector* V *is such that* $V=v.1_z$ *with* $v\in\mathbb{R}$, *then problem* 5.1.4 *can be solved if and only if the following three conditions are fulfilled*:

(a) $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ *has the* (m_s,n_s)-BKS *property, all* $s\in S$.

(b) $\mathcal{O}_\ell(s)$ *is optimal for player* ℓ *in the game* $[r(s,.,.)]$, $\ell=1,2$.

(c) $\mathrm{Val}\,(r(s,.,.)) = \frac{\alpha}{1+\alpha}.V_s = \frac{\alpha}{1+\alpha}.v$, all $s\in S$.

PROOF. For a discounted stochastic game with value $V^*=v.1_z$ it follows that $\Sigma_t p(t|s,\rho_s,\sigma_s).V_t^*=v$ for all s, ρ_s and σ_s. So

$$v = \mathrm{Val}\,(G_{s\alpha}(V^*)) = \mathrm{Val}(r(s,.,.)) + \frac{v}{1+\alpha}.$$

Moreover the sets of optimal actions for $[G_{s\alpha}(V^*)]$ are the same as those for $[r(s,.,.)]$. Hence, if there exists a map p which solves a problem 5.1.4 with $V=v.1_z$, then (a), (b) and (c) hold. On the other hand, if (a), (b) and (c) hold, then every map p is a solution to problem 5.1.4 as can be easily verified.

□

In the next theorem we give a sufficient condition for problem 5.1.4 to have a solution. This condition is rather strong. Note that under this condition we need no additional conditions on possible relations between the sets $\mathcal{O}_\ell(s)$ and the sets of optimal actions in the matrix games $[r(s,.,.)]$. This result suggests that such a relationship becomes effective in extreme cases as in theorem 5.1.9.

5.1.10. THEOREM. *If for problem* 5.1.4 *for each* $s\in S$ *the pair* $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ *has the* (m_s,n_s)-BKS *property, and if for each* $s\in S$ *and* $(i,j)\in A_s\times B_s$

$$r(s,i,j)+(1+\alpha)^{-1}.\underline{m} < V_s < r(s,i,j)+(1+\alpha)^{-1}.\bar{m},$$

then there exists a map p *which solves problem* 5.1.4.

PROOF. First, let the m_s,n_s- matrix game $[K_s]=[k_s(i,j)]$, for each $s\in S$ have the properties:
(a) $Val(K)=V_s$; (b) $\mathcal{O}_1(s)$ and $\mathcal{O}_2(s)$ are the sets of optimal actions for player 1 and player 2 respectively, and (c) for each $s\in S$ and $(i,j)\in A_s\times B_s$:

$$(5.1.3)\qquad r(s,i,j)+(1+\alpha)^{-1}.\underline{m} \leq k_s(i,j) \leq r(s,i,j)+(1+\alpha)^{-1}.\bar{m}.$$

(This is possible because the pair $(\mathcal{O}_1(s),\mathcal{O}_2(s))$ has the (m_s,n_s)-BKS property and in view of corollary A.1.12).
Let $\underline{s}$, $\bar{s}\in S$, such that $V_{\underline{s}}=\underline{m}$ and $V_{\bar{s}}=\bar{m}$. Then, by (5.1.3) there exists an $\varepsilon_{sij}\in[0,1]$ for each s, i and j, such that

$$(1+\alpha)(k_s(i,j)-r(s,i,j)) = \varepsilon_{sij}\underline{m}+(1-\varepsilon_{sij})\bar{m} = \varepsilon_{sij}V_{\underline{s}}+(1-\varepsilon_{sij})V_{\bar{s}}.$$

Now define p as:

$$\begin{aligned} p(\underline{s}|s,i,j) &= \varepsilon_{sij} \\ p(\bar{s}|s,i,j) &= 1-\varepsilon_{sij} \\ p(t|s,i,j) &= 0, \quad \text{if } t\neq\underline{s},\bar{s}. \end{aligned}$$

Obviously p is a solution to problem 5.1.4.

□

In the remainder of this section we restrict our attention to problem 5.1.4 for the subclass of stochastic games for which in each state one of the players is a dummy. A player is called a dummy in a state if he has only one action available in this state. In A2 it is pointed out that Markov decision problems can be seen as games for which one of the players is a dummy at each state. Now note that if for a stochastic game with value V^* one of the players is a dummy in state s, then the matrix game $[G_{s\alpha}(V^*)]$ is merely a row-vector game (in case player 1 is a dummy) or a column-vector game (in case player 2 is a dummy). Obviously in such a case the extreme optimal actions of the sets $\mathcal{O}_1(s)$ and $\mathcal{O}_2(s)$ are pure actions. Let $PO_\ell(s)$ the set of pure optimal actions for player ℓ in the matrix game $[G_{s\alpha}(V^*)]$. If in state s player 1 is a dummy then $PO_1(s)=\{1\}$ and if player 2 is a dummy then $PO_2(s)=\{1\}$. Then we have $\mathcal{O}_\ell(s)=P(PO_\ell(s))$, $\ell=1,2$.

Now fix a stochastic game Γ for which in each state either player 1 or player 2 is a dummy. Let V^* be the value of Γ. Let $\underline{m}=\min_s V_s^*$ and $\bar{m}=\max_s V_s^*$.

Suppose that player 2 is a dummy at state s. Then we may deduce from the theorems 5.1.7 and 5.1.8 that the following properties of $PO_1(s)$ must hold:

$OPT_s(1)$. If $i \in PO_1(s)$, then $\underline{m} \leq (1+\alpha)(V_s - r(s,i,1)) \leq \bar{m}$

$NOPT_s(1)$. If $i \notin PO_1(s)$, then $\underline{m} < (1+\alpha)(V_s - r(s,i,1))$

Likewise for a state s in which player 1 is a dummy, we have:

$OPT_s(2)$. If $j \in PO_2(s)$, then $\underline{m} \leq (1+\alpha)(V_s - r(s,1,j)) \leq \bar{m}$

$NOPT_s(2)$. If $j \notin PO_2(s)$, then $(1+\alpha)(V_s - r(s,1,j)) < \bar{m}$

For this class of games we have the following theorem, which shows that the conditions of the theorems 5.1.7 and 5.1.8 are also sufficient for problem 5.1.4 to have a solution in this case.

5.1.11. THEOREM. *Let a problem* 5.1.4 *be such that for each* $s \in S$ *either* $m_s = 1$ *or* $n_s = 1$ *(one of the players is a dummy). Then this problem can be solved if and only if*

(a) $O_\ell(s)$ *is the convex hull of a non-empty set of pure actions of player* ℓ, $\ell = 1,2$. *Denote this set by* $PO_\ell(s)$. *If player* ℓ *is a dummy in state* s, *then* $PO_\ell(s) = \{1\}$.

(b) *For each state* $s \in S$, *in which player* ℓ *is not a dummy, the properties* $OPT_s(\ell)$ *and* $NOPT_s(\ell)$ *hold,* $\ell = 1,2$.

PROOF. The "only if" part of the theorem is already shown above. So it remains to prove the " if" part, i.e. to choose a suitable map $\hat{p}$. Let $\underline{s}$, $\bar{s} \in S$ be such that $V_{\underline{s}} = \underline{m}$ and $V_{\bar{s}} = \bar{m}$. Fix $s \in S$. Then from the properties $OPT_s(1)$ and $OPT_s(2)$ (one of them holds) we see that for each $(i,j) \in PO_1(s) \times PO_2(s)$ there exists an $\varepsilon_{sij} \in [0,1]$ such that

$$(1+\alpha)(V_s - r(s,i,j)) = \varepsilon_{sij} V_{\underline{s}} + (1-\varepsilon_{sij}) V_{\bar{s}}$$

For this (i,j) define p as:

$$p(\underline{s}|s,i,j) = \varepsilon_{sij}$$
$$p(\bar{s}|s,i,j) = 1-\varepsilon_{sij}$$
$$p(t|s,i,j) = 0, \quad \text{if } t \neq \underline{s},\bar{s}.$$

Now let $(i,j) \notin PO_1(s) \times PO_2(s)$ and suppose that player ℓ, $\ell \in \{1,2\}$, is a dummy in state s. Then take

$$p(\underline{s}|s,i,j) = 2-\ell$$
$$p(\bar{s}|s,i,j) = \ell-1$$
$$p(t|s,i,j) = 0, \quad \text{if } t \neq \underline{s},\bar{s}.$$

It is easy to show (by theorem 4.2.4) that the map p defined in the above way, has the desired properties.

□

5.2. CHARACTERIZING PROPERTIES OF THE VALUE FUNCTIONS.

In this section we extend the axiomatic characterization of the value function of two-person zerosum games in normal form, presented by Vilkas (1963) and Tijs (1981), to the value function of discounted two-person zerosum stochastic games. This characterization can be indicated by the terms objectivity, monotonicity and sufficiency for the both players. The results of this section are outlined in Tijs & Vrieze (1981).

We wish to characterize the function $f^*: SG(S,\alpha) \to \mathbb{R}^Z$, where $f^*(\Gamma)$ equals the value of the game Γ for each $\Gamma \in SG(S,\alpha)$.

5.2.1. DEFINITION. *For a stochastic game Γ, with value V^*, we call an action $\hat{i} \in A_s$ for a state $s \in S$ superfluous if there exists a $\hat{\rho}_s \in P(A_s)$ with $\hat{\rho}_s(\hat{i})=0$, such that for each $j \in B_s$:*

$$r(s,\hat{i},j) + (1+\alpha)^{-1} \cdot \sum_{t=1}^{z} p(t|s,\hat{i},j).V_t^* \leq$$
$$r(s,\hat{\rho}_s,j) + (1+\alpha)^{-1} \cdot \sum_{t=1}^{z} p(t|s,\hat{\rho}_s,j).V_t^*.$$

An action $j \in B_s$ is called superfluous in an analogous way.

We now state four properties that a function $f: SG(S,\alpha) \to \mathbb{R}^Z$ could have. It will appear that these properties are necessary and sufficient for a function f to be the value function. Furthermore we prove that these axioms are independent.

5.2.2. AXIOM A1. *(Objectivity). If* $\Gamma \in SG(S,\alpha)$ *is such that for state* $s \in S$ *we have* $m_s=n_s=1$ *and* $p(s|s,1,1)=1$, *then* $(f(\Gamma))_s=\alpha^{-1}.(1+\alpha).r(s,1,1)$.

5.2.3. AXIOM A2. *(Monotonicity). If for two games* $\Gamma', \Gamma'' \in SG(S,\alpha)$ *the payoff functions* r' *and* r" *satisfy* $r' \leq r''$, *while the other game parameters are identical, then* $f(\Gamma') \leq f(\Gamma'')$.

5.2.4. AXIOM A3.ℓ. *(Sufficiency for player* ℓ, $\ell=1,2$*). If* $\Gamma' \in SG(S,\alpha)$ *results from* $\Gamma \in SG(S,\alpha)$ *by deleting a superfluous action of player* ℓ, *then* $f(\Gamma')=f(\Gamma)$.

5.2.5. THEOREM. *The value function* f^* *obeys the axioms* A1, A2, A3.1 *and* A3.2.

PROOF. If for a state $s \in S$ of a game Γ we have $m_s=n_s=1$ and $p(s|s,1,1)=1$, then for each pair of strategies (μ,ν):

$$V_{s\mu\nu} = \sum_{\tau=0}^{\infty} (1+\alpha)^{-\tau}.r(s,1,1) = \alpha^{-1}.(1+\alpha).r(s,1,1).$$

Hence $f_s^*(\Gamma)=\alpha^{-1}(1+\alpha).r(s,1,1)$ and so axiom A1 holds.

If for Γ' and Γ'' the payoff functions r' and r" satisfy $r' \leq r''$, while the other parameters are the same, then obviously for each μ and ν we have $V'(\mu,\nu) \leq V''(\mu,\nu)$. Hence $f^*(\Gamma') = \sup_{\mu} \inf_{\nu} V'(\mu,\nu) \leq \sup_{\mu} \inf_{\nu} V''(\mu,\nu) = f^*(\Gamma'')$. So Axiom A2 holds.

Now let for a state $s \in S$ of a game $\Gamma \in SG(S,\alpha)$ an action $\hat{i} \in A_s$ be superfluous in view of action $\hat{\rho}_s$. Let Γ' be the game which results from Γ after deleting action $\hat{i}$.
Let ρ_s^* be optimal for player 1 in the matrix game $[G_{s\alpha}(V^*)]$ corresponding to Γ. So for all $j \in B_s$:

$$(5.2.1) \qquad r(s,\rho_s^*,j) + (1+\alpha)^{-1}. \sum_{t=1}^{Z} p(t|s,\rho_s^*,j).V_t^* \geq V_s^*$$

Define the action $\tilde{\rho}_s$ as:

$$\tilde{\rho}_s(i) := \rho_s^*(i)+\rho_s^*(\hat{i}).\hat{\rho}_s(i) \quad \text{if } i\neq\hat{i}$$

$$\tilde{\rho}_s(\hat{i}) := 0$$

Now from (5.2.1) and definition 5.2.1 we see that for all $j\in B_s$:

$$\begin{aligned}(5.2.2)\qquad & r(s,\tilde{\rho}_s,j) + (1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\tilde{\rho}_s,j).V_t^* = \\ & r(s,\rho_s^*,j) + (1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\rho_s^*,j).V_t^* - \\ & \rho_s^*(\hat{i})\{r(s,\hat{i},j)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\hat{i},j).V_t^*-r(s,\hat{\rho}_s,j) - \\ & (1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\hat{\rho}_s,j).V_t^*\} \\ & \geq r(s,\rho_s^*,j)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\rho_s^*,j).V_t \geq V_s^*.\end{aligned}$$

Since $Val(G_{s\alpha}(V^*))=V_s^*$ it follows that in (5.2.2) the equality sign holds for some $\hat{j}\in B_s$. This means that with respect to the game Γ' we have $Val(G'_{s\alpha}(V^*))=V_s^*$. Since in the other states nothing changes we also have $Val(G'_{t\alpha}(V^*))=V_t^*$, $t\in S$ and $t\neq s$. Hence V^* is the fixed point of the map U_α, applied on game Γ' (cf. (4.2.4) and theorem 4.2.1). So by theorem 4.2.4 the value of the stochastic game Γ' equals V^*. This shows that also axiom A3.1 holds. Analogously one can show that the value function satisfies axiom A3.2.

□

So we have seen that the axioms A1, A2, A3.1 and A3.2 are necessary properties of a function to be the value function. In the next theorem we show that they are also sufficient.

5.2.6. THEOREM. *A function* f: $SG(S,\alpha)\to\mathbb{R}^z$ *satisfies the axioms* A1, A2, A3.1, A3.2 *if and only if* f *equals the value function.*

PROOF. The "if" part is proved in theorem 5.2.5. Suppose that f obeys the axioms A1, A2, A3.1 and A3.2. The proof of the "only if" part proceeds in two steps.

(a) First let Γ be a game with a state s and actions $(\hat{i},\hat{j})\in A_s\times B_s$, such that

$$p(s|s,i,j) = 1 \quad \text{if } i=\hat{i} \text{ or } j=\hat{j} \tag{5.2.3}$$

and such that

$$\inf_{j\in B_s} r(s,\hat{i},j) = r(s,\hat{i},\hat{j}) = \sup_{i\in A_s} r(s,i,\hat{j}) \tag{5.2.4}$$

We wish to show that $(f(\Gamma))_s=v_s^*$ for such a game Γ. Obviously $v_s^*=\alpha^{-1}.(1+\alpha).r(s,\hat{i},\hat{j})$, since for the stationary strategies $\hat{\rho}$ and $\hat{\sigma}$ with $\hat{\rho}_s(\hat{i})=1$ and $\hat{\sigma}_s(\hat{j})=1$ we have:

$$\inf_{\nu\in ST_2} V_{s\hat{\rho}\nu} = V_{s\hat{\rho}\hat{\sigma}} = \alpha^{-1}(1+\alpha).r(s,\hat{i},\hat{j}) = \sup_{\mu\in ST_1} V_{s\mu\hat{\sigma}}$$

Take a large number M and consider games $\Gamma',\Gamma''\in SG(S,\alpha)$, which differ from Γ only in the reward functions r' and r'' in the following way

$$r'(t,i,j) = r''(t,i,j)=r(t,i,j) \quad \text{if } t=s \text{ and } i=\hat{i} \text{ or } j=\hat{j}.$$

$$r'(t,i,j) = r(t,i,j)-M,\ r''(t,i,j) = r(t,i,j)+M \text{ elsewhere.}$$

From the monotonicity property of f we derive:

$$f(\Gamma') \le f(\Gamma) \le f(\Gamma''). \tag{5.2.5}$$

Now observe that in the game Γ' each action $i\in A_s\setminus\{\hat{i}\}$ is superfluous in view of action $\hat{i}$, when M is taken large enough. Hence, by axiom A3.1, the actions $i\neq\hat{i}$ may successively be deleted without disturbing the f-value. Let $\tilde{\Gamma}$ be the so obtained game, then $f(\Gamma')=f(\tilde{\Gamma})$. But in $\tilde{\Gamma}$ each action $j\in B_s\setminus\{\hat{j}\}$ is superfluous in view of action $\hat{j}$ (cf. (5.2.4)). This means that in $\tilde{\Gamma}$ the actions $j\neq\hat{j}$ may successively be deleted without disturbing the f-value. Let $\hat{\Gamma}$ be the remaining game. Then $f(\hat{\Gamma})=f(\tilde{\Gamma})=f(\Gamma')$. Now for game $\hat{\Gamma}$ in state s for both players only one action is left, namely $\hat{i}$ and $\hat{j}$ respectively. Moreover $p(s|s,\hat{i},\hat{j})=1$, so by axiom A1 we get:

$$\alpha^{-1}.(1+\alpha).r(s,\hat{i},\hat{j}) = (f(\hat{\Gamma}))_s = (f(\Gamma'))_s.$$

Analogously it can be shown that $(f(\Gamma''))_s=\alpha^{-1}(1+\alpha).r(s,\hat{i},\hat{j})$. Then by (5.2.5) we have $\alpha^{-1}(1+\alpha).r(s,\hat{i},\hat{j})\leq(f(\Gamma))_s\leq\alpha^{-1}(1+\alpha)r(s,\hat{i},\hat{j})$, which shows that $(f(\Gamma))_s=V_s^*$ for such a state s.

(b) Now take an arbitrary game $\Gamma\in SG(S,\alpha)$. Let $s\in S$. Consider the game $\Gamma'(s)\in SG(S,\alpha)$, which is constructed from Γ by adding in state s an (m_s+1)-th action for player 1 and an (n_s+1)-th action for player 2. r' and p' equal r and p respectively on $A_s\times B_s$ and for the extra entries we define:

$$r'(s,i,j) = \alpha(1+\alpha)^{-1}.V_s^* \quad \text{if } i=m_s+1 \text{ or } j=n_s+1 \tag{5.2.6}$$

and

$$p'(s|s,i,j) = 1 \quad \text{if } i=m_s+1 \text{ or } j=n_s+1. \tag{5.2.7}$$

Obviously $V^*(\Gamma)=V^*(\Gamma'(s))$. Moreover $\Gamma'(s)$ is a game, with respect to state s, of the type treated in step (a) of this proof. Hence

$$(f(\Gamma'(s)))_s = \alpha^{-1}.(1+\alpha).r'(s,m_s+1,n_s+1) = V_s^*(\Gamma). \tag{5.2.8}$$

If we could show that concerning game $\Gamma'(s)$ the actions m_s+1 for player 1 and n_s+1 for player 2 in state s are superfluous, then by the axioms A3.1 and A3.2 this would result in $(f(\Gamma'(s)))_s=(f(\Gamma))_s$. Combined with (5.2.8) this would give $(f(\Gamma))_s=V_s^*(\Gamma)$, as was to be proved.
Let $\hat{\rho}_s$ be optimal for player 1 in the matrix game $[G_{s\alpha}(V^*)]$; so for each $j\in B_s$:

$$r(s,\hat{\rho}_s,j) + (1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,\hat{\rho}_s,j).V_t^* \geq V_s^*. \tag{5.2.9}$$

From the definition of r' (cf. (5.2.6)) we see that (5.2.9) also holds for the game Γ', not only for $j\in B_s$ but also for $j=n_s+1$. Then

$$r'(s,\hat{\rho}_s,j) + (1+\alpha)^{-1}.\sum_{t=1}^{z} p'(t|s,\hat{\rho}_s,j).V_t^* \geq V_s^* \quad , j\in B_s\cup\{n_s+1\}. \tag{5.2.10}$$

By (5.2.6) and (5.2.7) we have:

$$r'(s,m_s+1,j) + (1+\alpha)^{-1}.\sum_{t=1}^{z} p'(t|s,m_s+1,j).V_t^* = V_s^*, \; j\in B_s\cup\{n_s+1\}. \tag{5.2.11}$$

Combining (5.2.10) and (5.2.11) gives that action m_s+1 is superfluous in view of $\hat{\rho}_s$ in game $\Gamma'(s)$. Likewise one can show that for the game remaining after deletion of action m_s+1 in $\Gamma'(s)$, the action n_s+1 is superfluous. □

Now we show that the four axioms A1, A2, A3.1 and A3.2 are independent. We do this as follows. For each triplet of them we give a function $f: SG(S,\alpha)\to\mathbb{R}^Z$ which satisfies these three axioms but not the fourth.

(a) (Objectivity). Let $f_1: SG(S,\alpha)\to\mathbb{R}^Z$ be the function with $f_1(\Gamma):=0_Z$ for all $\Gamma\in SG(S,\alpha)$. Obviously f_1 satisfies the axioms A2, A3.1 and A3.2 but not A1.

(b) (Sufficiency for player 1). Let $f_2: SG(S,\alpha)\to\mathbb{R}^Z$ be the function defined as:

$$(f_2(\Gamma))_s := \min_{i,j} \{r(s,i,j)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,i,j).V_t^*(\Gamma)\}.$$

Then f_2 obeys the axioms A1, A2 and A3.2 but not A3.1.

(c) (Sufficiency for player 2). Let $f_3: SG(S,\alpha)\to\mathbb{R}^Z$ be the function defined as:

$$(f_3(\Gamma))_s := \max_{i,j} \{r(s,i,j)+(1+\alpha)^{-1}.\sum_{t=1}^{z} p(t|s,i,j).V_t^*(\Gamma)\}.$$

Then f_3 obeys the axioms A1, A2 and A3.1 but not A3.2.

(d) (Monotonicity). More work has to be done, to show that axiom A2 is independent of the other axioms. First we pay some attention to matrix games. For an m,n-matrix game [K] let $NS_1(K)$ be the set of pure actions for player 1 which are not superfluous. Let $NS_2(K)$ have the analogous meaning. Let $NO_1(K):=\{i\,|\,i\in NS_1(K)$ and $\rho(i)=0$ for each $\rho\in O_1(K)\}$. Similarly $NO_2(K)$ is defined.

Suppose that a superfluous action i of player 1 is deleted from [K], resulting in a matrix game [K']. Then it can be verified that

(5.2.12) $\quad NO_1(K') = NO_1(K)$ and $NO_2(K') \subset NO_2(K)$.

If a superfluous action of player 2 is deleted, resulting in [K"], then analogously:

(5.2.13) $NO_2(K") = NO_2(K)$ and $NO_1(K") \subset NO_1(K)$.

Example 5.2.7 below shows that it may happen that $NO_2(K') \neq NO_2(K)$. Let $ANO_\ell(K) \subset NO_\ell(K)$ be such that $k \in ANO_\ell(K)$ if and only if $k \in NO_\ell(\tilde{K})$ for each matrix game $[\tilde{K}]$, which can be obtained from [K] by deleting superfluous rows and superfluous columns in any possible order. Note that for a matrix game $[\tilde{K}]$ which results from [K] after deleting a superfluous row or column, we have $ANO_\ell(\tilde{K}) = ANO_\ell(K)$, $\ell=1,2$. In example 5.2.7 the sets $ANO_\ell(K)$ are empty. This is not the case in example 5.2.8.

5.2.7. EXAMPLE. Let

$$[K] = \begin{bmatrix} 5 & 0 & 1 \\ 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}.$$

Then $NO_1(K)=\{2\}$ and $NO_2(K)=\{1,2\}$. If we delete the superfluous action 1 of player 1, then we obtain $[K']=\begin{bmatrix} 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}$. Now $NO_1(K')=\{1'\}=\{2\}=NO_1(K)$ and $NO_2(K')=\{1'\}=\{1\}\neq NO_2(K)$.

Moreover for this example we have $ANO_1(K)=ANO_2(K)=\emptyset$. To see this, first delete the superfluous row 1. Then column 2 becomes superfluous and is deleted. Next the superfluous row 2 and the superfluous column 1 are successively deleted. By doing so both players have only left their third action, which yields the matrix game $[\tilde{K}]=[3]$. Obviously $ANO_1(\tilde{K})=ANO_2(\tilde{K})=\emptyset$.

5.2.8. EXAMPLE. Let

$$[K] = \begin{bmatrix} 6 & 0 & 1 \\ 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}.$$

Then $ANO_1(K)=ANO_2(K)=\{1,2\}$ and no row or column of [K] is superfluous.

Now we return to stochastic games. For a game Γ with value V^* we denote the (i,j)-th element of the matrix game $[G_{s\alpha}(V^*)]$ by g(s,i,j). Then define f_4: $SG(S,\alpha)\to\mathbb{R}^Z$ as:

$$(f_4(\Gamma))_s = V_s^* - \sum_{i\in ANO_1(G_{s\alpha}(V^*))} \cdot \sum_{j\in ANO_2(G_{s\alpha}(V^*))} g(s,i,j)$$

It can be checked that f_4 satisfies the axioms A1, A3.1 and A3.2. We show that f_4 does not satisfy axiom A2.
Suppose that Γ is such that $[G_{s\alpha}(V^*)]$ is equal to the matrix game [K] of example 5.2.8 (such a game exists). Then $V_s^*(\Gamma)=3$, hence $(f_4(\Gamma))_s=3-6-0-0-6=-9$. Let Γ' be the stochastic game, which only differs from Γ in the fact that $r'(s,1,1)=r(s,1,1)+1$. Then $r'\geq r$ and $V_s^*(\Gamma')=V_s^*(\Gamma)$. But $(f_4(\Gamma'))_s=3-7-0-0-6=-10<(f_4(\Gamma))_s$. So f_4 does not satisfy the monotonicity axiom A2.

Now we wish to consider another interesting property of the value function, called symmetry. To that purpose we introduce the transpose of a stochastic game, which is the stochastic game that we obtain by interchanging the names of the players.

5.2.9. DEFINITION. *Let* $\Gamma\in SG(S,\alpha)$ *correspond to the stochastic game situation* $< S, \{A_s|s\in S\}, \{B_s|s\in S\}, r, p >$. *The transpose* Γ^T *of* Γ *is the discounted stochastic game associated with the stochastic game situation* $< \hat{S}, \{\hat{A}_s|s\in\hat{S}\}, \{\hat{B}_s|s\in\hat{S}\}, \hat{r}, \hat{p} >$, *where*:

$\hat{S} := S$,
$\hat{A}_s := B_s$ *and* $\hat{B}_s := A_s$ *for each* $s\in S$.
$\hat{r}(s,i,j) := -r(s,j,i)$ *and* $\hat{p}(t|s,i,j) = p(t|s,j,i)$ *for each* $s,t\in S$ *and each* $(i,j)\in\hat{A}_s\times\hat{B}_s$.

We call a function $f: SG(S,\alpha)\to\mathbb{R}^Z$ symmetric if the following property holds.

5.2.10. AXIOM A4. (*Symmetry*). $f(\Gamma^T)=-f(\Gamma)$ *for all* $\Gamma\in SG(S,\alpha)$.

It is straightforward to verify that the value function has the symmetry property. Furthermore it is simple to show that axiom A3.2 can be derived from the axioms A3.1 and A4. So we have the following alternative characterization of the value function.

5.2.11. THEOREM. *A function* $f: SG(S,\alpha)\to\mathbb{R}^Z$ *equals the value function if and only if* f *satisfies the axioms* A1, A2, A3.1 *and* A4.

Evidently in theorem 5.2.11 axiom A3.1 may be replaced by axiom A3.2.

We now introduce the concept of weakly superfluous actions. It appears that if we use this concept in characterizing the value function, then we do not need the monotonicity axiom anymore.

5.2.12. DEFINITION. *For a stochastic game* $\Gamma\in SG(S,\alpha)$ *with value* v^* *we call for a state* $s\in S$ *an action* $\hat{i}\in A_s$ *for player* 1 *weakly superfluous, if for each* $\rho_s\in P(A_s)$ *there exists an action* $\hat{\rho}_s\in P(A_s)$ *with* $\hat{\rho}_s(\hat{i})=0$ *and such that*

$$\inf_{j\in B_s}\{r(s,\hat{\rho}_s,j)+(1+\alpha)^{-1}\cdot\sum_{t=1}^{z}p(t|s,\hat{\rho}_s,j)v_t^*\}\geq$$

$$\inf_{j\in B_s}\{r(s,\rho_s,j)+(1+\alpha)^{-1}\cdot\sum_{t=1}^{z}p(t|s,\rho_s,j)v_t^*\}.$$

A weakly superfluous action of player 2 *is defined similarly.*

It is obvious that a superfluous action is also weakly superfluous, but the converse does not necessarily hold. This implies that the next axiom with respect to a function $f: SG(S,\alpha)\to\mathbb{R}^Z$, is stronger than axiom A3.ℓ, ℓ=1,2.

5.2.13. AXIOM A3.$\ell(w)$. (*Weak sufficiency for player* ℓ, ℓ=1,2). *If* $\Gamma'\in SG(S,\alpha)$ *results from* $\Gamma\in SG(S,\alpha)$ *by deleting a weakly superfluous action, then* $f(\Gamma')=f(\Gamma)$.

5.2.14. THEOREM. *A function* $f: SG(S,\alpha)\to\mathbb{R}^Z$ *equals the value function if and only if* f *satisfies the axioms* A1, A3.1(w) *and* A3.2(w).

PROOF. The "only if" part can easily be verified. Concerning the "if" part, in the proof of theorem 5.2.6 the monotonicity axiom is only used in the first step. There the game Γ was compared with the two games Γ' and Γ'', which were monotonic with respect to Γ, i.e. $r'\leq r\leq r''$. Next from Γ' and Γ'' superfluous actions could be deleted. However here we do not need games Γ' and Γ''. Suppose for a state s of a game Γ we have a saddle point as

determined by (5.2.3) and (5.2.4). Then at once in the game Γ all actions $i \neq \hat{i}$ and all actions $j \neq \hat{j}$ are weakly superfluous in view of action $\hat{i}$ and action $\hat{j}$ respectively. So they can successively be deleted.
The remainder of the proof proceeds analogously to the proof of theorem 5.2.6.

□

In Tijs & Vrieze (1979) the value function for more general classes of dynamic games is characterized in an axiomatic way. They showed that concerning the evaluation function a monotonicity assumption is the only essential condition to be able to give such a characterization.

5.3. PERTURBATION THEORY FOR DISCOUNTED STOCHASTIC GAMES.

In this section we study the effect on values and optimal stationary strategies of perturbations of the game parameters (payoff function, transition probabilities and the interest rate). Most of the results are first derived for two-person zerosum games in normal form and then transplanted to discounted stochastic games.

In this section we enlarge our game model. The state space may be a countable set and the action spaces are assumed to be compact metric. This section and the next section 5.4 are the only places of this monograph, where we abandon the finite two-person zerosum stochastic game model.

It should be noticed, that the question studied here is not only of theoretical importance, but also of practical value, because favourable answers to this question will give greater confidence in the use of game models in applications. Roughly speaking, "favourable" means here that small changes in the game parameters induce only small changes in the values, while good strategies in the original game are not too bad in a slightly perturbed game.

The results of this section are special cases of more general statements in Tijs & Vrieze (1980). For papers in the same spirit, but in a different context, we refer to Krabs (1977), Schweitzer (1968), Tijs (1976) and Whitt (1975).

By CSG(S) we mean the class of two-person zerosum stochastic games with countable state space S, compact metric action spaces for the players, a uniform bounded measurable payoff function and a measurable transition map. In addition we assume that the players are not allowed to randomize between their pure actions. Furthermore the immediate rewards are discounted according to some interest rate $\alpha \in (0,\infty)$. The measurability of the maps $r(s,.,.)$ and $p(t|s,.,.)$ is taken with respect to the product σ-algebra of $\mathcal{A}_s$ and $\mathcal{B}_s$, where $\mathcal{A}_s$ and $\mathcal{B}_s$ are the σ-algebra's generated by the Borel sets of the action spaces A_s for player 1 and B_s of player 2 respectively. Note that for a game $\Gamma \in CSG(S)$ the uniform boundedness of r and the measurability conditions on r and p ensure that the total discounted expected payoff exists for each pair of strategies.

5.3.1. REMARK. Let $\Gamma \in CSG(S)$. Suppose we would allow the players to randomize at each stage between their pure actions. Let $\mathcal{P}(A_s)$ and $\mathcal{P}(B_s)$ be the set of probability measures on the measure spaces $\langle A_s, \mathcal{A}_s \rangle$ and $\langle B_s, \mathcal{B}_s \rangle$ respectively, for each $s \in S$. Endow $\mathcal{P}(A_s)$ and $\mathcal{P}(B_s)$ with the weak topology, then it is known (e.g. Parthasarathy (1967), theorem 6.4, page 45), that $\mathcal{P}(A_s)$ and $\mathcal{P}(B_s)$ are compact and can be metrized.
Furthermore, it can be seen that the extensions of the maps $r(s,.,.)$ and $p(t|s,.,.)$ to $\hat{r}(s,.,.)$ and $\hat{p}(t|s,.,.)$ on $\mathcal{P}(A_s) \times \mathcal{P}(B_s)$ are measurable maps, where

$$\hat{r}(s,\rho_s,\sigma_s) := \int_{A_s} \int_{B_s} r(s,a,b)\, d\rho(a)\, d\sigma(b)$$

and

$$\hat{p}(t|s,\rho_s,\sigma_s) := \int_{A_s} \int_{B_s} p(t|s,a,b)\, d\rho(a)\, d\sigma(b).$$

Hence, one sees that the mixed extension of Γ, i.e. the game $\hat{\Gamma}$ where $\hat{S}:=S$, $\hat{A}_s := \mathcal{P}(A_s)$, $\hat{B}_s := \mathcal{P}(B_s)$ and $\hat{r}$ and $\hat{p}$ are as defined above, is also a member of the class CSG(S) and even a member with a specific property, namely that for each fixed $\tilde{\rho}_s \in \mathcal{P}(A_s)$ the maps $\hat{r}(s,\tilde{\rho}_s,\sigma_s)$ and $\hat{p}(t|s,\tilde{\rho}_s,\sigma_s)$ are affine in the variable σ_s. The same holds for a fixed $\tilde{\sigma}_s \in \mathcal{P}(B_s)$.

This reasoning implies that the class of games $RCSG(S) := \{\Gamma \mid \Gamma \in CSG(S)$, while in addition randomized actions are allowed$\}$ is a subset of the class $CSG(S)$. So assertions which hold for each member of $CSG(S)$ or each member of a subset of $CSG(S)$ enclosing $RCSG(S)$, also hold for the class of games $RCSG(S)$.

Now consider for a game $\Gamma \in CSG(S)$ for each $s \in S$ and $v \in \mathbb{R}^{|S|}$ the game in normal form $\langle A_s, B_s, G_s^{\alpha}(v) \rangle$ (cf. definition A.1.1), where for each $(a,b) \in A_s \times B_s$, $G_s^{\alpha}(v)(a,b) := r(s,a,b) + \frac{1}{1+\alpha} \sum_{t \in S} p(t \mid s,a,b).v_t$.
If the value of $\langle A_s, B_s, G_s^{\alpha}(v) \rangle$ exists (see definition A.1.3), then this value is denoted by $pVal(G_s^{\alpha}(v))$.
Let $VCSG(S)$ be the subset of $CSG(S)$ for which for each $s \in S$ and $v \in \mathbb{R}^{|S|}$ $pVal(G_s^{\alpha}(v))$ exists. Note that $SG(S,\alpha) \subset VCSG(S)$ for each $\alpha \in (0,\infty)$.

Now define for a game $\Gamma \in VCSG(S)$ the map $PU : \mathbb{R}^{|S|} \to \mathbb{R}^{|S|}$ as

$$PU_s(v) := pVal(G_s^{\alpha}(v)),$$

for each $s \in S$ and each $v \in \mathbb{R}^{|S|}$.

5.3.2. THEOREM. *If $\Gamma \in VCSG(S)$, then the value of Γ exists and equals the unique solution of the following set of functional equations in the variable $v \in \mathbb{R}^{|S|}$:*

$$PU_s(v) := pVal(G_s^{\alpha}(v)) = v_s \quad , s \in S$$

or equivalently

$$PU(v) = v.$$

Let V^* be this solution.
If, for $\varepsilon \geq 0$, the stationary strategies ρ^{ε} and σ^{ε} are such that for each $s \in S$, $\rho_s^{\varepsilon} \in A_s$ and $\sigma_s^{\varepsilon} \in B_s$ respectively are ε-optimal actions for player 1 and player 2 respectively in the game $\langle A_s, B_s, G_s^{\alpha}(V^*) \rangle$, then ρ^{ε} and σ^{ε} are $(\frac{1+\alpha}{\alpha}).\varepsilon$-optimal in the stochastic game Γ.

PROOF. Concerning the first part of the theorem, quite similar to the proof of theorem 4.2.4 it can be shown for fixed $\Gamma \in VCSG(S)$, that the map PU is a contraction map and therefore has a unique fixed point, say V^*.

Let for player 1 and $\varepsilon>0$ the stationary strategy ρ^{ε} be such that for each state $s\in S$ $\rho_s^{\varepsilon}\in A_s$ is an ε-optimal action in the game $\langle A_s,B_s,G_s^{\alpha}(V^*)\rangle$. Then for each stationary strategy σ of player 2 we have (in vector notation)

$$(5.3.1)\qquad r(\rho^{\varepsilon},\sigma) + \frac{1}{1+\alpha} P(\rho^{\varepsilon},\sigma).V^* + \varepsilon.1_{|S|} \geq V^*.$$

By repeated application of this inequality, we get

$$(5.3.2)\qquad V(\rho^{\varepsilon},\sigma) = \sum_{\tau=0}^{\infty} (\frac{1}{1+\alpha})^{\tau} P^{\tau}(\rho^{\varepsilon},\sigma).r(\rho^{\varepsilon},\sigma) \geq V^* - (\frac{1+\alpha}{\alpha}).\varepsilon 1_{|S|}$$

As in part I, chapter 3, it can be shown that corollary 3.5 also holds for the class of games VCSG(S). Hence (5.3.2) gives:

$$(5.3.3)\qquad \inf_{\nu\in ST_2} V(\rho^{\varepsilon},\nu) = \inf_{\sigma\in PSST_2} V(\rho^{\varepsilon},\sigma) \geq V^* - (\frac{1+\alpha}{\alpha}).\varepsilon 1_{|S|}$$

Analogously for a σ^{ε} with $\sigma_s^{\varepsilon}\in B_s$ ε-optimal for player 2 in $\langle A_s,B_s,G_s^{\alpha}(V^*)\rangle$ for each $s\in S$, we have

$$(5.3.4)\qquad \sup_{\mu\in ST_1} V(\mu,\sigma^{\varepsilon}) = \sup_{\rho\in PSST_1} V(\rho,\sigma^{\varepsilon}) \leq V^* + (\frac{1+\alpha}{\alpha}).\varepsilon 1_{|S|}$$

Then, using theorem 2.3.5, (5.3.3) and (5.3.4) show that V^* is the value of the game. Further, for any $\varepsilon\geq 0$ it follows that ρ^{ε} and σ^{ε} are $\alpha^{-1}(1+\alpha).\varepsilon$-optimal.

□

From now on we fix a state space S and action spaces A_s and B_s, $s\in S$. Let $\overline{VCSG}$ be the subset of VCSG(S) corresponding to these S, A_s and B_s, $s\in S$. A game $\Gamma\in\overline{VCSG}$ is completely characterized by a triplet $\langle r,p,\alpha\rangle$ and sometimes a game will be denoted by such a triplet.

We provide $\overline{VCSG}$ with the metric d defined by

$$d(\Gamma,\Gamma') := d(\langle r,p,\alpha\rangle,\langle r',p',\alpha'\rangle) := \max\{\|r-r'\|,\|p-p'\|,|\alpha-\alpha'|\},$$

where $\|p-p'\| := \sup_{s,a,b} \sum_{t\in S} |p(t|s,a,b)-p'(t|s,a,b)|$

and $\|r-r'\| = \sup_{s,a,b} |r(s,a,b)-r'(s,a,b)|$.

5.3.3. THEOREM. *The map* $V^*:\overline{VCSG}\to\mathbb{R}^{|S|}$, *where* $V^*(\Gamma)$ *is the value of the game* Γ, *is a continuous map (even pointwise Lipschitz continuous).*

PROOF. Let $M:=\sup_{s,a,b}|r(s,a,b)|$. Then

$$|V_s^*(\Gamma)| \le \frac{\alpha+1}{\alpha}.M \tag{5.3.5}$$

Let $\langle r,p,\alpha\rangle$ and $\langle r',p',\alpha'\rangle\in\overline{VCSG}$ and $\hat{M}=\max\{M,M'\}$.
First note that in view of lemma A.1.8 and theorem 4.2.4 we have

$$|V_s^*(\Gamma)-V_s^*(\Gamma')| = |pVal(G_s^{\alpha}(V^*(\Gamma)))-pVal(G_s^{\alpha'}(V^*(\Gamma')))| \le \tag{5.3.6}$$

$$\sup_{s,a,b}\Big|r(s,a,b)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,a,b)V_t^*(\Gamma)-r'(s,a,b)-\frac{1}{1+\alpha}\sum_{t\in S}p'(t|s,a,b)V_t^*(\Gamma')\Big|.$$

We use the following inequalities, which can easily be verified ($x\in\mathbb{R}$, $v,v'\in\mathbb{R}^{|S|}$).

$$\Big|\frac{x}{1+\alpha}-\frac{x'}{1+\alpha'}\Big| \le \frac{|x-x'|}{1+\alpha}+\frac{|\alpha-\alpha'|}{(1+\alpha)(1+\alpha')}|x| \le \frac{|x-x'|}{1+\alpha}+\frac{|\alpha-\alpha'|}{1+\alpha}|x| \tag{5.3.7}$$

$$\Big|\sum_{t\in S}p(t|s,a,b)v_t-\sum_{t\in S}p'(t\,s,a,b)v_t'\Big| \le \|v-v'\|+\|p-p'\|\,.\,\|v\| \tag{5.3.8}$$

Using (5.3.5), (5.3.7) and (5.3.8) we derive from (5.3.6):

$$|V_s^*(\Gamma)-V_s^*(\Gamma')| \le \|r-r'\|+\frac{\|V^*(\Gamma)-V^*(\Gamma')\|}{1+\alpha}+\frac{\|p-p'\|.M}{\alpha}+\frac{|\alpha-\alpha'|.M}{\alpha} \tag{5.3.9}$$

Rewriting (5.3.9) yields

$$\|V^*(\Gamma)-V^*(\Gamma')\| \le \frac{1+\alpha}{\alpha}\Big(\|r-r'\|+\|p-p'\|\frac{M}{\alpha}+|\alpha-\alpha'|\frac{M}{\alpha}\Big) \tag{5.3.10}$$

From (5.3.10) it follows that the map $V^*(.)$ is continuous. Putting $C_{r,\alpha}=\alpha^{-1}(1+\alpha)(1+2M.\alpha^{-1})$ leads to

$$\|V^*(\Gamma)-V^*(\Gamma')\| \le C_{r,\alpha}d(\Gamma,\Gamma'), \tag{5.3.11}$$

which shows the pointwise Lipschitz continuity of $V^*(.)$.

□

For a game $\Gamma=\langle r,p,\alpha\rangle\in\overline{VCSG}$ let $\mathcal{O}^{\varepsilon}_{\ell s}(\Gamma):=\mathcal{O}^{\varepsilon}_{\ell s}(r,p,\alpha)$ for $\varepsilon\geq 0$ be the set of ε-optimal actions for player ℓ, $\ell\in\{1,2\}$ in the dummy game $\langle A_s,B_s,G^{\alpha}_s(V^*(\Gamma))\rangle$. Then for $\varepsilon>0$, by theorem 5.3.2, the set $\underset{s\in S}{X}\mathcal{O}^{\varepsilon}_{\ell s}(r,p,\alpha)$ is a subset of the set $\alpha^{-1}(\alpha+1).\varepsilon$-optimal stationary strategies for player ℓ. For $\varepsilon=0$, the set $\underset{s\in S}{X}\mathcal{O}_{\ell s}(r,p,\alpha)$ can be identified with the set of optimal stationary strategies (the "if" part is proved in theorem 5.3.2, the "only if" part can be proved along the same lines as the proof of theorem 5.1.1).

5.3.4. THEOREM. *Let* $\varepsilon\geq 0$ *and* $\Gamma,\Gamma'\in\overline{VCSG}$ *such that*

$$d(\langle r,p,\alpha\rangle,\langle r',p',\alpha'\rangle)\leq\delta.$$

Then for each $s\in S$ *we have for* $\ell\in\{1,2\}$:

$$\mathcal{O}^{\varepsilon}_{\ell s}(r,p,\alpha)\subset\mathcal{O}^{\varepsilon+2C_{r\alpha}.\delta}_{\ell s}(r',q',\alpha'),$$

with $C_{r\alpha}$ *as defined in* (5.3.11).

The proof of this theorem is postponed until after the proof of the next lemma.

5.3.5. LEMMA. *Let* $\langle A,B,K\rangle$ *and* $\langle A,B,K'\rangle$ *be two games in normal form, which both have a value.*
If $d(K,K')=\sup^{a,b}|K(a,b)-K'(a,b)|\leq\delta$, *then* $\mathcal{O}^{\varepsilon}_{\ell}(K)\subset\mathcal{O}^{\varepsilon+2\delta}_{\ell}(K')$, *for each* $\varepsilon\geq 0$ *and* $\ell\in\{1,2\}$.

PROOF. We only show the inclusion for $\ell=1$.
Let $\tilde{a}\in\mathcal{O}^{\varepsilon}_1(K)$. The following three inequalities hold:

$$K'(\tilde{a},b)\geq K(\tilde{a},b)-\delta \qquad \text{for all } b\in B$$
$$K(\tilde{a},b)\geq \mathrm{pVal}(K)-\varepsilon \qquad \text{for all } b\in B$$
$$\mathrm{pVal}(K)\geq \mathrm{pVal}(K')-\delta$$

Combining these three inqualities yields:

$$K'(\tilde{a},b)\geq \mathrm{pVal}(K')-\varepsilon-2\delta, \qquad \text{for all } b\in B.$$

Hence $\tilde{a}\in\mathcal{O}^{\varepsilon+2\delta}_1(K')$.

□

PROOF of theorem 5.3.4.
$\mathcal{O}^{\varepsilon}_{\ell s}(r,p,\alpha)$ is derived from the game in normal form $\langle A_s,B_s,G^{\alpha}_s(V^*(\Gamma))\rangle$. If $d(\langle r,p,\alpha\rangle,\langle r',p',\alpha'\rangle)\leq\delta$, then from theorem 5.3.3 we get

$$d(G^{\alpha}_s(V^*(\Gamma)),\ G^{\alpha'}_s(V^*(\Gamma'))) \leq C_{r\alpha}.\delta.$$

So inserting this result in lemma 5.3.5 yields the theorem. □

5.3.6. DEFINITION. *We call a function* $K: A\times B\to\mathbb{R}$ *semi-continuous if for each* $\hat{b}\in B$ *the function* $K_{\hat{b}}: A\to\mathbb{R}$ *with* $K_{\hat{b}}(a):=K(a,\hat{b})$ *is upper semi-continuous and if for each* $\hat{a}\in A$ *the function* $K_{\hat{a}}: B\to\mathbb{R}$ *with* $K_{\hat{a}}(b):=K(\hat{a},b)$ *is lower semi-continuous. In that case the corresponding game in normal form* $\langle A,B,K\rangle$ *is called semi-continuous.*
A stochastic game $\Gamma\in VCSG(S)$ *is called semi-continuous if for each* $s\in S$ *the function* $r(s,.,.)$ *is semi-continuous on* $A_s\times B_s$ *and if for each uniformly bounded vector* $v\in\mathbb{R}^{|S|}$ *the function* $\sum_{t\in S} p(t|s,.,.)v_t$ *is semi-continuous on* $A_s\times B_s$.

SVCSG(S) denotes the subset of semi-continuous stochastic games of VCSG(S). Note that $SG(S,\alpha)\subset SVCSG(S)$ for each finite set S and each $\alpha\in(0,\infty)$.

Now if $\langle A,B,K\rangle$ is a semi-continuous game, for which the value exists, then we have that $\mathcal{O}^{\varepsilon}_1(K)$ is a closed subset of A. Namely $\mathcal{O}^{\varepsilon}_1(K):=f^{-1}([pVal(K)-\varepsilon,\infty))$, where f is the upper semi-continuous function on A defined by $f(a):=\inf_b K(a,b)$. Moreover, as A is compact, we have $\mathcal{O}_1(K)\neq\emptyset$ since $\mathcal{O}_1(K)=\cap_{\varepsilon>0}\mathcal{O}^{\varepsilon}_1(K)$. Analogously $\mathcal{O}^{\varepsilon}_2(K)$ is closed and $\mathcal{O}_2(K)\neq\emptyset$ if B is compact.

For a stochastic game $\Gamma\in SVCSG(S)$ we clearly have that the dummy games $\langle A_s,B_s,G^{\alpha}_s(V^*(\Gamma))\rangle$ are semi-continuous, so the above reasoning applies.

We now devote some attention to multifunctions. Let X and Y be Hausdorff spaces and let Y be compact. Let $f: X\to Y$ be a multifunction, assigning to each $x\in X$ a non-empty compact subset f(x) of Y. Following Berge ((1959), pp. 114,115), we will call such a multifunction upper semi-continuous if for each open set $\tilde{Y}\subset Y$, the set $\{x|x\in X \text{ and } f(x)\subset\tilde{Y}\}$ is an open subset of X.
If X and Y are metric spaces, then it is well-known that the multifunction $f: X\to Y$ is upper semi-continuous if and only if for each $(x,y)\in X\times Y$ and each

sequence (x_k,y_k), k=0,1,2,... converging to (x,y), while $y_k \in f(x_k)$, we have $y \in f(x)$. Especially this last characterization of upper semi-continuity of multifunctions indicates the usefulness of this concept for the problems in which we are involved. Namely X represents a family of games, Y is the set of actions (strategies) or action pairs (strategy pairs) and f assigns to each $x \in X$ the compact metric set of ε-optimal actions (strategies) of the game x.

Now we are ready to state the final theorem of this section. Again fix S, A_s and B_s, $s \in S$ and let $\overline{\text{SVCSG}}$ be the corresponding subset of SVCSG(S).

5.3.7. THEOREM. *For the class* $\overline{\text{SVCSG}}$ *the multifunctions*

$$O^{\varepsilon}_{1s}(r,p,\alpha): \overline{\text{SVCSG}} \to A_s$$

and

$$O^{\varepsilon}_{2s}(r,p,\alpha): \overline{\text{SVCSG}} \to B_s$$

are upper semi-continuous for all $s \in S$ *and all* $\varepsilon \geq 0$.

PROOF. We prove the theorem for $O^{\varepsilon}_{1s}(r,p,\alpha)$.
Let $\langle r_k,p_k,\alpha_k\rangle$, k=0,1,2,... be a sequence in $\overline{\text{SVCSG}}$ converging to $\langle r,p,\alpha\rangle \in \overline{\text{SVCSG}}$. Let $a_k \in O^{\varepsilon}_{1s}(r_k,p_k,\alpha_k)$ and suppose $\lim_{k\to\infty} a_k = a$. We need to prove that $a \in O^{\varepsilon}_{1s}(r,p,\alpha)$. Let $\delta_k = d(\langle r,p,\alpha\rangle, \langle r_k,p_k,\alpha_k\rangle)$ for each $k \in \mathbb{N}$.
Then by theorem 5.3.4 we have $a_k \in O^{\varepsilon+2C_{r\alpha}\cdot\delta_k}_{1s}(\Gamma)$.
Let k(h), with $h \in \mathbb{N}$, be such that $2C_{r\alpha}\cdot\delta_k \leq h^{-1}$ for each $k \geq k(h)$. This implies that

$$a_k \in O^{\varepsilon+h^{-1}}_{1s}(\Gamma) \quad \text{for all } k \geq k(h),$$

since $O^{\tilde{\varepsilon}}_{1s}(\Gamma) \subset O^{\hat{\varepsilon}}_{1s}(\Gamma)$ if $0 \leq \tilde{\varepsilon} \leq \hat{\varepsilon}$.

But then $a \in O^{\varepsilon+h^{-1}}_{1s}(\Gamma)$ as $\lim_{k\to\infty} a_k = a$ and $O^{\varepsilon+h^{-1}}_{1s}(\Gamma)$ is closed. So also $a \in O^{\varepsilon}_{1s}(\Gamma) = \bigcap_{h\in\mathbb{N}} O^{\varepsilon+h^{-1}}_{1s}(\Gamma)$.

□

In the next section the subclass of SVCSG(S) where both players have a unique optimal stationary strategy is considered.

We will end this section with an example, which shows that $\mathcal{O}^{\varepsilon}_{\ell s}$, $\varepsilon \geq 0$, is not necessarily a lower semi-continuous multifunction.

5.3.8. EXAMPLE. Consider the mixed extension of the row game $[K_\varepsilon]$, for given $\varepsilon \geq 0$, where $A=\{1\}$ and $B=[0,1]$ (an interval) and $K_\varepsilon(a,b)=1+\varepsilon b$ for $b \in [0,1]$ and $a=1$. So in fact player 1 is a dummy player. Then for $\varepsilon>0$ we have $\mathcal{O}_2(K_\varepsilon)=\{0\}$ and for $\varepsilon=0$ we have $\mathcal{O}_2(K_0)=[0,1]$. So for the sequence $[K_{k-1}]$, $k=1,2,\ldots$ converging to $[K_0]$ there exists an element $b \in \mathcal{O}_2(K_0)$ such that, it is not possible to construct a sequence b_k with $b_k \in \mathcal{O}_2(K_{k-1})$, such that $\lim_{k\to\infty} b_k = b$. This is equivalent to the fact that $\mathcal{O}_2(.)$ does not have the lower semi-continuity property in K_0.

5.4. UNIQUE OPTIMAL STRATEGIES.

This section is a continuation of the investigations of section 5.3. The notations introduced there are used again.

Most of the results of this section can be found in Tijs & Vrieze (1980). Let USVCSG(S) be the subset of SVCSG(S) for which both players have a unique optimal stationary strategy.

We remark that if a player has for some discounted stochastic game a unique optimal stationary strategy, then this strategy is his only optimal strategy, stationary or not. This observation does not hold for average reward games.

For a pair of action sets (A,B) we associate with each $(\hat{a},\hat{b}) \in A \times B$ a map $K_{\hat{a}\hat{b}}: A \times B \to \mathbb{R}$, such that the game $\langle A,B,K_{\hat{a}\hat{b}}\rangle$ has unique optimal actions, which are respectively $\hat{a}$ and $\hat{b}$.

Take $K_{\hat{a}\hat{b}}(a,b) := -1$ if $b=\hat{b}$ and $a \neq \hat{a}$

$:= 1$ if $a=\hat{a}$ and $b \neq \hat{b}$

$:= 0$ elsewhere.

Clearly $pVal(K_{\hat{a}\hat{b}})=0$ and $\hat{a}$ and $\hat{b}$ are the unique optimal actions.

5.4.1. REMARK. As the set of optimal stationary strategies of a game $\Gamma \in SVCSG(S)$ for player ℓ equals $\underset{s\in S}{X}\, \mathcal{O}_\ell(G^{\alpha}_s(V^*(\Gamma)))$, we see that Γ belongs to USVCSG(S) if and only if each game in normal form $\langle A_s,B_s,G^{\alpha}_s(V^*(\Gamma))\rangle$, has a pair of unique optimal actions.

5.4.2. THEOREM. *Let* $\Gamma\in$SVCSG(S) *and let* $(\hat\rho,\hat\sigma)\in\mathcal{O}_1(\Gamma)\times\mathcal{O}_2(\Gamma)$, *with* $\hat\rho=(\hat a_1,\hat a_2,\ldots)$ *and* $\hat\sigma=(\hat b_1,\hat b_2,\ldots)$. *Then the game* $\hat\Gamma$, *which results from* Γ *by replacing the payoff function* r *by* $\hat r(s,.,.):=r(s,.,.)+\varepsilon.K_{\hat a_s\hat b_s}(.,.)$, *for given* $\varepsilon>0$, *belongs to* USVCSG(S). *Moreover* $\hat\rho$ *and* $\hat\sigma$ *are the unique optimal stationary strategies for the respective players.*

PROOF. For the game $\langle A_s,B_s,\hat G_s^{\alpha}(V^*(\Gamma))\rangle$, $s\in S$ we have for $(a,b)\in A_s\times B_s$, in view of the optimality of $\hat a_s$ and $\hat b_s$:

$$(5.4.1)\qquad r(s,a,\hat b_s)+\varepsilon K_{\hat a_s\hat b_s}(a,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,a,\hat b_s).V_t^*(\Gamma)\le$$

$$r(s,\hat a_s,\hat b_s)+\varepsilon K_{\hat a_s\hat b_s}(\hat a_s,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\hat a_s,\hat b_s).V_t^*(\Gamma)\le$$

$$r(s,\hat a_s,b)+\varepsilon K_{\hat a_s\hat b_s}(\hat a_s,b)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\hat a_s,b).V_t^*(\Gamma).$$

Hence $(\hat a_s,\hat b_s)$ is a saddle point for the game $\langle A_s,B_s,\hat G_s(V^*(\Gamma))\rangle$. So by theorem A.1.4 the value of this game in normal form equals:

$$r(s,\hat a_s,\hat b_s)+\varepsilon K_{\hat a_s\hat b_s}(\hat a_s,\hat b_s)+(1+\alpha)^{-1}\sum_{t\in S}p(t|s,\hat a_s,\hat b_s)V_t^*(\Gamma)=V_s^*(\Gamma).$$

But this means that for the game $\hat\Gamma$, $V^*(\Gamma)$ is the unique solution of the equation v=PU(v), which by theorem 5.3.2 shows, that $V^*(\Gamma)$ equals the value of $\hat\Gamma$. From the inequalities (5.4.1) we now deduce that $\hat\rho$ and $\hat\sigma$ are optimal in the game $\hat\Gamma$, since $\hat a_s\in\mathcal{O}_{1s}(\hat\Gamma)$ and $\hat b_s\in\mathcal{O}_{2s}(\hat\Gamma)$ for each $s\in S$.
Suppose $\tilde\rho\in\mathcal{O}_1(\hat\Gamma)$ and $\tilde\rho\neq\hat\rho$, say $\tilde a_s\neq\hat a_s$. Then (using theorem 5.3.2)

$$(5.4.2)\qquad V_s^*(\Gamma)=\varepsilon K_{\hat a_s\hat b_s}(\hat a_s,\hat b_s)+r(s,\hat a_s,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\hat a_s,\hat b_s)V_t^*(\Gamma)=$$

$$\varepsilon K_{\hat a_s\hat b_s}(\tilde a_s,\hat b_s)+r(s,\tilde a_s,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\tilde a_s,\hat b_s).V_t^*(\Gamma).$$

On the other hand by definition of $K_{\hat a\hat b}$ and the optimality of $\hat\sigma$ in Γ:

$$\varepsilon K_{\hat a_s\hat b_s}(\tilde a_s,\hat b_s)+r(s,\tilde a_s,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\tilde a_s,\hat b_s).V_t^*(\Gamma)<$$

$$r(s,\tilde a_s,\hat b_s)+\frac{1}{1+\alpha}\sum_{t\in S}p(t|s,\tilde a_s,\hat b_s).V_t^*(\Gamma)\le V_s^*(\Gamma),$$

which in combination with (5.4.2) leads to a contradition. So $\hat\Gamma\in$USVCSG(S).

□

5.4.3. REMARK. If in theorem 5.4.2 we replace $(\hat{\rho},\hat{\sigma})\in\mathcal{O}_1(\Gamma)\times\mathcal{O}_2(\Gamma)$ by $(\hat{\mathcal{O}}_1(\Gamma)\times\hat{\mathcal{O}}_2(\Gamma))\subset\mathcal{O}_1(\Gamma)\times\mathcal{O}_2(\Gamma)$, and if $K_{\hat{\mathcal{O}}_{1s}\hat{\mathcal{O}}_{2s}}$ on $A_s\times B_s$ is defined as

$$\begin{aligned} K_{\hat{\mathcal{O}}_{1s}\hat{\mathcal{O}}_{2s}}(a,b) &= -1 \quad \text{if } b\in\hat{\mathcal{O}}_{2s} \text{ and } a\notin\hat{\mathcal{O}}_{1s} \\ &= 1 \quad \text{if } a\in\hat{\mathcal{O}}_{1s} \text{ and } b\notin\hat{\mathcal{O}}_{2s} \\ &= 0 \quad \text{elsewhere,} \end{aligned}$$

then a similar statement can be made.

5.4.4. REMARK. Concerning m,n-matrix games it is not possible to construct for each element of $P(\mathbb{N}_m)\times P(\mathbb{N}_n)$ a matrix game, such that this element corresponds to the unique optimal actions. This follows at once from the dimension theorem for matrix games (see theorem A.1.11). If $(\rho,\sigma)\in P(\mathbb{N}_m)\times P(\mathbb{N}_n)$ is the unique saddle point for a matrix game, then ρ and σ must have the same number of components unequal to zero. This means, that for action spaces A and B, where $A=P(\mathbb{N}_m)$ and $B=P(\mathbb{N}_n)$, $m,n\in\mathbb{N}$, it is in general not possible to construct for each $\hat{a}\in A$ and $\hat{b}\in B$ a function $K_{\hat{a}\hat{b}}$, where $\hat{a}$ and $\hat{b}$ are the unique optimal actions for the game $\langle A,B,K_{\hat{a}\hat{b}}\rangle$ and where additionally $K_{\hat{a}\hat{b}}$ has the affine property mentioned in remark 5.3.1.
If and only if $\hat{a}$ and $\hat{b}$, seen as elements of $P(\mathbb{N}_m)$ and $P(\mathbb{N}_n)$, have the same number of components unequal to zero, such a matrix game $[K_{\hat{a}\hat{b}}]$ exists. A procedure to carry out this construction can be found in Karlin (1959, p. 70).

For fixed S, A_s and B_s, $s\in S$ let $\overline{\text{USVCSG}}$ be the subset of $\overline{\text{SVCSG}}$ for which both players have a unique optimal stationary strategy.

The next theorem is an immediate consequence of theorem 5.3.7 and the proof of it will be omitted.

5.4.5. THEOREM. *Let* $\Gamma_k\in\overline{\text{SVCSG}}$, k=1,2,..., *such that* $\lim_{k\to\infty}\Gamma_k=\Gamma$, *with* $\Gamma\in\overline{\text{USVCSG}}$. *If* $\rho^k\in\mathcal{O}_1(\Gamma_k)$ *and* $\sigma^k\in\mathcal{O}_2(\Gamma_k)$, *then* $\lim_{k\to\infty}\rho^k=\rho$ *and* $\lim_{k\to\infty}\sigma^k=\sigma$, *where* ρ *and* σ *respectively are the unique optimal stationary strategies of player* 1 *and player* 2 *in* Γ.

5.4.6. THEOREM (a). *The restrictions of the maps* $\mathcal{O}_{1s}\colon \overline{\text{SVCSG}}\to A_s$ *and* $\mathcal{O}_{2s}\colon \overline{\text{SVCSG}}\to B_s$ *to the subset* $\overline{\text{USVCSG}}$ *are continuous maps for each* $s\in S$.
(b) $\overline{\text{USVCSG}}$ *is a dense subset of* $\overline{\text{SVCSG}}$.

PROOF. (a) follows from the fact that a single-valued map, which is upper semi-continuous in the multi-valued sense, is continuous.
(b) Let $\Gamma\in\overline{SVCSG}$ and $\varepsilon>0$. We have to prove that there exists a $\hat{\Gamma}\in\overline{USVCSG}$ such that $d(\Gamma,\hat{\Gamma})<\varepsilon$. Take $(\hat{\rho},\hat{\sigma})\in\mathcal{O}_1(\Gamma)\times\mathcal{O}_2(\Gamma)$. Then by theorem 5.4.2 we can construct for any $\delta>0$ a game $\hat{\Gamma}\in\overline{USVCSG}$ which only differs from Γ in the payoff function: $\hat{r}(s,.,.):=r(s,.,.)+\delta K_{\hat{\rho}_s\hat{\sigma}_s}(.,.)$. Obviously $d(\Gamma,\hat{\Gamma})<\varepsilon$ for $0<\delta<\varepsilon$, which proves the theorem. □

Also for the next theorem the functions $K_{\hat{\rho}_s\hat{\sigma}_s}$ are of valuable use.

5.4.7. THEOREM. *For fixed* S, A_s *and* B_s, $s\in S$ *the class of games* $\overline{USVCSG}$ *is connected if and only if* A_s *and* B_s *are connected for all* $s\in S$.

PROOF. (a) First suppose that, say A_s is not connected. Then let A_{s1} and A_{s2} be two disjunct non-empty open subsets of A_s with $A_s=A_{s1}\cup A_{s2}$.
Let $US_k:=\{\Gamma\,|\,\Gamma\in\overline{USVCSG};\ \mathcal{O}_{1s}(\Gamma)\subset A_{sk}\}$ for k=1,2. It is obvious that $\overline{USVCSG}=US_1\cup US_2$ and that $US_1\cap US_2=\emptyset$.
Further, US_1 and US_2 are open sets in $\overline{USVCSG}$, because $\mathcal{O}_1$ and $\mathcal{O}_2$ are upper semi-continuous multifunctions.
If we can show that $US_1\neq\emptyset$ and $US_2\neq\emptyset$, then we have proved that $\overline{USVCSG}$ is not connected when A_s is not connected. This will prove the necessity part of the theorem.
Take $\hat{a}_{s1}\in A_{s1}$ and $\hat{a}_{s2}\in A_{s2}$ and $\hat{b}_s\in B$. Let Γ_k, k=1,2 be a stochastic game, which concerning state s is defined as: $r_k(s,.,.):=K_{\hat{a}_{sk}\hat{b}_s}(.,.)$ and $p_k(t\,|\,s,a,b)=1$ if t=s and $p_k(t\,|\,s,a,b)=0$ if $t\neq s$. Define for the states unequal to s, Γ_k in such a way that $\Gamma_k\in\overline{USVCSG}$. Then clearly $\Gamma_k\in US_k$, k=1,2. So both US_1 and US_2 are non-empty.

(b) Now, we suppose that the sets A_s and B_s, $s\in S$ are connected sets. Suppose that US_1 and US_2 are disjunct open subsets of $\overline{USVCSG}$ such that $\overline{USVCSG}=US_1\cup US_2$.
If we can show that $US_1=\emptyset$ or $US_2=\emptyset$, then this would imply that $\overline{USVCSG}$ is connected. First associate with each pair $(\hat{\rho},\hat{\sigma})\in(\underset{s\in S}{X}A_s)\times(\underset{s\in S}{X}B_s)$, where $\hat{\rho}=(\hat{a}_1,\hat{a}_2,\ldots)$ and $\hat{\sigma}=(\hat{b}_1,\hat{b}_2,\ldots)$, the game $\hat{\Gamma}_{\hat{\rho}\hat{\sigma}}$ defined by $\hat{r}(s,a,b):=d(b,\hat{b}_s)-d(a,\hat{a}_s)$, all $(a,b)\in A_s\times B_s$ and $\hat{p}(t\,|\,s,a,b):=1$ for t=s and $p(t\,|\,s,a,b)=0$ if $t\neq s$ for all $(a,b)\in A_s\times B_s$ and all $s,t\in S$.

It can be checked that $\hat{\Gamma}_{\hat{\rho}\hat{\sigma}} \in \overline{USVCSG}$ with value $0.1^{|S|}$ and that $\mathcal{O}_1(\hat{\Gamma}_{\hat{\rho}\hat{\sigma}})=\{\hat{\rho}\}$ and $\mathcal{O}_2(\hat{\Gamma}_{\hat{\rho}\hat{\sigma}})=\{\hat{\sigma}\}$.
Let $F: (\underset{s\in S}{X} A_s)\times(\underset{s\in S}{X} B_s)\to\overline{USVCSG}$ be the map defined by $F(\rho,\sigma)=\hat{\Gamma}_{\rho\sigma}$. It is straightforward to show that

$$\|F(\tilde{\rho},\tilde{\sigma}) - F(\tilde{\tilde{\rho}},\tilde{\tilde{\sigma}})\| \leq d(\tilde{\rho},\tilde{\tilde{\rho}}) + d(\tilde{\sigma},\tilde{\tilde{\sigma}})$$

for all $(\tilde{\rho},\tilde{\sigma})$ and $(\tilde{\tilde{\rho}},\tilde{\tilde{\sigma}})$. Hence, F is a continuous map from the connected set $(X_{s\in S}A_s)\times(X_{s\in S}B_s)$ into $\overline{USVCSG}$. This implies that either $\hat{\Gamma}_{\rho\sigma}\in US_1$ for each ρ and σ or $\hat{\Gamma}_{\rho\sigma}\in US_2$ for each ρ and σ.
Without loss of generality, we may suppose that

$$\hat{\Gamma}_{\rho\sigma}\in US_1, \quad \text{all } (\rho,\sigma)\in(X_{s\in S}A_s)\times(X_{s\in S}B_s).$$

Now take an arbitrary $\Gamma\in\overline{USVCSG}$. Let $\mathcal{O}_1(\Gamma)\times\mathcal{O}_2(\Gamma)=\{(\hat{\rho},\hat{\sigma})\}$.
Define for each $\varepsilon\in[0,1]$ the game $\Gamma_\varepsilon := \varepsilon\Gamma+(1-\varepsilon)\hat{\Gamma}_{\hat{\rho}\hat{\sigma}}$ as $r(s,a,b) := \varepsilon.r(s,a,b)+(1-\varepsilon)(d(b,\hat{b}_s)-d(a,\hat{a}_s))$ and $p_\varepsilon(t|s,a,b) := \varepsilon p(t|s,a,b)+(1-\varepsilon)\delta_{ts}$, where $\delta_{ts}=1$ if $t=s$ and $\delta_{ts}=0$ for $t\neq s$.
Along the same lines as the proof of theorem 5.4.2 it now can be shown that
(i) $\Gamma_\varepsilon\in\overline{USVCSG}$, (ii) $p\,Val(\Gamma_\varepsilon)=\varepsilon.v^*(\Gamma)$ and (iii) $\mathcal{O}_1(\Gamma_\varepsilon)\times\mathcal{O}_2(\Gamma_\varepsilon)=\{(\hat{\rho},\hat{\sigma})\}$.
Let $FF: [0,1]\to\overline{USVCSG}$ be the map defined as $FF(\varepsilon) := \Gamma_\varepsilon$, then $FF(0)=\hat{\Gamma}_{\hat{\rho}\hat{\sigma}}$ and $FF(1)=\Gamma$. Furthermore

$$\|FF(\varepsilon_1)-FF(\varepsilon_2)\| \leq |\varepsilon_1-\varepsilon_2|(\|r\|+1+\sup_{s,a,b}|d(b,\hat{b}_s)-d(a,\hat{a}_s)|+1)$$

for each ε_1, $\varepsilon_2\in[0,1]$. Hence F is continuous. But since $[0,1]$ is connected and $F(0)\in US_1$, it follows that $F(1)\in US_1$. So $\Gamma\in US_1$. As $\Gamma\in\overline{USVCSG}$ was arbitrary, this proves $\overline{USVCSG}=US_1$. Hence $US_2=\emptyset$, which completes the proof of the theorem.

□

2.4.8. REMARK. The framework of the model, studied in this section is principally the stochastic game. However in the proofs of the theorems the dummy games stand central, i.e. two-person zerosum games in normal form. Clearly all the theorems of this section and those of the preceding one, also hold for the class of two-person zerosum games in normal form. So perhaps it would have been more logical if we had first

deduced the results of these two sections for this class of games and next had extended them to stochastic games. Such an approach is chosen in Tijs & Vrieze (1980). However, since this monograph is concerned primarily with stochastic games, we have chosen for the reverse approach.

5.4.9. REMARK. The set $\overline{\text{USVCSG}}$ is not necessarily an open subset of $\overline{\text{SVCSG}}$ as the following example shows. Take a stochastic game with one state, with $A_1 := [0,1]$ and $B_1 := \{1\}$. Then, for $\hat{\Gamma}$ defined by $r(1,a,b)=a$, $p(1|1,a,b)=1$, $(a,b)\in A_1 \times B_1$ and $\alpha \in (0,\infty)$ arbitrarily we have $\hat{\Gamma} \in \overline{\text{USVCSG}}$. For each $\varepsilon>0$, the ε-neighbourhood of $\hat{\Gamma}$ contains the game $\Gamma_{\frac{1}{2}\varepsilon}$ defined by $r_{\frac{1}{2}\varepsilon}(1,a,b) := \min\{a, 1-\frac{1}{2}\varepsilon\}$. Clearly $\Gamma_{\frac{1}{2}\varepsilon} \notin \overline{\text{USVCSG}}$, hence $\overline{\text{USVCSG}}$ is not open.

5.4.10. REMARK. Concerning matrix games, Bohnenblust, Karlin & Shapley (1950) proved, that the set U_{mn} of those m,n-matrix games ($m,n \in \mathbb{N}$), for which the mixed extension has a unique saddle point is an open and dense subset of the set M_{mn} of all m,n-matrix games. With some labour, one can prove that U_{mn} is not connected for all $(m,n) \neq (1,1)$. We will not do this here, but remark that for the case where $(m,n)=(1,2)$ we have

$$U_{12} = \{[r(1,1)\ r(1,2)] \,|\, r(1,1) \neq r(1,2)\}$$

$$= \{[r(1,1)\ r(1,2)] \,|\, r(1,1) > r(1,2)\} \cup$$

$$\{[r(1,1)\ r(1,2)] \,|\, r(1,1) < r(1,2)\}.$$

Note that this phenomenon is due to the fact mentioned in remark 5.4.4, namely the fact that not each point of the Cartesian product of the mixed action spaces can serve as a unique pair of optimal actions for some appropriate matrix game.

Now consider the class of games $SG(S,\alpha)$ (i.e. finite state and action spaces). Let $USG(S,\alpha)$ be the subset of stochastic games, for which both players have a unique optimal stationary strategy. Now let $\Gamma \in SG(S,\alpha)$, with value $V^*(\Gamma)$. Then for each $s \in S$ the m_s,n_s-matrix game $[G_{s\alpha}(V^*(\Gamma))]$ has value $V_s^*(\Gamma)$.

By the above mentioned result (remark 5.4.10) of Bohnenblust, Karlin & Shapley, it follows that for each $\varepsilon>0$, there exists a matrix game

$[K_\varepsilon] \in U_{m_s n_s}$ with value $v_s^*(\Gamma)$, such that $d(G_{s\alpha}(v^*(\Gamma)), K_\varepsilon) < \varepsilon$. Hence we may write $K_\varepsilon(i,j) = r_\varepsilon(s,i,j) + (1+\alpha)^{-1} \sum_{t=1}^{z} p(t|s,i,j) v_t^*(\Gamma)$, where $r_\varepsilon(s,i,j)$, $(i,j) \in A_s \times B_s$, can be chosen such that $\|r(s,.,.) - r_\varepsilon(s,.,.)\| < \varepsilon$.
But this means, that for the game $\hat{\Gamma}_\varepsilon$ which follows from Γ by replacing r by $\hat{r}(s,.,.) := r_\varepsilon(s,.,.)$, it holds that $v^*(\Gamma)$ is the solution of the set of equations $v_s = \mathrm{Val}(\hat{G}_{s\alpha}(v))$, $s \in S$. So $v^*(\hat{\Gamma}_\varepsilon) = v^*(\Gamma)$ and as $[\hat{G}_{s\alpha}(v^*(\hat{\Gamma}_\varepsilon))] \in U_{m_s n_s}$ we have $\hat{\Gamma}_\varepsilon \in \mathrm{USG}(S,\alpha)$. Moreover $d(\Gamma, \hat{\Gamma}_\varepsilon) < \varepsilon$, which shows that the result of Bohnenblust, Karlin & Shapley can be extended to stochastic games. This fact is stated in the next theorem.

5.4.11. THEOREM. *The set* $\mathrm{USG}(S,\alpha)$ *is an open and dense subset of* $\mathrm{SG}(S,\alpha)$.

6. Algorithms for discounted stochastic games.

In this chapter some algorithms for solving finite two-person zerosum stochastic games are considered. In section 6.1 we give a brief survey of existing algorithms in this field. In section 6.2 we treat an extension to discounted stochastic games of the Brown-Robinson procedure of matrix games. Finally in section 6.3 an algorithm for a subclass of stochastic games is presented, namely the class of games for which in each state only one of the players governs the transitions, although not necessarily in each state the same player.

6.1. SOME ALGORITHMS.

In the first place we mention the algorithm which in a natural way arises from Shapley's proof of the existence of the value of a discounted stochastic game (cf. Shapley (1953)). This algorithm can be seen as the standard method of successive approximation.

6.1.1. ALGORITHM (Shapley)

(*i*) *Choose* $v_0 \in \mathbb{R}^z$ *arbitrarily*, $\tau := 0$

(*ii*) $v_{\tau+1}(s) := Val(G_{s\alpha}(v_\tau))$

(*iii*) *Choose* $\rho^\tau = (\rho_1^\tau, \ldots, \rho_z^\tau)$ *and* $\sigma^\tau = (\sigma_1^\tau, \ldots, \sigma_z^\tau)$ *such that for each* $s \in S$:

$$v_{\tau+1}(s) = \min_{j \in B_s} \{r(s, \rho_s^\tau, j) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s, \rho_s^\tau, j) v_\tau(t)\}$$

and

$$v_{\tau+1}(s) = \max_{i \in A_s} \{r(s, i, \sigma_s^\tau) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s, i, \sigma_s^\tau) . v_\tau(t)\}.$$

Hence ρ_s^τ *and* σ_s^τ *are optimal in* $[G_{s\alpha}(v_\tau)]$.

(*iv*) $\tau := \tau+1$, *go to step* (*ii*).

From the contraction property of the value operator one can derive for this algorithm that

(a) $\lim_{\tau\to\infty} v_\tau = V^*(\Gamma)$

(b) $V(\mu,\sigma^\tau) \leq v_\tau + \frac{1}{\alpha}\bar{m}_\tau.1_z$ for all $\mu\in ST_1$ and all τ

(c) $V(\rho^\tau,\nu) \geq v_\tau + \frac{1}{\alpha}\underline{m}_\tau.1_z$ for all $\nu\in ST_2$ and all τ

(d) $v_\tau + \frac{\underline{m}_\tau}{\alpha}.1_z \leq V^*(\Gamma) \leq v_\tau + \frac{\bar{m}_\tau}{\alpha}.1_z$, for all τ,

where $\bar{m}_\tau := \max_{s\in S}\{v_\tau(s)-v_{\tau-1}(s)\}$ and $\underline{m}_\tau := \min_{s\in S}\{v_\tau(s)-v_{\tau-1}(s)\}$.

Hence v_τ approximates the value of the game and for each $\varepsilon>0$, ρ^τ and σ^τ are ε-optimal stationary strategies for τ large enough. The algorithm is stopped as soon as the bounds in (d) are tight enough.

The second algorithm we describe, is the discounted version of the algorithm of Hoffman & Karp (1966). Originally they presented their algorithm for average reward stochastic games, but their procedure can also be applied to discounted stochastic games.

6.1.2. ALGORITHM (Hoffman & Karp).

(*i*) *Choose* $v_0\in\mathbb{R}^z$ *arbitrarily,* $\tau:=0$.

(*ii*) *Determine for player* 2 *a stationary strategy* $\sigma^\tau\in SST_2$, *such that* σ^τ_s *is optimal in* $[G_{s\alpha}(v_\tau)]$ *for each* $s\in S$.

(*iii*) *Solve for player* 1 *the Markov decision problem which results when player* 2 *fixes* σ^τ *(see theorem* 3.4*).*
Let $v_{\tau+1}$ *be the optimal value of* MDP(σ^τ).

(*iv*) $\tau := \tau+1$, *go to step* (*ii*).

This algorithm is a generalization of Howard's policy iteration method (1960), combined however with successive approximation ideas. So it can be seen as value oriented policy iteration. The way in which this algorithm

approximates the value of the game and produces ε-optimal stationary strategies is given by the following properties (cf. Van der Wal (1977)).

(a) $\lim_{\tau\to\infty} v_\tau = V^*(\Gamma)$

(b) $\mathrm{Val}(G_{s\alpha}(v_\tau)) + \frac{\underline{m}_\tau}{\alpha}.1_Z \le V_s^*(\Gamma) \le \mathrm{Val}(G_{s\alpha}(v_\tau)) + \frac{\bar{m}_\tau}{\alpha}$, $s\in S$

where $\underline{m}_\tau := \min_{s\in S}\{\mathrm{Val}(G_{s\alpha}(v_\tau)) - v_\tau(s)\}$ and $\bar{m}_\tau := \max_{s\in S}\{\mathrm{Val}(G_{s\alpha}(v_\tau)) - v_\tau(s)\}$.

(c) If $\hat{\rho}^\tau$ and $\hat{\sigma}^\tau$ are such that for each $s\in S$, $\hat{\rho}_s^\tau$ and $\hat{\sigma}_s^\tau$ are optimal actions for player 1 and player 2 respectively in the matrix game $[G_{s\alpha}(v_\tau)]$, then $\hat{\rho}^\tau$ and $\hat{\sigma}^\tau$ are ε-optimal stationary strategies if $\bar{m}_\tau - \underline{m}_\tau \le \varepsilon.\alpha$.

We now mention the algorithm of Pollatschek & Avi-Itzhak (1969), which can be seen as a straightforward generalisation of Howard's policy iteration method.

6.1.3. ALGORITHM (Pollatschek & Avi-Itzhak).

(*i*) *Choose* $v_0\in\mathbb{R}^Z$, $\tau := 0$.

(*ii*) *Determine stationary strategies* $\rho^\tau=(\rho_1^\tau,\ldots,\rho_Z^\tau)$ *and* $\sigma^\tau=(\sigma_1^\tau,\ldots,\sigma_Z^\tau)$ *such that for each* $s\in S$, ρ_s^τ *and* σ_s^τ *are optimal actions in the matrix game* $[G_{s\alpha}(v_\tau)]$.

(*iii*) *Put* $v_{\tau+1} := V(\rho^\tau,\sigma^\tau)$.

(*iv*) $\tau := \tau+1$ *and go to step* (*ii*).

Pollatschek & Avi-Itzhak have shown that their algorithm always converges to the solution of the game if the following condition holds:

$$\max_{s\in S} \sum_{t=1}^{Z} \left(\max_{i,j} p(t|s,i,j) - \min_{a,b} p(t|s,i,j)\right) < \alpha,$$

with α the interest rate. Further for $\alpha>2$ (or discount factor $\beta<1/3$) the algorithm always converges for any transition map.

Pollatschek & Avi-Itzhak noticed in their paper that their algorithm may also converge under less restrictive conditions. Rao, Chandrasekaran & Nair (1973) claimed that the algorithm 6.1.3 always converges, but their proof is incorrect and an example provided by Van der Wal (1977) shows that the proof cannot be repaired.

Van der Wal (1977) presented a set of algorithms depending on an integer $\lambda \in \mathbb{N} \cup \{\infty\}$. It turns out that for $\lambda=1$ the corresponding algorithm coincides with Shapley's algorithm and for $\lambda=\infty$ the algorithm of Hoffman & Karp is obtained.

6.1.4. ALGORITHM (λ) (Van der Wal).

(*i*) *Choose* $v_0 \in \mathbb{R}^Z$ *arbitrarily,* $\tau := 0$.

(*ii*) *Determine for player 2 a stationary strategy* $\sigma^\tau = (\sigma_1^\tau, \ldots, \sigma_Z^\tau)$ *such that* σ_s^τ *is an optimal action in the matrix game* $[G_{s\alpha}(v_\tau)]$ *for each* $s \in S$.

(*iii*) *Put* $v_{\tau+1} := ML_{\sigma^\tau}^\lambda(v_\tau)$, *where* $ML_{\sigma^\tau}^\lambda$ *is the* λ^{th} *iterate of the map* $ML_{\sigma^\tau}: \mathbb{R}^Z \to \mathbb{R}^Z$ *defined by*

$$(ML_{\sigma^\tau}(v_\tau))_s := \max_{i \in A_s} \{r(s,i,\sigma^\tau) + \frac{1}{1+\alpha} \sum_{t=1}^{Z} p(t|s,i,\sigma^\tau).v_\tau(t)\}$$

(*iv*) $\tau := \tau+1$ *and go to step* (*ii*).

Van der Wal (1977) has shown that the properties (a)-(e) mentioned below algorithm 6.1.2, hold for every algorithm (λ), $\lambda \in \mathbb{N} \cup \{\infty\}$.

Van der Wal extensively studied generalizations of the method of successive approximation. For that purpose he introduced the concept of stopping times in stochastic games. We refer for further details to Van der Wal (1981).

Parthasarathy & Raghavan (1981) considered the class of games for which one player governs the transitions. For this special case they proved the orderfield property, i.e. the property that the value of the

game and the extreme optimal stationary strategies of both players lie in the same Archimedean field as the game parameters. This result holds as well for discounted stochastic games as for average reward games.

For discounted stochastic games they proposed the following linear programming algorithm, which is in fact a finite algorithm. It is assumed that in each state the transition probabilities are controlled by player 2.

6.1.5. ALGORITHM (Parthasarathy & Raghavan).

$$\max \sum_{s=1}^{z} v_s$$

$$\textit{subject to} \quad \sum_{i \in A_s} r(s,i,j).x_s(i) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,j).v_t \geq v_s, \quad j \in B_s, \quad s \in S$$

$$x_s(i) \geq 0, \qquad i \in A_s, \quad s \in S$$

$$\sum_{i \in A_s} x_s(i) = 1, \qquad s \in S.$$

Note that the NLP 4.4.1 reduces to this LP problem when $p(t|s,i,j)$ is independent of i.

Parthasarathy & Raghavan (1981) proved that an optimal solution (v^*,x^*) to this LP corresponds to a solution of the stochastic game, in the sense that v^* is the value of the game and ρ^* defined as $\rho_s^*(i)=x_s^*(i)$ proves to be an optimal stationary strategy for player 1. An optimal stationary strategy for player 2 can be obtained by solving the dual program of the above LP.

6.2. FICTITIOUS PLAY AS AN ITERATIVE METHOD FOR SOLVING DISCOUNTED STOCHASTIC GAMES.

Brown (1949;1951) suggested a method, called fictitious play, for solving a matrix game. Robinson (1950) proved the validity of that method, while Shapiro (1958) provided for the Brown-Robinson scheme an a priori estimate of the rate of convergence.

Extensions of the Brown-Robinson method to infinite zerosum games are given in Danskin (1954) and Van den Akker (1976), while Miyasawa (1961) studied an extension to 2×2-bimatrix games. Shapley (1964) has given examples showing that the natural generalization of the Robinson theorem

to arbitrary bimatrix games is not valid. A systematic study of this phenomenon was done by Rosenmüller (1971).

The method of fictitious play of Brown and Robinson can be seen as an infinite stage learning process associated with a matrix game [K]. Here at each stage the players choose a pure action which, among the pure actions, is a best answer with respect to the collection of past choices of the other player.

The purpose of this section is to extend the ideas of Brown and Robinson to discounted stochastic games. However before we can do so, we first study the consequences of fictitious play applied to a situation in which the matrix game [K] is not exactly known in advance, but where at each decision epoch τ an approximation $[K_\tau]$ is given such that $\lim_{\tau\to\infty} K_\tau = K$. This investigation is of independent interest, as it will appear that fictitious play can be used for a converging sequence of matrix games. Next we describe an iterative method for solving a discounted stochastic game with finite state and action spaces.
The results of this section are distilled from Vrieze & Tijs (1982).

Now let us consider a converging sequence $K_1, K_2, K_3, \ldots$ of m,n-matrices with $\lim_{\tau\to\infty} K_\tau = K$, i.e. $\lim_{\tau\to\infty} r_\tau(i,j) = r(i,j)$ for all $(i,j)\in \mathbb{N}_m \times \mathbb{N}_n$. In the following we denote by $\tau(\varepsilon)$, for any $\varepsilon>0$, the smallest integer such that

$$(6.2.1) \qquad |r_\tau(i,j) - r(i,j)| \leq \varepsilon, \quad \text{all } (i,j) \text{ and all } \tau \geq \tau(\varepsilon)$$

We also use the following notation.
For a vector $x=(x_1,x_2,\ldots,x_k)\in\mathbb{R}^k$, $\max\{x_1,\ldots,x_k\}$ is denoted by $\max(x)$ and $\min\{x_1,\ldots,x_k\}$ by $\min(x)$. Note that in this setting for a matrix K the value of the matrix game [K] equals

$$\mathrm{Val}(K) = \max_{\rho\in P(\mathbb{N}_m)} \min(\rho^T.K) = \min_{\sigma\in P(\mathbb{N}_n)} \max(K.\sigma).$$

C_τ^j and C^j denote the j-th column of K_τ and K respectively; similarly R_τ^i and R^i denote the i-th row.
Furthermore, let $M := \sup\{ |r_\tau(i,j)| \mid (i,j)\in\mathbb{N}_m\times\mathbb{N}_n,\ \tau\in\mathbb{N} \}$.

6.2.1. DEFINITION. *We call a pair of sequences*

$$x(0), x(1), x(2), \ldots \quad \text{in } \mathbb{R}^n$$
$$y(0), y(1), y(2), \ldots \quad \text{in } \mathbb{R}^m$$

a vector system for the sequence $K_1, K_2, K_3, \ldots$ *if*

(v1) $\min(x(0)) = \max(y(0))$

and if for each $\tau \in \mathbb{N}$

(v2) $x(\tau) := x(\tau-1) + R_\tau^{i(\tau)}$, *where* $i(\tau) \in \mathbb{N}_m$ *satisfies* $y_{i(\tau)}(\tau-1) = \max(y(\tau-1))$

(v3) $y(\tau) := y(\tau-1) + C_\tau^{j(\tau)}$, *where* $j(\tau) \in \mathbb{N}_n$ *satisfies* $x_{j(\tau)}(\tau-1) = \min(x(\tau-1))$.

It will be obvious how such a vector system can be formed recursively from given $x(0)$ and $y(0)$ satisfying (v1).
We wish to show that

$$\lim_{\tau\to\infty} \tau^{-1} \max(y(\tau)) = \lim_{\tau\to\infty} \tau^{-1} \min(x(\tau)) = \mathrm{Val}(K).$$

In proving this, we proceed as far as possible along the same lines as in Robinson (1950), who studied the situation in which $K_\tau = K$ for each τ.

6.2.2. LEMMA.

$$\limsup_{\tau\to\infty} \tau^{-1} \min(x(\tau)) \le \mathrm{Val}(K) \le \liminf_{\tau\to\infty} \tau^{-1} \max(y(\tau)).$$

PROOF. Take $\varepsilon>0$ and $k>\tau(\varepsilon)$. Then

$$x(k) = x(\tau(\varepsilon)) + \sum_{\tau=\tau(\varepsilon)+1}^{k} R_\tau^{i(\tau)} \le x(\tau(\varepsilon)) + \sum_{\tau=\tau(\varepsilon)+1}^{k} R^{i(\tau)} + \varepsilon.(k-\tau(\varepsilon)).1_n$$

Let $\rho(k,i)$ be the number of times that action $i \in \mathbb{N}_m$ appears in the sequence $i(\tau(\varepsilon)+1), i(\tau(\varepsilon)+2), \ldots, i(k)$.
Then $\rho(k) = (k-\tau(\varepsilon))^{-1}(\rho(k,1), \ldots, \rho(k,m))$ is a mixed action for player 1 in the matrix game $[K]$ and furthermore

$$(6.2.2) \qquad \rho^T(k).K = (k-\tau(\varepsilon))^{-1}. \sum_{\tau=\tau(\varepsilon)+1}^{k} R^{i(\tau)}.$$

Hence

$$x(k) \leq x(\tau(\varepsilon))+(k-\tau(\varepsilon)).\rho^{T}(k).K+\varepsilon(k-\tau(\varepsilon)).1_{n},$$

which implies that:

$$k^{-1}\min(x(k)) \leq k^{-1}\max(x(\tau(\varepsilon)))+k^{-1}(k-\tau(\varepsilon))\min(\rho^{T}(k).K)+$$
$$\varepsilon k^{-1}(k-\tau(\varepsilon))$$

$$\leq k^{-1}\max(x(\tau(\varepsilon)))+k^{-1}(k-\tau(\varepsilon)).\mathrm{Val}(K)+\varepsilon k^{-1}(k-\tau(\varepsilon)).$$

Hence, $\lim\sup_{k\to\infty} k^{-1}\min(x(k)) \leq \mathrm{Val}(K)+\varepsilon$ for each $\varepsilon>0$, from which the first inequality in the lemma follows.
The second inequality can be proved analogously.

□

6.2.3. DEFINITION. *If* $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ *is a vector system for the sequence of matrices* $K_1,K_2,K_3,\ldots$, *then the* i-th *pure action of player* 1 *is said to be* <u>*eligible*</u> *for the vector system in the interval* $[\tau,\tau']$ *when there exists a* $\tau_1\in[\tau,\tau']$ *such that* $y_i(\tau_1)=\max(y(\tau_1))$. *Eligibility of a pure action* j *for player* 2 *is defined analogously.*

6.2.4. LEMMA. *If for* $k,\tau\in\mathbb{N}$ *all pure actions of both players are eligible in* $[k,k+\tau]$, *then*

$$\max(x(k+\tau))-\min(x(k+\tau)) \leq 2\tau M$$

$$\max(y(k+\tau))-\min(y(k+\tau)) \leq 2\tau M.$$

PROOF. The lemma follows by modifying in an obvious way the proof of lemma 2 in Robinson (1950), using the definition of M.

□

6.2.5. LEMMA. *If, for given* $k\geq\tau(\varepsilon)$ *all pure actions of both players are eligible in* $[k,k+\tau]$, *then*

(6.2.3) $$\max(y(k+\tau))-\min(x(k+\tau)) \leq 4\tau M+2\varepsilon(k+\tau)+2M\tau(\varepsilon).$$

PROOF. In view of lemma 6.2.4.

$$\max(y(k+\tau))-\min(x(k+\tau)) \le 4\tau M+\min(y(k+\tau))-\max(x(k+\tau)). \tag{6.2.4}$$

Let $\rho(k+\tau)$ be the mixed action for player 1, as defined in the proof of lemma 6.2.2. The inequality (cf. (6.2.2))

$$x(k+\tau) \ge x(\tau(\varepsilon))+(k+\tau-\tau(\varepsilon)).\rho^{T}(k+\tau).K-\varepsilon(k+\tau-\tau(\varepsilon)).1_{n}$$

then implies,

$$\max(x(k+\tau)) \ge \min(x(\tau(\varepsilon)))+(k+\tau-\tau(\varepsilon))\mathrm{Val}(K^{T})-\varepsilon(k+\tau-\tau(\varepsilon)) \tag{6.2.5}$$

where K^{T} is the transpose of the matrix K.
Similarly, one can show that

$$\min(y(k+\tau)) \le \max(y(\tau(\varepsilon)))+(k+\tau-\tau(\varepsilon))\mathrm{Val}(K^{T})+\varepsilon(k+\tau-\tau(\varepsilon)). \tag{6.2.6}$$

Note further, that by (v1) and the definition of M,

$$\max(y(\tau(\varepsilon)))-\min(x(\tau(\varepsilon))) \le 2M\tau(\varepsilon). \tag{6.2.7}$$

Combining the inequalities (6.2.4)-(6.2.7) yields the assertion of the lemma.

□

For a matrix K we denote by K^{-i} the matrix, which is obtained from K by deleting the i-th row. K^{-j} is the matrix obtained from K by deleting the j-th column. When writing $K^{-\ell}$ it will be clear from the context whether ℓ is an action of player 1 (delete row ℓ) or an action of player 2 (delete column ℓ).
For a vector $y=(y_1,y_2,\ldots,y_m)\in\mathbb{R}^m$, let y^{-i} be the vector $(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m)$.

6.2.6. LEMMA. *Let* $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ *be a vector system for the sequence* $K_1,K_2,K_3,\ldots$ *of* m,n*-matrices, converging to* K. *Suppose that, in the interval* $[k,k+\hat{\tau}]$, *pure action* i *of player* 1 *is not eligible for the vector system. For* $\tau\in\{0,1,\ldots,\hat{\tau}\}$, *let*

$$x'(\tau) := x(k+\tau)+(\max(y(k))-\min(x(k))).1_n \text{ and } y'(\tau)=y^{-i}(k+\tau).$$

Then $\langle x'(\tau),y'(\tau);\ \tau\in\{0,1,\ldots,\hat{\tau}\}\rangle$ *is the first part of a vector system for the converging sequence of* m-1,n*-matrices* $K_{k+1}^{-i},K_{k+2}^{-i},K_{k+3}^{-i},\ldots$. *Furthermore*

$$\max(y(k+\hat{\tau}))-\min(x(k+\hat{\tau})) = (\max(y'(\hat{\tau}))-\min(x'(\hat{\tau})))+$$
$$(\max(y(k))-\min(x(k))).$$

PROOF. Obviously, $\min(x'(0))=\max(y'(0))$, since pure action i is not eligible at k. Because $\langle x(\tau),y(\tau)\rangle$ is a vector system and i is not eligible in $[k,k+\hat{\tau}]$, we have for $i'(\tau)=i(k+\tau)\in\{1,\ldots,i-1,i+1,\ldots,m\}$ and $j'(\tau)=j(k+\tau)\in\mathbb{N}_n$, and $\tau\in\{1,\ldots,\hat{\tau}\}$, that:

$$x(k+\tau)=x(k+\tau-1)+R_{k+\tau}^{i'(\tau)}, \quad y_{i'(\tau)}(k+\tau-1)=\max(y(k+\tau-1))$$
$$y(k+\tau)=y(k+\tau-1)+C_{k+\tau}^{j'(\tau)}, \quad x_{j'(\tau)}(k+\tau-1)=\min(x(k+\tau-1))$$

These inequalities imply for $\tau\in\{1,\ldots,\hat{\tau}\}$:

$$x'(\tau)=x'(\tau-1)+R_{k+\tau}^{i'(\tau)}, \quad y'_{i'(\tau)}(\tau-1)=\max(y'(\tau-1))$$
$$y'(\tau)=y'(\tau-1)+C_{k+\tau}^{j'(\tau)}, \quad x'_{j'(\tau)}(\tau-1)=\min(x'(\tau-1))$$

Consequently the first assertion in the lemma is proved.
Furthermore

$$\max(y(k+\hat{\tau}))-\min(x(k+\hat{\tau}))=\max(y'(\tau))-(\min(x'(\tau))-(\max(y(k))-\min(x(k)))),$$

which finishes the proof.

□

Analogously one can formulate a "player 2 version" of lemma 6.2.6, involving a non-eligible action j of player 2. Both versions will be used in the proof of the following lemma.

6.2.7. LEMMA. *Let* $K_1,K_2,K_3,\ldots$ *be a converging sequence of matrices and let* $\varepsilon>0$. *Then there exists a non-negative real number* $T(\varepsilon)$ *such that for each* $k\in\mathbb{N}_0$ *and each vector system* $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ *of the sequence* $K_{k+1},K_{k+2},K_{k+3},\ldots$ *we have*:

$$\max(y(\tau))-\min(x(\tau)) \leq \varepsilon\tau \text{ for all } \tau\geq T(\varepsilon).$$

PROOF. The lemma is proved by induction with respect to the size of the matrices (the number of rows plus the number of columns) under consideration.
For sequences of 1,1-matrices (of size 2) we can take $T(\varepsilon)=0$, since $y(\tau)=x(\tau)\in\mathbb{R}$ for all $\tau\in\mathbb{N}_0$.
Now suppose that the statement is true for converging sequences of $\hat{m},\hat{n}$-matrices, with size $\hat{m}+\hat{n}<m+n$.
Let $K_1,K_2,K_3,\ldots$ be a converging sequence of m,n-matrices and let $\varepsilon>0$.
By applying the induction hypothesis to the finite number of converging sequences $K_1^{-i},K_2^{-i},K_3^{-i},\ldots$, $i\in\mathbb{N}_m$ and $K_1^{-j},K_2^{-j},K_3^{-j},\ldots$, $j\in\mathbb{N}_n$ of size m+n-1, we may conclude that there exists a $\hat{T}(\varepsilon)$ such that for each $k\in\mathbb{N}_0$ and each vector system $\langle x'(\tau),y'(\tau)\rangle$ for any of the sequences $K_{k+1}^{-i},K_{k+2}^{-i},K_{k+3}^{-i},\ldots$ or $K_{k+1}^{-j},K_{k+2}^{-j},K_{k+3}^{-j}$ we have

(6.2.8) $$\max y'(\tau)-\min x'(\tau) \leq \varepsilon\tau \quad \text{for all } \tau\geq\hat{T}(\varepsilon).$$

Take a $k\in\mathbb{N}_0$ and an arbitrary vector system $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ for $K_{k+1},K_{k+2},K_{k+3},\ldots$. Let $\tau\geq\tau(\varepsilon)+\hat{T}(\varepsilon)$. Then there is a $h\in\mathbb{N}$ and $q\in[0,1)$, such that $\tau=\tau(\varepsilon)+(h+q).\hat{T}(\varepsilon)$.
We distinguish two cases.

Case 1. Suppose that there is an integer $\tilde{h}\in\{1,2,\ldots,h\}$ such that all actions of both players are eligible in the interval $[\tau(\varepsilon)+(q+\tilde{h}-1)\hat{T}(\varepsilon),\tau(\varepsilon)+(q+\tilde{h})\hat{T}(\varepsilon)]$. Let $\hat{h}$ be the largest integer in $\{1,2,\ldots,h\}$ with this property.
Then in each interval $[\tau(\varepsilon)+(q+d-1)\hat{T}(\varepsilon),\tau(\varepsilon)+(q+d)\hat{T}(\varepsilon)]$ with $d\in\{\hat{h}+1,\ldots,h\}$ at least one action for one of the players is not eligible. Repeated application of lemma 6.2.6 with respect to these $h-\hat{h}$ intervals yields

in view of (6.2.8):

$$(6.2.9)\qquad \max(y(\tau))-\min(x(\tau)) = (\max(y(\tau(\varepsilon)+(q+\hat{h}).\hat{T}(\varepsilon)))- \min(x(\tau(\varepsilon)+(q+\hat{h}).\hat{T}(\varepsilon))))+(h-\hat{h}).\varepsilon.\hat{T}(\varepsilon).$$

Since all actions are eligible in the interval $[\tau(\varepsilon)+(q+\hat{h}-1)\hat{T}(\varepsilon),\tau(\varepsilon)+(q+\hat{h}).\hat{T}(\varepsilon)]$, the first term in the right hand side of (6.2.9) is by lemma 6.2.5 at most

$$4\hat{T}(\varepsilon).M+2\varepsilon(\tau(\varepsilon)+(q+\hat{h})\hat{T}(\varepsilon))+2M\tau(\varepsilon) \le 2\varepsilon\tau+4\hat{T}(\varepsilon).M+2M\tau(\varepsilon).$$

And the second term is at most $\varepsilon\tau$. Hence

$$\max(y(\tau))-\min(x(\tau)) \le 3\varepsilon\tau+\gamma$$

where $\gamma=M(4\hat{T}(\varepsilon)+2\tau(\varepsilon))$. Consequently

$$(6.2.10)\qquad \max(y(\tau))-\min(x(\tau)) \le 4\varepsilon\tau, \text{ if } \tau \ge \max\{\tau(\varepsilon)+\hat{T}(\varepsilon), \varepsilon^{-1}\gamma\}$$

Case 2. If there is no such an integer $\tilde{h}$ with the property described in case 1, then lemma 6.2.6 (and its player 2 version) can be applied h times, yielding in view of (6.2.8):

$$\max(y(\tau))-\min(x(\tau)) \le \max(y(\tau(\varepsilon)+q\hat{T}(\varepsilon)))-\min(x(\tau(\varepsilon)+q\hat{T}(\varepsilon)))+ h\varepsilon\hat{T}(\varepsilon) \le 2M(\tau(\varepsilon)+\hat{T}(\varepsilon))+\varepsilon\tau \le \varepsilon\tau+\gamma.$$

This implies for $\tau\ge\max\{\tau(\varepsilon)+\hat{T}(\varepsilon), \frac{1}{3}\varepsilon^{-1}\gamma\}$ that

$$(6.2.11)\qquad \max(y(\tau))-\min(x(\tau)) \le 4\varepsilon\tau.$$

Combining the two cases we have by (6.2.10) and (6.2.11):

$$\max(y(\tau))-\min(x(\tau)) \le 4\varepsilon\tau \quad \text{for all } \tau\ge T(4\varepsilon),$$

when we take $T(4\varepsilon)=\max\{\tau(\varepsilon)+\hat{T}(\varepsilon),\varepsilon^{-1}\gamma\}$. This completes the proof of the induction step, since $T(4\varepsilon)$ does not depend on k and on the vector system $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$.

We now can state the main result of this section.

□

6.2.8. THEOREM. *Let* $K_1,K_2,K_3,\ldots$ *be a sequence of matrices with* $\lim_{\tau\to\infty}K_\tau=K$. *Then for each vector system* $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ *of this sequence, we have*

$$\lim_{\tau\to\infty}\tau^{-1}\max(y(\tau)) = \lim_{\tau\to\infty}\tau^{-1}\min(x(\tau)) = \mathrm{Val}(K).$$

PROOF. The proof is a direct consequence of lemmas 6.2.2 and 6.2.7.

□

For a vector system $\langle x(\tau),y(\tau);\ \tau\in\mathbb{N}_0\rangle$ of the sequence $K_1,K_2,K_3,\ldots$ converging to K, let for each $\tau\in\mathbb{N}$ $\hat{\rho}(\tau)$ and $\hat{\sigma}(\tau)$ be the mixed actions of players 1 and 2 respectively for which $\tau\hat{\rho}(\tau,i)$ equals the number of times i appears in $i(1),i(2),\ldots,i(\tau)$ and $\tau\hat{\sigma}(\tau,j)$ equals the number of times j appears in $j(1),j(2),\ldots,j(\tau)$.

6.2.9. THEOREM. *Each limit point of the sequence* $\hat{\rho}(1),\hat{\rho}(2),\hat{\rho}(3),\ldots$ *is an optimal mixed action of player* 1 *in the matrix game* [K]. *Each limit point of* $\hat{\sigma}(1),\hat{\sigma}(2),\hat{\sigma}(3),\ldots$ *is optimal for player* 2 *in* [K].

PROOF. We only prove the first assertion. Let $\tilde{\rho}$ be a limit point of $\hat{\rho}(1),\hat{\rho}(2),\hat{\rho}(3),\ldots$. Without loss of generality, we suppose that $\lim_{\tau\to\infty}\hat{\rho}(\tau)=\tilde{\rho}$. (Otherwise consider a properly chosen subsequence). Let $\varepsilon>0$ and let $\tau>\tau(\varepsilon)$. Then

$$\begin{aligned} x(\tau) &= x(0) + \sum_{k=1}^{\tau(\varepsilon)} R_k^{i(k)} + \sum_{k=\tau(\varepsilon)+1}^{\tau} R_k^{i(k)} \\ &\le x(0) + \sum_{k=1}^{\tau} R^{i(k)} + 2M\tau(\varepsilon).1_n + \varepsilon(\tau-\tau(\varepsilon)).1_n \\ &= x(0) + \tau.\hat{\rho}^T(\tau).K + 2M\tau(\varepsilon).1_n + \varepsilon(\tau-\tau(\varepsilon)).1_n. \end{aligned}$$

By taking limits we get

$$\lim_{\tau\to\infty}\tau^{-1}\min(x(\tau)) \le \min(\tilde{\rho}^T.K)+\varepsilon \quad \text{for each } \varepsilon>0.$$

But then by theorem 6.2.8

$$\mathrm{Val}(K) \leq \min(\tilde{\rho}^T.K) \leq \mathrm{Val}(K)$$

Consequently, $\tilde{\rho}$ is optimal for player 1 in the matrix game [K]. □

6.2.10. REMARK. If the sequence $K_1, K_2, K_3, \ldots$ converges to a matrix K with $[K] \in U_{mn}$ (see remark 5.4.10), then the sequences $\hat{\rho}(1), \hat{\rho}(2), \hat{\rho}(3), \ldots$ and $\hat{\sigma}(1), \hat{\sigma}(2), \hat{\sigma}(3), \ldots$ as defined above converge and the limits equal the unique optimal actions in the game [K].

Now we return to stochastic games. In the remainder of this section a stochastic game $\Gamma \in SG(S,\alpha)$ is supposed to be fixed and let v^* be the value of this game.
First we recall the notation $[G_{s\alpha} v)]$ for $v \in \mathbb{R}^z$, which has been defined as the matrix game, where the (i,j)-th entry is equal to $r(s,i,j)+(1+\alpha)^{-1}\sum_{t=1}^{z} p(t|s,i,j)v_t$. Then $\rho_s^T.G_{s\alpha}(v)$ for $\rho_s \in P(A_s)$ is defined as the vector in $\mathbb{R}^n$ whose j-th coördinate equals

$$r(s,\rho_s,j) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,\rho_s,j)v_t.$$

Analogously for a mixed action $\sigma_s \in P(B_s)$ of player 2 the vector $G_{s\alpha}(v).\sigma_s$ is defined.

For $i \in A_s$, e_i denotes the element ρ_s of $P(A_s)$ with $\rho_s(i)=1$ and $\rho_s(\hat{i})=0$, $\hat{i} \neq i$. Similarly, $e_j \in P(B_s)$ is defined. As we always use the notation i for an action of player 1 and the notation j for an action of player 2, no confusion will arise. Then $e_i^T.G_{s\alpha}(v)$ equals the i-th row of $G_{s\alpha}(v)$ and $G_{s\alpha}(v).e_j$ equals the j-th column of $G_{s\alpha}(v)$.

We now describe the algorithm for the discounted stochastic game. In the algorithm the following sequences are recursively defined.

$$\begin{aligned}
&V(0), V(1), V(2), \ldots && \in \mathbb{R}^z \\
&x(s,0), x(s,1), x(s,2), \ldots && \in \mathbb{R}^{n_s}, \quad s \in S \\
&y(s,0), y(s,1), y(s,2), \ldots && \in \mathbb{R}^{m_s}, \quad s \in S \\
&\rho_s(0), \rho_s(1), \rho_s(2), \ldots && \in P(A_s), \quad s \in S \\
&\sigma_s(0), \sigma_s(1), \sigma_s(2), \ldots && \in P(B_s), \quad s \in S.
\end{aligned}$$

6.2.11. ALGORITHM.

(i) *Choose* $x(s,0)$ *and* $y(s,0)$ *such that*

$$\min(x(s,0)) = \max(y(s,0)) \text{ and } \min(y(s,0)) \geq V_s^*, \quad s\in S.$$

Put $V_s(0) := \max(y(s,0)), \quad s\in S.$

Choose $\rho_s(0)\in P(A_s)$ *and* $\sigma_s(0)\in P(B_s)$ *arbitrarily,* $s\in S$, $\tau:=0$.

(ii) *Take for each* $s\in S$

$$i(s,\tau)\in A_s \text{ such that } y_{i(s,\tau)}(s,\tau-1) = \max(y(s,\tau-1))$$

and

$$j(s,\tau)\in B_s \text{ such that } x_{j(s,\tau)}(s,\tau-1) = \min(x(s,\tau-1)).$$

(iii) *Set for each* $s\in S$

$$V_s(\tau) := \min\{\max(\tau^{-1}y(s,\tau-1)),\ V_s(\tau-1)\}$$

$$x(s,\tau) := x(s,\tau-1)+e^T_{i(s,\tau)}.G_{s\alpha}(V(\tau))$$

$$y(s,\tau) := y(s,\tau-1)+G_{s\alpha}(V(\tau)).e_{j(s,\tau)}$$

$$\rho_s(\tau) := \frac{(\tau-1).\rho_s(\tau)+e_{i(s,\tau)}}{\tau}$$

$$\sigma_s(\tau) := \frac{(\tau-1).\sigma_s(\tau)+e_{j(s,\tau)}}{\tau}.$$

(iv) $\tau := \tau+1$ *and go to step* (ii).

We next show the following theorems.

6.2.12. THEOREM. *For each* $s\in S$:

$$\lim_{\tau\to\infty} \tau^{-1} \max(y(s,\tau)) = \lim_{\tau\to\infty} \tau^{-1} \min(x(s,\tau)) = \lim_{\tau\to\infty} V_s(\tau) = V_s^*.$$

6.2.13. THEOREM. *For each* $s\in S$ *let* $\hat{\rho}_s$ *be a limit point of the sequence* $\rho_s(1),\rho_s(2),\rho_s(3),\ldots$ *and* $\hat{\sigma}_s$ *a limit point of* $\sigma_s(1),\sigma_s(2),\sigma_s(3),\ldots$. *Then* $\hat{\rho}=(\hat{\rho}_1,\ldots,\hat{\rho}_z)$ *and* $\hat{\sigma}=(\hat{\sigma}_1,\ldots,\hat{\sigma}_z)$ *are optimal stationary strategies for player* 1 *and player* 2 *respectively.*

To prove these theorems, we need the following lemma.

6.2.14. LEMMA. $\lim_{\tau\to\infty} V(\tau)$ *exists.*

PROOF. By definition of $V_s(\tau)$, for each $s\in S$, the sequence $V_s(0),V_s(1),V_s(2),\ldots$ is decreasing. Hence, we have only to show that the sequence is bounded from below. We prove by induction that for each $\tau\in\mathbb{N}_0$

$$V_s(\tau) \geq V_s^* \quad \text{for each } s\in S. \tag{6.2.12}$$

For $\tau=0$ we have

$$V_s(0) = \max(y(s,0)) \geq \min(y(s,0)) \geq V_s^* \quad \text{for each } s\in S.$$

Suppose now that $V_s(\tau)\geq V_s^*$ for all $\tau\leq k$ and all $s\in S$, where $k\in\mathbb{N}$. Then

$$\begin{aligned} y(s,k) &= y(s,0) + \sum_{\tau=1}^{k} G_{s\alpha}(V(\tau)).e_{j(s,\tau)} \\ &\geq V_s^* + \sum_{\tau=1}^{k} G_{s\alpha}(V^*).e_{j(s,\tau)}. \end{aligned}$$

For the mixed action $\sigma_s(k)$, we have (see also (6.2.2))

$$\sum_{\tau=1}^{k} G_{s\alpha}(V^*).e_{j(s,\tau)} = k.G_{s\alpha}(V^*).\sigma_s(k).$$

Consequently,

$$\max(y(s,k)) \geq V_s^* + k \max(G_{s\alpha}(V^*).\sigma_s(k));$$

but the seond term in the right hand side of this inequality is at least $k.V_s^*$ by theorem 4.2.4. Hence

$$V_s(k+1) = \min\{\max((k+1)^{-1}y(s,k)),V_s(k)\} \geq \min\{V_s^*,V_s(k)\} = V_s^*,$$

for each $s\in S$. This completes the proof of the lemma.

□

PROOF of theorem 6.2.12. Take $s \in S$. Let $\hat{V}=\lim_{\tau\to\infty} V(\tau)$. This limit exists in view of lemma 6.2.14.
For $\tau \in \mathbb{N}$, let $[K_s(\tau)]$ be the m_s,n_s-matrix game $[G_{s\alpha}(V(\tau))]$.
Then $\lim_{\tau\to\infty} K_s(\tau)$ exists by lemma 6.2.14 and equals $K_s := G_{s\alpha}(\hat{V})$.
It is obvious that $\langle x(s,\tau), y(s,\tau);\ \tau \in \mathbb{N}_0 \rangle$ is a vector system for the sequence $K_s(1), K_s(2), K_s(3), \ldots$. Hence by theorem 6.2.8 we have

$$\lim_{\tau\to\infty} \tau^{-1} \max(y(s,\tau)) = \lim_{\tau\to\infty} \tau^{-1} \min(x(s,\tau)) = \mathrm{Val}(K_s).$$

Taking the limit in the definition of $V_s(\tau)$ and using the above equality we obtain

$$\hat{V}_s = \min\{\mathrm{Val}(K_s), \hat{V}_s\} \text{ or } \hat{V}_s \leq \mathrm{Val}(K_s) = \mathrm{Val}(G_{s\alpha}(\hat{V})).$$

This inequality holds for each $s \in S$. From lemma 4.2.3 we then infer that $\hat{V} \leq V^*$. Conversely from the proof of lemma 6.2.14, we obtain $\hat{V}=\lim_{\tau\to\infty} V(\tau) \geq V^*$. So $\hat{V}=V^*$. This implies by theorem 4.2.4 that $\mathrm{Val}(G_{s\alpha}(V^*)) = \mathrm{Val}(K_s) = V^*_s$, which proves theorem 6.2.12.

□

PROOF of theorem 6.2.13. By theorem 6.2.9, $\hat{\rho}_s$ and $\hat{\sigma}_s$ are optimal mixed actions in the matrix game $[K_s]=[G_{s\alpha}(V^*)]$ for player 1 and player 2 respectively. This holds for each $s \in S$. The theorem follows now from theorem 4.2.4.

□

6.2.15. REMARK. The vectors $V(\tau)$ of the algorithm approximate V^* from above. If one wants also an approximation from below, then one can modify the iteration procedure in the following way. Start with vectors $x(s,0)$, $y(s,0)$ and $V'_s(0)$ with

$$V'_s(0) = \min(x(s,0)) = \max(y(s,0)) \leq V^*_s$$

and define

$$V'_s(\tau) := \max\{\min(\tau^{-1} x(s,\tau)),\ V'_s(\tau-1)\}, \quad s \in S.$$

Then $[V'_s(\tau), V_s(\tau)]$ is an estimation interval around V^*_s whose length shrinks to zero, when τ increases to infinity.

6.2.16. REMARK. In section 5.4 we had defined by $USG(S,\alpha)$ the class of stochastic games with state space S and interest rate α and for which both players have a unique optimal stationary strategy. From theorem 4.4.11 we have seen that $USG(S,\alpha)$ is an open and dense subset of $SG(S,\alpha)$. For games belonging to the class $USG(S,\alpha)$ we can sharpen theorem 6.2.13 as follows (cf. remark 6.2.10).
If $\Gamma\in USG(S,\alpha)$, then the sequences $\rho(\tau)=(\rho_1(\tau),\ldots,\rho_z(\tau))$ and $\sigma(\tau)=(\sigma_1(\tau),\ldots,\sigma_z(\tau))$, $\tau\in\mathbb{N}$ converge to the unique optimal stationary strategies of player 1 and player 2 respectively.

6.2.17. REMARK. Consider a stochastic game for which in each state one player, say player 2, is a dummy, i.e. $B_s=\{1\}$ for each $s\in S$. Then the resulting problem constitutes a Markov decision problem.
Without loss of generality we may suppose that all payoffs are negative.
Take $y(s,0)=0_{m_s}\in\mathbb{R}^{m_s}$ and $x(s,0)=0\in\mathbb{R}$.
Then the algorithm yields ($\tau\geq 1$):

$$V_s(\tau) = \max(\tau^{-1}y(s,\tau-1))$$

$$\tau^{-1}y_i(s,\tau) = r(s,i,1) + \frac{1}{1+\alpha}\sum_{t=1}^{z} p(t|s,i,1).\tilde{V}_t(\tau),\ i\in\{1,\ldots,m_s\}$$

where $\tilde{V}_t(\tau) = \tau^{-1}\sum_{k=1}^{\tau} V_t(k)$.
This scheme shows similarity with the successive approximation method for solving Markov decision problems (cf. Howard (1960)), but the convergence rate may be slower.
In Markov decision theory, the following scheme is used:

$$V_s(\tau) = \tau^{-1}\max(y(s,\tau))$$

$$\tau^{-1}y_i(s,\tau) = r(s,i) + \frac{1}{1+\alpha}\sum_{t=1}^{z} p(t|s,i).V_t(\tau-1),\ i\in\{1,\ldots,m_s\}$$

which assures a geometric rate of convergence.
In a similar way as for that algorithm, it can be shown for our algorithm when applied on Markov decision problems, that for τ sufficiently large only actions $i(s,\tau)$ are chosen for which $\rho(\tau)=(i(1,\tau),\ldots,i(z,\tau))$ is an optimal stationary strategy. Moreover, if we change our iteration scheme somewhat by taking

$$V_s(\tau) := \min\{\max(\tau^{-1}y(s,\tau-1)), V_s(\tau-1), \max(G_{s\alpha}(V(\tau-1)).e_{j(s,\tau)})\},$$

then, in applications to Markov decision problems our algorithm and the above mentioned algorithm for Markov decision problems give the same iteration scheme.

For stochastic games where player 1 has an optimal pure stationary strategy, this modification of the iteration scheme also gives an improvement of the convergence rate. In that case the convergence rate becomes geometric. For general stochastic games no improvement may be expected.

6.2.18. REMARK. Shapiro (1958) obtained an a priori estimate of the convergence rate of $O(\tau^{-1/(m+n+2)})$ of the Brown-Robinson scheme for solving an m,n-matrix game.
In a similar way as Shapiro did, we can prove for a stochastic game that for each $s \in S$

$$\max(\tau^{-1}y(s,\tau)) - \min(\tau^{-1}x(s,\tau)) \leq M_\alpha 2^{m_s+n_s}.\tau^{-1/(m_s+n_s-2)},$$

where $M_\alpha = (\alpha+1)\alpha^{-1} \max_{s,i,j} r(s,i,j)$.

Hence, in spite of the fact, that during the iteration also the limit matrix is approximated, the a priori convergence rate is no worse than in the Brown-Robinson procedure. This shows that our algorithm may be a competitor of the usual successive approximation algorithms (cf. section 6.1). The next table indicates that the possibilities of our iteration scheme are better than the possibilities of the Brown-Robinson procedure of being a competitor of linear programming in solving a matrix game. Namely our iteration scheme needs z times the number of computations of the Brown-Robinson procedure, while the successive approximation method needs z times the number of iterations times the number of computations of solving an LP problem.

	LP/succ.approx.	Brown-Robinson/our algorithm
matrix game of size m,n	to solve 1 dual pair of LP problems of size m,n	at each step to compute the maximum of an m-vector and the minimum of an n-vector
stochastic game of size m_s,n_s in state s, $s\in\{1,\ldots,z\}$	at each step, to solve a dual pair of LP problems of size m_s,n_s, for each $s\in\{1,\ldots,z\}$	at each step to compute the maximum of an m_s-vector and the minimum of an n_s-vector for all $s\in\{1,\ldots,z\}$.

6.2.19. REMARK. The algorithm can also be applied when we are dealing with a stochastic game which is not exactly known in advance, but for which at each stage $\tau\in\mathbb{N}$ an approximation $<r^\tau,p^\tau,\alpha^\tau>$ is given, with $\lim_{\tau\to\infty}<r^\tau,p^\tau,\alpha^\tau>=<r,p,\alpha>$.
At stage τ the approximation $<r^\tau,p^\tau,\alpha^\tau>$ should be used. Also in this case the algorithm will converge to the value of the game $<r,p,\alpha>$. Furthermore, taking limits, optimal stationary strategies for both players are obtained.

The reason why also in this case the algorithm can be applied, is due to the fact that the value function on $SG(S,\alpha)$ is a continuous one (theorem 4.3.4) and the fact that the multifunctions $\mathcal{O}_1$ and $\mathcal{O}_2$ are upper semi-continuous on $SG(S,\alpha)$ (cf. theorem 4.3.8).

6.3.3. A FINITE ALGORITHM FOR THE DISCOUNTED SWITCHING CONTROL STOCHASTIC GAME.

In this section we describe an algorithm for the switching control stochastic game, i.e. a game for which in each state only one of the players governs the transitions, where not necessarily in each state the same player governs the transition. This model is an extension of the model of Parthasarathy & Raghavan (1981). They considered stochastic games for which in each state the same player governs the transitions.

The algorithm presented here, consists of a finite sequence of linear programming problems. The linear programs involved in each step, correspond

to the linear program of algorithm 6.1.5 which solves a "one player control" stochastic game.

In proving that this algorithm yields the value and optimal stationary strategies for the both players the orderfield property arises in a natural way. The validity of the orderfield property for this model is first shown by Filar (1981).

Further, in the proof the matrix lemma of Parthasarathy & Raghavan (1981) plays a crucial role.

6.3.1. DEFINITION. *A discounted switching control stochastic game is a game* Γ *for which in each state* $s \in S$ *either*

(1) $p(t|s,i,j)=p(t|s,i,\hat{j})$ *for all* $(j,\hat{j}) \in B_s \times B_s$, *all* $t \in S$ *and* $i \in A_s$

or

(2) $p(t|s,i,j)=p(t|s,\hat{i},j)$ *for all* $(i,\hat{i}) \in A_s \times A_s$, *all* $t \in S$ *and* $j \in B_s$

In case (1) *we say that player* 1 *controls the transitions and in case* (2) *player* 2 *controls the transitions.*

The transition probabilities for a state s where player 1 (player 2) controls them are denoted by $p(t|s,i)$ $(p(t|s,j))$.
The set of states where player 1 (player 2) controls the transitions is denoted by S_1 (S_2).

Before we give the algorithm, we investigate what happens if a player fixes the part of a stationary strategy corresponding to the states where he controls the transitions.
Let player 1 fix $\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S_1\}$. Then associate with $\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S_1\}$ the game $\hat{\Gamma}(\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S_1\})$ in the following way: $\hat{S}:=S$; for $s \in S_1$, $\hat{A}_s:=\{1\}$, $\hat{B}_s:=B_s$, $\hat{r}(s,1,j):= \sum_{i \in A_s} r(s,i,j)\rho_s(i)=r(s,\rho_s,j)$ and $\hat{p}(t|s,1,j):=p(t|s,\rho_s)$; for $s \in S_2$, $\hat{A}_s:=A_s$, $\hat{B}_s=B_s$, $\hat{r}(s,i,j):=r(s,i,j)$ and $\hat{p}(t|s,i,j)=p(t|s,j)$. So $\hat{\Gamma}(\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S_1\})$, sometimes abbreviated to $\hat{\Gamma}(\{.\})$, is a "player 2 control" stochastic game as discussed in algorithm 6.1.5.

Observe that if, after fixing $\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S_1\}$, the players play the game $\hat{\Gamma}(\{.\})$, they restrict themselves in their possibilities of choosing strategies. Namely their strategies in $\hat{\Gamma}(\{.\})$ cannot depend on the actual outcomes of the chance experiments corresponding to $\{\rho_s | \rho_s \in \mathcal{P}(A_s),\ s \in S\}$.

However, along the same lines as the proof of theorem A.2.3 it can be shown that, after fixing $\{\rho_s | \rho_s \in P(A_s),\ s \in S_1\}$, neither player 1 nor player 2 can do better than concentrating upon $\hat{\Gamma}(\{.\})$.

We now state the algorithm and then show that this algorithm solves in a finite number of iterations the discounted switching control stochastic game.

6.3.2. ALGORITHM.

(i) *Put* $\tau=0$ *and choose* $\rho_s(0) \in P(A_s)$, $s \in S_1$ *as an extreme optimal action of the matrix game* $[G_{s\alpha}(o_z)]$ (*cf. theorem* A.1.9).

(ii) *Put* $\tau := \tau+1$.

Solve $\hat{\Gamma}(\{\rho_s(\tau-1) | \rho_s(\tau-1) \in P(A_s),\ s \in S_1\})$ *with the following* LP *problem* (*variables* v_s, $s \in S$ *and* $x_s(i)$, $i \in A_s$ *and* $s \in S_2$)

$$\max \sum_{s=1}^{z} v_s$$

subject to

$$r(s,\rho_s(\tau-1),j) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,\rho_s(\tau-1)).v_t \geq v_s, \quad j \in B_s,\ s \in S_1$$

$$\sum_{i \in A_s} r(s,i,j)x_s(i) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,j).v_t \geq v_s, \quad j \in B_s,\ s \in S_2$$

$$x_s(i) \geq 0, \quad i \in A_s,\ s \in S_2$$

$$\sum_{i \in A_s} x_s(i) = 1, \quad s \in S_2.$$

Define $v(\tau)$ *as the optimal value of this program.*

(iii) *If* $\mathrm{Val}(G_{s\alpha}(v(\tau))) = v_s(\tau)$ *for each* $s \in S_1$, *then the algorithm stops; else for* $s \in S_1$, *choose* $\rho_s(\tau)$ *such that* $\rho_s(\tau)$ *is an extreme optimal action of the matrix game* $[G_{s\alpha}(v(\tau))]$ *for player* 1. *Return to* (ii).

We will prove that algorithm 6.3.2 stops after a finite number of iterations. For this purpose the following lemma is needed.

6.3.3. LEMMA. *For* $\tau=1,2,\ldots$ *we have* $v(\tau)\leq v(\tau+1)$ *Furthermore, if* $\mathrm{Val}(G_{\hat{s}\alpha}(v(\tau)))\neq v_{\hat{s}}(\tau)$ *for some* $\hat{s}\in S_1$, *then* $v(\tau)<v(\tau+1)$.

PROOF. Since the LP problem in step 2 coïncides with algorithm 6.1.5 we have by the result of Parthasarathy & Raghavan (1981) that

$$v(\tau) = \mathrm{Val}(\hat{\Gamma}(\{\rho_s(\tau-1) \mid \rho_s(\tau-1)\in P(A_s),\ s\in S_1\})), \text{ so}$$

$$v_s(\tau) = \mathrm{Val}(G_{s\alpha}(v(\tau))), \quad s\in S_2 \tag{6.3.1}$$

and

$$v_s(\tau) = \min_{j\in B_s}\{r(s,\rho_s(\tau-1),j) + \frac{1}{1+\alpha}\sum_{t=1}^{z} p(t|s,\rho_s(\tau-1)).v_t(\tau)\},\ s\in S_1 \tag{6.3.2}$$

From the definition of $\rho_s(\tau)$, we get by (6.3.2):

$$v_s(\tau) \leq \min_{j\in B_s}\{r(s,\rho_s(\tau),j) + \frac{1}{1+\alpha}\sum_{t=1}^{z} p(t|s,\rho_s(\tau)).v_t(\tau)\}, \quad s\in S_1. \tag{6.3.3}$$

Let for $s\in S_2$, $\hat{\rho}_s$ be optimal in $G_{s\alpha}(v(\tau))$. Let the stationary strategy $\tilde{\rho}$ be such that $\tilde{\rho}_s=\hat{\rho}_s$, $s\in S_2$ and $\tilde{\rho}_s=\rho_s(\tau)$, $s\in S_1$.
Then from (6.3.1) and (6.3.3) it follows that for every pure stationary strategy σ^p of player 2 we have:

$$v(\tau) \leq r(\tilde{\rho},\sigma^p) + \frac{1}{1+\alpha}.P(\tilde{\rho},\sigma^p).v(\tau)$$

which by lemma 4.2.3 implies:

$$V(\tilde{\rho},\sigma^p) \geq v(\tau), \quad \text{all } \sigma^p\in PSST_2. \tag{6.3.4}$$

But the "S_2-part" of $\tilde{\rho}$ is a strategy for player 1 in the game $\hat{\Gamma}(\{\rho_s(\tau) \mid \rho_s(\tau)\in P(A_s),\ s\in S_1\})$, for which moreover (6.3.1) and (6.3.3) hold, so

$$v(\tau+1) = \mathrm{Val}(\hat{\Gamma}(\{\rho_s(\tau) \mid \rho_s(\tau)\in P(A_s),\ s\in S_1\})) \geq \inf_{\nu\in ST_2} V(\tilde{\rho},\nu) =$$

$$\min_{\sigma^P\in PSST_2} V(\tilde{\rho},\sigma^P) \geq v(\tau),$$

which proves the first part of the lemma.

If $\mathrm{Val}(G_{\hat{s}\alpha}(v(\tau))) > v_{\hat{s}}(\tau)$ for some $\hat{s}\in S_1$, then in (6.3.4) the inequality sign holds (also by lemma 4.2.3), so in this case $v(\tau+1) > v(\tau)$.

□

6.3.4. THEOREM. *Algorithm* 6.3.2 *stops after a finite number of iterations.*

PROOF. If at τ the algorithm does not stop, then by lemma 6.3.3

$$\mathrm{Val}(\hat{\Gamma}(\{\rho_s(\tau) \mid \rho_s(\tau)\in P(A_s),\ s\in S_1\})) = v(\tau+1) >$$

$$v(\tau) = \mathrm{Val}(\hat{\Gamma}(\{\rho_s(\tau-1) \mid \rho_s(\tau-1)\in P(A_s),\ s\in S_1\})).$$

But this implies that for each $\tau,k\in\mathbb{N}$ with $\tau \neq k$:

$$(6.3.5) \qquad \{\rho_s(\tau) \mid \rho_s(\tau)\in P(A_s),\ s\in S_1\} \neq \{\rho_s(\tau+k) \mid \rho_s(\tau+k)\in P(A_s),\ s\in S_1\}.$$

Now we invoke the matrix lemma of Parthasarathy & Raghavan (1981, lemma 4.1). This lemma says that an extreme optimal action for player 1 in a matrix game of payoff type $[r_{ij}+f_i]$ with pure action sets A and B respectively, is also an extreme optimal action for player 1 in some subgame $[r_{ij}]$ with pure action sets $\hat{A}$ and B, with $\hat{A}\subset A$. Application of the matrix lemma to step (iii) of our algorithm means that for each state $s\in S$ and each $\tau\in\mathbb{N}$ an extreme optimal action will be chosen of some matrix subgame $[r(s,.,.)]$ with pure action sets $\hat{A}_s(\tau)$ and B_s respectively, where $\hat{A}_s(\tau)\subset A_s$. But, since there are a finite number of different subsets $\hat{A}_s(\tau)$ of A_s and since a matrix game has a finite number of extreme optimal actions (Shapley & Snow (1950), cf. theorem A.1.9), we see that for any $s\in S$ and $\tau\geq 1$ the action $\rho_s(\tau)$ is chosen from the same finite set. Combination of this observation with (6.3.5) yields the theorem.

□

We now prove that when the algorithm stops, we have a solution of the switching control stochastic game.

6.3.5. THEOREM. *If algorithm* 6.3.2 *stops at the* τ-th *iteration, then* $v(\tau)$ *equals the value of the game.*

PROOF. Let the algorithm stop at the τ-th iteration.
Then from (iii), we derive $v_s(\tau)=\mathrm{Val}(G_{s\alpha}(v(\tau)))$, $s\in S_1$.
Since from the LP problem corresponding to $\hat{\Gamma}(\{\rho_s(\tau-1) \mid \rho_s(\tau-1)\in P(A_s), s\in S_1\})$, we already have $v_s(\tau)=\mathrm{Val}(G_{s\alpha}(v(\tau)))$, $s\in S_2$, we see that $v_s(\tau)=\mathrm{Val}(G_{s\alpha}(v(\tau)))$, all $s\in S$. Hence by theorem 4.2.4 $v(\tau)$ equals the value of the game.

□

6.3.6. REMARK. Let the algorithm stop at the τ-th iteration step, Then clearly optimal stationary strategies can be constructed from the matrix games $[G_{s\alpha}(v(\tau))]$, $s\in S$ (cf. theorem 4.2.4). For player 1 an optimal stationary strategy can directly be derived from the LP of step (ii) at iteration step τ. Namely, let $(\hat{x},v(\tau))$ be an optimal solution of this LP; then it follows that the stationary strategy $\hat{\rho}$, defined by $\hat{\rho}_s=\rho_s(\tau-1)$, $s\in S_1$ and $\hat{\rho}_s=(\hat{x}_s(1),\ldots,\hat{x}_s(m_s))$, $s\in S_2$, is optimal for player 1, because $\hat{\rho}_s$ is an optimal action in the matrix game $[G_{s\alpha}(v(\tau))]$, $s\in S$ and $v(\tau)$ is the value of the game.

6.3.7. REMARK. The orderfield property for the class of discounted switching control stochastic games, as shown by Filar (1981), can be alternatively proved by the algorithm. It follows by a result of Weyl (1950) that both the extreme optimal actions of a matrix game and the value of that game are in the same ordered field as the entries of the matrix. This clearly holds also for the solution of a feasible linear programming problem with a bounded solution. Then we can show by induction, that for each τ both $v(\tau)$ and each $\rho_s(\tau)$, $s\in S_1$, are in the same ordered field as the parameters of the game. Hence by the theorems 6.3.4 and 6.3.5 this results in the orderfield property.

Part III. *Average reward stochastic games.*

7. Introduction and preliminaries.

7.1. HISTORICAL REVIEW.

In definition 2.3.2 we defined the two-person zerosum stochastic game with the average reward as criterion. The notation introduced in part I will again be used.

In part III we consider only two-person zerosum stochastic games with finite state spaces and finite action spaces for the two players, and when we speak of a stochastic game we tacitly mean a game of this kind.

Bewley & Kohlberg (1978) have indicated that there are several ways of defining the average payoff, corresponding to a pair of strategies μ and ν and a starting state s. For instance when player 1 takes account of his worst case, one could take

$$\liminf_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r(Z_{\mu\nu}^{\tau}, X_{\mu\nu}^{\tau}, Y_{\mu\nu}^{\tau})\}$$

or

$$\mathbb{E}_s\{\liminf_{k\to\infty} \frac{1}{k+1} \sum_{\tau=0}^{k} r(Z_{\mu\nu}^{\tau}, X_{\mu\nu}^{\tau}, Y_{\mu\nu}^{\tau})\}$$

or

$$\liminf_{\alpha\to 0} \mathbb{E}_s\{\frac{\alpha}{1+\alpha} \sum_{\tau=0}^{\infty} (\frac{1}{1+\alpha})^{\tau} r(Z_{\mu\nu}^{\tau}, X_{\mu\nu}^{\tau}, Y_{\mu\nu}^{\tau})\}$$

or

$$\mathbb{E}_s\{\liminf_{\alpha\to 0} \frac{\alpha}{1+\alpha} \sum_{\tau=0}^{\infty} (\frac{1}{1+\alpha})^{\tau} r(Z_{\mu\nu}^{\tau}, X_{\mu\nu}^{\tau}, Y_{\mu\nu}^{\tau})\}$$

Obviously there is no reason why lim inf should not be replaced by lim sup or any convex combination thereof. Fortunately, Mertens & Neyman (1981) have shown that for any stochastic game with finite state and action spaces, both players have, for each $\varepsilon>0$, ε-optimal strategies with respect to each "reasonable" definition of the average reward payoff, including those given above.

In defining the average payoff, the zerosum aspect of the game should not be disturbed, in the sense that if $W_{s\mu\nu}$ is defined as the average payoff to player 1, then the payoff to player 2 should be defined as $-W_{s\mu\nu}$. This implies that, in general, the definition of the average payoff is asymmetric with respect to the players. A symmetric definition may lead to the strange situation that both players may profit from cooperating though the immediate payoffs are zerosum. For instance, suppose that the following matrix game $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is repeatedly played. Suppose that the players 1 and 2 respectively play the strategies μ and ν which are both defined as: play n^n times action 1, followed by $(n+1)^{n+1}$ times action 2, etc. Then it can be verified that

$$\limsup_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})\} = 1$$

and

$$\liminf_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})\} = -1.$$

Let $r_j(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})$ be the stochastic variable denoting the payoff to player j, $j\in\{1,2\}$, at decision epoch τ. Then the use of the symmetric definition "lim sup" for both players would lead to:

$$W_{s\mu\nu}(1) := \limsup_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r_1(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})\} = 1$$

and

$$W_{s\mu\nu}(2) := \limsup_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r_2(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})\} = 1.$$

This shows firstly that the zerosum property has vanished in the limit case, and secondly that both players realise profit in cooperating.

As already defined in section 2.3 we adopt the following definition of the average reward

$$(7.1.1) \qquad W_{s\mu\nu} := \liminf_{k\to\infty} \mathbb{E}_s\{\frac{1}{k+1} \sum_{\tau=0}^{k} r(Z^\tau_{\mu\nu}, X^\tau_{\mu\nu}, Y^\tau_{\mu\nu})\}.$$

which equals the payoff to player 1. The payoff to player 2 equals by definition $-W_{s\mu\nu}$.

Two-person zerosum stochastic games with the average reward as criterion (often called undiscounted stochastic games) were introduced by Gillette (1957). He considered games with perfect information (in each state one of the players has only one action available), and irreducible stochastic games (games where for each pair of stationary strategies the associated stochastic matrix $P(\rho,\sigma)$ has a single ergodic class and no transient states). However, Gillette used an incorrect extension of the Hardy-Littlewood theorem, in showing that for these models both players possess optimal stationary strategies. This was pointed out by Liggett & Lippman (1969).

Blackwell & Ferguson (1968) used a slightly modified version of an example of Gillette to show for undiscounted stochastic games that, in general, the players need not possess optimal stationary strategies. What is more, for this example called the big match, one of the players has no ε-optimal strategy within the class of semi-Markov strategies for $\varepsilon>0$ small enough (cf. example 3.3).

For a long time it was an open question whether average reward stochastic games with finite state and action spaces always have a value. Only about 1980 this question was answered in the affirmative, independently by Monash (1979) and Mertens & Neyman (1981). Before that, results for special cases of undiscounted stochastic games were obtained by several authors. The emphasis was laid mainly on the existence of optimal stationary strategies for the players. Gillette's paper has already been discussed. Hoffman & Karp (1966) have also treated irreducible stochastic games. Their approach is based on results of Markov decision theory.

Kohlberg (1974) treated so-called "repeated games with absorbing states". These are games where all but one of the states are absorbing, and where the remaining state is transient or recurrent, depending on the strategies played. The big match (example 3.3) belongs to this class of games. Kohlberg showed that these games have a value which can be found by considering the τ-step game and letting τ tend to infinity. It was later discovered that this paper of Kohlberg indicated the way in which in the general case the existence of the value can be shown.

Stern (1975) has proved the existence of the value and the existence of optimal stationary strategies for both players for games having the property that a state $s\in S$ exists such that for each pair of stationary strategies this state can be reached from any other state. Stern also considered games for which only one of the players governs the transitions.

He was able to prove the existence of the value and of optimal semi-Markov strategies. In a nice way this last model is also studied by Parthasarathy & Raghavan (1978, 1981). They showed that the value of such a game lies in the same Archimedean field as the parameters of the game, and that both players have optimal stationary strategies also lying in that field (these results do not hold in general). Furthermore the optimal stationary strategy of the player controlling the transitions can be chosen in such a way that it is also optimal for the discounted stochastic game with interest rate α for all values of α sufficiently close to 0.

A variant of the model of Parthasarathy & Raghavan is studied by Filar (1979). He treated the switching control case, i.e. games for which in each state one of the players governs the transitions but not necessarily the same player has this privilege in every state. Filar obtained results analogous to the above mentioned results of Parthasarathy & Raghavan.

Federgruen (1978) has extended methods used in Markov decision theory to stochastic games. One class of games he studied is the class of games where, for each stationary strategy of one of the players, the other player has a stationary strategy such that the associated Markov matrix of one-step transition probabilities is irreducible. This condition is an extension of the communicatingness property as treated in Bather (1973). Another class of games considered by Federgruen consists of games where for each pair of stationary strategies ρ and σ the associated stochastic matrix $P(\rho,\sigma)$ has the same number of ergodic classes. In both cases he proved that the value exists and that both players have optimal stationary strategies. He obtained his results by showing that $\lim_{\alpha\to 0} \alpha(1+\alpha)^{-1} V_\alpha$ exists where V_α is the value of the $(1+\alpha)^{-1}$-discounted game, and using this fact in an appropriate way. In fact Federgruen proved his results for the N-person game. Bewley & Kohlberg (1976a, 1976b, 1978) exposed in an elegant way some of the relationships between the discounted game, the τ-step game and the undiscounted game. In the next section we shall explain their use of the field of real Puiseux series and mention some of their results, which will be used in several sections of part III of this monograph. Regarding extensions of the model of the two-person zerosum games to N-person stochastic games we mention the work of Rogers (1969), Sobel (1971) and Federgruen (1978).

7.2. THE LIMIT DISCOUNT EQUATION.

In this section we first introduce the field of real Puiseux series. As Bewley & Kohlberg (1976a, 1978) have shown this field appears to be a useful tool for analysing stochastic games.
Formally, let for a positive integer M

$$F_M := \{ \sum_{k=-\infty}^{K} c_k \theta^{k/M} \mid K \text{ is an integer, } c_k \in \mathbb{R} \text{ and such that the series } \sum_{k=-\infty}^{K} c_k \tau^{k/M} \text{ converges for all sufficiently large real numbers } \tau \}.$$

Here θ represents an arbitrarily large real number.
Thus the members of F_M are power series in $\theta^{1/M}$.
Addition and multiplication in F_M are defined in a way that corresponds to the same operations on power series. The ordering on F_M reflects the notion that θ represents an arbitrarily large real number. To be more specific

$$\sum_{k=-\infty}^{K_1} c_k \theta^{k/M} + \sum_{k=-\infty}^{K_2} d_k \theta^{k/M} = \sum_{k=-\infty}^{\max(K_1,K_2)} (c_k+d_k)\theta^{k/M}$$

$$\sum_{k=-\infty}^{K_1} c_k \theta^{k/M} \cdot \sum_{k=-\infty}^{K_2} d_k \theta^{k/M} = \sum_{k=-\infty}^{K_1+K_2} \Big(\sum_{i+j=k} c_i d_j \Big) \theta^{k/M}$$

$\sum_{k=-\infty}^{K} c_k \theta^{k/M} > 0$ if and only if $c_{k^*} > 0$, where k^* is the largest integer k, such that $c_k \neq 0$.

Once can easily verify by elementary analysis that F_M is an ordered field.

Let $F := \bigcup_{M=1}^{\infty} F_M$, then F is also an ordered field and F is called the field of real Puiseux series. If $w = \sum_{k=-\infty}^{K} c_k \theta^{k/M} \in F$, then $\phi_\tau(w)$, for $\tau \in \mathbb{R}^+$, denotes the sum $\sum_{k=-\infty}^{K} c_k \tau^{k/M}$. Sometimes we write $w(\theta)$ for an element of F, in which case $w(\tau)$ represents $\phi_\tau(w)$.

The following facts are clear: if $w \in F$, then $\phi_\tau(w)$ is well defined for sufficiently large τ, and $w>0$ if and only if $\phi_\tau(w)>0$ for all sufficiently large τ.

For $w \in F_M$ the valuation of w is defined as $\Phi(w) := \frac{k^*}{M}$, where k^* is the largest integer k such that $c_k \neq 0$.

The expression $o(\theta^\gamma)$ and $O(\theta^\gamma)$ will be used to denote an element of F of respective valuations less than γ and at most γ.
F^z denotes the z-fold Cartesian product of F.

7.2.1. DEFINITION. *For a two-person zerosum stochastic game with finite state and action spaces the set of equations*

$$(7.2.1) \qquad x_s = \operatorname*{Val}_{A_s \times B_s}\left(r(s,.,.) + \frac{1}{1+\theta^{-1}} \sum_{t=1}^{z} p(t|s,.,.)x_t\right), \quad s \in S$$

where $x=(x_1,x_2,\ldots,x_z) \in F^z$, *is called the limit discount equation.*

Weyl (1950) showed that if the elements of a matrix belong to a certain ordered field, then the value of the corresponding matrix game also belongs to this field. Hence, if $x \in F^z$ then the right hand side of (7.2.1) is an element of F. The equations (7.2.1) will be abbreviated to

$$(7.2.2) \qquad x_s = \operatorname*{Val}_{A_s \times B_s}(G_{s\theta}(x)), \quad s \in S.$$

Note that for $\theta=\tau$, (7.2.1) represents the optimality equation for the discounted stochastic game, with interest rate $1/\tau$. Bewley & Kohlberg (1976a) have shown that for a matrix game $[h(.,.)]$ with entries belonging to F it holds that

$$\phi_\tau(\operatorname{Val}(h(.,.))) = \operatorname{Val}(\phi_\tau(h(.,.))).$$

So a solution to (7.2.1) would solve the discounted stochastic game for all sufficiently small interest rates.

Further one should note that a mixed action for player 1 in an m,n-matrix game with entries in F is of the form

$\sum_{k=-\infty}^{0} f_k \theta^{k/M}$, where $M \in \mathbb{N}$, $f_0 \in P(\mathbb{N}_m)$, $f_k \in \mathbb{R}^m$ such that $\sum_{i=1}^{m} f_k(i)=0$ for each $-k \in \mathbb{N}$, and for each $i \in \mathbb{N}_m$ and each $-K \in \mathbb{N}$: $\sum_{k=K}^{0} f_k(i)\theta^{k/M} \geq 0$.

Analogous properties hold for a mixed action $\sum_{k=-\infty}^{0} g_k \theta^{k/M}$ of player 2.

In the notation $\underset{A_s \times B_s}{\text{Val}}\,(h(.,.))$, where $h(.,.)$ has entries in F and A_s and B_s are finite sets, we implicitly assume that the value is taken with respect to the Puiseux mixed actions as mentioned above. f will denote such an action for player 1 and g for player 2.

Bewley & Kohlberg (1976a) proved the following lemma .

7.2.2. LEMMA. *Equations* (7.2.1) *have a unique solution* $x^* \in F^Z$, *where for each* $s \in S$, $\Phi(x_s^*) \leq 1$. *Furthermore* $\phi_\tau(x^*) := (\phi_\tau(x_1^*), \ldots, \phi_\tau(x_Z^*))$ *is the value of the discounted game with interest rate* $1/\tau$ *for all sufficiently large* τ. *If for* $s \in S$ f_s *is an optimal action for player* 1 *in the matrix game* $[G_{s\theta}(x^*)]$, *then* $\rho_\tau = (\phi_\tau(f_1), \ldots, \phi_\tau(f_Z))$ *is an optimal stationary strategy for player* 1 *in the discounted game with interest rate* $1/\tau$, *for all sufficiently large* τ. *Similar results hold for player* 2.

For a stochastic game let $FV(\tau) \in \mathbb{R}^Z$, with $\tau \in \mathbb{N}$, be the value of the finite horizon game with $D_\tau = \{0,1,\ldots,\tau-1\}$ as set of decision epochs. As Shapley (1953) already remarked, we have

$$(7.2.3) \qquad FV_s(\tau+1) = \underset{A_s \times B_s}{\text{Val}}\,\Big(r(s,.,.) + \sum_{t=1}^{z} p(t|s,.,.).FV_t(\tau)\Big)$$

The matrix game in the right hand side of (7.2.3) will be abbreviated to $[G_s(FV(\tau))]$.

The following result of Bewley & Kohlberg (1976a) is essential in the theory of stochastic games.

7.2.3. LEMMA. $\lim_{\tau\to\infty} \tau^{-1}\phi_\tau(x_s^*)$ *and* $\lim_{\tau\to\infty} \tau^{-1} FV_s(\tau)$ *exist and are equal for all* $s \in S$.

Clearly the limit in the above lemma is an obvious candidate to be the value of the average reward game. As already mentioned, Mertens & Neyman (1981) proved that indeed this limit equals the value of the average reward stochastic game, while Monash (1979) showed this fact for a weaker definition of the average reward.

7.2.4. LEMMA. *Let* $g_s=\lim_{\tau\to\infty} \tau^{-1}FV_s(\tau)=\lim_{\tau\to\infty} \tau^{-1}\phi_\tau(x_s^*)$. *Then* $g=(g_1,\ldots,g_z)$ *equals the value of the average reward stochastic game.*

Using the lemma's 7.2.2 and 7.2.3 Bewley and Kohlberg were able to prove nearly all the earlier results for stochastic games. The following lemma stands central in their proofs (Bewley & Kohlberg (1978)).

7.2.5. LEMMA. *Let* $x^*\in F^z$ *be the solution of the limit discount equation. If players* 1 *and* 2 *both have real actions* $\rho_s\in P(A_s)$ *and* $\sigma_s\in P(B_s)$ *respectively, which guarantee them* $\mathrm{Val}(G_{s\theta}(x^*))+o(\theta^0)$ *in the matrix games* $[G_{s\theta}(x^*)]$, *for all* $s\in S$, *then the stationary strategies* $\rho=(\rho_1,\rho_2,\ldots,\rho_z)$ *and* $\sigma=(\sigma_1,\sigma_2,\ldots,\sigma_z)$ *are optimal in the undiscounted stochastic game.*

In part III x^* or $x^*(\theta)$ will always denote the solution of the limit discount equation.

8. Structural properties of undiscounted stochastic games.

8.1. STOCHASTIC GAMES AND OPTIMAL STATIONARY STRATEGIES.

In this section we characterize the class of stochastic games where both players have optimal stationary strategies for the average reward criterion. Most of the results have been derived from Vrieze (1979a, 1981b).

First we define the Cesaro limit of a stochastic matrix. For a two-person zerosum stochastic game with finite state and action spaces, let ρ and σ be stationary strategies for the respective players and let $P(\rho,\sigma)$ be the corresponding stochastic matrix. Then the Cesaro limit $Q(\rho,\sigma)$ of $P(\rho,\sigma)$ is defined as

$$(8.1.1) \qquad Q(\rho,\sigma) := \lim_{k\to\infty} \frac{1}{k+1} \sum_{\tau=0}^{k} P^{\tau}(\rho,\sigma),$$

where $P^{\circ}(\rho,\sigma)=I_{zz}$ and $P^{\tau}(\rho,\sigma)=P(\rho,\sigma)(P^{\tau-1}(\rho,\sigma))$ for $\tau\geq 1$.

It is well-known (e.g. Kemeny & Snell (1961)) that $Q(\rho,\sigma)$ exists for each pair (ρ,σ), and that $Q(\rho,\sigma)$ has the following properties:

$$(8.1.2) \qquad Q(\rho,\sigma).P(\rho,\sigma) = P(\rho,\sigma)Q(\rho,\sigma) = Q(\rho,\sigma).Q(\rho,\sigma) = Q(\rho,\sigma).$$

Observe that the (s,t)-th element of $Q(\rho,\sigma)$ equals the mean number of times that the system is in state t when state s is the starting state and the players choose ρ and σ as their strategies respectively.

Furthermore, by a suitable renumbering of the states, $P(\rho,\sigma)$ can be written as:

$$(8.1.3) \qquad P(\rho,\sigma) = \left(\begin{array}{ccc|c} P_{11}(\rho,\sigma) & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & & P_{\gamma\gamma}(\rho,\sigma) & 0 \\ \hline P_{\gamma+1\ 1}(\rho,\sigma) & \cdots & P_{\gamma+1\ \gamma}(\rho,\sigma) & P_{\gamma+1\ \gamma+1}(\rho,\sigma) \end{array}\right)$$

In (8.1.3) $P_{nn}(\rho,\sigma)$, $n\in\{1,\ldots,\gamma\}$, is a square matrix corresponding to the n-th ergodic class of the Markov chain associated with $P(\rho,\sigma)$.

The rows below the dotted line correspond to the transient states, and $(I_{\gamma+1\ \gamma+1}-P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}$ exists and equals $\sum_{\tau=0}^{\infty} P^{\tau}_{\gamma+1\ \gamma+1}(\rho,\sigma)$, hence this matrix has non-negative elements. Then $Q(\rho,\sigma)$ has the form:

$$(8.1.4)\qquad Q(\rho,\sigma) = \left(\begin{array}{ccc|c} Q_{11}(\rho,\sigma) & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & Q_{\gamma\gamma}(\rho,\sigma) & 0 \\ \hline Q_{\gamma+1\ 1}(\rho,\sigma) & \cdots & Q_{\gamma+1\ \gamma}(\rho,\sigma) & 0 \end{array}\right)$$

where $Q_{nn}(\rho,\sigma)$, $n\in\{1,2,\ldots,\gamma\}$, is a square matrix corresponding to the n^{th} ergodic class of the Markov chain associated with $P(\rho,\sigma)$. Each row of $Q_{nn}(\rho,\sigma)$ is identical and equals the unique invariant distribution of the Markov subchain corresponding to it. Further each element of $Q_{nn}(\rho,\sigma)$ is strictly positive. The rows below the dotted line correspond to the transient states. The matrices $Q_{\gamma+1\ n}(\rho,\sigma)$, $n\in\{1,\ldots,\gamma\}$, reflect the probabilities with which the system vanishes from the transient states into the different states of ergodic class n.

It is well-known (cf. Kemeny & Snell (1961)) that

$$(8.1.5)\qquad Q_{\gamma+1\ n}(\rho,\sigma) = (I_{\gamma+1\ \gamma+1}-P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}P_{\gamma+1\ n}(\rho,\sigma).Q_{nn}(\rho,\sigma).$$

Observe that the (s,t)-th element of the matrix $(I_{\gamma+1\ \gamma+1}-P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}.P_{\gamma+1\ n}(\rho,\sigma)$ equals the probability that, starting in the transient state s, the system ever reaches ergodic class n with state t as entry state.

As introduced in section 2.3, we denote the average reward for a pair of strategies (μ,ν) by $W_{\mu\nu}$ (a z-vector). Clearly, for a pair of stationary strategies (ρ,σ) we have:

$$(8.1.6)\qquad W_{\rho\sigma} = \lim_{k\to\infty}\inf \frac{1}{k+1}\sum_{\tau=0}^{k} P^{\tau}(\rho,\sigma).r(\rho,\sigma) = Q(\rho,\sigma).r(\rho,\sigma).$$

Observe by 8.1.4. that $W_{\rho\sigma}$ is constant on each ergodic set of $P(\rho,\sigma)$.

8.1.1. LEMMA. *If* $g\in\mathbb{R}^z$, $w\in\mathbb{R}^z$ *and a pair of stationary strategies are such that*:

$$(8.1.7)\qquad P(\rho,\sigma).g \geq g$$

and for each state s *recurrent with respect to* $P(\rho,\sigma)$:

$$(8.1.8)\qquad r(s,\rho_s,\sigma_s) + \sum_{t=1}^{z} p(t|s,\rho_s,\sigma_s)w_t \geq w_s+g_s$$

then

(a) In (8.1.7) *the equality sign holds in the components corresponding to the recurrent states of* $P(\rho,\sigma)$ *and* g *is constant on each ergodic set.*

(b) $W_{\rho\sigma} \geq g$ *and* $W_{\rho\sigma} = g$ *if and only if* $P(\rho,\sigma).g=g$ *and in* (8.1.8) *for each recurrent state* s *the equality sign holds.*

The lemma remains true when all inequality signs are reversed.

PROOF. (a) Assume that $P(\rho,\sigma)$ has the form (8.1.3). Then, with respect to the n-th ergodic class, (8.1.7) can be written as:

$$(8.1.9)\qquad P_{nn}(\rho,\sigma).g(n) \geq g(n),$$

where g(n) equals the part of g corresponding to the states of the n-th ergodic class.
Now suppose that in (8.1.9) the inequality sign holds in at least one component. Then after multiplying (8.1.9) by $Q_{nn}(\rho,\sigma)$ we would obtain $Q_{nn}(\rho,\sigma).g(n)>Q_{nn}(\rho,\sigma).g(n)$, which obviously is a contradiction. (The inequality mentioned before is a consequence of the following arguments: $Q_{nn}(\rho,\sigma).P_{nn}(\rho,\sigma)=Q_{nn}(\rho,\sigma)$, each element of $Q_{nn}(\rho,\sigma)$ is strictly positive and $Q_{nn}(\rho,\sigma)$ has identical rows.). Hence in (8.1.9) the equality sign holds, which proves the first part of (a).

From $P_{nn}(\rho,\sigma).g(n)=g(n)$ we deduce $P^2_{nn}(\rho,\sigma).g(n)=P_{nn}(\rho,\sigma).g(n)=g(n)$, and next $P^{\tau}_{nn}(\rho,\sigma).g(n)=g(n)$ for each $\tau\in\mathbb{N}$. But then also $Q_{nn}(\rho,\sigma).g(n)=g(n)$ (see (8.1.1)). Since $Q_{nn}(\rho,\sigma)$ has rowsum 1 and only strictly positive elements, it follows from this last equation that g(n) has identical components.

(b) Using part (a) of this lemma we obtain after multiplying (8.1.8) with respect to the n-th ergodic class, by $Q_{nn}(\rho,\sigma)$:

$$(8.1.10)\qquad W_{\rho\sigma}(n) := Q_{nn}(\rho,\sigma).r_n(\rho,\sigma) \geq Q_{nn}(\rho,\sigma).g(n) = g(n).$$

Here $r_n(\rho,\sigma)$ is the part of $r(\rho,\sigma)$ corresponding to the n-th ergodic class. For the transient states we have (cf. (8.1.5)):

$$(8.1.11)\qquad W_{\rho\sigma}(\gamma+1) := \sum_{n=1}^{\gamma} Q_{\gamma+1\ n}(\rho,\sigma).r_n(\rho,\sigma)$$

$$= \sum_{n=1}^{\gamma} (I_{\gamma+1\ \gamma+1} - P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}.P_{\gamma+1\ n}(\rho,\sigma).Q_{nn}(\rho,\sigma).r_n(\rho,\sigma)$$

$$\geq (I_{\gamma+1\ \gamma+1} - P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}.\sum_{n=1}^{\gamma} P_{\gamma+1\ n}(\rho,\sigma).g(n)$$

$$\geq (I_{\gamma+1\ \gamma+1} - P_{\gamma+1\ \gamma+1}(\rho,\sigma))^{-1}.(I_{\gamma+1\ \gamma+1} - P_{\gamma+1\ \gamma+1}(\rho,\sigma)).g(\gamma+1)$$

$$= g(\gamma+1).$$

In the last step of (8.1.11) we have used the fact that with respect to the transient states (8.1.7) can be written as:

$$\sum_{n=1}^{\gamma} P_{\gamma+1\ n}(\rho,\sigma).g(n) \geq (I_{\gamma+1\ \gamma+1} - P_{\gamma+1\ \gamma+1}(\rho,\sigma)).g(\gamma+1).$$

Now (8.1.10) and (8.1.11) show the first part of (b).

Concerning the second part of (b), suppose $W_{\rho\sigma}=g$. Now, if in (8.1.8) the inequality sign holds for some state s belonging to the n-th ergodic class then this is also the case in (8.1.10), which is a contradiction. Furthermore, from $W_{\rho\sigma}=Q(\rho,\sigma).r(\rho,\sigma)$, $P(\rho,\sigma).Q(\rho,\sigma)=Q(\rho,\sigma)$ and $W_{\rho\sigma}=g$ it follows that $P(\rho,\sigma).g=g$. On the other hand, suppose that $P(\rho,\sigma).g=g$ and that in (8.1.8) the equality sign holds for each recurrent state s. Then one can verify that in (8.1.10) and (8.1.11) the equality signs also hold. This proves part (b) of the lemma.

□

8.1.2. COROLLARY. *If* $g\in\mathbb{R}^z$ *and* $w\in\mathbb{R}^z$ *are such that*

$$(8.1.12)\qquad g_s \leq \operatorname*{Val}_{A_s\times B_s}\Big(\sum_{t=1}^{z} p(t|s,.,.)g_t\Big)$$

and

$$(8.1.13)\qquad w_s+g_s \leq \operatorname*{Val}_{E_{1s}\times B_s}\Big(r(s,.,.)+\sum_{t=1}^{z} p(t|s,.,.)w_t\Big)$$

then for some stationary strategy $\tilde{\rho}$ *of player* 1 *we have* $\inf_{\nu} W_{\tilde{\rho}\nu} \geq g$. *(Here* E_{1s} *is the finite set of extreme optimal actions of player* 1 *for the matrix game mentioned in* (8.1.12).)

PROOF. Let $\tilde{\rho}_s$ be optimal for player 1 in the matrix game $[r(s,.,.)+\sum_{t=1}^{z} p(t|s,.,.)w_t]$ on $E_{1s}\times B_s$. Let $\tilde{\rho}=(\tilde{\rho}_1,\ldots,\tilde{\rho}_z)$ and let σ^p be an arbitrary pure stationary strategy for player 2. Then lemma 8.1.1. can be applied to g, w and $(\tilde{\rho},\sigma^p)$ yielding $W_{\tilde{\rho}\sigma^p}\geq g$. Now corollary 3.5 proves the corollary. □

If $x^*(\theta) = \sum_{k=-\infty}^{M} c_k\theta^{k/M}$ is the solution of the limit discount equation, with $c_k=(c_{k1},c_{k2},\ldots,c_{kz})$, then in section 7.2 we have seen that c_M equals the value of the average reward stochastic game. In the sequel this value will be denoted by g and the expression $\sum_{k=-\infty}^{M} c_k\theta^{k/M}$ will always denote the solution of the limit discount equation.

$[G_s(v)]$ with $v\in\mathbb{R}^z$ denotes the matrix game $[r(s,.,.)+\sum_{t=1}^{z} p(t|s,.,.)v_t]$.

The following lemma is well-known (cf. Federgruen (1978)). We give a proof without using the Puiseux series expansion of the solution of the limit discount equation.

8.1.3. LEMMA. *For a two-person zerosum stochastic game with finite state and action spaces the following two assertions are equivalent:*

(i) $g=g^*.1_z$, *with* $g^*\in\mathbb{R}$, *and both players have optimal stationary strategies.*

(ii) there exist $w\in\mathbb{R}^z$ *and* $g^*\in\mathbb{R}$, *such that*

$$w_s+g^* = \operatorname{Val}_{A_s\times B_s}(G_s(w)), \quad s\in S.$$

PROOF. Suppose that (i) is true. Let $\tilde{\rho}$ be optimal for player 1. Then by the theorems 3.4 and A.2.6 it follows that there exists a $w(1)\in\mathbb{R}^z$ such that for each $s\in S$: $g^*+w_s(1)=\min_j \{r(s,\tilde{\rho}_s,j)+\sum_{t=1}^{z} p(t|s,\tilde{\rho}_s,j)w_t(1)\}$. But then

$$g^*+w_s(1) \leq \operatorname{Val}_{A_s\times B_s}(G_s(w(1))). \tag{8.1.14}$$

Using an optimal stationary strategy for player 2 gives analogously the existence of a vector $w(2)\in\mathbb{R}^z$ such that

$$g^*+w_s(2) \geq \operatorname{Val}_{A_s\times B_s}(G_s(w(2))). \tag{8.1.15}$$

Since also $w(1)+\delta.1_z$, with $\delta\in\mathbb{R}$, satisfies (8.1.14) we may assume $w(1)\leq w(2)$. Then for each $s\in S$:

(8.1.16) $\quad g^*+w_s(1) \le \mathrm{Val}(G_s(w(1))) \le \mathrm{Val}(G_s(w(2))) \le g^*+w_s(2).$

Since the Val-operator is monotonous and Lipschitz continuous with constant 1 (cf. Lemma A.1.8) the inequalities (8.1.16) guarantee the existence of a vector w with $w(1) \le w \le w(2)$ such that (ii) holds.

Now suppose that (ii) holds. Trivially $g^* = \mathrm{Val}_{A_s \times B_s} (\sum_{t=1}^{z} p(t|s,.,.)g^*)$ and all actions are optimal in this matrix game. Then we may apply corollary 8.1.2 two times, giving the existence of stationary strategies $\tilde{\rho}$ and $\tilde{\sigma}$ such that

$$\sup_{\mu} W_{\mu\tilde{\sigma}} \le g^*.1_z \le \inf_{\nu} W_{\tilde{\rho}\nu}.$$

Hence theorem 2.3.4 shows that (i) holds.

□

We now prove the following lemma.

8.1.4. LEMMA.

$$g_s = \mathrm{Val}_{A_s \times B_s} (\sum_{t=1}^{z} p(t|s,.,.)g_t), \quad s \in S.$$

PROOF. From the solution of the limit discount equation we derive:

$$(8.1.17) \qquad \sum_{k=-\infty}^{M} c_{ks}\tau^{k/M} = \mathrm{Val}\,(r(s,.,.)+(1+\tau^{-1})^{-1} \sum_{t=1}^{z} p(t|s,.,.). \sum_{k=-\infty}^{M} c_{kt}\tau^{k/M}),$$

for each sufficiently large τ and each $s \in S$.
Dividing both sides of (8.1.17) by τ, letting τ tend to infinity and using the fact that the Val-operator is continuous, we obtain the assertion of the lemma.

□

In the following $O_{\ell s}$, $\ell=1,2$, will denote the set of optimal actions for player ℓ in the matrix game $[\sum_{t=1}^{z} p(t|s,.,.).g_t]$ on $A_s \times B_s$, and $O_\ell := \mathop{X}_{s=1}^{z} O_{\ell s}$. It is well-known that $O_{\ell s}$ is the convex hull of a finite set (cf. theorem A.1.9). $E_{\ell s}$ will denote this finite set of extreme optimal actions and $E_\ell := \mathop{X}_{s \in S} E_{\ell s}$, $\ell \in \{1,2\}$.

8.1.5. LEMMA. *If* ρ^* *is an optimal stationary strategy for player* 1, *then* $\rho_s^* \in O_{1s}$, $s \in S$. *Similarly for player* 2.

PROOF. By assumption $\inf_\nu W_{\rho^*\nu} = g$, then by the theorems 3.4 and A.2.6 we obtain:

$$g_s = \min_{j \in B_s} \{ \sum_{t=1}^{z} p(t|s,\rho_s^*,j).g_t \},$$

which shows that $\rho_s^* \in O_{1s}$. □

With a stochastic game Γ we associate two stochastic games, called $\Gamma(1)$ and $\Gamma(2)$, which are defined by:

(a) for $\ell \in \{1,2\}$, in $\Gamma(\ell)$ the set of states and the set of pure actions for player $3-\ell$ remain as in Γ, and the set of pure actions for player ℓ in state s equals $E_{\ell s}$.

(b1) for $\Gamma(1)$ the immediate payoffs and the transitions are defined for each $(\rho_s, j) \in E_{1s} \times B_s$ as:

$$r_1(s,\rho_s,j) := \sum_{i \in A_s} r(s,i,j)\rho_s(i) - g_s$$

and

$$p_1(t|s,\rho_s,j) := \sum_{i \in A_s} p(t|s,i,j).\rho_s(i).$$

(b2) for $\Gamma(2)$ the immediate payoffs and the transitions are defined for each $(i,\sigma_s) \in A_s \times E_{2s}$ as:

$$r_2(s,i,\sigma_s) := \sum_{j \in B_s} r(s,i,j)\sigma_s(j) - g_s$$

and

$$p_2(t|s,i,\sigma_s) := \sum_{j \in B_s} p(t|s,i,j).\sigma_s(j).$$

8.1.6. THEOREM. *If for a stochastic game* Γ *both players have optimal stationary strategies with respect to the average reward criterion, then*

(a) $\Gamma(1)$ *and* $\Gamma(2)$ *both have average reward value* O_z.

(b) the sets of optimal stationary strategies for player ℓ *coincide for* Γ *and* $\Gamma(\ell)$, $\ell \in \{1,2\}$.

PROOF. We shall prove only the assertions concerning $\Gamma(1)$, since those concerning $\Gamma(2)$ can be shown in an analogous way.

Suppose that ρ^* and σ^* are optimal for players 1 and 2 respectively in the game Γ. Then by corollary 3.5, the fact that g is the value of the game Γ, and lemma 8.1.5, we have:

$$(8.1.17)\qquad \min_{\sigma\in SST_2} Q(\rho^*,\sigma).r(\rho^*,\sigma)=g=\max_{\rho\in SST_1} Q(\rho,\sigma^*).r(\rho,\sigma^*)=\max_{\rho\in\mathcal{O}_1} Q(\rho,\sigma^*).r(\rho,\sigma^*)$$

A two-fold application of theorem A.2.8 on (8.1.17) gives:

$$(8.1.18)\qquad \min_{\sigma\in SST_2} Q(\rho^*,\sigma)(r(\rho^*,\sigma)-g)=0_z=\max_{\rho\in\mathcal{O}_1} Q(\rho,\sigma^*)(r(\rho,\sigma^*)-g).$$

Since $\rho^*\in\mathcal{O}_1$ (lemma 8.1.5), part (a) of the theorem follows from (8.1.18) as a consequence of corollary 3.5. Furthermore we see from (8.1.18) that ρ^* and σ^* are optimal in $\Gamma(1)$, so half of part (b) is proved.

Now suppose that $\tilde{\rho}$ is optimal for player 1 in $\Gamma(1)$. Then for each stationary strategy σ:

$$Q(\tilde{\rho},\sigma).(r(\tilde{\rho},\sigma)-g) \geq 0_z, \text{ or}$$

$$(8.1.19)\qquad Q(\tilde{\rho},\sigma).r(\tilde{\rho},\sigma) \geq Q(\tilde{\rho},\sigma).g.$$

Since $\tilde{\rho}_s\in\mathcal{O}_{1s}$, we have $P(\tilde{\rho},\sigma).g\geq g$ and then with induction $P^\tau(\tilde{\rho},\sigma).g\geq g$ for all τ which implies $Q(\tilde{\rho},\sigma).g\geq g$. Insertion of this result in (8.1.19) shows by the fact that σ is arbitrarily chosen and corollary 3.5, that $\tilde{\rho}$ is also optimal in Γ.

□

It is obvious that player 2 may have optimal stationary strategies in $\Gamma(1)$ which are not optimal in Γ since such strategies may be beaten by a strategy of player 1 outside the set $\mathcal{O}_1$.

The following theorem is a generalization of theorem 8.1.6, part (a).

8.1.7. THEOREM. *Suppose that for the undiscounted stochastic game Γ both players have optimal stationary strategies. Then for each z pairs of finite sets $(\tilde{A}_s,\tilde{B}_s)\subset P(A_s)\times P(B_s)$, $s\in\{1,\ldots,z\}$, such that $P(\tilde{A}_s)\supset\mathcal{O}_{1s}$ and $P(\tilde{B}_s)\supset\mathcal{O}_{2s}$, there exists a $v\in\mathbb{R}^z$ such that*

$$v_s+g_s = \operatorname{Val}_{\tilde{A}_s\times\tilde{B}_s}\Big(r(s,.,.)+\sum_{t=1}^{z} p(t|s,.,.).v_t\Big), \qquad s\in S.$$

PROOF. Define the game $\Gamma(3)$, where the sets of pure actions in state s for the players are $\tilde{A}_s$ and $\tilde{B}_s$ respectively, and where

$$r_3(s,\rho_s,\sigma_s) := \sum_{i\in A_s} \sum_{j\in B_s} r(s,i,j).\rho_s(i).\sigma_s(j) - g_s$$

and

$$p_3(t|s,\rho_s,\sigma_s) := \sum_{i\in A_s} \sum_{j\in B_s} p(t|s,i,j).\rho_s(i).\sigma_s(j)$$

for each $(\rho_s,\sigma_s)\in\tilde{A}_s\times\tilde{B}_s$.
Then, analogous to the first part of the proof of theorem 8.1.6 we derive that the average reward value of $\Gamma(3)$ equals 0_z and thus the theorem is a consequence of lemma 8.1.3.

□

Theorem 7.3.3 part (a) of Federgruen (1978) is a special case (namely $\tilde{A}_s=E_{1s}$, $\tilde{B}_s=E_{2s}$) of theorem 8.1.7. Federgruen gave a counter example (page 174), showing that the existence of a v satisfying

$$v_s+g_s = \operatorname*{Val}_{E_{1s}\times E_{2s}} (G_s(v)), \quad s\in S,$$

is not sufficient for the existence of optimal stationary strategies.

In the next theorem we give a necessary and sufficient condition for the existence of optimal stationary strategies.

8.1.8. THEOREM. *For a two-person zerosum undiscounted stochastic game with finite state and action spaces, the following two assertions are equivalent.*

(i) The value of the game is g *and both players have optimal stationary strategies.*

(ii) (a)

$$g_s = \operatorname*{Val}_{A_s\times B_s} \Big(\sum_{t=1}^{z} p(t|s,.,.)g_t\Big), \quad s\in S \tag{8.1.20}$$

(b) There exist vectors $v_1\in\mathbb{R}^z$ *and* $v_2\in\mathbb{R}^z$ *such that*

$$v_{1s}+g_s = \operatorname*{Val}_{E_{1s}\times B_s} (G_s(v_1)), \quad s\in S \tag{8.1.21}$$

and

$$(8.1.22)\qquad v_{2s}+g_s = \operatorname*{Val}_{A_s\times E_{2s}}\left(G_s(v_2)\right), \quad s\in S.$$

PROOF. Suppose that (i) is true. Part (a) of (ii) is generally valid and has been proved in lemma 8.1.4.

Consider the game $\Gamma(1)$ associated with Γ, as defined above. Then from theorem 8.1.6 part (a) and lemma 8.1.3 the existence of a v_1 satisfying (8.1.21) is obvious. Similarly via $\Gamma(2)$ the existence of a v_2 satisfying (8.1.22) follows.

Now suppose that (ii) is true.

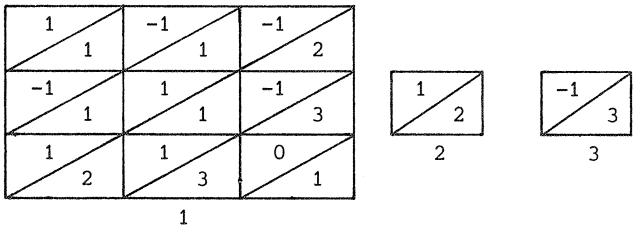
Then, by corollary 8.1.2 applied to (8.1.20) and (8.1.21) and by the player 2 version of corollary 8.1.2 applied to (8.1.20) and (8.1.22), we have for some stationary strategies $\tilde{\rho}$ and $\tilde{\sigma}$:

$$\inf_{\nu} W_{\tilde{\rho}\nu} \geq g \geq \sup_{\mu} W_{\mu\tilde{\sigma}}.$$

Finally by theorem 2.3.4 we obtain the derived result.

□

If both players possess optimal stationary strategies, it is generally not possible to choose $v_1=v_2$ in theorem 8.1.8, as the following example shows.



Here $g=(0,1,-1)$, $O_{11}=E_{11}=(\frac{1}{2},\frac{1}{2},0)$, $O_{21}=E_{21}=(\frac{1}{2},\frac{1}{2},0)$ and for both players the respective strategies ρ^* and σ^* are the unique optimal strategies if $\rho^*_1=\sigma^*_1=(\frac{1}{2},\frac{1}{2},0)$.

Let $v=(v_1,v_2,v_3)$ satisfy both (8.1.21) and (8.1.22). Then from (8.1.21) we deduce $v_1\leq\frac{1}{2}(v_2+v_3)-1$, while from (8.1.22) we get $v_1\geq\frac{1}{2}(v_2+v_3)+1$, which yields contradiction.

In Markov decision problems, once one has a solution to the functional equations (cf. theorem A.2.6), there is a clear rule whether a stationary strategy is optimal or not. In stochastic games, this problem is much more complicated. The next theorem gives a characterization of the optimality of a fixed stationary strategy.
The example thereafter shows that the results for the MDP cannot be extended in a straightforward manner to stochastic games, and furthermore it can be seen that theorem 8.1.10 can hardly be strengthened.

8.1.9. REMARK. Along the same lines as the proof of theorem 8.1.8 we can show the following:

Knowing that the value of a stochastic game equals g, *the existence of a vector* v_1 *such that*

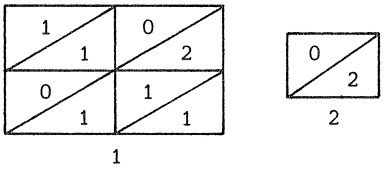
$$(8.1.23) \qquad v_{1s}+g_s \leq \operatorname*{Val}_{E_{1s}\times B_s}(G_s(v_1)), \quad s\in S$$

is a necessary and sufficient condition for player 1 *to possess an optimal stationary strategy. Also the existence of a vector* v_2 *such that*

$$v_{2s}+g_s \geq \operatorname*{Val}_{A_s\times E_{2s}}(G_s(v_2)), \quad s\in S,$$

is a necessary and sufficient condition for player 2 *to possess an optimal stationary strategy.*

The following example shows that the inequality signs in the above equations cannot generally be replaced by equality signs.



Clearly g=(0,0), $E_{11}=A_1$, each strategy of player 1 is optimal and player 2 has no optimal stationary strategy. However for $v=(v_1,v_2)$ we have

$$\delta := \operatorname{Val}_{E_{11}\times B_s}(G_1(v)) = \operatorname{Val}\begin{bmatrix} 1+v_1 & v_2 \\ v_1 & 1+v_1 \end{bmatrix}$$

If $v_2 \geq 1+v_1$, then $\delta = 1+v_1 > v_1 = v_1+g_1$.

If $v_2 < 1+v_1$, then $\delta = \dfrac{(1+v_1)^2 - v_1 v_2}{v_1+2-v_2} > v_1 = v_1+g_1$.

Hence for each v satisfying (8.1.23) the inequality sign holds for state 1.

For a stationary strategy ρ of player 1, we define $R(\rho) := \{s \mid s \in S$, s is recurrent with respect to $P(\rho,\sigma)$ for some optimal reply σ of player 2 against ρ in the stochastic game$\}$.

8.1.10. THEOREM. *For a stochastic game with value* g, *a stationary strategy* $\tilde{\rho}$ *for player* 1 *is optimal if and only if there exists a* $v \in \mathbb{R}^z$, *such that:*

(a) $\tilde{\rho}_s \in \mathcal{O}_{1s}$, $\quad s \in S$

(b) $v_s+g_s \leq \operatorname{Val}_{E_{1s}\times B_s}(G_s(v))$,

such that for each state $s \in R(\tilde{\rho})$, *the action* $\tilde{\rho}_s$ *guarantees player* 1 v_s+g_s *in the matrix game* $[G_s(v)]$ *on* $E_{1s}\times B_s$.

PROOF. Let $\tilde{\rho}$ be optimal. Part (a) is proved in lemma 8.1.5.
By looking at the Markov decision situation $MDS(\tilde{\rho})$ corresponding to fixing $\tilde{\rho}$, we may conclude from theorem 3.4 and theorem A.2.6 that there exists a $v \in \mathbb{R}^z$, such that

$$v_s+g_s = \min_{j \in B_s} \{r(s,\tilde{\rho}_s,j) + \sum_{t=1}^{z} p(t \mid s,\tilde{\rho}_s,j).v_t\}, \text{ each } s \in S,$$

which implies part (b).

Now suppose that (a) and (b) of the theorem hold. Let $\tilde{\sigma}$ be an optimal reply to $\tilde{\rho}$. Then we have the following inequalities:

(8.1.24) $\quad g \leq P(\tilde{\rho},\tilde{\sigma}).g$

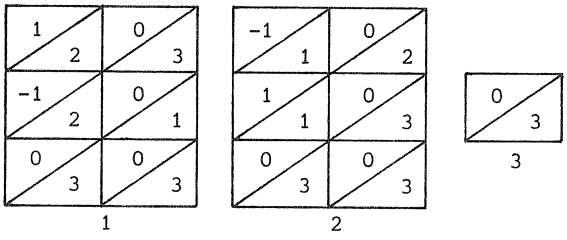
and

(8.1.25) $\quad v_s+g_s \leq r(s,\tilde{\rho}_s,\tilde{\sigma}_s)+\sum_{t=1}^{z} p(t|s,\tilde{\rho}_s,\tilde{\sigma}_s).v_t$

for each state s which is recurrent with respect to $P(\tilde{\rho},\tilde{\sigma})$. Inequalities (8.1.24) and (8.1.25) imply by lemma 8.1.1 that $g \leq Q(\tilde{\rho},\tilde{\sigma}).r(\tilde{\rho},\tilde{\sigma})$. Since $\tilde{\sigma}$ was assumed to be an optimal reply to $\tilde{\rho}$, this shows that $\tilde{\rho}$ is optimal for player 1 in the stochastic game.

□

8.1.11. EXAMPLE.



The following facts can be derived.

(a) the value equals (0,0,0); both players have optimal stationary strategies.

(b) g=(0,0,0) and v=(d,d,d) with $d \in \mathbb{R}$ are the only solutions to the functional equations

$$g_s = \mathrm{Val}\left(\sum_{t=1}^{z} p(t|s,.,.)g_t\right), \quad s \in S$$

and

$$v_s+g_s = \mathrm{Val}_{E_{1s} \times B_s}(G_s(v)), \quad s \in S$$

(Here $E_{1s}=A_s$, $s \in S$).

(c) For the optimal stationary strategy $\rho^*=(\rho_1^*,\rho_2^*,\rho_3^*)$, with $\rho_1^*=(1,0,0)$ and $\rho_2^*=(1,0,0)$, it holds that ρ_2^* is not optimal in the matrix game $[G_s(v)]$ for v=(d,d,d).

(d) For ρ^* as in (c), the pair g=(0,0,0) and v=(d,d-1,d) has the properties of theorem 8.1.10.

(e) There exists no pair g=(0,0,0) and $v \in \mathbb{R}^3$ such that, for each optimal stationary strategy the properties mentioned in theorem 8.1.10 hold. (The optimal stationary strategy ρ^* as in (c) asks for $v_1 \geq v_2+1$,

while the optimal stationary strategy $\tilde{\rho}$ with $\tilde{\rho}_1=(0,1,0)$ and $\tilde{\rho}_2=(0,1,0)$ asks for $v_2 \geq v_1+1$.

If for each pair of pure stationary strategies (ρ,σ), all states are recurrent with respect to $P(\rho,\sigma)$, then the sets of optimal stationary strategies can be fully characterized. We have

8.1.12. THEOREM. *If for a two-person zerosum stochastic game, for each pair of stationary strategies (ρ,σ) all states are recurrent with respect to $P(\rho,\sigma)$, then a stationary strategy $\tilde{\rho}$ is optimal for player 1 if and only if*

(a) $\tilde{\rho}_s \in \mathcal{O}_{1s}$, *all* $s \in S$.

(b) for each $v \in \mathbb{R}^Z$, *such that*

$$v_s+g_s = \operatorname{Val}_{E_{1s} \times B_s} (G_s(v)), \quad s \in S$$

the action $\tilde{\rho}_s$ *is optimal in the matrix game* $[G_s(v)]$.

PROOF. The "sufficiency" follows at once from the recurrency assumption and theorem 8.1.10.

The "necessity" can be shown as follows.
If $\tilde{\rho}_s \notin \mathcal{O}_{1s}$ for some s, then $\tilde{\rho}$ cannot be optimal by lemma 8.1.5. So we may assume $\tilde{\rho}_s \in \mathcal{O}_{1s}$, each $s \in S$. Suppose for some $\tilde{s} \in S$ that $\tilde{\rho}_{\tilde{s}}$ is not optimal in the matrix game $[G_{\tilde{s}}(v)]$ on $E_{1\tilde{s}} \times B_{\tilde{s}}$. Then there exists a $\tilde{\sigma}$ such that

$$v+g \geq r(\tilde{\rho},\tilde{\sigma})+P(\tilde{\rho},\tilde{\sigma}).v,$$

where the inequality sign holds at least in component $\tilde{s}$.
Since each state is recurrent with respect to $P(\tilde{\rho},\tilde{\sigma})$, lemma 8.1.1(b) now shows that $W_{\tilde{\rho}\tilde{\sigma}} < g$. Hence $\tilde{\rho}$ cannot be optimal.

□

8.2. THE ASYMPTOTIC BEHAVIOUR OF $\|FV(\tau)-\tau g\|$.

In this section we shall study the asymptotic behaviour of $\|FV(\tau)-\tau g\|$. Here $FV(\tau)$ equals the value of the τ-step game and g the value of the average reward game. Firstly we shall characterize the asymptotic behaviour $\|FV(\tau)-\tau g\| \leq B \log \tau$ for some $B \in \mathbb{R}$ and secondly we shall analyse the asymptotic behaviour $\|FV(\tau)-\tau g\| \leq C$ for some $C \in \mathbb{R}$.

Bewley & Kohlberg (1976b) have shown that, in general, a vector funtion $y(\tau) = \sum_{k=1}^{K} d_k \tau^{k/K}$, with $K \in \mathbb{N}$, can be constructed such that $\|FV(\tau)-y(\tau)\| \leq D \log \tau$ for some $D \in \mathbb{R}$. Furthermore they give an example which shows that, in general, an expansion in fractional powers of τ, which is more precise than the above $y(\tau)$, need not exist.

Most of the results of this section can also be found in Vrieze (1979b). As before, $x^*(\theta) = \sum_{k=-\infty}^{M} c_k \theta^{k/M}$ denotes the solution of the limit discount equation and $g := c_M$ equals the value of the average reward game.

8.2.1. THEOREM. *For a stochastic game the following assertions are equivalent:*

(i) $\|FV(\tau)-\tau g\| \leq B \log \tau$ *for each* $\tau \in \mathbb{N}$ *and some* $B \in \mathbb{R}$.

(ii) $c_{M-1}=c_{M-2}=\ldots=c_1=0_Z$.

(iii) there exists a $v \in \mathbb{R}^Z$, *such that*

$$\operatorname{Val}_{A_s \times B_s} (G_s(g\theta+v)) = g_s(\theta+1)+v_s+O(\theta^{-1}), \quad s \in S.$$

(iv) there exists a $v \in \mathbb{R}^Z$, *such that*

$$\operatorname{Val}_{E_{1s} \times E_{2s}} (G_s(v)) = g_s+v_s, \quad s \in S$$

where $E_{\ell s}$, $\ell=1,2$, *equals the finite set of extreme optimal actions of player* ℓ *in the matrix game* $[\sum_{t=1}^{Z} p(t|s,.,.) g_t]$.

The proof of this theorem will be built up in a number of lemma's, which respectively prove: (i)⇒(ii), (ii)⇒(iii), (iii)⇔(iv) and (iii)⇒(i).

8.2.2. LEMMA. *If for some* $B \in \mathbb{R}$ $\|FV(\tau)-\tau g\| \leq B \log \tau$ *for all* τ *holds, then* $c_{M-1}=c_{M-2}=\ldots=c_1=0_Z$.

PROOF. Let $y(\tau)=\sum_{k=1}^{K} d_k \tau^{k/K}$ and $C\in\mathbb{R}$ be such that $\|FV(\tau)-y(\tau)\| \leq C \log \tau$ for all τ and suppose that $y(\tau)$ is constructed as indicated by Bewley & Kohlberg (1976b, statements 6.9-6.12, page 326).
Obviously the assumption of the lemma implies:

$$d_K = g \text{ and } d_{K-1} = d_{K-2} = \ldots = d_1 = 0_z.$$

However, considering the way in which the vectors $d_{K-1}, d_{K-2}, \ldots, d_1$ are constructed in Bewley & Kohlberg (1976b, statement 6.7), it can be seen that, if one of the vectors $c_{M-1}, c_{M-2}, \ldots, c_1$ does not equal 0_z then at least one of the vectors $d_{K-1}, d_{K-2}, \ldots, d_1$ does not equal 0_z. This proves the lemma. □

8.2.3. LEMMA. *If for the solution of the limit discount equation we have* $c_{M-1}=c_{M-2}=\ldots=c_1=0_z$, *then there exists a* $v\in\mathbb{R}^z$ *such that*

$$\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v)) = g_s(\theta+1)+v_s+O(\theta^{-1}), \quad s\in S.$$

PROOF. Take $v=c_0-g$. Since $c_{M-1}=c_{M-2}=\ldots=c_1=0_z$, we derive from the limit discount equation that

$$\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v)) = g_s(\theta+1)+v_s+o(\theta^{0}), \quad s\in S. \tag{8.2.1}$$

Weyl (1950) proved that, if the elements of a matrix game belong to a linearly ordered field, then also the value of this game belongs to that field. Using this result here implies that $\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v))$ is of the form $\sum_{k=-\infty}^{M} d_{ks}\theta^k$, with $d_{ks}\in\mathbb{R}$, which combined with (8.2.1) shows the lemma. □

8.2.4. LEMMA. *For a stochastic game the following two assertions are equivalent:*

(iii) there exist $v\in\mathbb{R}^z$ *and* $g\in\mathbb{R}^z$ *such that*

$$\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v)) = g_s(\theta+1)+v_s+O(\theta^{-1}), \quad s\in S$$

(iv) there exist $v\in\mathbb{R}^z$ *and* $g\in\mathbb{R}^z$ *such that*

$$g_s = \operatorname*{Val}_{A_s \times B_s} \left(\sum_{t=1}^{z} p(t|s,.,.) g_t \right), \quad s \in S$$

and

$$g_s + v_s = \operatorname*{Val}_{E_{1s} \times E_{2s}} (G_s(v)), \quad s \in S.$$

PROOF. Suppose (iii) to be true. Then for some $C_s \in \mathbb{R}$ and sufficiently large τ

$$(8.2.2) \qquad -C_s \tau^{-1} \leq \operatorname*{Val}_{A_s \times B_s} \left(r(s,.,.) + \sum_{t=1}^{z} p(t|s,.,.)(g_t \tau + v_t) \right) - g_s(\tau+1) - v_s \leq C_s \tau^{-1}.$$

Dividing each term of (8.2.2) by τ and letting τ tend to infinity gives by the continuity property of the val-operator:

$$(8.2.3) \qquad \operatorname*{Val}_{A_s \times B_s} \left(\sum_{t=1}^{z} p(t|s,.,.) g_t \right) = g_s$$

i.e. the first part of (iv).

Consider now the following limit:

$$d_s := \lim_{\tau \to \infty} \left\{ \operatorname*{Val}_{A_s \times B_s} (G_s(g\tau + v)) - \operatorname*{Val}_{A_s \times B_s} \left(\sum_{t=1}^{z} p(t|s,.,.) g_t \tau \right) \right\}$$

$$= \lim_{\tau \to \infty} \frac{\left\{ \operatorname*{Val}_{A_s \times B_s} \left(\sum_{t=1}^{z} p(t|s,.,.) g_t + G_s(v) . \tau^{-1} \right) - \operatorname*{Val}_{A_s \times B_s} \left(\sum_{t=1}^{z} p(t|s,.,.) g_t \right) \right\}}{\tau^{-1}}$$

Mills (1953) has called such a limit a marginal value, and has shown that the limit exists and equals $\operatorname*{Val}_{E_{1s} \times E_{2s}} (G_s(v))$.

On the other hand, after substitution of the equality in (iii) and of (8.2.3) into the definition of d_s we obtain:

$$d_s = \lim_{\tau \to \infty} \{ g_s(\tau+1) + v_s + O(\tau^{-1}) - g_s \tau \} = g_s + v_s$$

hence showing that (iii) implies (iv).

Now suppose that (iv) is true.

In the proof of lemma 8.2.3 we already argued that:

$$(8.2.4)\qquad \operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v)) = \sum_{k=-\infty}^{1} d_{ks}\theta^k, \text{ with } d_{ks}\in\mathbb{R}.$$

So we need to show $d_{1s}=g_s$ and $d_{0s}=v_s+g_s$.

From (8.2.4) we derive for each τ large enough:

$$\operatorname{Val}_{A_s\times B_s}\Big(\sum_{t=1}^{z} p(t|s,.,.)g_t+G_s(v).\tau^{-1}\Big) = \sum_{k=-\infty}^{0} d_{k+1\,s}\tau^k$$

Taking the limit as $\tau\to\infty$ yields

$$(8.2.5)\qquad \operatorname{Val}_{A_s\times B_s}\Big(\sum_{t=1}^{z} p(t|s,.,.)g_t\Big) = d_{1s},$$

hence $d_{1s}=g_s$.

As in the first part of the proof we have

$$\lim_{\tau\to\infty}\Big\{\operatorname{Val}_{A_s\times B_s}(G_s(g\tau+v)) - \operatorname{Val}_{A_s\times B_s}\Big(\sum_{t=1}^{z} p(t|s,.,.)g_t\tau\Big)\Big\} = \operatorname{Val}_{E_{1s}\times E_{2s}}(G_s(v)).$$

On the other hand, after substitution of (8.2.4) and (8.2.5), this limit equals d_{0s}. But, by assumptions, $\operatorname{Val}_{E_{1s}\times E_{2s}}(G_s(v))=g_s+v_s$, which shows that $d_{0s}=g_s+v_s$. □

8.2.5. LEMMA. *If, for a stochastic game, there exists a* $v\in\mathbb{R}^z$ *such that* $\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+v))=g_s(\theta+1)+v+O(\theta^{-1})$, $s\in S$, *then there exists a* $B\in\mathbb{R}$ *such that* $\|FV(\tau)-\tau g\|\le B\log\tau$, *all* $\tau\in\mathbb{N}$.

PROOF. By assumption there exists a $C\in\mathbb{R}$ such that for large τ:

$$\Big|\operatorname{Val}_{A_s\times B_s}(G_s(\tau g+v))-g_s(\tau+1)-v_s\Big| \le C.\tau^{-1}, \quad s\in S.$$

Then

$$(8.2.6)\qquad |FV_s(\tau+1)-g_s(\tau+1)-v_s| = \Big|\operatorname*{Val}_{A_s\times B_s}(G_s(FV(\tau)))-g_s(\tau+1)-v_s\Big| =$$

$$\Big|\operatorname*{Val}_{A_s\times B_s}(G_s(FV(\tau)))-\operatorname*{Val}_{A_s\times B_s}(G_s(g\tau+v))+\operatorname*{Val}_{A_s\times B_s}(G_s(g\tau+v))-g_s(\tau+1)-v_s\Big| \le$$

$$\Big|\operatorname*{Val}_{A_s\times B_s}(G_s(FV(\tau)))-\operatorname*{Val}_{A_s\times B_s}(G_s(g\tau+v))\Big|+\Big|\operatorname*{Val}_{A_s\times B_s}(G_s(g\tau+v))-g_s(\tau+1)-v_s\Big| \le$$

$$\|FV(\tau)-g\tau-v\| + C.\tau^{-1}.$$

The last inequality follows from the fact that the val-operator is Lipschitz continuous with constant 1. By repeated application of (8.2.6) we obtain for each τ and a fixed K sufficiently large:

$$\|FV(\tau+K)-g(\tau+K)-v\| \le \|FV(K)-gK-v\| + C.\sum_{k=1}^{\tau}(k+K-1)^{-1}.$$

Since $\sum_{k=1}^{\tau} k^{-1}=O(\log \tau)$ for each τ, the assertion in the lemma follows. □

Summarizing the lemma's 8.2.2-8.2.5 proves theorem 8.2.1.

Federgruen (1978) called the set of equations mentioned above in (iv) the natural extension to stochastic games of the set of functional equations for average reward Markov decision problems (cf. theorem A.2.6). For MDP's this set of equations characterizes the optimal gain and the set of optimal stationary strategies. In section 8.1 we have seen that it makes more sense to regard the set of functional equations mentioned in theorem 8.1.8 (ii) as the natural extention. With these equations both the value and the existence of optimal stationary strategies can be characterized. However, for the set of equations mentioned above in (iv), theorem 8.2.1 leads to the following theorem.

8.2.6. THEOREM. *If for a stochastic game there exist* $\tilde{g}\in\mathbb{R}^Z$ *and* $v\in\mathbb{R}^Z$, *such that*

$$\tilde{g}_s = \operatorname*{Val}_{A_s\times B_s}\Big(\sum_{t=1}^{z} p(t|s,.,.)\tilde{g}_t\Big) \text{ and } \tilde{g}_s+v_s = \operatorname*{Val}_{E_{1s}\times E_{2s}}(G_s(v)), \quad s\in S$$

then $\tilde{g}$ *is uniquely determined, namely* $\tilde{g}=g=c_M$.

PROOF. The proof uses some of the arguments of the previous lemma's. From (iv) with $\tilde{g}$ instead of g we obtain (iii) with $\tilde{g}$ instead of g (lemma 8.2.4). From (iii) with $\tilde{g}$ instead of g we get (i) with $\tilde{g}$ instead of g (lemma 8.2.5). But this implies $\lim_{\tau\to\infty} \frac{FV(\tau)}{\tau} = \tilde{g}$, which by lemma 7.2.4 shows the theorem. □

We now shall studie conditions under which $||FV(\tau)-g\tau|| \leq B$ for some $B\in\mathbb{R}$. For a strategy μ of player 1, we denote by $FV_{\mu 2}(\tau)$ the minimum expected payoff in the τ-step game which player 2 can guarantee himself, knowing that player 1 plays μ. Similarly $FV_{\nu 1}(\tau)$ is defined for a strategy ν of player 2. Obviously for each μ and ν and each $\tau\in\mathbb{N}$:

$$(8.2.7) \qquad FV_{\mu 2}(\tau) \leq FV(\tau) \leq FV_{\nu 1}(\tau).$$

For a vector $w\in\mathbb{R}^Z$, we shall use $\nabla w := \max_s w_s$ and $\Delta w := \min_s w_s$.
By MDP(ρ) we denote the Markov decision problem that results when player 1 fixes the stationary strategy ρ.
The following lemma will be used.

8.2.7. LEMMA. *For a stationary strategy ρ of player 1, there exists a number B such that for each $\tau\in\mathbb{N}$:*

$$\nabla(FV_{\rho 2}(\tau)-\tau g) \leq B+\tau\nabla(g(\rho)-g) \leq B.$$

where $g(\rho)$ and g equal the average reward value of MDP(ρ) and of the original game respectively.

PROOF. From theorem 3.4 we know that when player 1 announces that he will play the stationary strategy ρ, then the best player 2 can do is to solve MDP(ρ). For Markov decision problems it is known that the value of the τ-step problem minus τ times the average reward value is uniformly bounded in τ, cf. Brown (1965). This, applied to MDP(ρ) yields for some $B\in\mathbb{R}$ and each $\tau\in\mathbb{N}$:

$$||FV_{\rho 2}(\tau)-\tau.g(\rho)|| \leq B.$$

From this inequality the first inequality of the lemma can be deduced, while the second inequality follows from the fact $g(\rho)\leq g$.

□

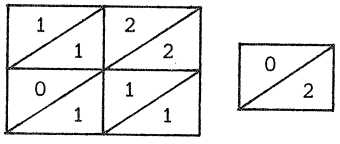
A similar lemma can be given for player 2.

Bewley & Kohlberg (1978) mentioned a strategy μ for player 1 uniformly τ-stage optimal if $\nabla(FV(\tau)-FV_{\mu 2}(\tau))\leq B$ for all τ and some number B. (Similarly for player 2 with reversal of the inequality sign.) They showed that if a player has, for each $s\in S$, a real action which guarantees him in the game $[G_{s\theta}(x^*)]$ on $A_s\times B_s$ a payoff $\underset{A_s\times B_s}{\text{Val}}\,(G_{s\theta}(x^*))+o(\theta^0)$, then the stationary strategy corresponding to these actions is uniformly τ-stage optimal.
The first part of the following theorem is obvious (cf. lemma 7.2.4). The second part is a consequence of lemma 8.2.7.

8.2.8. THEOREM. *A uniformly τ-stage optimal stationary strategy is optimal with respect to the average reward criterion. For a stochastic game where $FV(\tau)-\tau g$ is not bounded from above, player 1 has no stationary strategy which is uniformly τ-stage optimal. Analogously for player 2.*

The case mentioned in theorem 8.2.8 occurs in the following example.

8.2.9. EXAMPLE.



1 2

The average payoff value of this game equals (0,0). Player 2 has no optimal stationary strategy, while each (stationary) strategy of player 1 is optimal. For state 2, $FV_2(\tau)=0$ for each τ and for state 1:

$$FV_1(\tau+1) = \text{Val}\begin{bmatrix} 1+FV_1(\tau) & 2 \\ FV_1(\tau) & 1+FV_1(\tau) \end{bmatrix} = FV_1(\tau) + \frac{1}{FV_1(\tau)}$$

It can be verified that this relation implies $FV_1(\tau)\geq\tau^{\frac{1}{2}}$, and thus $FV_1(\tau)-\tau g_1$ is not bounded from above. Therefore, for this game, it may be wise for player 1 to play a Markov strategy which is optimal for a τ-step game for some large τ, instead of playing an optimal stationary strategy.

8.2.10. THEOREM. *For a stochastic game where* $\nabla(FV(\tau)-\tau g)\leq B$ *for all* τ *and some* $B\in\mathbb{R}$, *a stationary strategy* ρ *for player* 1 *is optimal with respect to the average payoff if and only if this strategy is uniformly* τ*-stage optimal.*

PROOF. Theorem 8.2.8 states that a uniformly τ-stage optimal stationary strategy is also optimal in the average reward game, which proves the sufficiency of the theorem.
Let ρ be optimal. Consider MDP(ρ), then $g(\rho)=g$. But since by Brown (1965) $\|FV_{\rho 2}(\tau)-\tau.g(\rho)\|\leq C$ for some $C\in\mathbb{R}$ and by assumption $\nabla(FV(\tau)-\tau g)\leq B$, we then have $\nabla(FV(\tau)-FV_{\rho 2}(\tau))\leq D$ for some $D\in\mathbb{R}$ and each $\tau\in\mathbb{N}$, i.e. ρ is uniformly τ-stage optimal.

□

An analogous statement can be made for player 2, in which case the inequality sign in theorem 8.2.10 must be reversed and ∇ must be replaced by Δ.

The following theorem may be of computational importance.

8.2.11. THEOREM. *If both players have optimal stationary strategies, then* $\|FV(\tau)-\tau g\|\leq B$ *for some* $B\in\mathbb{R}$ *and each* $\tau\in\mathbb{N}$.

PROOF. Follows at once from theorem 8.2.10.

□

The converse of this theorem is not true. The "big match" of Blackwell & Ferguson (1968) (cf. example 3.3) presents a counterexample. Namely $FV_1(\tau)=\frac{1}{2}\tau=g_1.\tau$ and player 1 possesses no optimal stationary strategy. However, for stochastic games, where g does not depend on the initial state, the converse of theorem 8.2.11 proves to be true.

8.2.12. THEOREM. *The following two assertions are equivalent:*

(i) $g=g^*.1_z$ *and both players have optimal stationary strategies.*

(ii) $\|FV(\tau)-\tau g\|\leq B$ *with* $g=g^*.1_z$ *and* $B\in\mathbb{R}$.

PROOF. Theorem 8.2.11 shows that (i) implies (ii).

Now suppose that (ii) is true. Then theorem 8.2.1 (the equivalence of (i) and (iv)) implies that there exists a $v\in\mathbb{R}$ such that

$$(8.2.8)\qquad g^*+v_s = \operatorname*{Val}_{E_{1s}\times E_{2s}}(G_s(v)), \quad s\in S$$

But since $g=g^*.1_z$ it can be seen that $E_{1s}=A_s$ and $E_{2s}=B_s$, since each entry of the matrix game $[\sum_{t=1}^{z} p(t|s,.,.)g_t]$ equals g^*. Then from lemma 8.1.3 we may conclude that both players possess optimal stationary strategies. □

One may ask under what properties of the game parameters it holds that $\|FV(\tau)-\tau g\|\leq B$. At this point we wish to express the following conjecture.

8.2.13. CONJECTURE. *$\|FV(\tau)-\tau g\|\leq B$ for all $\tau\in\mathbb{N}$ and some $B\in\mathbb{R}$ if and only if the solution of the limit discount equation can be expressed as a Laurent series expansion, i.e.* $x^*(\theta)=\sum_{k=-\infty}^{1} c_k\theta^k$.

However we have not yet been able to prove this assertion. The following theorem gives a sufficient condition. In this theorem we use the limit recursion equation, as formulated by Bewley & Kohlberg (1976a). We now define this limit recursion equation. To this end let, for $w=\sum_{k=-\infty}^{K} c_k\theta^{k/M}\in F$, $\kappa(w)$ be the element of F defined as $\kappa(w):=\sum_{k=-\infty}^{K} c_k(\theta+1)^{k/M}$, where $(\theta+1)^{k/M}$ is defined to be

$$(8.2.9)\qquad (\theta+1)^{k/M} := \theta^{k/M} + \frac{k}{M}\theta^{k/M-1} + \tfrac{1}{2}\frac{k}{M}\left(\frac{k}{M}-1\right)\theta^{k/M-2} + \ldots$$

As $\phi_\tau(\kappa(w))=\phi_{\tau+1}(w)$, obviously $\kappa(w)\in F$.

8.2.14. DEFINITION. *The following set of equations in the variable $x=(x_1,x_2,\ldots,x_z)\in F^z$ is referred to as the limit recursion equation:*

$$(8.2.10)\qquad \kappa(x_s) = \operatorname*{Val}_{A_s\times B_s}(G_s(x)), \quad s\in S.$$

Recall that $FV_s(\tau+1)=\mathrm{Val}(G_s(FV(\tau)))$, for each $s\in S$. Suppose that $x\in F^z$ describes an asymptotic expression for $FV(\tau)$ in the sense that $\phi_\tau(x)=FV(\tau)$ for large τ. Then $\phi_{\tau+1}(x_s)=\mathrm{Val}(G_s(\phi_\tau(x)))$ for large τ. Now the limit recursion equation can be seen as the replacement of this sequence of equations by a single equation.

The following lemma is an immediate consequence of the fact that $1/(1+\theta^{-1})=1-\theta^{-1}+\theta^{-2}-\ldots$ and the fact that the solution of the limit discount equation has valuation at most 1 (cf. Bewley & Kohlberg (1976a)).

8.2.15. LEMMA. *If* $x^* \in F^Z$ *solves the limit discount equation, then* $y^*=x^*/(1+\theta^{-1})$ *satisfies:*

$$\kappa(y^*_s) = \operatorname*{Val}_{A_s \times B_s}(G_s(y^*))+o(\theta^0)$$

8.2.16. THEOREM. *If for a stochastic game the limit recursion equation has a solution* $y=\sum_{k=-\infty}^{M} d_k \theta^{k/M}$, *with* $d_{M-1}=d_{M-2}=\ldots=d_1=0$, *then* $\|FV(\tau)-\tau g\| \leq B$ *for each* $\tau \in \mathbb{N}$ *and some* $B \in \mathbb{R}$.

PROOF. From $\phi_{\tau+1}(y_s)=\operatorname*{Val}_{A_s \times B_s}(G_s(\phi_\tau(y)))$ for large τ, and $FV_s(\tau+1)=\operatorname*{Val}_{A_s \times B_s}(G_s(FV(\tau)))$ for each s and each τ, we derive by means of the Lipschitz continuity of the val-operator: $\|\phi_{\tau+1}(y)-FV(\tau+1)\| \leq \|\phi_\tau(y)-FV(\tau)\|$ for large τ. Thus $\|\phi_{\tau+k}(y)-FV(\tau+k)\| \leq \|\phi_k(y)-FV(k)\|$ for each $\tau \in \mathbb{N}$ and each $k \in \mathbb{N}$. This implies that $\phi_\tau(y)-FV(\tau)$ is bounded uniformly in τ, which in view of $\lim_{\tau\to\infty} \frac{FV(\tau)}{\tau} = g$ leads to $y=g.\theta+d_0+o(\theta^0)$. Hence it follows that $\|FV(\tau)-\tau g\|$ is bounded uniformly in τ. □

We conclude this section with the following lemma, which might be useful in showing the reverse of theorem 8.2.16 which we suspect to be true.

8.2.17. LEMMA. *If* $y=g\theta+d_0+o(\theta^0) \in F^Z$ *is such that*

$$\kappa(y_s) = \operatorname*{Val}_{A_s \times B_s}(G_s(y))+o(\theta^{-1}), \quad \textit{each } s \in S,$$

then the limit recursion equation has a solution

$$\tilde{y} = g\theta+d_0+o(\theta^0) \in F^Z.$$

PROOF. Suppose that $y \in F^Z_M$ is such that $\kappa(y_s)=\operatorname*{Val}_{A_s \times B_s}(G_s(y))+O(\theta^{-\frac{M+1}{M}})$. Let $v=D.\theta^{-1/M}.1_Z$ with $D \in \mathbb{R}$.

Since $\kappa\theta^{-1/M}=\theta^{-1/M}+\frac{1}{M}.\theta^{-\frac{M+1}{M}}+O(\theta^{-\frac{2M+1}{M}})$, we obtain for each $s\in S$:

$$\kappa(y_s+v_s)=\operatorname{Val}_{A_s\times B_s}(G_s(y_s+v_s))+(C+\tfrac{D}{M})\theta^{-\frac{M+1}{M}}+o(\theta^{-\frac{M+1}{M}}) \text{ for some } C\in\mathbb{R}.$$

Then by choosing D appropriately we see that there exist $\bar{y}$ and $\underline{y}\in F_M^Z$, with $\bar{y}\geq\underline{y}$, such that, for each $s\in S$:

$$(8.2.11)\qquad \kappa(\bar{y}_s)\geq \operatorname{Val}_{A_s\times B_s}(G_s(\bar{y}))\geq \operatorname{Val}_{A_s\times B_s}(G_s(\underline{y}))\geq \kappa(\underline{y}_s).$$

Since both the operator κ and the val-operator are monotone and continuous, (8.2.11) assures the existence of an $\tilde{y}$, with $\bar{y}\geq\tilde{y}\geq\underline{y}$, such that

$\kappa(\tilde{y}_s)=\operatorname{Val}_{A_s\times B_s}(G_s(\tilde{y}))$, $s\in S$. □

From the solution of the limit discount equation, we see that if $x^*=g\theta+c_0+O(\theta^{-1/M})$, then

$$(8.2.12)\qquad \kappa(g_s\theta+c_{0s}-g_s)=\operatorname{Val}_{A_s\times B_s}(G_s(g\theta+c_0-g))+O(\theta^{-1}),\quad \text{all } s\in S,$$

which by theorem 8.2.1 proves to be equivalent to $\|FV(\tau)-\tau g\|\leq B \log \tau$. Now the condition of lemma 8.2.17 appears to be the step "one-stronger" than (8.2.12). So in view of lemma 8.2.17 we have the feeling that the condition of theorem 8.2.16 is also a necessary one for the property $\|FV(\tau)-\tau g\|\leq B$. However we have not yet been able to provide either a proof or a counterexample of this conjecture.

8.3. GAMES WITH A VALUE INDEPENDENT OF THE INITIAL STATE.

In this section we consider the class of stochastic games for which the average reward value g has the form $g=g^*.1_Z$. Some of the results of this section can also be found in Vrieze (1979a).

8.3.1. THEOREM. *For a two-person zerosum stochastic game with finite state and action spaces, the following assertions are equivalent.*

(i) $g=g^*.1_Z$.

(ii) there exists a $g^*\in\mathbb{R}$ *such that, for each* $\varepsilon>0$, *there exists a* $v(\varepsilon)\in\mathbb{R}^Z$ *with*

(8.3.1) $$\left|v_s(\varepsilon)+g^*-(\operatorname{Val}_{A_s\times B_s} G_s(v(\varepsilon)))\right| \le \varepsilon, \quad \textit{each } s\in S.$$

PROOF. Suppose that (i) is true. From the solution of the limit discount equation we derive

(8.3.2) $$\operatorname{Val}_{A_s\times B_s} \left(G_s\Big(\sum_{k=0}^{M} c_k\theta^{k/M}-c_M\Big)\right) = \sum_{k=0}^{M} c_k\theta^{k/M}+O(\theta^{-1/M})$$

By assumption, $c_M=g^*.1_z$. Hence, we see from (8.3.2) that, for fixed $\varepsilon>0$, $v(\varepsilon)=\sum_{k=0}^{M} c_k\tau^{k/M}$, for sufficiently large τ, satisfies assertion (ii) of the theorem.

Now suppose that (ii) is true. Fix $\varepsilon>0$ and let $v(\varepsilon)$ satisfy (8.3.1). Let $\rho_s(\varepsilon)$ be an optimal action for player 1 in the matrix game $[G_s(v(\varepsilon))]$ and let $\rho(\varepsilon)$ be the stationary strategy $\rho(\varepsilon)=(\rho_1(\varepsilon),\ldots,\rho_z(\varepsilon))$. Then for each pure stationary strategy σ^p of player 2:

(8.3.3) $$v(\varepsilon)+g^*.1_z-\varepsilon.1_z \le r(\rho(\varepsilon),\sigma^p)+P(\rho(\varepsilon),\sigma^p).v(\varepsilon).$$

Multiplying (8.3.3) from the left by $Q(\rho(\varepsilon),\sigma^p)$ gives:

(8.3.4) $$g^*.1_z-\varepsilon.1_z \le Q(\rho(\varepsilon),\sigma^p).r(\rho(\varepsilon),\sigma^p) = W_{\rho(\varepsilon)\sigma^p}.$$

But (8.3.4) implies by corollary 3.5, and the fact that $\varepsilon>0$ is arbitrary:

(8.3.5) $$g^*.1_z \le \sup_{\mu}\inf_{\nu} W_{\mu\nu}.$$

Similarly we can derive

(8.3.6) $$g^*.1_z \ge \inf_{\nu}\sup_{\mu} W_{\mu\nu}.$$

Since for each function $f(\mu,\nu)$ it holds that $\inf_{\nu}\sup_{\mu} f(\mu,\nu)\ge\sup_{\mu}\inf_{\nu} f(\mu,\nu)$, the combination of (8.3.5) and (8.3.6) shows that assertion (i) is true.

□

8.3.2. THEOREM. *If for a stochastic game* $g=g^*.1_z$, *then*

(a) Both players possess ε*-optimal stationary strategies for each* $\varepsilon>0$.

(b) Both players have optimal Markov strategies.

PROOF. (a) This follows from the proof of the preceding theorem; especially (8.3.4) shows by corollary 3.5 that $\rho(\varepsilon)$ is an ε-optimal stationary strategy.
Concerning (b), an optimal Markov strategy can be constructed by making use of the fact that $\frac{FV_s(\tau)}{\tau}$ converges to g^* for each state $s\in S$ (a scheme for such a Markov strategy can be found in Bewley & Kohlberg (1978, page 117)).

□

8.3.3. REMARK. ε-optimal stationary strategies can be derived from the solution of the limit discount equation. Namely, let $\rho_s(\tau)$ be an optimal action for player 1 in the matrix game $[G_{s\tau}(x^*(\tau))]$ which has value $x_s^*(\tau)$. Then the stationary strategy $\rho(\tau)=(\rho_1(\tau),\ldots,\rho_z(\tau))$, which is optimal for the discounted stochastic game with interest rate τ^{-1}, is, for sufficiently large τ, ε-optimal for the average reward game. This can be shown along the same lines as the relations (8.3.2), (8.3.3) and (8.3.4) were established.

Federgruen has given a characterization for the class of games with a value independent of the initial state and for which in addition both players possess optimal stationary strategies. In fact his result equals lemma 8.1.3. In the next lemma we extend this characterization with three further equivalencies.

8.3.4. LEMMA. *The next assertions are equivalent.*

(i) $g=g^*.1_z$ *and both players have optimal stationary strategies.*

(ii) *there exist a* $w\in\mathbb{R}^z$ *and a* $g^*\in\mathbb{R}$ *such that*

$$w_s+g^* = \operatorname{Val}_{A_s\times B_s}(G_s(w)), \quad \textit{each } s\in S.$$

(iii) $x^*(\theta)$ *has the property that* $c_M=g^*.1_z$ *and* $c_{M-1}=c_{M-2}=\ldots=c_1=0_z$.

(iv) *there exists a* $g^*\in\mathbb{R}$ *such that* $\bigcap_{\varepsilon>0} F(\varepsilon)\neq\emptyset$, *where* $F(\varepsilon)$ *for* $\varepsilon>0$ *equals the set of vectors* $v(\varepsilon)$ *satisfying* (8.3.1).

(v) *there exists a* $g^*\in\mathbb{R}$ *such that* $\|FV(\tau)-\tau g^*.1_z\| \leq B$, *with* $B\in\mathbb{R}$ *and each* $\tau\in\mathbb{N}$.

PROOF. The equivalence of (i) and (ii) is already proved in lemma 8.1.3, while the equivalence of (ii) and (iii) is a special case of theorem

8.2.1 (since $g=g^*.1_z$ it follows that $E_{1s}=A_s$ and $E_{2s}=B_s$). From theorem 8.3.1 we infer that (ii) and (iv) are equivalent assertions. Finally theorem 8.2.12 states the equivalence of (i) and (v).

□

In the following theorem we give a sufficient and necessary condition for one player to have an optimal stationary strategy (cf. remark 8.1.9).

8.3.5. THEOREM. *For a stochastic game with average payoff value* $g^*.1_z$, *player* 1 *has an optimal stationary strategy if and only if there exists a vector* $v\in\mathbb{R}^z$, *such that*

$$(8.3.7)\qquad v_s+g^* \leq \operatorname{Val}_{A_s\times B_s}(G_s(v)), \quad \textit{each } s\in S.$$

PROOF. Suppose that (8.3.7) holds. Let the stationary strategy $\tilde{\rho}$ be such that $\tilde{\rho}_s$ is optimal for player 1 in $[G_s(v)]$. Then inequalities (8.3.3) and (8.3.4) hold with $\varepsilon=0$ and $\rho(\varepsilon)$ replaced by $\tilde{\rho}$. Then corollary 3.5 shows that $\tilde{\rho}$ is optimal.

Now suppose that ρ^* is an optimal stationary strategy for player 1. By considering MDP(ρ^*) it follows, from the minimizing version of theorem A.2.6, that there exists a $\tilde{v}\in\mathbb{R}^z$ with

$$\tilde{v}_s+g^* = \min_{j\in B_s}\{r(s,\rho_s^*,j)+\sum_{t=1}^{z} p(t|s,\rho_s^*,j).\tilde{v}_t\}, \quad s\in S.$$

But since

$$\min_{j\in B_s}\{r(s,\rho_s^*,j)+\sum_{t=1}^{z} p(t|s,\rho_s^*,j).\tilde{v}_t\} \leq \operatorname{Val}_{A_s\times B_s}(G_s(\tilde{v})),$$

we see that $\tilde{v}$ satisfies (8.3.7).

□

Obviously an analogous statement of theorem 8.3.5 can be made concerning player 2. Note that, if in (8.3.7) for some v the equality sign holds for each $s\in S$, then by lemma 8.3.4 also player 2 has an optimal stationary strategy.

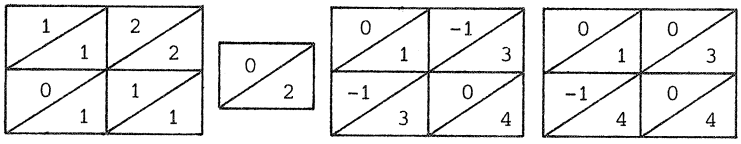
We finish this section by demonstrating a curious phenomenon. In lemma 8.3.4 we have seen that if the solution of the limit discount equation has properties (a) $c_M=g^*.1_z$ and (b) $\sum_{k=1}^{M-1} c_k\theta^{k/M}=0_z$, then both players have

optimal stationary strategies. A stochastic game for which the solution of the limit discount equation has the properties (a) above and (b') $\sum_{k=1}^{M-1} c_k \theta^{k/M} \geq 0_z$ appears more favourable and easier to play from the viewpoint of player 1 than that with properties (a) and (b). Thus one would be inclined to believe that (a) and (b') would imply the existence of an optimal stationary strategy for player 1. The next example shows that in general the above suggestion need not be true. However the converse of this statement does hold i.e.

if player 1 has an optimal stationary strategy, then $\sum_{k=1}^{M-1} c_k \theta^{k/M} \geq 0_z$.

Further, this property holds for arbitrary c_M. This can be seen by observing that $x^*(\theta) \geq V_{\rho^*}(\theta)$, if ρ^* is an optimal stationary strategy where $V_{\rho^*}(\theta)$ equals the solution of the limit discount equation associated with MDP(ρ^*). Since $V_{\rho^*}(\theta) = c_M \theta + O(\theta^O)$ the assertion follows.

8.3.6. EXAMPLE.



Let $x^* = (x_1^*, x_2^*, x_3^*, x_4^*)$ be the solution of the limit discount equation. The states 1 and 2 can be considered apart.

Solving the limit discount equation for these two states results in $x_1^* = \theta^{\frac{1}{2}}(1+\theta^{-1})$ and $x_2^* = 0$, so $g_1 = 0$ and $g_2 = 0$. Obviously each stationary strategy of player 1 is optimal in the states 1 and 2.

Now consider the states 3 and 4. Define for $0 < \varepsilon < 1$ the stationary strategy ρ^ε as follows: ρ_1^ε and ρ_2^ε are arbitrary, ρ_3^ε and ρ_4^ε both put weight ε on the first action and weight $1-\varepsilon$ on the second. Then for the states 3 and 4, ρ^ε guarantees player 1 an average payoff $-\varepsilon$. On the other hand player 2 can assure himself an average payoff 0 for the states 3 and 4 by playing his second action in these states. Conclusion: the average reward value of this game equals (0,0,0,0). Furthermore it can be checked that player 1 has no optimal stationary strategy. So if we can show that $\sum_{k=1}^{M-1} c_{ks} \theta^{k/M} \geq 0$ for the states s=3 and s=4 then we are ready (for s=1 and s=2 this inequality

follows from x_1^* and x_2^*).

Consider the game with interest rate τ^{-1}. Let $\rho(\tau)$ be as follows: $\rho_1(\tau)$ and $\rho_2(\tau)$ are optimal in the τ^{-1} discounted game, both $\rho_3(\tau)$ and $\rho_4(\tau)$ are such that the first row is chosen with probability $\tau^{-\frac{1}{2}}$ and the second row with probability $1-\tau^{-\frac{1}{2}}$. Let $y_3(\tau)$ and $y_4(\tau)$ be the τ^{-1}-discounted expected payoff that the strategy $\rho(\tau)$ guarantees player 1 in state 3 and state 4 respectively.

Then by theorem A.2.5:

$$(8.3.8)\qquad y_3(\tau) = \min\{\tau^{-\frac{1}{2}}\cdot\frac{\tau^{\frac{1}{2}}(1+\tau^{-1})}{1+\tau^{-1}} - (1-\tau^{-\frac{1}{2}}) + \frac{(1-\tau^{-\frac{1}{2}})\cdot y_3(\tau)}{1+\tau^{-1}},$$

$$- \tau^{-\frac{1}{2}} + \frac{\tau^{\frac{1}{2}}\cdot y_3(\tau) - (1-\tau^{-\frac{1}{2}})y_4(\tau)}{1+\tau^{-1}}\}$$

and

$$(8.3.9)\qquad y_4(\tau) = \min\{\tau^{-\frac{1}{2}}\cdot\frac{\tau^{\frac{1}{2}}(1+\tau^{-1})}{1+\tau^{-1}} - (1-\tau^{-\frac{1}{2}}) + \frac{(1-\tau^{-\frac{1}{2}})\cdot y_4(\tau)}{1+\tau^{-1}},$$

$$\frac{\tau^{-\frac{1}{2}}\cdot y_3(\tau) + (1-\tau^{-\frac{1}{2}})\cdot y_4(\tau)}{1+\tau^{-1}}\}.$$

Suppose in (8.3.8) that the minimum is attained by the first component, then:

$$(1+\tau^{-1})y_3(\tau) = \tau^{-\frac{1}{2}}(1+\tau^{-1}) + (1-\tau^{-\frac{1}{2}})\cdot y_3(\tau),$$

or

$$(8.3.10)\qquad y_3(\tau) = \frac{1+\tau^{-1}}{1+\tau^{-\frac{1}{2}}}.$$

Similarly if in (8.3.9) the minimum is attained by the first component, then

$$(8.3.11)\qquad y_4(\tau) = \frac{1+\tau^{-1}}{1+\tau^{-\frac{1}{2}}}.$$

Now suppose that as well in (8.3.8) as in (8.3.9) the second component yields the minimum, then from (8.3.9) we get $y_3(\tau)=y_4(\tau)(1+\tau^{-\frac{1}{2}})$. Substitution of this relation in (8.3.8) with respect to the second component results in:

$$y_4(\tau)(1+\tau^{-\frac{1}{2}}) = -\tau^{-\frac{1}{2}} + \frac{\tau^{-\frac{1}{2}}(1+\tau^{-\frac{1}{2}})y_4(\tau)+(1-\tau^{-\frac{1}{2}})y_4(\tau)}{1+\tau^{-1}}$$

or

(8.3.12) $\quad y_4(\tau) = -1$ and then $y_3(\tau) = -(1+\tau^{-\frac{1}{2}})$.

The above analysis shows, by (8.3.10), (8.3.11) and (8.3.12) that in both (8.3.8) and (8.3.9) the minimum is attained by the second component, and that (8.3.12) represents the solution. Obviously $x_3^*(\tau)\geq y_3(\tau)=-(1+\tau^{-\frac{1}{2}})$ and $x_4^*(\tau)\geq y_4(\tau)=-1$ for sufficiently large τ. Hence we may conclude that $\sum_{k=1}^{M-1} c_{ks}\theta^{ks}\geq 0$ for s=3 and s=4. In fact $\sum_{k=1}^{M-1} c_{ks}\theta^{ks}=0$ for s=3 and s=4. Namely if in the states 3 and 4 player 2 always chooses his second action, then, for each discount factor, this strategy assures him a discounted expected payoff of at most 0 for the initial states s=3 or s=4.

Summarizing, example 8.3.6 is a stochastic game with average payoff value $c_M=g^*.1_z$, for which $\sum_{k=1}^{M-1} c_k\theta^{k/M}\geq 0_z$, but for which player 1 has no optimal stationary strategy.

8.4. ON THE EXISTENCE OF EASY INITIAL STATES.

In this section we show that there exist non-empty subsets, S^*, $S^{**} \subset S$ such that, at least for the states belonging to S^* (S^{**}), player 1 (player 2) can assure himself the average payoff value by choosing an appropriate stationary strategy. A state with this property is called an easy initial state for player 1 (player 2).

The results of this section are based on Tijs & Vrieze (1986).
In section 7.2 we have mentioned the important result of Bewley & Kohlberg, that the solution of the limit discount equation is of the form

$$(8.4.1) \qquad x^*(\theta) = \Sigma_{k=-\infty}^{M} c_k \theta^{k/M}, \text{ where } x^*(\theta)=(x_1^*,\ldots,x_z^*),\ c_k \in \mathbb{R}^z$$

$$\text{and } x_s^* = \Sigma_{k=-\infty}^{M} c_{ks} \theta^{k/M}.$$

It follows from a well-known result of Weyl (1950) that, for the matrix game

$$[G_{s\theta}(x^*)] := [r(s,.,.) + \frac{\sum_{t=1}^{z} p(t|s,.,.)x_t^*}{1+\theta^{-1}}] \quad \text{on } A_s \times B_s,$$

the players possess optimal actions, say $f_s(\theta)$ and $g_s(\theta)$ respectively, such that $f_s(\theta) \in F_M^{m_s}$ and $g_s(\theta) \in F_M^{n_s}$. From now on we fix such an optimal action $\tilde{f}_s(\theta)$, for each $s \in S$, for player 1.
$\mathbb{N}^-$ will denote the set $\{-1,-2,\ldots\}$.
We recall that

$$(8.4.2) \qquad \tilde{f}_s(\theta) = \sum_{k=-\infty}^{0} \tilde{f}_{ks} \theta^{k/M}$$

where for each $s \in S$ and each $k \in \mathbb{N}^-$:

$$(8.4.3) \qquad \tilde{f}_{ks} \in \mathbb{R}^{m_s}, \quad \tilde{f}_{0s} \in P(A_s), \quad \sum_{i \in A_s} \tilde{f}_{0s}(i)=1, \quad \sum_{i \in A_s} \tilde{f}_{ks}(i)=0$$

and where for each $s \in S$, $i \in A_s$ and $k \in \mathbb{N}^-$:

$$(8.4.4) \qquad \tilde{f}_{0s}(i) \geq 0$$

$$(8.4.5) \qquad \tilde{f}_{ks}(i) \geq 0 \quad \text{if } \tilde{f}_{\ell s}(i)=0,\ \ell=k+1,k+2,\ldots,0.$$

Let $\tilde{\rho}$ be the stationary strategy for player 1 with $\tilde{\rho}_s := \tilde{f}_{0s}$. Note that $\tilde{\rho}_s = \lim_{\tau\to\infty} \tilde{f}_s(\tau)$. By lemma 7.2.2 $(\tilde{f}_1(\tau),\ldots,\tilde{f}_z(\tau))$ is optimal in the discounted stochastic game with interest rate τ^{-1}, hence we see that $\tilde{\rho}$ can be considered as the limit of optimal stationary strategies for player 1 in discounted games with interest rate α, for $\alpha\downarrow 0$. We shall show in this section that this strategy $\tilde{\rho}$ for player 1 is optimal for all specific plays starting in a non-empty subset S^*. Here S^* is defined as follows. Let

$$(8.4.6) \qquad \sum_{k=1}^{M} d_k \theta^{k/M} := \max_{s\in S} \left(\sum_{k=1}^{M} c_{ks} \theta^{k/M} \right)$$

(the maximum is taken with respect to the order in F; note that d_k is a scalar). Then

$$(8.4.7) \qquad S^* := \{s\in S \mid \sum_{k=1}^{M} d_k \theta^{k/M} = \sum_{k=1}^{M} c_{ks} \theta^{k/M}\}.$$

If we call $\Sigma_{k=1}^{M} c_{ks} \theta^{k/M}$ of expression (8.4.1) the main part of x_s^*, then we can say that S^* consists precisely of those states in S with maximal main part. Observe that $\Sigma_{k=1}^{M} c_{ks} \theta^{k/M}$ corresponds to the positive powers of θ. Theorem 8.4.1 below states that S^* consists of easy initial states for player 1.

8.4.1. THEOREM. *The states in S^* are easy states for player* 1. *Namely if the initial state* s *of a specific game belongs to the set S^*, then player* 1 *can guarantee himself the average payoff value c_{Ms} by means of the stationary strategy $\tilde{\rho}$.*

A similar theorem can be formulated for player 2. The elements of the set S^{**}, consisting of those states s for which the main part of x_s^* is minimal, are easy states for player 2.

The proof of theorem 8.4.1 is postponed until we have proved a number of preliminary lemma's.

From now on fix an arbitrarily chosen pure stationary strategy σ^P, which puts weight 1 on some action $j_s \in B_s$, $s\in S$. Since $\tilde{f}_s(\theta)$ is an optimal action in the matrix game $[G_{s\theta}(x^*)]$, we have for each $s\in S$:

$$(8.4.8)\qquad -x_s^*+R_s+(1+\theta^{-1})^{-1}.\sum_{t\in S}P_{st}(x_t^*)\geq 0,$$

in which

$$(8.4.9)\qquad R_s := \sum_{k=-\infty}^{0} R_{ks}\theta^{k/M}$$

is the expected immediate reward in state s, and

$$(8.4.10)\qquad P_{st} := \sum_{k=-\infty}^{0} P_{kst}\theta^{k/M}$$

can be interpreted as the probability that the system jumps to state $t\in S$, if in state s player 1 uses the mixed action $\tilde{f}_s(\theta)$ and player 2 chooses j_s. For the coefficients in (8.4.9) and (8.4.10) we have for all $k\in\{0,-1,-2,\ldots\}$ and $s\in S$:

$$(8.4.11)\qquad R_{ks} = \sum_{i\in A_s} r(s,i,j_s)\tilde{f}_{ks}(i)$$

and

$$(8.4.12)\qquad P_{kst} = \sum_{i\in A_s} p(t|s,i,j_s)\tilde{f}_{ks}(i).$$

For further use, we note that in view of (8.4.3), (8.4.4) and (8.4.5) we have, for all $k\in\mathbb{N}^-$, that:

$$(8.4.13)\qquad \sum_{t\in S} P_{0st} = 1, \qquad \sum_{t\in S} P_{kst} = 0$$

$$(8.4.14)\qquad P_{0st}\geq 0$$

$$(8.4.15)\qquad P_{kst}\geq 0 \quad \text{if } P_{\ell st}=0,\ \ell=k+1,k+2,\ldots,0.$$

Denote the left side of (8.4.8) by the Puiseux series $\sum_{k=-\infty}^{M} y_{ks}\theta^{k/M}$. Then for each $k<M$:

$$(8.4.16)\qquad y_{Ms}\geq 0 \quad \text{and } y_{ks}\geq 0 \quad \text{if } y_{\ell s}=0,\ \ell=k+1,k+2,\ldots,M.$$

In the following, we are especially interested in the expressions for the coefficients corresponding to non-negative powers. For them we obtain, as a result of (8.4.8), (8.4.9), (8.4.10) and (8.4.1):

$$(8.4.17)\qquad y_{ks} = -c_{ks} + \sum_{t\in S} \sum_{\ell=-(M-k)}^{0} P_{\ell st} c_{k-\ell t}, \qquad k=M,M-1,\ldots,1$$

$$(8.4.18)\qquad y_{0s} = -c_{0s} + R_{0s} + \sum_{t\in S} \sum_{\ell=-M}^{0} P_{\ell st} c_{-\ell t} - \sum_{t\in S} P_{0st} c_{Mt}$$

The following subsets of S play a rule. For $k\in\{M,M-1,\ldots,1\}$, let

$$(8.4.19)\qquad S_k := \{s\in S \mid y_{\ell s}=0,\ \ell\in\{k,k+1,\ldots,M\}\}$$

$$(8.4.20)\qquad T_k := \{s\in S \mid c_{\ell s}=d_\ell,\ \ell\in\{k,k+1,\ldots,M\}\}$$

(cf. (8.4.6) for the definition of d_ℓ).
The elements of T_k correspond to the set of states s for which the M-k+1 leading coefficients of the Puiseux series $x_s^* = \sum_{k=-\infty}^{M} c_{ks}\theta^{k/M}$ are equal to the M-k+1 leading coefficients of the maximal main part $\sum_{k=1}^{M} d_k\theta^{k/M}$. Further, S_k can be interpreted as the set of states s for which the action j_s of player 2 is a best answer to the optimal action $\tilde{f}_s(\theta)$ of player 1 in the matrix game $[G_s(\theta)]$ with respect to the M-k+1 highest powers of θ. Define also $S_{M+1}:=S$ and $T_{M+1}:=S$. Observe that

$$(8.4.21)\qquad S_{k-1}\subset S_k,\ T_{k-1}\subset T_k \quad \text{for } k\in\{M+1,M,\ldots,2\}$$

$$(8.4.22)\qquad S^* = T_1.$$

From (8.4.16) we obtain

$$(8.4.23)\qquad y_{k-1\ s} \geq 0 \quad \text{for } k\in\{M+1,M,\ldots,1\} \text{ and } s\in S_k$$

$$(8.4.24)\qquad y_{k-1\ s} > 0 \quad \text{for } k\in\{M+1,M,\ldots,2\} \text{ and } s\in S_k\setminus S_{k-1}$$

and from (8.4.6)

$$(8.4.25)\qquad c_{k-1\ s} \leq d_{k-1} \quad \text{for } k\in\{M+1,M,\ldots,2\} \text{ and } s\in T_k$$

(8.4.26) $\quad c_{k-1\ s} < d_{k-1} \quad$ for $k\in\{M+1,M,\ldots,2\}$ and $s\in T_k \setminus T_{k-1}$

In lemma 8.4.4 we shall show that the following two properties hold for each $k\in\{M,M-1,\ldots,1\}$.

Property (Y_k): $T_k \subset S_k$

Property (Z_k): $P_{\ell st} = 0$ for all $s\in T_k$, $t\in S\setminus T_{k-\ell}$
$\ell\in\{0,-1,\ldots,-(M-k)\}$.

8.4.2. LEMMA. *If property* (Z_k) *holds for some* $k\in\{M,M-1,\ldots,2\}$, *then for each* $s\in T_{k-1}$ *and each* $\ell\in\{-1,-2,\ldots,-(M-k+1)\}$.

(a) $P_{\ell st} \geq 0$, *for all* $t\in T_{k-\ell}\setminus T_{k-1-\ell}$

(b) $\sum_{t\in T_{k-\ell}\setminus T_{k-1-\ell}} P_{\ell st} = -\sum_{t\in T_{k-1-\ell}} P_{\ell st}$

PROOF. By (Z_k) and (8.4.21) we have for $\ell\in\{-1,-2,\ldots,-(M-k+1)\}$, $0 \geq\tilde{\ell}>\ell$, and $t\in T_{k-\ell}\setminus T_{k-1-\ell}$: $P_{\tilde{\ell} st}=0$, which by (8.4.15) implies part (a) of the lemma. Furthermore by (Z_k) and (8.4.21)

$$P_{\ell st} = 0, \quad \text{if } t\notin T_{k-\ell},$$

which in view of (8.4.13) proves part (b) of the lemma. □

Let for fixed $k\in\{M,M-1,\ldots,2\}$

(8.4.27) $\quad u_{k0s} := -c_{k-1\ s} + \sum_{t\in S} P_{0st}\, c_{k-1\ t}$

(8.4.28) $\quad u_{k\ell s} := \sum_{t\in S} P_{\ell st} c_{k-1-\ell t}$, $\ell\in\{-1,-2,\ldots,-(M-k+1)\}$

Then from the definition of $y_{k-1\ s}$ (cf. 8.4.17) it follows that

(8.4.29) $\quad y_{k-1\ s} = \sum_{\ell=-(M-k+1)}^{0} u_{k\ell s}$.

8.4.3. LEMMA. *If the properties* (Y_k) *and* (Z_k) *hold for some* $k\in\{M,M-1,\ldots,2\}$, *then* $u_{k\ell s} = 0$ *for each* $s \in T_{k-1}$ *and each* $\ell \in \{0,-1,\ldots,-(M-k+1)\}$.

PROOF. In view of $s\in T_{k-1}\subset T_k$, (8.4.20), (Z_k), (8.4.25) and (8.4.13) we have

$$(8.4.30)\qquad u_{k0s} = -d_{k-1} + \sum_{t\in T_k} P_{0st}\cdot c_{k-1\ t} \le d_{k-1}(-1+ \sum_{t\in T_k} P_{0st}) = 0$$

And by (Z_k), (8.4.20), lemma 8.4.2(a) and (b) and (8.4.25) we have for each $\ell\in\{-1,-2,\dots,-(M-k+1)\}$

$$(8.4.31)\qquad u_{k\ell s} = \sum_{t\in T_{k-\ell}} P_{\ell st} c_{k-1-\ell\ t} =$$

$$= \sum_{t\in T_{k-\ell}\setminus T_{k-1-\ell}} P_{\ell st} c_{k-1-\ell\ t} + \sum_{t\in T_{k-1-\ell}} P_{\ell st} d_{k-1-\ell}$$

$$= \sum_{t\in T_{k-\ell}\setminus T_{k-1-\ell}} P_{\ell st}(c_{k-1-\ell\ t} - d_{k-1-\ell}) \le 0.$$

Now observe that $s\in T_{k-1}\subset T_k\subset S_k$ (since (Y_k) holds) implies with (8.4.23) that

$$(8.4.32)\qquad y_{k-1\ s} \ge 0.$$

But then the combination of (8.4.29)-(8.4.32) gives the assertion of the lemma.

□

8.4.4. LEMMA. *The properties* (Y_k) *and* (Z_k) *hold for each* $k\in\{M,M-1,\dots,1\}$.

PROOF. We prove the lemma by induction with respect to k. For k=M, we have by (8.4.16), (8.4.17) and (8.4.25) and $s\in T_M$:

$$0 \le y_{Ms} = -c_{Ms} + \sum_{t\in S} P_{0st}\cdot c_{Mt} \le d_M(-1 + \sum_{t\in S} P_{0st}) = 0.$$

Then clearly $c_{Mt}=d_M$ for t such that $P_{0st}>0$ and this proves both (Y_M) and (Z_M).

Now take $k\in\{M,M-1,\dots,2\}$. We shall show that (Y_k) and (Z_k) imply (Y_{k-1}) and (Z_{k-1}), which terminates the proof of the lemma. If we could prove that for each $s\in T_{k-1}$ the following three statements are true:

$$(8.4.33)\qquad y_{k-1\ s} = 0,$$

$$(8.4.34)\qquad P_{0st} = 0 \quad \text{if } t\in T_k\setminus T_{k-1},$$

(8.4.35) $P_{\ell st} = 0$ if $t \in T_{k-\ell} \setminus T_{k-1-\ell}$ and $\ell \in \{-1,-2,\ldots,-(M-k+1)\}$,

then we can combine (Y_k) and (8.4.33), using (8.4.21) to conclude that (Y_{k-1}) holds; and we can combine (Z_k), (8.4.34) and (8.4.35) to conclude that (Z_{k-1}) holds.
So the only assertions to prove are (8.4.33)-(8.4.35).
Fix $s \in T_{k-1}$. Firstly from (8.4.29) and lemma 8.4.3 we derive (8.4.33). Next observe from (8.4.25) and (8.4.26) that $u_{k0s}=0$ if and only if (8.4.34) holds; hence lemma 8.4.3, taking $\ell=0$, shows that (8.4.34) is true. Finally from (8.4.31), lemma 8.4.2(a) and (8.4.26) we see that $u_{k\ell s}=0$ if and only if (8.4.35) holds, but then lemma 8.4.3, taking successively $\ell=-1,-2,\ldots,-(M-k+1)$, guarantees (8.4.35).

□

8.4.5. LEMMA. *For* $s \in S^* = T_1$ *we have*

(a) $$\sum_{t \in S} \sum_{\ell=-M}^{-1} P_{\ell st} \cdot c_{-\ell t} \leq 0$$

(b) $$\sum_{t \in S} P_{0st}\, c_{Mt} = d_M .$$

PROOF. From lemma 8.4.4 (property (Z_k) for $k=1$ and $\ell=0$) and (8.4.20) we infer for $s \in S^* = T_1$:

$$\sum_{t \in S} P_{0st} c_{Mt} = \sum_{t \in T_1} P_{0st} c_{Mt} = d_M,$$

which proves part (b).
Fix $s \in S^* = T_1$.

Take $\ell \in \{-1,-2,\ldots,-M\}$. For each $\tilde{\ell} \in \{0,-1,\ldots,-\ell+1\}$ we have in view of lemma 8.4.4, taking $k=1$:

(8.4.36) $P_{\ell st} = 0$ if $t \notin T_{1-\ell}$

(8.4.37) $P_{\tilde{\ell} st} = 0$ if $t \in T_{1-\ell} \setminus T_{-\ell}$

But then by (8.4.15) and (8.4.37):

(8.4.38) $P_{\ell st} \geq 0$ if $t \in T_{1-\ell} \setminus T_{-\ell}$

And by (8.4.13) and (8.4.36):

$$(8.4.39)\qquad \sum_{t\in T_{1-\ell}\setminus T_{-\ell}} P_{\ell st} = -\sum_{t\in T_{-\ell}} P_{\ell st}.$$

Now the combination of (8.4.36), (8.4.6), (8.4.39) and (8.4.38) yields for each $\ell\in\{-1,-2,\ldots,-M\}$:

$$\sum_{t\in S} P_{\ell st}c_{-\ell t} = \sum_{t\in T_{1-\ell}\setminus T_{-\ell}} P_{\ell st}c_{-\ell t} + \sum_{t\in T_{-\ell}} P_{\ell st}c_{-\ell t} \le$$

$$\sum_{t\in T_{1-\ell}\setminus T_{-\ell}} P_{\ell st}c_{-\ell t} + \sum_{t\in T_{-\ell}} P_{\ell st}d_{-\ell} =$$

$$\sum_{t\in T_{1-\ell}\setminus T_{-\ell}} P_{\ell st}c_{-\ell t} - \sum_{t\in T_{1-\ell}\setminus T_{-\ell}} P_{\ell st}d_{-\ell} \le 0,$$

which proves part (a) of the lemma.

□

8.4.6. LEMMA. *For each* $s\in S^*$, *we have*

(a) $\quad 0 \le -c_{0s} + \sum_{t\in S^*} P_{0st}\cdot c_{0t} + R_{0s} - d_M$

(b) $\quad P_{0st} = 0 \quad$ *for* $t\in S\setminus S^*$.

PROOF. Since $S^*=T_1$, part (b) is already proved in lemma 8.4.4 (property (Z_k) for k=1 and ℓ=0).

Take $s\in S^*$. Then by property (Y_1) of lemma 8.4.4 we see that $s\in S_1$. Hence by (8.4.23) and (8.4.18):

$$(8.4.40)\qquad 0 \le y_{0s} = -c_{0s} + R_{0s} + \sum_{t\in S}\sum_{\ell=-M}^{0} P_{\ell st}c_{-\ell t} - \sum_{t\in S} P_{0st}\cdot c_{Mt}$$

Inserting lemma 8.4.5(a) and (b) into (8.4.40) proves the lemma.

□

We are now ready to prove theorem 8.4.1.

PROOF of theorem 8.4.1.

Suppose that player 1 uses the stationary strategy $\tilde{\rho}$ and player 2 the pure stationary strategy σ^p. Then by lemma 8.4.6 part (b), we see that the subset of states S^* can be treated separately, since the system cannot jump out of S^*.

Let c_0^*, P^*, Q^* and r^* be the parts of c, P, Q and r respectively that refer to S^*.
Then lemma 8.4.6(a), written in vector notation, gives (cf. (8.4.11) and (8.4.12)):

$$(8.4.41) \qquad 0 \leq c_0^* + r^*(\tilde{\rho},\sigma^P) + P^*(\tilde{\rho},\sigma^P).c_0^* - d_M.1_{|S^*|}.$$

Multiplying (8.4.41) by $Q^*(\tilde{\rho},\sigma^P)$ gives then

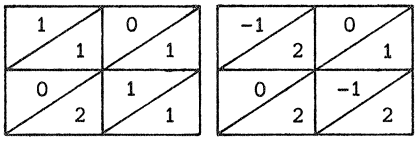
$$(8.4.42) \qquad Q^*(\tilde{\rho},\sigma^P).r^*(\tilde{\rho},\sigma^P) \geq d_M.1_{|S^*|}.$$

Thus by (8.4.42) we have proved that for the states belonging to S^*, player 1 can assure himself the average reward value by playing $\tilde{\rho}$ against each pure stationary strategy of player 2. But then by corollary 3.5 this is also the case against any strategy of player 2, by which theorem 8.4.1 is proved.

□

We conclude this section with two examples. In the first example the set of easy initial states of player 1 (player 2) coïncides with S^* (S^{**}) and $S=S^* \cup S^{**}$, $S^* \cap S^{**}=\emptyset$. In the second example S^* (S^{**}) is a proper subset of the set of easy initial states of player 1 (player 2) and there is a state which is easy for neither player.

8.4.7. EXAMPLE.

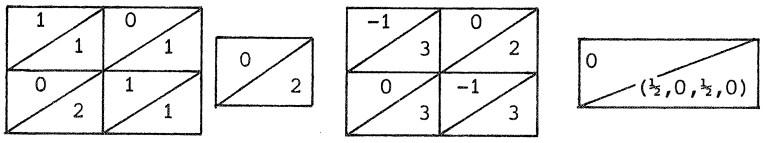


1 2

Obviously by symmetry we have $x^*(\theta)=(x_1^*,-x_1^*)$, and by solving the limit discount equation we obtain $x_1^*=\frac{1}{2}\sqrt{(2\theta+1)}-\frac{1}{2}$. Hence $S^*=\{1\}$ and $S^{**}=\{2\}$. The value of the undiscounted stochastic game is (0,0), and it can be verified that only state 1 is easy for player 1 and only state 2 is easy for player 2.
For the strategy $\tilde{\rho}$, as defined at the beginning of this section, we have

that $\tilde{\rho}$ prescribes the choice of the first row in both states ($\tilde{\rho}$ is unique in this game). This strategy is optimal for player 1 if the game starts in state 1.

8.4.8. EXAMPLE.



1 2 3 4

Here $x^*(\theta)=(x_1^*,0,-x_1^*,0)$, and by solving the limit discount equation we obtain

$$x_1^* = (1+\theta^{-1})(\sqrt{(\theta+1)}-1)$$

This implies that $S^*=\{1\}$, $S^{**}=\{3\}$ and that the value of the undiscounted game is (0,0,0,0).

State 4 is not easy for either player, and state 1 (state 3) is not an easy initial state for player 2 (player 1), while state 2 is easy for both players.

9. Algorithms for undiscounted stochastic games.

In section 9.2 we shall show how stochastic games in which one player controls the transitions can be solved by a linear programming problem. In section 9.3 we consider the switching control stochastic game, by which we mean a game such that in each state only one of the players, but not in each state necessarily the same one, controls the transitions. It results that this class of games can be solved by a finite sequence of linear programming problems.

In section 9.1 we give a review of well-known algorithms for undiscounted stochastic games.

9.1. SOME KNOWN ALGORITHMS.

The first algorithm we describe is the algorithm of Hoffman & Karp (1966). It can be applied to irreducible stochastic games, i.e. games for which, for each pair of pure stationary strategies, the corresponding stochastic matrix has a single ergodic class and no transient states. In part II, algorithm 6.1.2, we have given the discounted version of this algorithm.

9.1.1. ALGORITHM (Hoffman & Karp).

(i) Choose $v^0 \in \mathbb{R}^z$ such that $v_1^0=0$; let $\tau:=0$.

(ii) Determine a stationary strategy $\sigma^\tau=(\sigma_1^\tau,\ldots,\sigma_z^\tau)$, such that σ_s^τ is an optimal action for player 2 in the matrix game $[G_s(v^\tau)]$.

(iii) Solve the Markov decision problem, MDP(σ^τ), which results when player 2 fixes σ^τ. This solution corresponds to the unique solution of the following set of equations:

$$g^{\tau+1}+v_s^{\tau+1} = \max_{i\in A_s} \{r(s,i,\sigma_s^\tau)+ \sum_{t=1}^{z} p(t|s,i,\sigma_s^\tau).v_t^{\tau+1}\}, \quad s\in S$$

$$v_1^{\tau+1} = 0.$$

(iv) $\tau:=\tau+1$ and go to step (ii).

Hoffman & Karp have shown that the sequence (g^τ, v^τ), $\tau=1,2,\ldots$ has a limit, say (g^*, v^*), for which

$$(9.1.1) \qquad g^*+v^*_s = \operatorname{Val}_{A_s \times B_s} (G_s(v^*)), \quad s\in S.$$

From lemma 8.1.3 we recall that (9.1.1) implies that the value equals $g^*.1_z$, while optimal actions in $[G_s(v^*)]$ for the players provide optimal stationary strategies.

Federgruen (1978) has given two algorithms for undiscounted stochastic games. The first one can be applied succesfully whenever the stochastic game has the properties (a) both players have optimal stationary strategies, and (b) the average payoff value is independent of the initial state. This algorithm proceeds as follows.

9.1.2. ALGORITHM (Federgruen)

(i) Fix a sequence α_τ, $\tau=1,2,\ldots$, which satisfies

$$(1-\alpha_\tau)(1-\alpha_{\tau-1})\ldots(1-\alpha_1) \to 0 \quad \text{as } \tau\to\infty$$

$$\sum_{\ell=2}^{\tau} (1-\alpha_\tau)\ldots(1-\alpha_\ell)\,|\alpha_\ell^{1/M}-\alpha_{\ell-1}^{1/M}| \to 0 \quad \text{as } \tau\to\infty$$

(for example the choice $\alpha_\tau=\tau^{-\gamma}$ with $0<\gamma\leq 1$ satisfies these relations)

(ii) Choose $v^1\in\mathbb{R}^z$ such that $v^1_1=0$.

(iii) Calculate recursively for $\tau=1,2,\ldots$

$$g^{\tau+1} = \operatorname{Val}_{A_1\times B_1} \left(r(1,.,.)+(1+\alpha_\tau)^{-1} \sum_{t=1}^{z} p(t|1,.,.)v_t^\tau\right)$$

$$v_s^{\tau+1} = \operatorname{Val}_{A_s\times B_s} \left(r(s,.,.)+(1+\alpha_\tau)^{-1} \sum_{t=1}^{z} p(t|s,.,.)v_t^\tau\right)-g^{\tau+1}, \quad s\in S.$$

In fact this algorithm is an extension of the modified value-iteration method of Hordijk & Tijms (1975) to stochastic games.

The way in which this algorithm approximates the value $g^*.1_z$ of the game and produces ε-optimal stationary strategies is given by the following properties. Let

$$\underline{m}_\tau = \min_{s\in S} \{v_s^{\tau+1}+g^{\tau+1}-(1+\alpha_\tau)^{-1}v_s^\tau\}$$

$$\bar{m}_\tau = \max_{s\in S} \{v_s^{\tau+1}+g^{\tau+1}-(1+\alpha_\tau)^{-1}v_s^\tau\}.$$

Then

(a) $\lim_{\tau\to\infty} \underline{m}_\tau = \lim_{\tau\to\infty} \bar{m}_\tau = \lim_{\tau\to\infty} g^\tau = g^*$.

(b) $\lim_{\tau\to\infty} v^\tau$ exists and equals v^* (say), where

$$g^*+v_s^* = \operatorname{Val}_{A_s\times B_s} (G_s(v^*)), \quad s\in S$$

(c) $\underline{m}_\tau \le g^* \le \bar{m}_\tau$ and $\underline{m}_\tau \le g^\tau \le \bar{m}_\tau$

(d) If $\bar{m}_\tau - \underline{m}_\tau < \varepsilon$, then g^τ is an ε-approximation of g^*, and a stationary strategy for a player built up from optimal actions of the matrix games $[r(s,.,.)+(1+\alpha_\tau)^{-1}\sum_{t=1}^{z} p(t|s,.,.)v_t^\tau]$ is ε-optimal for that player.

The second algorithm of Federgruen (1978) can be applied to stochastic games for which for each pair of pure stationary strategies, the corresponding Markov chain is unichained (transient states are allowed). In addition it is assumed that the data-transformation of Schweitzer (1971) is carried out in order to ensure the strong aperiodicity property, i.e. $p(t|s,i,j)>0$ for all t,s,i,j. For this class of games the data-transformation leaves unchanged both the value of the game and the sets of optimal stationary strategies of the players. Furthermore the aperiodicity property ensures the convergence of the sequence of vectors $w^\tau-\tau.g^*.1_z$, $\tau=0,1,\ldots$, where w^0 is arbitrary and

$$(9.1.2) \qquad w_s^\tau := \operatorname{Val}_{A_s\times B_s} (G_s(w^{\tau-1})) \quad \text{for } \tau\ge 1 \text{ and } s\in S.$$

This leads to the following algorithm, which bears a resemblance to algorithm 9.1.2.

9.1.3. ALGORITHM (Federgruen).

(i) Choose $v^1 \in \mathbb{R}^Z$, such that $v_1^1 = 0$.

(ii) Calculate recursively for $\tau = 1,2,\ldots$

$$g^{\tau+1} = \operatorname*{Val}_{A_1 \times B_1} \left(r(1,.,.) + \sum_{t=1}^{z} p(t|1,.,.) v_t^\tau \right)$$

$$v_s^{\tau+1} = \operatorname*{Val}_{A_s \times B_s} \left(r(s,.,.) + \sum_{t=1}^{z} p(t|s,.,.) v_t^\tau \right) - g^{\tau+1} \qquad s \in S.$$

Federgruen has shown that for this algorithm the same relations hold as for algorithm 9.1.2, mentioned under (a)-(d) (now $\alpha_\tau = 0$ in the definitions of $\underline{m}_\tau$ and $\bar{m}_\tau$). In addition, the convergence of $\underline{m}_\tau$ and $\bar{m}_\tau$ is monotone. Furthermore the convergence rate is geometric as a consequence of the data-transformation.

Finally we wish to mention some of the results of Van der Wal (1981). His algorithm agrees with the standard method of successive approximations, which is represented by equation (9.1.2). A scheme which reflects his algorithm would be algorithm 9.1.3 above. Van der Wal shows that this algorithm can be applied to two classes of stochastic games.

The first one is the same class as Federgruen treats with algorithm 9.1.3. Van der Wal obtained similar results as Federgruen. Moreover he proved that under the strong aperiodicity assumption it holds that

$$w^{\tau+1} - w^\tau = g^* . 1_z + O((1-\gamma)^{\tau/(z-1)})$$

if $\tau \to \infty$, where γ is a constant, $0 < \gamma \leq 1$, depending on the transition probabilities of the game.

In the second case Van der Wal (1981) considers the class of games for which the set of equations

$$v_s + g = \operatorname*{Val}_{A_s \times B_s} (G_s(v)), \qquad s \in S$$

has a solution. If in addition the data transformation is carried out, then he shows that the following holds. Putting $g_s^\tau := w_s^{\tau+1} - w_s^\tau$, $s \in S$ (w^τ is defined as in (9.1.2)) and $\bar{m}_\tau := \max_{s \in S} g_s^\tau$ and $\underline{m}_\tau := \min_{s \in S} g_s^\tau$, then both $\bar{m}_\tau$ and $\underline{m}_\tau$ converge

monotonically to g^*, where $g^*.1_z$ is the average payoff value of the game. This demonstrates that in this case also the method of standard successive approximation yields an ε-band around the value of the game and nearly optimal stationary strategies for the players.

9.2. STOCHASTIC GAMES WHERE ONE PLAYER CONTROLS THE TRANSITIONS.

In this section a linear programming problem is elaborated which solves the class of games for which only one of the players governs the transitions. A consequence of our analysis is a constructive proof of the orderfield property for this class of games (cf. Parthasarathy & Raghavan (1981)).

A stochastic game in which the transition probabilities are controlled by one player is a special case of a switching control stochastic game as defined in definition 6.3.1.

9.2.1. DEFINITION. *A stochastic game in which player* 2 *governs the transitions is a game for which* $p(t|s,i,j)=p(t|s,\hat{i},j)$ *for all* $(i,\hat{i})\in A_s\times A_s$, *each* $j\in B_s$ *and each* $(s,t)\in S\times S$. *This probability is abbreviated to* $p(t|s,j)$.

Stern (1975) proved that such games have an average payoff value. Bewley & Kohlberg (1978) and Parthasarathy & Raghavan (1978) independently showed that both players possess optimal stationary strategies. In addition Parthasarathy & Raghavan have proved the orderfield property for this class of games, i.e. the property that a solution of the game lies in the same Archimedean field as its parameters. By a solution we mean the value plus optimal stationary strategies. Given this property it was to be expected that a solution could be found by a finite procedure. Filar & Raghavan (1979) have given such a procedure. However their algorithm is rather cumbersome, since many computations have to be done. For instance for each pair of pure stationary strategies the average payoff must be computed. From these payoffs z matrix games are constructed. The values of these matrix games correspond to the value of the stochastic game. However to find optimal stationary strategies still further calculations must be carried out.

Below we will formulate a linear programming problem, whose solution gives in one blow the value of the game and optimal stationary strategies for both players. The results of this section are deduced from Vrieze (1981a).

Independently Hordijk & Kallenberg (1981b) have proposed a similar algorithm. They analysed their linear programming problem in detail and have stated some interesting properties of it.

We first state the algorithm, then show that this linear programming problem has a solution and lastly prove that this solution corresponds to a solution of the stochastic game.

Consider the following linear programming problem in the variables $(g_1,\ldots,g_z)$, $(v_1,\ldots,v_z)$, $x_s(i)$, $s\in S$, $i\in A_s$.

9.2.2. ALGORITHM.

LP1:

$$\max \sum_{s=1}^{z} g_s, \text{ subject to}$$

(i) $$g_s - \sum_{t=1}^{z} p(t|s,j).g_t \leq 0 \quad , s\in S,\ j\in B_s$$

(ii) $$g_s + v_s - \sum_{i\in A_s} x_s(i).r(s,i,j) - \sum_{t=1}^{z} p(t|s,j).v_t \leq 0 \quad , s\in S,\ j\in B_s$$

(iii) $$\sum_{i\in A_s} x_s(i)=1,\ x_s(i)\geq 0 \quad , s\in S,\ i\in A_s.$$

The dual linear programming problem in the variables $(w_1,\ldots,w_z)$, $y_s(j)$, $z_s(j)$, $s\in S$, $j\in B_s$ is:

DLP1:

$$\min \sum_{s=1}^{z} w_s \text{ subject to}$$

(j) $$\sum_{s=1}^{z} \sum_{j\in B_s} (\delta_{st} - p(t|s,j))y_s(j) + \sum_{j\in B_t} z_t(j) = 1 \quad , t\in S$$

(jj) $$\sum_{s=1}^{z} \sum_{j\in B_s} (\delta_{st} - p(t|s,j))z_s(j) = 0 \quad , t\in S$$

(jjj) $$-\sum_{j\in B_s} z_s(j).r(s,i,j) + w_s \geq 0 \quad , s\in S,\ i\in A_s$$

(jv) $$y_s(j),\ z_s(j)\geq 0 \quad , s\in S,\ j\in B_s.$$

(Here $\delta_{st} = 1$ if $s=t$ and $\delta_{st} = 0$ else.)

Note that the set $X:=\{x \mid x:=\{x_s(i);s\in S,\ i\in A_s\},\ x$ satisfies (iii)$\}$ corresponds to the set of stationary strategies of player 1 in a one to one manner in the following way. If $x\in X$, then define

(9.2.1) $\quad \rho^x=(\rho_1^x,\ldots,\rho_z^x)$ by $\rho_s^x(i):=x_s(i)$,

which obviously is a stationary strategy for player 1 in view of condition (iii). On the other hand for a stationary strategy ρ define

(9.2.2) $\quad x^\rho$ by $x_s^\rho(i):=\rho_s(i)$.

Then $x^\rho\in X$ and clearly $\rho^{(x^\rho)}=\rho$ and $x^{(\rho^x)}=x$.

Intuitively, for any state s the numbers $z_s(j)$, $j\in B_s$ are proportional to the probabilities of player 2 choosing his pure actions at state s.

9.2.3. LEMMA. *Both linear programming problems are feasible and have bounded solutions.*

PROOF. Consider the primal problem. Observe that if $\hat{g}_s=\min_{(t,i,j)} r(t,i,j)$, each $s\in S$ and $\hat{v}_s=0$, $s\in S$ and $\hat{x}\in X$ arbitrary, then $(\hat{g},\hat{v},\hat{x})$ satisfies (i)-(iii), and thus the primal problem is feasible.

Now let (g,v,x) be a feasible solution and let σ^p be an arbitrary pure stationary strategy of player 2. Then from (i) and (ii) we obtain (in vector notation):

(9.2.3) $\quad g\leq P(\sigma^p).g$

(9.2.4) $\quad g+v\leq r(\rho^x,\sigma^p)+P(\sigma^p).v$

(Since, for a pair of stationary strategies ρ and σ the transition probability matrix depends only on σ we will write $P(\sigma)$ instead of $P(\rho,\sigma)$.) From (9.2.3) and (9.2.4) we derive by lemma 8.1.1:

(9.2.5) $\quad g\leq Q(\sigma^p).r(\rho^x,\sigma^p)$

Now (9.2.5) shows that g is bounded from above (e.g. by $\max_{(s,i,j)} r(s,i,j)$) uniformly in the feasible solutions (g,v,x).
Hence a finite optimal solution to the primal problem exists.

From the duality theorem for linear programming problems it follows that the dual problem also is feasible and has a bounded solution.

□

9.2.4. LEMMA. *Let* (g,v,x) *be a feasible solution to the primal problem. Then*

$$\min_{\nu} W_{\rho} x_{\nu} \geq g.$$

PROOF. This result can be deduced immediately from (9.2.5) by using corollary 3.5.

□

In the following we will frequently use some properties of the Cesaro limit Q of a stochastic matrix P. These properties are outlined in section 8.1. We recall that if P is a stochastic matrix corresponding to an ergodic Markov chain, then $Q:=\lim_{T\to\infty}(T+1)^{-1}\Sigma_{\tau=0}^{T}P^{\tau}$ has the properties: (a) each row of Q is identical and equal to the law vector q (say), (b) q is the invariant distribution of P, i.e. the unique solution to qP=q, $\Sigma_{s\in S} q_s=1$, and (c) each component of q is strictly positive.

Associated with a feasible solution (w,y,z) to the dual program, we define a number of quantities:

(9.2.6) $\quad u_s := \Sigma_{j\in B_s} z_s(j), \quad s\in S$ and $u:=(u_1,\ldots,u_z)$

(9.2.7) $\quad S_0:=\{s \mid s\in S \text{ and } u_s=0\}$

(9.2.8) $\quad \tilde{z}_s(j):=z_s(j)/u_s, \quad s\in S\setminus S_0$ and $j\in B_s$

(9.2.9) $\quad d_s := \Sigma_{j\in B_s} (y_s(j)+z_s(j)), \quad s\in S$

(9.2.10) $\quad \tilde{y}_s(j):=(y_s(j)+z_s(j))/d_s, \quad s\in S,\ j\in B_s$

Observe that from condition (j) we have $\Sigma_{j\in B_t} y_t(j) > 0$ if $\Sigma_{j\in B_t} z_t(j) = 0$ thus $d_s > 0$ for each $s\in S$.

Further, stationary strategies $\tilde{\sigma}$ and σ^* for player 2 are defined as:

(9.2.11) $\quad \tilde{\sigma}_s(j) := \tilde{y}_s(j), \quad s \in S,\ j \in B_s$

(9.2.12) $\quad \sigma^*_s(j) := \tilde{z}_s(j),\ s \in S \setminus S_0,\ j \in B_s$ and $\sigma^*_s(j) := \tilde{y}_s(j),\ s \in S_0,\ j \in B_s$

It will result that σ^* is optimal for player 2 if it is associated with an optimal solution (w,y,z) of DLP1.

9.2.5. REMARK. *Observe from the conditions* (j) *and* (jj) *that*:

$$\sum_{s=1}^{z} \sum_{j \in B_s} (\delta_{st} - p(t|s,j))(y_s(j) + z_s(j)) + \sum_{j \in B_t} z_t(j) = 1, \quad t \in S$$

which in view of (9.2.6), (9.2.9) *and* (9.2.11) *is equivalent to*

(9.2.13) $\quad d_t - \sum_{s=1}^{z} p(t|s,\tilde{\sigma}_s).d_s + u_t = 1, \quad t \in S$

In the following we assume that, for a stationary strategy σ of player 2, after suitable rearranging of the states, $P(\sigma)$ has the form as in (8.1.3). So

$$P(\sigma) = \left(\begin{array}{ccc|c} P_{11}(\sigma) & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & P_{\gamma\gamma}(\sigma) & 0 \\ \hline P_{\gamma+1\,1}(\sigma) & \cdots & P_{\gamma+1\,\gamma}(\sigma) & P_{\gamma+1\,\gamma+1}(\sigma) \end{array}\right)$$

$P_{nn}(\sigma)$ corresponds to the n-th ergodic class of $P(\sigma)$, whose set of states is denoted by $E_n(\sigma)$, $n \in \{1,\ldots,\gamma\}$, and $P_{\gamma+1\,\gamma+1}(\sigma)$ corresponds to the transient states of $P(\sigma)$. This set of transient states is denoted by $T(\sigma)$. In the following $\tilde{\gamma}$ corresponds to $\tilde{\sigma}$ and γ^* to σ^*.
If $P=(p(t|s))$ corresponds to a stochastic matrix, then for each $c \in \mathbb{R}^z$:

(9.2.14) $\quad \sum_{t=1}^{z} (c_t - \sum_{s=1}^{z} p(t|s)c_s) = 0$

as a consequence of $\sum_{t=1}^{z} p(t|s) = 1$.

9.2.6. LEMMA. (a) $u=u.P(\sigma^*)$.

(b) *the transient states of* $P(\sigma^*)$ *are exactly the states* S_0.

PROOF. Part (a) follows immediately after inserting definitions (9.2.6) and (9.2.12) into condition (jj). Concerning part (b), note first that summing up condition (j) over $t\in S$ yields

$$\sum_{s=1}^{z} u_s = z > 0 \tag{9.2.15}$$

Now, by (9.2.15) and part (a) of the lemma, it follows from the theory of Markov chains that u can be written as a linear combination of the invariant distributions of $P(\sigma^*)$:

$$u=\lambda_1(q_1:0:\ldots:0)+\lambda_2(0:q_2:0:\ldots:0)+\ldots+\lambda_{\gamma^*}(0:\ldots:0:q_{\gamma^*}:0),$$

with $\lambda_n\geq 0$, $\sum_{n=1}^{\gamma^*}\lambda_n=z$ and where q_n equals the unique invariant distribution of $P_{nn}(\sigma^*)$, $n\in\{1,\ldots,\gamma^*\}$. Note that the vector q_n is strictly positive. Hence if $u_s>0$, then $s\in E_{\tilde{n}}(\sigma^*)$ for some $\tilde{n}\in\{1,\ldots,\gamma^*\}$, and moreover $u_t>0$ for each $t\in E_{\tilde{n}}(\sigma^*)$. Therefore, if we wish to show that S_0 is exactly the set of transient states of $P(\sigma^*)$, it suffices to show that no ergodic class lies entirely within S_0.

Suppose then for some $n\in\{1,\ldots,\gamma^*\}$ that $E_n(\sigma^*)\subset S_0$.

Summing up (9.2.13) over $t\in E_n(\sigma^*)$ yields in view of (9.2.14) (here $|E_n(\sigma^*)|$ equals the number of states of ergodic class n):

$$-\sum_{t\in E_n(\sigma^*)}\ \sum_{s\notin E_n(\sigma^*)}\ \sum_{j\in B_s} p(t|s,j).y_s(j) = |E_n(\sigma^*)|$$

As $y_s(j)\geq 0$ the left hand side is non-positive, but then we have a contradiction since the right hand side is strictly positive. This shows that the assumption $E_n(\sigma^*)\subset S_0$ is wrong, which proves the lemma. □

9.2.7. COROLLARY. u *can be written as*

$$u=\lambda_1(q_1:0:\ldots:0)+\lambda_2(0:q_2:0:\ldots:0)+\ldots+\lambda_{\gamma^*}(0:\ldots:0:q_{\gamma^*}:0)$$

with $\lambda_n>0$, $n\in\{1,\ldots,\gamma^*\}$, $\sum_{n=1}^{\gamma^*}\lambda_n=z$, *and where* q_n *equals the invariant distribution of* $P_{nn}(\sigma^*)$.

9.2.8. COROLLARY. *Let* ρ^p *be a pure stationary strategy for player* 1. *Then for each* $n\in\{1,\ldots,\gamma^*\}$:

$$(9.2.16)\qquad (a)\quad \sum_{s\in E_n(\sigma^*)} \sum_{j\in B_s} z_s(j).r(s,\rho^p_s,j) = \lambda_n W_{\rho^p\sigma^*}(n)$$

$$(9.2.17)\qquad (b)\quad \sum_{s=1}^{z} \sum_{j\in B_s} z_s(j).r(s,\rho^p_s,j) = \sum_{n=1}^{\gamma^*} \lambda_n . W_{\rho^p\sigma^*}(n)$$

(Here $W_{\rho^p\sigma^*}(n)$ *equals the expected payoff for the pair of strategies* (ρ^p,σ^*) *when the starting state of the specific play belongs to* $E_n(\sigma^*)$; *remember* (cf. (8.1.4) *and* (8.1.6)) *that* $W_{\rho^p\sigma^*}$ *is constant on an ergodic set).*

Corollary 9.2.8 can be verified by writing $z_s(j) = \tilde{z}_s(j).u_s$, $s\in S\setminus S_0$, next inserting the expression for u of corollary 9.2.7 in the left hand sides of (9.2.16) and (9.2.17), and by realizing that

$$\sum_{s\in E_n(\sigma^*)} q_n(s).r(s,\rho^p_s,\sigma^*_s) = W_{\rho^p\sigma^*}(n).$$

From now on (g,v,x) and (w,y,z) are assumed to correspond to a dual pair of optimal solutions.

9.2.9. LEMMA. (a) $w_s=0$ *for* $s\in S_0$

$$(b)\qquad \sum_{s\in E_n(\sigma^*)} w_s = \lambda_n \max_{\rho^p} W_{\rho^p\sigma^*}(n)$$

PROOF. Part (a) follows directly from $u_s=0$, $s\in S_0$, condition (jjj), and the fact that the dual LP is a minimization problem.
Part (b) follows from Corollary 9.2.8 (a).

□

9.2.10. LEMMA. $P(\sigma^*).g=g$ *and* $P(\tilde{\sigma}).g=g$.

PROOF. From $P(\sigma^*).g\geq g$ (condition (i)) it follows that the equality sign holds for the components belonging to the set $S\setminus S_0$ of recurrent states of $P(\sigma^*)$ (see lemma 8.1.1(a)). Hence we have for $s\in S$ and $j\in B_s$:

(9.2.18) if $z_s(j)>0$, then $\Sigma_{t=1}^{z} p(t|s,j).g_t = g_s$.

From the complementary slackness property we obtain for $s\in S$ and $j\in B_s$:

(9.2.19) if $y_s(j)>0$, then $\Sigma_{t=1}^{z} p(t|s,j).g_t = g_s$.

(9.2.18) and (9.2.19) together with the definitions of σ^* and $\tilde{\sigma}$ (see (9.2.11) and (9.2.12)) give the lemma.

□

9.2.11. COROLLARY.

(a) *For each* $n\in\{1,\ldots,\gamma^*\}$, g *is constant on* $E_n(\sigma^*)$.

(b) *For each* $n\in\{1,\ldots,\tilde{\gamma}\}$, g *is constant on* $E_n(\tilde{\sigma})$.

In the following g(n), for $n\in\{1,\ldots,\gamma^*\}$, denotes the constant value of g on $E_n(\sigma^*)$. Similarly for $n\in\{1,\ldots,\tilde{\gamma}\}$.

9.2.12. LEMMA. *For* $(n_1,n_2)\in\{1,\ldots,\tilde{\gamma}\}\times\{1,\ldots,\gamma^*\}$ *we have either*

$E_{n_1}(\tilde{\sigma})\cap E_{n_2}(\sigma^*) = E_{n_2}(\sigma^*)$, *or*

$E_{n_1}(\tilde{\sigma})\cap E_{n_2}(\sigma^*) = \emptyset$.

PROOF. Let $s\in E_{n_1}(\tilde{\sigma})\cap E_{n_2}(\sigma^*)$ and let $t\in E_{n_2}(\sigma^*)$.
Then s and t communicate under $P(\sigma^*)$. But since $\tilde{y}_s(j)=\tilde{\sigma}_s(j)>0$ if $\tilde{z}_s(j)=\sigma^*_s(j)>0$ (see 9.2.10), it follows that s and t also communicate under $P(\tilde{\sigma})$, so $t\in E_{n_1}(\tilde{\sigma})$.

□

Lemma 9.2.12 implies that the following are sensible definitions. Let

$$D_{\tilde{n}} := \{n \mid E_{\tilde{n}}(\tilde{\sigma})\supset E_n(\sigma^*)\}, \quad \tilde{n}\in\{1,\ldots,\tilde{\gamma}\}$$

$$T_{\tilde{n}} := E_{\tilde{n}}(\tilde{\sigma})\cap T(\sigma^*) \quad , \tilde{n}\in\{1,\ldots,\tilde{\gamma}\}$$

$$T := \{n \mid T(\tilde{\sigma})\supset E_n(\sigma^*)\}$$

$$TT := T(\tilde{\sigma})\cap T(\sigma^*)$$

Obviously

$$(9.2.20)\qquad \{1,\ldots,\gamma^*\} = \Big(\bigcup_{\tilde{n}=1}^{\tilde{\gamma}} D_{\tilde{n}}\Big) \cup T \text{ and } S_0 = \Big(\bigcup_{\tilde{n}=1}^{\tilde{\gamma}} T_{\tilde{n}}\Big) \cup TT$$

For a finite set C we mean by $|C|$ the number of elements of C.

9.2.13. LEMMA. *For* $\tilde{n}\in\{1,\ldots,\tilde{\gamma}\}$ *we have*

$$\sum_{t\in E_{\tilde{n}}(\tilde{\sigma})} u_t = \sum_{n\in D_{\tilde{n}}} \lambda_n =$$

$$= \sum_{n\in D_{\tilde{n}}} |E_n(\sigma^*)| + |T_{\tilde{n}}| + \sum_{t\in E_{\tilde{n}}(\tilde{\sigma})} \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}).d_s$$

PROOF. The first equality follows from corollary 9.2.7. The second can be checked by summing up (9.2.13) over $t\in E_{\tilde{n}}(\tilde{\sigma})$ and using 9.2.14 (with $P=P_{nn}(\tilde{\sigma})$).

□

From (9.2.13) we can also infer:

$$(9.2.21)\qquad d_t = 1 + \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma})d_s \quad \text{for } t\in TT.$$

And after summing up (9.2.13) over $t\in E_n(\sigma^*)$ with $n\in T$ and using corollary 9.2.7 we obtain

$$(9.2.22)\qquad \lambda_n = |E_n(\sigma^*)| - \sum_{t\in E_n(\sigma^*)} d_t + \sum_{t\in E_n(\sigma^*)} \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}).d_s$$

9.2.14. LEMMA.

$$\max_{\rho P} W_{\rho P \sigma^*}(n) = g(n), \qquad n\in\{1,\ldots,\gamma^*\}$$

PROOF. From the duality theorem for linear programs, lemma 9.2.9(a) and (b) and (9.2.20) we obtain:

$$(9.2.23)\qquad \sum_{s\in S_{\tilde{\gamma}}} g_s = \sum_{s\in S} w_s = \sum_{n=1}^{\gamma^*} \lambda_n . \max_{\rho P} W_{\rho P \sigma^*}(n)$$

$$= \sum_{\tilde{n}=1}^{\tilde{\gamma}} \sum_{n\in D_{\tilde{n}}} \lambda_n \max_{\rho P} W_{\rho P \sigma^*}(n) + \sum_{n\in T} \lambda_n . \max_{\rho P} W_{\rho P \sigma^*}(n).$$

Now observe that, using Corollary 3.5, lemma 9.2.4, and the fact that $W_{\rho\sigma^*}$ is constant on $E_n(\sigma^*)$ for each ρ:

$$\max_{\rho P} W_{\rho P_{\sigma^*}}(n) \geq W_{\rho x_{\sigma^*}}(n) \geq g(n), \tag{9.2.24}$$

for all $n\in\{1,\ldots,\gamma^*\}$.

Substituting inequality (9.2.24) into (9.2.23) and recalling that $g(n)=g(\tilde{n})$ for $n\in D_{\tilde{n}}$ yields:

$$\sum_{s\in S} g_s \geq \sum_{\tilde{n}=1}^{\tilde{\gamma}} g(\tilde{n}) \cdot \sum_{n\in D_{\tilde{n}}} \lambda_n + \sum_{n\in T} \lambda_n . g(n). \tag{9.2.25}$$

Inserting the expression for $\sum_{n\in D_{\tilde{n}}} \lambda_n$ of lemma 9.2.13 and also relation (9.2.22) into (9.2.25) leads to:

$$\begin{aligned}
\sum_{s\in S} g_s \geq{}& \sum_{\tilde{n}=1}^{\tilde{\gamma}} g(\tilde{n}) \sum_{n\in D_{\tilde{n}}} |E_n(\sigma^*)| + \sum_{\tilde{n}=1}^{\tilde{\gamma}} g(\tilde{n}) |T_{\tilde{n}}| \\
&+ \sum_{\tilde{n}=1}^{\tilde{\gamma}} g(\tilde{n}) \sum_{t\in E_{\tilde{n}}(\tilde{\sigma})} \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}) d_s + \sum_{n\in T} g(n) . |E_n(\sigma^*)| \\
&- \sum_{n\in T} g(n) \sum_{t\in E_n(\sigma^*)} d_t + \sum_{n\in T} g(n) \sum_{t\in E_n(\sigma^*)} \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}) d_s \\
={}& \sum_{s\in S\setminus S_0} g_s \quad \text{(the first + the fourth term)} \\
&+ \sum_{s\in S_0 \cap (S\setminus T(\tilde{\sigma}))} g_s \quad \text{(the second term)} \\
&+ \sum_{s\in S_0 \cap T(\tilde{\sigma})} g_s \quad \text{(extra term added)} \\
&- \sum_{t\in TT} g_t d_t + \sum_{t\in TT} \sum_{s\in T(\tilde{\sigma})} g_t . p(t|s,\tilde{\sigma}) . d_s \\
&\qquad \text{(in view of (9.2.21), the same extra term subtracted)} \\
&+ \sum_{t\in S\setminus TT} g_t \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}) d_s \quad \text{(the third and the sixth term)} \\
&- \sum_{n\in TT} g(n) \sum_{t\in E_n(\sigma^*)} d_t \quad \text{(the fifth term)} \\
={}& \sum_{s\in S} g_s + \sum_{t\in S} g_t \sum_{s\in T(\tilde{\sigma})} p(t|s,\tilde{\sigma}) . d_s - \sum_{t\in T(\tilde{\sigma})} g_t d_t .
\end{aligned} \tag{9.2.26}$$

But since $\sum_{t\in S} p(t|s,\tilde{\sigma}).g_t=g_s$ by lemma 9.2.10, it follows that the second and the third term of the last expression of (9.2.26) sum up to zero. Then (9.2.26) results in the inequality:

$$\sum_{s\in S} g_s \geq \sum_{s\in S} g_s$$

Hence in (9.2.25) the equality sign must hold, and since $\lambda_n>0$ for each $n\in\{1,\ldots,\tau^*\}$ this implies that the equality sign holds in (9.2.24) also, which proves the lemma.

□

Now we can state the main theorem of this section.

9.2.15. THEOREM. *Algorithm* 9.2.2 *provides a solution method for undiscounted stochastic games in which player 2 controls the transitions. If* (g,v,x) *and* (w,y,z) *is a dual pair of optimal solutions to this program, then* g *equals the value of the game,* ρ^x *(cf.* (9.2.1)*) is an optimal stationary strategy for player* 1 *and* σ^* *(cf.* (9.2.12)*) is an optimal stationary strategy for player* 2.

PROOF. From lemma 9.2.14 we obtain

$$\max_{\rho^p} W_{s\rho^p\sigma^*} = g_s \quad \text{if } s\in S\setminus S_0. \tag{9.2.27}$$

$P(\sigma^*)$ is independent of ρ^p, so for each ρ^p we have the same configuration of recurrent and transient states, namely $S\setminus S_0$ is the set of recurrent states and S_0 is the set of transient states (see lemma 9.2.6).
By lemma 9.2.10, $P(\sigma^*)g=g$. From this equality and from (9.2.27) we derive along the same lines as (8.1.11), with inequality signs replaced by equality signs, that

$$\max_{\rho^p} W_{s\rho^p\sigma^*} = g_s \quad \text{for } s\in S_0.$$

Thus

$$\max_{\rho^p} W_{\rho^p\sigma^*} = g.$$

But then by corollary 3.5 and lemma 9.2.4 we see that for all strategies μ and ν

$$W_{\mu\sigma^*} \leq W_{\rho^x\sigma^*} = g \leq W_{\rho^x\nu}.$$

These inequalities prove theorem 9.2.15 as a consequence of theorem 2.3.4. □

9.2.16. REMARK. If in each state player 1 has only one action, then the game reduces to a minimizing Markov decision problem. In this case our algorithm results in an algorithm presented by Hordijk & Kallenberg (1979). Parts of their proofs could be projected on our problem; in particular the fact that S_0 is exactly the set of transient states for $P(\sigma^*)$ could be proved for both cases in an analogous way. The problem of proving the optimality of σ^* is essentially different. Following their line of argument lead to the result that σ^* is "optimal" against all ρ^p, such that $\rho^p \in \overset{z}{\underset{s=1}{X}} \tilde{A}_s$, where $\tilde{A}_s := \{i \mid i \in A_s \text{ and } \rho_s^x(i) > 0\}$.
Clearly this is not enough for showing the optimality of σ^*.

9.2.17. REMARK. For the important case in which, for each pure stationary strategy σ^p, the probability matrix $P(\sigma^p)$ has a single ergodic class and no transient states, both the algorithm and the proofs can be considerably simplified. The algorithm now becomes:

maximize the scalar g, subject to

$$\text{(i)} \qquad g + v_s - \sum_{i \in A_s} x_s(i).r(s,i,j) - \sum_{t=1}^{z} p(t \mid s,j).v_t \leq 0 \quad s \in S,\ j \in B_s$$

$$\text{(ii)} \qquad \sum_{i \in A_s} x_s(i) = 1 \text{ and } x_s(i) \geq 0 \quad , s \in S,\ i \in A_s$$

The dual of this linear programming problem is:

$$\min \sum_{s=1}^{z} w_s, \quad \text{subject to}$$

$$\text{(j)} \qquad \sum_{s=1}^{z} \sum_{j \in B_s} z_s(j) = 1$$

$$\text{(jj)} \qquad \sum_{s=1}^{z} \sum_{j \in B_s} (\delta_{st} - p(t \mid s,j)) z_s(j) = 0 \quad , t \in S.$$

(jjj) $\quad -\sum_{j\in B_s} z_s(j).r(s,i,j)+w_s \geq 0 \quad , s\in S,\ i\in A_s$

(jv) $\quad z_s(j) \geq 0 \quad , s\in S,\ j\in B_s$

In this case the stationary strategy σ^* with $\sigma^*_s(j)=z_s(j)/\sum_{j\in B_s} z_s(j)$ for each j and s is optimal for player 2 when the $z_s(j)$'s belong to an optimal solution of the dual program.

In the rest of this section we give some further properties of a player 2 control stochastic game. First we obtain two results which will be used in the next section.
For a player 2 control stochastic game Γ, $R(\Gamma)$ denotes the set of states s for which player 2 has an optimal stationary strategy σ, such that state s is recurrent under $P(\sigma)$.

9.2.18. LEMMA. *Let* g *be the value of a player* 2 *control stochastic game* Γ. *Then*

(a) $\quad g_s=\min_{j\in B_s} \sum_{t=1}^{z} p(t|s,j)g_t \quad , s\in S.$

Let $E_{2s}=\{j|j\in B_s$ *and* $g_s=\sum_{t=1}^{z} p(t|s,j).g_t\}$.

(b) *If* σ *is optimal for player* 2, *then*

$$\sigma_s\in P(E_{2s}) \text{ and } g=P(\sigma).g.$$

(c) *Let* $v\in\mathbb{R}^z$ *be such that*

(9.2.28) $\quad g_s+v_s \leq \operatorname{Val}_{A_s\times E_{2s}} (G_s(v)), \quad$ *all* $s\in S$

then the equality sign holds in (9.2.28) *for each* $s\in R(\Gamma)$.

PROOF. Part (a) holds in view of lemma 8.1.4. Part (b) is proved in a more general context in lemma 8.1.5, while the equality $g=P(\sigma).g$ is a consequence of part (a) and the definition of E_{2s}.

Considering part (c), take $\tilde{s}\in R(\Gamma)$ and let $\tilde{\sigma}$ be an optimal stationary strategy for player 2 such that state $\tilde{s}$ is recurrent under $P(\tilde{\sigma})$. Let $\tilde{E}(\tilde{\sigma})$ be the ergodic set to which $\tilde{s}$ belongs. Now suppose that the inequality (9.2.28) is strict for $\tilde{s}$. Then there exists a stationary strategy $\tilde{\rho}$ for player 1 such that

$$g+v \leq r(\tilde{\rho},\tilde{\sigma})+P(\tilde{\sigma}).v,$$

with strict inequality at least for component $\tilde{s}\in\tilde{E}(\tilde{\sigma})$. Application of the second part of lemma 8.1.1(b) shows that $\tilde{\sigma}$ cannot be optimal. Hence we have a contradiction and therefore the equality sign holds in (9.2.28) for $\tilde{s}$.

□

Note that the existence of a v satisfying (9.2.28) is quaranteed by both theorem 8.1.8 and by the existence of an optimal solution (g,v,x) to LP1 (cf. theorem 9.2.15).

With a player 2 control stochastic game we associate another linear programming problem, called LP2. Because we use this program for games with payoffs of the type $\bar{r}(s,i,j)=r(s,i,j)-g_s$ (g is the average reward value), it is convenient to incorporate this special form here.
$g_1,\ldots,g_z$ in LP2 are not variables like in LP1, but given numbers.
Intuïtively in this game the average reward value is identical to zero.

9.2.19. ALGORITHM.

LP2: variables $u=(u_1,\ldots,u_z)$, $x=\{x_s(i)\,|\,s\in S,\ i\in A_s\}$

$$\max \sum_{s\in S} u_s, \text{ subject to}$$

(i) $$u_s - \sum_{i\in A_s} \bar{r}(s,i,j).x_s(i) - \sum_{t\in S} p(t|s,j)u_t \leq 0, \quad s\in S,\ j\in B_s$$

(ii) $$\sum_{i\in A_s} x_s(i)=1 \text{ and } x_s(i)\geq 0 \quad , s\in S,\ i\in A_s.$$

The dual LP is:

DLP2: variables $d=(d_1,\ldots,d_z)$, $y=\{y_s(j)\,|\,s\in S,\ j\in B_s\}$.

$$\min \sum_{s\in S} d_s, \text{ subject to}$$

$$\text{(j)} \qquad \sum_{s\in S}\sum_{j\in B_s}(\delta_{st}-p(t|s,j))y_s(j)=1 \quad , \ t\in S$$

$$\text{(jj)} \qquad -\sum_{j\in B_s}\bar{r}(s,i,j).y_s(j)+d_s \geq 0, \ s\in S, \ i\in A_s$$

$$\text{(jjj)} \qquad y_s(j)\geq 0 \quad \text{for all } s\in S, \ j\in B_s.$$

Hordijk & Kallenberg (1981a) have shown that for the transient case (i.e. the case where $\lim_{\tau\to\infty} P^{\tau}(\sigma^p)=0_{zz}$ for all pure stationary strategies σ^p) LP2 is feasible. For such games the solution to LP2 corresponds to the total payoff value of the game. We need an extension of their result to what we call a semi-transient player 2 control stochastic game. This is a game for which (a) $\sum_{t\in S} p(t|s,j)\leq 1$, each $j\in B_s$, $s\in S$, (b) the average payoff value equals 0_z and (c) player 2 has a stationary strategy σ such that $P(\sigma)$ is transient.

9.2.20. LEMMA. *For a semi-transient player 2 control stochastic game with payoffs of the form* $\bar{r}(s,i,j)=r(s,i,j)-g_s$ *the corresponding linear program* LP2 *is feasible and has a bounded optimal solution* u^* *for which*

$$u_s^* = \operatorname*{Val}_{A_s\times B_s}\left(r(s,.,.)-g_s+\sum_{t=1}^{z}p(t|s,.)u_t^*\right) \quad , \ s\in S$$

PROOF. Add a state z+1, where both players have one action denoted by the scalar 1 and such that $p(z+1|z+1,1)=1$, $p(z+1|s,j)=1-\sum_{t\in S}p(t|s,j)$, $s\in S$, and $r(z+1,1,1)=0$. Then we obtain a stochastic game with non-stopping transition probabilities which obviously has also average payoff value 0. But this means (see lemma 8.1.3) that there exists a vector $v\in\mathbb{R}^{z+1}$ such that

$$\text{(9.2.29)} \qquad v_s = \operatorname*{Val}_{A_s\times B_s}\left(r(s,.,.)-g_s+\sum_{t=1}^{z+1}p(t|s,.)v_t\right), \quad s\in\{1,2,\ldots,z+1\}$$

As in (9.2.1) and(9.2.2) there exists a one-to-one correspondence between the set of stationary strategies for player 1 and the set of all x satisfying condition (ii) of LP2. Let ρ be such that ρ_s is an optimal action for player 1 in (9.2.29) for each $s\in S$. Then it can be checked that the pair (u,x^{ρ}) satisfies conditions (i) and (ii) of LP2, when $u_s=v_s-v_{z+1}$, $s\in S$. So LP2 is feasible.

Next let (u,x) be an arbitrary feasible pair. Let σ be such that $P(\sigma)$ is transient. Then condition (i) implies

$$(9.2.30)\qquad u \le r(\rho^x,\sigma)-g+P(\sigma).u \text{ and also } u_s \le \operatorname{Val}_{A_s\times B_s}(r(s,.,.)-g_s+\sum_{t=1}^{z} p(t|s,.)u_t)$$

By iterating the first inequality we obtain in view of $P(\sigma)$ being transient:

$$(9.2.31)\qquad u \le \sum_{\tau=0}^{\infty} P^\tau(\sigma)(r(\rho^x,\sigma)-g) \le \sup_{\rho} \sum_{\tau=0}^{\infty} P^\tau(\sigma)(r(\rho,\sigma)-g).$$

Since $P(\sigma)$ is transient the right hand side of (9.2.31) is bounded. Hence for any feasible solution (u,x) to LP2, $\sum_{s=1}^{z} u_s$ is uniformly bounded from above. This implies that LP2 has a finite optimal solution.

Now let (u^*,x^*) be any optimal solution of LP2. In view of the second inequality of 9.2.30 it remains to show, that for no $\tilde{s} \in S$:

$$(9.2.32)\qquad u^*_{\tilde{s}} < \operatorname{Val}_{A_{\tilde{s}}\times B_{\tilde{s}}}(r(\tilde{s},.,.)-g_{\tilde{s}}+\sum_{t=1}^{z} p(t|\tilde{s},.)u^*_t).$$

Let $\tilde{\rho}_{\tilde{s}}$ be an optimal action for player 1 for the matrix game in the right hand side of (9.2.32). Then, for sufficiently small $\varepsilon>0$, it follows that

$$u_{\tilde{s}}+\varepsilon \le \min_j \{ \sum_{i\in A_{\tilde{s}}} (r(\tilde{s},i,j)-g_{\tilde{s}}).\tilde{\rho}_{\tilde{s}}(i)+ \sum_{t\in S\setminus\{\tilde{s}\}} p(t|\tilde{s},j)u^*_t + p(\tilde{s}|\tilde{s},j)(u^*_{\tilde{s}}+\varepsilon)\}.$$

This inequality implies that the pair $(\bar{u},\bar{x})$, with $\bar{u}_s=u^*_s$, $\bar{x}_s(i)=x^*_s(i)$ if $s\neq\tilde{s}$ and $\bar{u}_{\tilde{s}}=u_{\tilde{s}}+\varepsilon$, $\bar{x}_{\tilde{s}}(i)=\tilde{\rho}_{\tilde{s}}(i)$, is feasible for LP2 and $\Sigma_{s\in S}\bar{u}_s>\Sigma_{s\in S}u^*_s$. But this is in contradiction with the optimality of (u^*,x^*) for LP2. Hence the lemma is proved.

□

We conclude this section with a theorem which states that the solution of the limit discount equation for games in which one player governs the transitions is a rational function of θ.

9.2.21. THEOREM. *The solution of the limit discount equation can be expressed as a Laurent series expansion if the game is such that one player controls the transitions.*

PROOF. Suppose that player 2 governs the transitions. Parthasarathy & Raghavan (1981), theorem 4.2, have shown that, for such a game, player 2 has a stationary strategy which is uniformly discount optimal, i.e. optimal for each interest rate close enough to zero.
Bewley & Kohlberg (1978), theorem 6.1, proved that a stationary strategy σ is uniformly discount optimal if and only if it is optimal in the limit discount equation, i.e. if and only if, for each $s \in S$, the real action σ_s is an optimal action in the matrix game

$$[r(s,.,.) + \frac{\sum_{s=1}^{z} p(s|t,.)x_t^*(\theta)}{1+\theta^{-1}}]$$

with entries in the field of real Puiseux series.
But then, for such an uniformly discount optimal stationary strategy σ, the solution of the limit discount equation satisfies the following relations:

$$x_s(\theta) = \max_{i \in A_s} \{r(s,i,\sigma_s) + \frac{\sum_{t=1}^{z} p(t|s,\sigma_s)x_t(\theta)}{1+\theta^{-1}}\} \quad , s \in S$$

Now this set of equations is nothing else than the limit discount equation for MDP(σ), i.e. the Markov decision problem which results when player 2 fixes σ. And it is well-known that the unique solution of the limit discount equation for Markov decision problems has a Laurent series expansion.

□

Observe that the kind of reasoning used in the proof of theorem 9.2.21 generalizes theorem 6.4 of Bewley & Kohlberg (1978). They state that $x^*(\theta)$ is a rational function of θ if both players have uniformly discount optimal stationary strategies, while in our case only one of the players possesses such a strategy.

9.3. A FINITE ALGORITHM FOR THE SWITCHING CONTROL STOCHASTIC GAME.

In this section we show how the switching control stochastic game can be solved with the aid of a finite sequence of linear programming problems. A switching control stochastic game has already been defined in section 6.3,

where we gave an algorithm for the discounted version. We adopt the notation as introduced there.

In his Ph.D. dissertation, Filar (1979) proved that for switching control undiscounted stochastic games the orderfield property holds. This indicates that a finite algorithm should exist for this class of games. A first attempt to find such an algorithm was made by Filar & Raghavan (1980).

This section provides an efficient algorithm for finding the solution of the undiscounted version of the switching control stochastic game. The results of this section are based on Vrieze, Tijs, Raghavan & Filar (1983).

The part of a stationary strategy ρ of player 1 which refers to the set S_1 is denoted by ρ^c. Thus ρ^c corresponds to a set $\{\rho_s^c | \rho_s^c \in P(A_s), s \in S_1\}$. As already mentioned in section 6.3, if a particular ρ^c is fixed then the remaining game is a player 2 control stochastic game. This game will be denoted by $\hat{\Gamma}(\rho^c)$.

Thus $\hat{\Gamma}(\rho^c) = \langle \hat{S}, \{\hat{A}_s | s \in \hat{S}\}, \{\hat{B}_s | s \in S\}, \hat{r}, \hat{p} \rangle$, where $\hat{S} = S = S_1 \cup S_2$, where

for $s \in S_1$: $\hat{A}_s := \{1\}$, $\hat{B}_s := B_s$, $\hat{r}(s,1,j) := \sum_{i \in A_s} r(s,i,j)\rho_s^c(i)$,

$\hat{p}(t|s,j) := \sum_{i \in A_s} p(t|s,i)\rho_s^c(i)$, and

for $s \in S_2$: $\hat{A}_s := A_s$, $\hat{B}_s = B_s$, $\hat{r}(s,i,j) = r(s,i,j)$, $\hat{p}(t|s,j) := p(t|s,j)$.

The corresponding LP1 of algorithm 9.2.2 for this game will be denoted by $LP1(\hat{\Gamma}(\rho^c))$.

Now fix for a moment a subset $S_0 \subset S$, vectors g, $w \in \mathbb{R}^Z$, a particular ρ^c and for each $s \in S_0$ a non-empty subset E_{2s} of B_s. Then corresponding to Γ and the five parameters S_0, g, w, ρ^c and $\{E_{2s} | s \in S_0\}$ we introduce the following player 2 control stochastic game:

$\bar{\Gamma}(S_0, g, w, \rho^c, \{E_{2s} | s \in S_0\}) = \langle \bar{S}, \{\bar{A}_s | s \in \bar{S}\}, \{\bar{B}_s | s \in \bar{S}\}, \bar{r}, \bar{p} \rangle$, where $\bar{S} = S_0$ and where

for $s \in \bar{S} \cap S_1$: $A_s := \{1\}$, $\bar{B}_s := E_{2s}$, $\bar{r}(s,i,j) := -g_s + \sum_{i \in A_s} (r(s,i,j) + \sum_{t \in S \setminus S_0} p(t|s,i)w_t)\rho_s^c(i)$,

$\bar{p}(t|s,j) := \sum_{i \in A_s} p(t|s,i)\rho_s^c(i)$ for $t \in S_0 = \bar{S}$, and

for $s \in \bar{S} \cap S_2$: $\bar{A}_s := A_s$, $\bar{B}_s := E_{2s}$, $\bar{r}(s,i,j) := -g_s + r(s,i,j) + \sum_{t \in S \setminus S_0} p(t|s,j)w_t$

and $\bar{p}(t|s,j) = p(t|s,j)$ for $t \in S_0 = \bar{S}$.

It will result that this game is a semi-transient player 2 control stochastic game as introduced in the preceding section.

The corresponding LP2-program of algorithm 9.2.19 for this game is denoted by $LP2(\bar{\Gamma}(S_0,g,w,\rho^c,\{E_{2s}|s\in S_0\}))$.

Now we have enough tools to establish our algorithm.

9.3.1. ALGORITHM.

Step 1. Take $\tau=0$ and choose $g(0)=(-M,\ldots,-M)$, where $M=\max_{s,i,j}|r(s,i,j)|$. Choose $w(0)=0_z$, $S(0)=\emptyset$ and $\rho^c(0)$ such that for each $s\in S_1$ the action $\rho^c_s(0)$ is an extreme optimal action for player 1 in the matrix game $[r(s,i,j)]$ on $A_s\times B_s$.

Step 2. Consider the current value of τ and the associated values of the entities $g(\tau)$, $w(\tau)$, $S(\tau)$, $\rho^c(\tau)$. Determine for each $s\in S_1$:

$$E_{1s}(\tau+1) := \{i\in A_s \mid \sum_{t\in S} p(t|s,i)g_t(\tau) = \max_{\tilde{i}\in A_s}\{\sum_{t\in S} p(t|s,\tilde{i}).g_t(\tau)\}$$

and for each $s\in S_2$:

$$E_{2s}(\tau+1) := \{j\in B_s \mid \sum_{t\in S} p(t|s,j)g_t(\tau)=g_s(\tau)\}.$$

Proceed to step 3.

Step 3. Choose $\rho^c(\tau+1)$ such that, for each $s\in S_1$, $\rho^c_s(\tau+1)$ is an extreme optimal action for player 1 in the matrix game

$$\Lambda_{1s}(\tau) := [r(s,i,j)+\sum_{t=1}^{z} p(t|s,i)w_t(\tau)] \quad \text{on } E_{1s}(\tau+1)\times B_s.$$

However if $Car(\rho^c_s(\tau))\subset E_{1s}(\tau+1)$, and if both $g_s(\tau)+w_s(\tau)=Val(\Lambda_{1s}(\tau))$ and $\rho^c_s(\tau)$ is an optimal action for player 1 in the game $\Lambda_{1s}(\tau)$, then put $\rho^c_s(\tau+1):=\rho^c_s(\tau)$.

Step 4. Obtain $g(\tau+1)$, $v(\tau+1)$ by solving LP1 $(\hat{\Gamma}(\rho^c(\tau+1)))$.

Step 5. If $g(\tau+1)\neq g(\tau)$, then put $w(\tau+1):=v(\tau+1)$, $S(\tau+1)=\emptyset$ and return to step 2, taking $\tau:=\tau+1$. If $g(\tau+1)=g(\tau)$, then continue to step 6.

Step 6. Let

$$G_1(\tau+1) := \{s\in S_1 \mid g_s(\tau)+w_s(\tau)<Val(\Lambda_{1s}(\tau))\}$$

$$G_2(\tau+1) := \{s\in S_2 \mid g_s(\tau)+w_s(\tau)<Val(\Lambda_{2s}(\tau))\}$$

where $\Lambda_{2s}(\tau) := [r(s,i,j) + \sum_{t=1}^{z} p(t|s,j).w_t(\tau)]$ on $A_s \times E_{2s}(\tau+1)$.
Put $G(\tau+1) := G_1(\tau+1) \cup G_2(\tau+1)$.
If $G(\tau+1) = \emptyset$, then go to step 9. Otherwise put $S(\tau+1) := G(\tau+1) \cup S(\tau)$ and go to step 7.

Step 7. Put $E_{2s}(\tau+1) := B_s$ for $s \in S(\tau+1) \cap S_1$.
Find $u_s(\tau+1)$ for each $s \in S(\tau+1)$ by solving for a semi-transient player 2 control stochastic game the LP problem

$$LP2(\bar{\Gamma}(S(\tau+1),\ g(\tau+1),\ w(\tau),\ \rho^c(\tau+1),\ \{E_{2s}(\tau+1) \mid s \in S(\tau+1)\})).$$

Step 8. Put $w_s(\tau+1) := w_s(\tau)$ if $s \notin S(\tau+1)$
$w_s(\tau+1) := u_s(\tau+1)$ if $s \in S(\tau+1)$.
Return to step 2 with $\tau := \tau + 1$.
Step 9. The algorithm is stopped. The vector $g(\tau)$ is the value vector for the undiscounted switching control stochastic game. Moreover ρ^* and σ^* are optimal stationary strategies if they are chosen as follows:
for $s \in S_1$, ρ_s^* and σ_s^* are optimal in the matrix game $\Lambda_{1s}(\tau)$ and for $s \in S_2$, ρ_s^* and σ_s^* are optimal in the matrix game $\Lambda_{2s}(\tau)$.

In proving that in step 9 we indeed obtain a solution of the game, we show that at each stage $\tau = 0,1,2,\ldots$ the properties below are valid. Here $g(-1)$ is chosen such that $g(-1) < g(0)$.
We recall that, for a player 2 control stochastic game Γ, $R(\Gamma)$ is defined as the set of states s for which player 2 has an optimal stationary strategy σ such that state s is recurrent under $P(\sigma)$.
Consider the following properties.

$A_1(\tau)$: $g_s(\tau) \le \sum_{t=1}^{z} p(t|s,\rho_s^c(\tau)) g_t(\tau)$, all $s \in S_1$

$A_2(\tau)$: $g_s(\tau) \le \sum_{t=1}^{z} p(t|s,j).g_t(\tau)$, all $s \in S_2$, $j \in B_s$

$B_1(\tau)$: $g_s(\tau) + w_s(\tau) \le r(s,\rho_s^c(\tau),j) + \sum_{t=1}^{z} p(t|s,\rho_s^c(\tau)) w_t(\tau)$, all $s \in S_1$, $j \in B_s$

$B_2(\tau)$: $g_s(\tau) + w_s(\tau) \le \mathrm{Val}(\Lambda_{2s}(\tau))$, all $s \in S_2$

(Although, in step 6, $\Lambda_{2s}(\tau)$ is only defined for the case $g(\tau+1) = g(\tau)$ this definition is extended to general τ.)

$C(\tau)$: $g(\tau) \ge g(\tau-1)$

$D(\tau)$: If $g(\tau) = g(\tau-1)$, then $R(\hat{\Gamma}(\rho^c(\tau))) \subset R(\hat{\Gamma}(\rho^c(\tau-1)))$ and $\rho_s^c(\tau) = \rho_s^c(\tau-1)$ for each $s \in R(\hat{\Gamma}(\rho^c(\tau))) \cap S_1$.

E(τ): $S(\tau)\cap R(\hat{\Gamma}(\rho^c(\tau)))=\emptyset$

F(τ): If $g(\tau)=g(\tau-1)$ and $G(\tau)\neq\emptyset$, then $w(\tau)\geq w(\tau-1)$, with strict inequality in at least one component.

Since $g(-1)<g(0)$ and by definition $S(0)=\emptyset$, it follows that $A_1(0)$, $A_2(0)$, $B_1(0)$, $B_2(0)$, $C(0)$, $D(0)$, $E(0)$ and $F(0)$ hold. By induction on τ we wish to prove that $A_1(\tau),\ldots,F(\tau)$ hold for each $\tau\in\{0,1,\ldots\}$. For this purpose we need a string of lemma's.

9.3.2. LEMMA. *Suppose* $g_s(\tau)=\max_{i\in A_s}\sum_{t=1}^{z} p(t|s,i).g_t(\tau)$ *for some* $s\in S_1$. *Then* $Car(\rho_s^c(\tau))\subset E_{1s}(\tau+1)$. *Furthermore, if property* $B_1(\tau)$ *holds for this state* $s\in S$, *then for all* $j\in B_s$:

$$g_s(\tau)+w_s(\tau)\leq r(s,\rho_s^c(\tau+1),j)+\sum_{t\in S} p(t|s,\rho_s^c(\tau+1)).w_t(\tau).$$

PROOF. Condition (i) of $LP1(\hat{\Gamma}(\rho^c(\tau)))$ yields $g_s(\tau)\leq\sum_{t=1}^{z} p(t|s,\rho_s^c(\tau)).g_t(\tau)$, which in combination with the assumption in the lemma implies that $g_s(\tau)=\sum_{t=1}^{z} p(t|s,\tilde{i}).g_t(\tau)$ for each $\tilde{i}\in Car(\rho_s^c(\tau))$. Hence $Car(\rho_s^c(\tau))\subset E_{1s}(\tau+1)$. Now this fact in combination with $B_1(\tau)$ implies

$$g_s(\tau)+w_s(\tau)\leq\min_j\{r(s,\rho_s^c(\tau),j)+\sum_{t=1}^{z} p(t|s,\rho_s^c(\tau)).w_t(\tau)\}\leq$$

$$\leq Val(\Lambda_{1s}(\tau))=\min_j\{r(s,\rho_s^c(\tau+1),j)+\sum_{t=1}^{z} p(t|s,\rho_s^c(\tau+1)).w_t(\tau)\}.$$

□

9.3.3. LEMMA. *Properties* $A_1(\tau)$ *and* $A_2(\tau)$ *hold for all* $\tau\geq 0$.

PROOF. This is an immediate consequence of condition (i) of $LP1(\hat{\Gamma}(\rho^c(\tau+1)))$.

□

9.3.4. LEMMA. *Suppose that* $A_1(\tau)$, $A_2(\tau)$, $B_1(\tau)$ *and* $B_2(\tau)$ *hold. Then* $C(\tau+1)$ *holds.*

PROOF. Choose the stationary strategy $\hat{\rho}$ for player 1 as follows. If $s\in S_1$, then $\hat{\rho}_s:=\rho_s^c(\tau+1)$ and if $s\in S_2$, let $\hat{\rho}_s$ be an optimal action in the matrix game $\Lambda_{2s}(\tau)$. Let σ^p be an arbitrary pure stationary strategy for player 2. Since

$\hat{\rho}$ is a feasible strategy for the game $\hat{\Gamma}(\rho^c(\tau+1))$, it is sufficient to show that $W_{\hat{\rho}\sigma^p} \geq g(\tau)$. By $A_1(\tau)$, $A_2(\tau)$ and the formulae for $\rho^c(\tau+1)$ we have

(9.3.1) $\quad g(\tau) \leq P(\hat{\rho},\sigma^p).g(\tau).$

Let $R(\hat{\rho},\sigma^p)$ be the set of recurrent states for $P(\hat{\rho},\sigma^p)$. Then by lemma 8.1.1.(a) the equality sign in (9.3.1) holds for each component $s \in R(\hat{\rho},\sigma^p)$. For $s \in R(\hat{\rho},\sigma^p) \cap S_1$ this yields

$$g_s(\tau) = \sum_{t=1}^{z} p(t|s,\rho_s^c(\tau+1))g_t(\tau) = \max_{i \in A_s} \sum_{t=1}^{z} p(t|s,i).g_t(\tau).$$

So we may apply lemma 9.3.2, obtaining

(9.3.2) $\quad g_s(\tau)+w_s(\tau) \leq r(s,\hat{\rho}_s,\sigma^p) + \sum_{t \in S} p(t|s,\hat{\rho}_s).w_t(\tau).$

For $s \in R(\hat{\rho},\sigma^p) \cap S_2$ we have by $B_2(\tau)$, by the choice of $\hat{\rho}$, and by noting that $Car(\sigma_s^p) \subset E_{2s}(\tau+1)$:

(9.3.3) $\quad g_s(\tau)+w_s(\tau) \leq r(s,\hat{\rho}_s,\sigma^p) + \sum_{t=1}^{z} p(t|s,\sigma_s^p).w_t(\tau).$

Then, by lemma 8.1.1(b), the inequalities (9.3.1), (9.3.2) and (9.3.3) imply

$$g(\tau) \leq Q(\hat{\rho},\sigma^p).r(\hat{\rho},\sigma^p) = W_{\hat{\rho}\sigma^p}$$

And since σ^p is arbitrary we obtain:

$$g(\tau+1) = \max_{\rho} \min_{\sigma^p} W_{\rho\sigma^p} \geq g(\tau),$$

where the maximum is taken with respect to each stationary strategy ρ admissible in the game $\hat{\Gamma}(\rho^c(\tau+1))$. □

9.3.5. LEMMA. *Suppose* $A_1(\tau)$, $B_1(\tau)$ *and* $B_2(\tau)$ *hold. Then* $D(\tau+1)$ *holds.*

PROOF. Suppose $g(\tau+1)=g(\tau)$. Observe that in $\hat{\Gamma}(\rho^c(\tau+1))$ in the states belonging to S_1 player 2 has no influence on the transition probabilities. Then by lemma 9.2.18: $g_s(\tau) = \sum_{t=1}^{z} p(t|s,\rho_s^c(\tau+1)).g_t(\tau)$ for $s \in S_1$. Since $\rho_s^c(\tau+1) \in P(E_{1s}(\tau+1))$ this implies:

$$(9.3.4)\qquad g_s(\tau) = \max_{i\in A_s} \sum_{t=1}^{z} p(t|s,i)g_t(\tau), \quad \text{for all } s\in S_1.$$

Hence by lemma 9.3.2 for all $s\in S_1$:

$$(9.3.5)\qquad g_s(\tau)+w_s(\tau)\leq \min_{j\in B_s} \{r(s,\rho_s^c(\tau+1),j)+ \sum_{t=1}^{z} p(t|s,\rho_s^c(\tau+1)).w_t(\tau)\}.$$

Since $g(\tau+1)=g(\tau)$ equals the value of $\hat{\Gamma}(\rho^c(\tau+1))$, lemma 9.2.18 can be applied to (9.3.5) and $B_2(\tau)$ (which together form the assumption in part (c) of lemma 9.2.18). This implies that for $s\in R(\hat{\Gamma}(\rho^c(\tau+1)))$ the equality sign holds in the respective inequalities. Since $Car(\rho_s^c(\tau))\subset E_{1s}(\tau+1)$ (cf. (9.3.4)) we conclude that, for $s\in R(\hat{\Gamma}(\rho^c(\tau+1)))\cap S_1$, $Val(\Lambda_{1s}(\tau))=g_s+w_s$.

Further, by $B_1(\tau)$, $\rho_s^c(\tau)$ is optimal in $\Lambda_{1s}(\tau)$. So by step 3 of the algorithm

$$(9.3.6)\qquad \rho_s^c(\tau+1) = \rho_s^c(\tau) \quad \text{for all } s\in R(\hat{\Gamma}(\rho^c(\tau+1)))\cap S_1,$$

which proves the second part of $D(\tau+1)$.

Fix $s\in R(\hat{\Gamma}(\rho^c(\tau+1)))$ and let σ be optimal for player 2 in $\hat{\Gamma}(\rho^c(\tau+1))$, such that state s is recurrent under $\hat{P}(\sigma)$. Then (9.3.6) and $g(\tau+1)=g(\tau)$ imply that, for the ergodic set to which s belongs, σ is also optimal in $\hat{\Gamma}(\rho^c(\tau))$, and obviously state s remains recurrent in $\hat{\Gamma}(\rho^c(\tau))$. This shows that $R(\hat{\Gamma}(\rho^c(\tau+1)))\subset R(\hat{\Gamma}(\rho^c(\tau)))$.

□

9.3.6. LEMMA. *Suppose* $A_1(\tau)$, $B_1(\tau)$, $B_2(\tau)$ *and* $E(\tau)$ *hold. Then* $E(\tau+1)$ *holds.*

PROOF. If $g(\tau+1)>g(\tau)$, then $S(\tau+1)=\emptyset$ and hence $E(\tau+1)$ is true. Thus suppose $g(\tau+1)=g(\tau)$. From $E(\tau)$ and $R(\hat{\Gamma}(\sigma^c(\tau+1)))\subset R(\hat{\Gamma}(\sigma^c(\tau)))$ (lemma 9.3.5) it follows that $S(\tau)\cap R(\hat{\Gamma}(\rho^c(\tau+1)))=\emptyset$. In view of the definition of $S(\tau+1)$ it then suffices to show that

$$(9.3.7)\qquad G(\tau+1)\cap R(\hat{\Gamma}(\rho^c(\tau+1)))=\emptyset.$$

In the proof of lemma 9.3.5 it has been shown that

$$g_s(\tau)+w_s(\tau)=Val(\Lambda_{1s}(\tau)) \quad \text{for } s\in R(\hat{\Gamma}(\rho^c(\tau+1)))\cap S_1$$

and by lemma 9.2.18,

$$g_s(\tau)+w_s(\tau) = \mathrm{Val}(\Lambda_{2s}(\tau)) \quad \text{for } s\in R(\hat{\Gamma}(\rho^c(\tau+1)))\cap S_2.$$

Hence by these two equations it follows from the definition of $G(\tau+1)$ in step 6 that (9.3.7) holds.

□

9.3.7. LEMMA. *Suppose* $A_1(\tau)$, $B_1(\tau)$, $B_2(\tau)$ *and* $E(\tau)$ *hold. Then* $F(\tau+1)$ *holds.*

PROOF. Suppose $g(\tau+1)=g(\tau)$ and $G(\tau+1)\neq\emptyset$. From relation (9.3.5) in the proof of lemma 9.3.5 we have that, for $s\in S(\tau+1)\cap S_1$,

$$(9.3.8) \qquad g_s(\tau)+w_s(\tau)\leq \min_{j\in B_s}\{r(s,\rho_s^c(\tau+1),j)+\sum_{t=1}^{z} p(t|s,\rho_s^c(\tau+1)).w_t(\tau)\},$$

and, for $s\in S(\tau+1)\cap S_2$, we see from $B_2(\tau)$ that:

$$(9.3.9) \qquad g_s(\tau)+w_s(\tau)\leq \mathrm{Val}(\Lambda_{2s}(\tau)).$$

Since the value of $\hat{\Gamma}(\rho^c(\tau+1))$ equals $g(\tau+1)=g(\tau)$ and since $S(\tau+1)\cap R(\hat{\Gamma}(\rho^c(\tau+1)))=\emptyset$ (lemma 9.3.6) it can be verified that the game $\bar{\Gamma}(S(\tau+1),g(\tau+1),w(\tau),\rho^c(\tau+1),\{E_{2s}(\tau+1)\,|\,s\in S(\tau+1)\})$ is a semi-transient player 2 control stochastic game. Namely:

(a) $\sum_{t\in S(\tau+1)} \bar{p}(t|s,j)\leq 1$;

(b) The $S(\tau+1)$-part of an optimal stationary strategy σ of player 2 in the game $\hat{\Gamma}(\rho^c(\tau+1))$ gives, when applied to $\bar{\Gamma}(.,.,.,.,.,.)$, a transient stochastic matrix; and

(c) the average reward value equals 0_z. To see (c), by (b) we have that the value is at most 0_z and if σ is such that some states of $S(\tau+1)$ are recurrent, then σ is disadvantageous for player 2 in view of $S(\tau+1)\cap R(\hat{\Gamma}(\rho^c(\tau+1)))=\emptyset$. Hence the best player 2 can do is playing a transient stationary strategy, resulting in value 0_z.

Furthermore, putting $\hat{x}_s(1)=1$ if $s\in S_1\cap S(\tau+1)$ and $\hat{x}_s(i)=\hat{\rho}_s(i)$, $i\in \bar{A}_s$ if $s\in S_2\cap S(\tau+1)$, where $\hat{\rho}_s$ is optimal for player 1 in $\Lambda_{2s}(\tau)$, it can be seen that the pair $(\{w_s(\tau)\,|\,s\in S(\tau+1)\}, \{\hat{x}_s(i)\,|\,s\in S(\tau+1), i\in \bar{A}_s\})$ satisfies conditions (i) and (ii) of LP2$(\bar{\Gamma}(.,.,.,.,.,.))$. But since $G(\tau+1)\neq\emptyset$ in (9.3.8) or (9.3.9) at least one strict inequality sign holds. Hence, by lemma 9.2.20, we obtain for the solution $\{u_s\,|\,s\in S(t+1)\}$ of this LP2 problem that $u_s\geq w_s(\tau)$, all $s\in S(\tau+1)$, with the inequality sign holding for at least one coordinate.

□

9.3.8. LEMMA. *Suppose* $A_1(\tau)$, $B_1(\tau)$, $B_2(\tau)$ *and* $E(\tau)$ *hold. Then* $B_1(\tau+1)$ *and* $B_2(\tau+1)$ *hold.*

PROOF. If $g(\tau+1)\neq g(\tau)$, then $B_1(\tau+1)$ and $B_2(\tau+1)$ follow from condition (ii) of $LP1(\hat{\Gamma}(\rho^C(\tau+1)))$. Suppose now $g(\tau+1)=g(\tau)$. From $F(\tau+1)$ (lemma 9.3.7) we obtained $w_s(\tau+1)\geq w_s(\tau)$ for each $s\in S(\tau+1)$. By definition, $w_s(\tau+1)=w_s(\tau)$ for each $s\in S\setminus S(\tau+1)$.
In the first part of the proof of lemma 9.3.5 (cf. (9.3.4)), it has been shown that the condition of lemma 9.3.2 is satisfied for each $s\in S_1$. So using $B_1(\tau)$, lemma 9.3.2 and $B_2(\tau)$ we have that $B_1(\tau+1)$ and $B_2(\tau+1)$ hold for $s\in S\setminus S(\tau+1)$.
Further, by condition (i) of $LP2(\bar{\Gamma}(.,.,.,.,.,.))$, it follows that $B_1(\tau+1)$ and $B_2(\tau+1)$ also hold for $s\in S(\tau+1)$.

□

Now, combining the lemma's 9.3.3-9.3.8 we conclude that from the assumption "$A_1(\tau)$, $A_2(\tau)$, $B_1(\tau)$, $B_2(\tau)$, $C(\tau)$, $D(\tau)$, $E(\tau)$ and $F(\tau)$ hold" it follows that "$A_1(\tau+1)$, $A_2(\tau+1)$, $B_1(\tau+1)$, $B_2(\tau+1)$, $C(\tau+1)$, $D(\tau+1)$, $E(\tau+1)$ and $F(\tau+1)$ hold".
Hence we have proved.

9.3.9. THEOREM. *For each* $\tau\in\{0,1,2,\ldots\}$ *the properties* $A_1(\tau)$, $A_2(\tau)$, $B_1(\tau)$, $B_2(\tau)$, $C(\tau)$, $D(\tau)$, $E(\tau)$ *and* $F(\tau)$ *hold.*

The following is an important theorem.

9.3.10. THEOREM. *Algorithm* 9.3.1 *stops after a finite number of iterations.*

PROOF. Parthasarathy & Raghavan (1981) have shown that an extreme optimal action for player 1 in a matrix game of payoff type $[f(i,j)+h(i)]$ on $A\times B$ is also an extreme optimal action for player 1 in some subgame $[f(i,j)]$ on $\alpha\times B$ with $\alpha\subset A$ (cf. Parthasarathy & Raghavan (1981), lemma 4.1 p. 381). Applied to step 3 of our algorithm, this means that, for each state $s\in S_1$, at any stage τ an extreme optimal action $\rho_s^C(\tau)$ of player 1 for some matrix game $[r(k,i,j)]$ on $\alpha_s(\tau)\times B_s$, with $\alpha_s(\tau)\subset A_s$, is chosen. Shapley & Snow (1950) have shown that a matrix game has only a finite number of extreme optimal actions. Furthermore, a matrix game has a finite number of submatrices. Since there are a finite number of states, this implies that

(9.3.10) the set from which $\rho^C(\tau)$, $\tau\geq 0$, is chosen is a finite one.

It remains to show that no cycles can occur, i.e. that no strategy repeats itself infinitely often.

By the properties $C(\tau)$ and $F(\tau)$ we deduce that for each τ exactly one of the following events occurs:

H1: $g(\tau)>g(\tau-1)$

H2: $g(\tau)=g(\tau-1)$, $\rho^C(\tau)\neq\rho^C(\tau-1)$, $G(\tau)\neq\emptyset$, $w(\tau)>w(\tau-1)$

H3: $g(\tau)=g(\tau-1)$, $\rho^C(\tau)=\rho^C(\tau-1)$, $G(\tau)\neq\emptyset$, $w(\tau)>w(\tau-1)$

H4: $g(\tau)=g(\tau-1)$, $\rho^C(\tau)=\rho^C(\tau-1)$, $G(\tau)=\emptyset$.

Since $\hat{\Gamma}(\rho^C(\tau))$ depends only on $\rho^C(\tau)$ we have in view of $C(\tau)$:

(9.3.11) if H1 occurs then $\rho^C(k)\neq\rho^C(\ell)$,
$k\in\{\tau,\tau+1,\ldots\}$ and $\ell\in\{\tau-1,\tau-2,\ldots,0\}$.

Now suppose that, from stage τ, H2 repeats itself infinitely often. Since $|S|$ is finite we may assume without loss of generality that $S(\tau)=S(\tau+1)=S(\tau+2)=\ldots$. But then observe that the optimal value of $LP2(\bar{\Gamma}(S(\tau-1+\ell),g(\tau-1+\ell), w(\tau-2+\ell), \rho^C(\tau-1+\ell), \{E_{2s}(\tau-1+\ell)\,|\,s\in S(\tau-1+\ell)\}))$ in step 7 of the algorithm depends only on $\{\rho^C_s(\tau-1+\ell)\,|\,s\in S(\tau-1+\ell)=S(\tau)\}$, $\ell=1,2,\ldots$, since the other parameters do not change. But since $w(\tau-1+\ell)>w(\tau-2+\ell)$ we find $\rho^C(k+\ell)\neq\rho^C(k)$, for $\ell=1,2,\ldots$ and $k=\tau-1,\tau,\tau+1,\ldots$. But then in view of (9.3.10)

(9.3.12) H2 cannot repeat itself infinitely often.

Let k be the first time that H2 does not occur. Then either $S(k)=\emptyset$, in which case H1 occurs, or H4, or possibly H3 occurs.
If H3 occurs, then, by the construction of $G_1(\tau)$ and $G_2(\tau)$, and by the equality in the assertion of lemma 9.2.20, we see that $G(\tau)\cap S(\tau-1)=\emptyset$. Hence

(9.3.13) if H3 occurs then $S(\tau)$ strictly includes $S(\tau-1)$.

As final statement we have

(9.3.14) if H4 occurs then the algorithm stops.

Now by (9.3.12) and (9.3.13) we see, in view of the finite number of states, that a sequence in which only the events H2 and H3 occur cannot happen. But then in view of (9.3.11) and (9.3.10), H4 must occur within a finite number of iterations, which by (9.3.14) proves the theorem. □

9.3.11. THEOREM. *Step 9 of the algorithm is reached after a finite number of iterations and provides a solution to the game, i.e.* $g(\tau)$ *equals the value of the game and* ρ^* *and* σ^* *are optimal stationary strategies.*

PROOF. By theorem 9.3.10 step 9 is reached after a finite number of iterations.

From $g(\tau+1)=g(\tau)$, $\rho^c(\tau+1)=\rho^c(\tau)$, $G(\tau+1)=\emptyset$ and the definitions of ρ^* and σ^* (observe that $\rho_s^* \in P(E_{1s}(\tau+1))$, for $s \in S_1$ and $\sigma_s^* \in P(E_{2s}(\tau+1))$, for $s \in S_2$), we conclude that for any two pure stationary strategies ρ^p and σ^p

$$g(\tau) \leq P(\rho^*,\sigma^p).g(\tau) \tag{9.3.15}$$

and for $s \in R(\rho^*,\sigma^p) \cap S_1$

$$g_s(\tau)+w_s(\tau) \leq r(s,\rho_s^*,\sigma_s^p)+\sum_{t=1}^{z} p(t|s,\rho_s^*).w_t(\tau) \tag{9.3.16}$$

and for $s \in R(\rho^*,\sigma^p) \cap S_2$

$$g_s(\tau)+w_s(\tau) \leq r(s,\rho_s^*,\sigma_s^p)+\sum_{t=1}^{z} p(t|s,\sigma_s^p).w_t(\tau). \tag{9.3.17}$$

Similarly

$$g(\tau) \geq P(\rho^p,\sigma^*).g(\tau) \tag{9.3.18}$$

and for $s \in R(\rho^p,\sigma^*) \cap S_1$

$$g_s(\tau)+w_s(\tau) \geq r(s,\rho_s^p,\sigma_s^*)+\sum_{t=1}^{z} p(t|s,\rho_s^p).w_t(\tau) \tag{9.3.19}$$

and for $s \in R(\rho^p,\sigma^*) \cap S_2$

$$g_s(\tau)+w_s(\tau) \geq r(s,\rho_s^p,\sigma_s^*)+\sum_{t=1}^{z} p(t|s,\sigma_s^*).w_t(\tau). \tag{9.3.20}$$

Now (9.3.15)-(9.3.20) imply by lemma 8.1.1(b) that

$$W_{\rho^*\sigma^p} = Q(\rho^*,\sigma^p).r(\rho^*,\sigma^p) \geq g(\tau) = Q(\rho^*,\sigma^*).r(\rho^*,\sigma^*) \geq$$

$$\geq Q(\rho^p,\sigma^*).r(\rho^p,\sigma^*) = W_{\rho^p\sigma^*}.$$

By corollary 3.5 and theorem 2.3.4 these inequalities yield the theorem. □

9.3.12. REMARK. For the class of stochastic games for which, for each pair of pure stationary strategies (ρ^p,σ^p), the Markov chain associated with $P(\rho^p,\sigma^p)$ has no transient states algorithm 9.3.1 can be considerably simplified. Only the steps 1-5 are necessary in this case (cf. also remark 9.2.17). Furthermore, as soon as $g(\tau+1)=g(\tau)$ the algorithm can stop since the value of the game is reached. These facts follow immediately from the properties $D(\tau)$ and $E(\tau)$.

We conclude this section with the observation that our algorithm provides a constructive proof of the existence of the value and of optimal stationary strategies for both players in the switching control stochastic game. Also the result of Filar (1979) that player 1 has an optimal stationary strategy ρ^* such that, for each $s \in S_1$, ρ^*_s is an extreme optimal action in a matrix game $[r(s,i,j)]$ on $\alpha_s \times B_s$ with $\alpha_s \subset A_s$, can be derived from our algorithm. Similarly for player 2. Furthermore the finiteness of the algorithm gives a constructive proof of the ordered field property.

Appendix.

A.1. MATRIX GAMES.

In this section we give a number of well-known concepts in matrix game theory. Also some results will be mentioned which are used in this monograph.

A.1.1. DEFINITION. *A two-person zerosum game in normal form is an ordered triplet* $< A, B, K >$, *where* A *and* B *are non-empty sets and* $K: A \times B \to \mathbb{R}$ *is a real-valued function on the Cartesian product of* A *and* B. *The sets* A *and* B *are called the action spaces of player* 1 *and player* 2 *respectively. The elements of* A *and* B *are called actions and* K *is the payoff function. A matrix game is a two-person zerosum game in which both* A *and* B *are finite sets.*

Since we only consider games in normal form, and not games in extensive form, we usually omit the qualification "in normal form".

Such a (non-cooperative) game is played as follows. The players 1 and 2 choose, independently of one another, an action $a \in A$ and an action $b \in B$ respectively; subsequently player 2 pays player 1 the amount $K(a,b)$. (If $K(a,b)$ is negative, then player 2 receives $-K(a,b)$ from player 1). We call $K(a,b)$ the payoff of the play. Clearly player 1 wishes to maximize, and player 2 to minimize, this payoff.

According to definition A.1.1 the players are not allowed to randomize their actions, i.e. to select a (pure) action with the aid of a chance experiment. However in non-cooperative game theory it is the custom to permit the players to use lotteries. This results in the mixed extension of a game (cf. e.g. Luce & Raiffa (1957)). Since in this monograph we only consider games where the players have finite action spaces, the next definition is restricted to that case. Such games are called matrix games.

A.1.2. DEFINITION. *A mixed extension of a matrix game* $< A,B,K >$ *is a two-person zerosum game* $< P(A), P(B), \tilde{K} >$ *where* $P(A)$ *is the family of all probability measures on the finite set* A. *If* A *consists of* m *elements, then* A *is identified with the set* $\mathbb{N}_m := \{1,2,\ldots,m\}$; *so* $P(A)$ *corresponds with the* $(m-1)$*-dimensional simplex* $\{x \mid x=(x_1,\ldots,x_m),\ x_i \geq 0$ *and* $\sum_{i=1}^{m} x_i = 1\}$. *Likewise* $P(B) = \{y \mid y=(y_1,\ldots,y_n),\ y_i \geq 0$ *and* $\Sigma_{j=1}^{n} y_j = 1\}$ *if* B *has* n *elements. Furthermore* $\tilde{K}(x,y) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} K(i,j).x_i.y_j$ *for each* $(x,y) \in P(A) \times P(B)$.

When we speak of an action in a matrix game, this action may either be pure or mixed. Concerning matrix games, the variable i always denotes a pure action for player 1 and j a pure action for player 2. Though it might cause ambiguity, we usually write $K(x,y)$ instead of $\tilde{K}(x,y)$. By $K(x,j)$ and $K(i,y)$ we mean $K(x,e_j)$ and $K(e_i,y)$, where e_j respectively e_i corresponds to the probability measure that puts weight 1 on action $j \in B$ and $i \in A$ respectively.

Note that a mixed extension of a matrix game itself is a two-person game in normal form. The mixed extension of a game is played as follows. The players 1 and 2 choose independently of one another an action $x \in \mathcal{P}(A)$ and $y \in \mathcal{P}(B)$ respectively; then for each player a chance experiment according to the probability measures x and y respectively is carried out, in order to select pure actions $i \in A$ and $j \in B$ respectively. Subsequently player 2 pays player 1 the amount $K(i,j)$. Thus, if the players play x and y, the expected outcome of the game equals $K(x,y)$.

In this monograph the mixed extension of a matrix game $< A,B,K >$ is abbreviated to $[K]$. When we speak of a matrix game, the mixed version is intended. However, when we speak of a two-person zerosum game $< A,B,K >$, the "pure" version is intended.

A.1.3. DEFINITION. *A two-person zerosum game is said to have a value* V^*, *if*

$$\sup_{a \in A} \inf_{b \in B} K(a,b) = \inf_{b \in B} \sup_{a \in A} K(a,b) = V^*.$$

For a game with given value V^*, *the actions* a^ε *and* b^ε *are called* ε*-optimal, with* $\varepsilon \geq 0$, *for player* 1 *and player* 2 *respectively, if*

$$\inf_{b \in B} K(a^\varepsilon,b) \geq V^* - \varepsilon \text{ and } \sup_{a \in A} K(a,b^\varepsilon) \leq V^* + \varepsilon.$$

Zero-optimal actions are named optimal.

For a good understanding of the value concept one should note that when playing optimally player 1 (player 2) can guarantee himself a payoff of at least (at most) the value of the game, whatever action the other player chooses. The value of a two-person zerosum game $<A,B,K>$ is denoted by $pVal(K)$. The "p" in pVal reflects the notion that only pure actions are allowed.

A.1.4. THEOREM. *For a two-person zerosum game* $< A,B,K >$ *the following assertions are equivalent.*

(i) *The game has value* V^* *and* $\max \inf_{a\in A} \inf_{b\in B} K(a,b)$ *and* $\min \sup_{b\in B} \sup_{a\in A} K(a,b)$ *exist.*

(ii) *There exist* $V^*\in\mathbb{R}$, $\hat{a}\in A$ *and* $\hat{b}\in B$, *such that for all* $(a,b)\in A\times B$

$$K(a,\hat{b}) \leq K(\hat{a},\hat{b}) = V^* \leq K(\hat{a},b).$$

PROOF. Suppose (i) is true. Let $V^* = \max \inf_{a\in A} \inf_{b\in B} K(a,b) = \min \sup_{b\in B} \sup_{a\in A} K(a,b)$ be the value of the game. Obviously there exist $\hat{a}\in A$ and $\hat{b}\in B$ such that $\inf_{b\in B} K(\hat{a},b) = V^* = \sup_{a\in A} K(a,\hat{b})$. Hence (ii) is true. Suppose (ii) is true. Then

(A.1.1)
$$\sup_{a\in A} \inf_{b\in B} K(a,b) \geq \inf_{b\in B} K(\hat{a},b) = V^* =$$
$$\sup_{a\in A} K(a,\hat{b}) \geq \inf_{b\in B} \sup_{a\in A} K(a,b).$$

But for each $(a,b)\in A\times B$ we have

$$K(a,b) \leq \sup_{a\in A} K(a,b).$$

So for each $a\in A$ we have

$$\inf_{b\in B} K(a,b) \leq \inf_{b\in B} \sup_{a\in A} K(a,b),$$

which implies

(A.1.2)
$$\sup_{a\in A} \inf_{b\in B} K(a,b) \leq \inf_{b\in B} \sup_{a\in A} K(a,b).$$

Now (A.1.2) implies that the equality signs must hold throughout (A.1.1). This results in

$$\max \inf_{a\in A} \inf_{b\in B} K(a,b) = \inf_{b\in B} K(\hat{a},b) = V^* = \sup_{a\in A} K(a,\hat{b}) = \min \sup_{b\in B} \sup_{a\in A} K(a,b).$$

So (i) is true.

□

Note that if $\hat{a}$ and $\hat{b}$ obey (ii) of theorem A.1.4, then $\hat{a}$ is optimal for player 1 and $\hat{b}$ is optimal for player 2.
The following theorem is well-known (e.g. Tijs (1977)).

A.1.5. THEOREM. *If for a game* $< A,B,K >$, *for each* $\varepsilon>0$, *there exists* $(a^\varepsilon,b^\varepsilon)\in A\times B$, *such that for each* $(a,b)\in A\times B$:

$$K(a,b^\varepsilon)-\varepsilon \leq K(a^\varepsilon,b^\varepsilon) \leq K(a^\varepsilon,b)+\varepsilon,$$

then the value of the game exists and equals $\lim_{\varepsilon\downarrow 0} K(a^\varepsilon,b^\varepsilon)$.

The class of matrix games with m rows and n columns is denoted by M_{mn}. Already J. von Neumann (1928) proved the following theorem concerning matrix games.

A.1.6. THEOREM. *For all* $m,n\in\mathbb{N}$ *each matrix game* $[K]\in M_{mn}$ *has a value and both players possess optimal actions.*

The value of a matrix game $[K]$ is denoted by $\mathrm{Val}(K)$. The set of ε-optimal actions for player ℓ, $\ell\in\{1,2\}$, is denoted by $O_\ell^\varepsilon(K)$ for $\varepsilon>0$ and by $O_\ell(K)$ for $\varepsilon=0$.
In the next lemma, three well-known properties of the value operator are stated. Here J_{mn} denotes an m,n-matrix for which each element equals 1.

A.1.7. LEMMA. *If* $[K_1]$ *and* $[K_2]\in M_{mn}$ *with* $K_1\geq K_2$ *and if* $(\hat{x},\hat{y})\in P(A)\times P(B)$, *then*

(a) $\inf_{y\in P(B)} K_1(\hat{x},y) = \min_{j\in B} K_1(\hat{x},j)$ *and* $\sup_{x\in P(A)} K_1(x,\hat{y}) = \max_{i\in A} K_1(i,\hat{y})$

(b) $\mathrm{Val}(K_1+cJ_{mn}) = \mathrm{Val}(K_1)+c$ *for any* $c\in\mathbb{R}$

(c) $\mathrm{Val}(K_1) \geq \mathrm{Val}(K_2)$.

PROOF. (a) We only prove the first equality; the second can be shown analogously. Clearly

$$\inf_{y\in P(B)} K_1(\hat{x},y) \leq \min_{j\in B} K_1(\hat{x},j). \tag{A.1.3}$$

However, since $y(j) \geq 0$ and $\Sigma_{j=1}^{n} y(j) = 1$, we have for all y:

$$K_1(\hat{x},y) = \sum_{j=1}^{n} K_1(\hat{x},j)y(j) \geq \sum_{j=1}^{n} y(j) \min_{j\in B} K_1(\hat{x},j) = \min_{j\in B} K_1(\hat{x},j).$$

So in (A.1.3) the equality sign holds.

(b)

$$\mathrm{Val}(K_1+cJ_{mn}) = \sup_x \inf_y (K_1(x,y)+(x^T.J_{mn}y).c)$$

$$= \sup_x \inf_y K_1(x,y) + c = \mathrm{Val}(K_1)+c.$$

(c) Let $x^* \in O_1(K_2)$. Then

$$\mathrm{Val}(K_1) \geq \min_{j\in B} K_1(x^*,j) \geq \min_{j\in B} K_2(x^*,j) = \mathrm{Val}(K_2).$$

□

The next lemma states the Lipschitz continuity property of the value operator.

A.1.8. LEMMA. *If* $[K_1]$ *and* $[K_2] \in M_{mn}$, *then*

$$|\mathrm{Val}(K_1)-\mathrm{Val}(K_2)| \leq d(K_1,K_2).$$

PROOF. From $K_1-d(K_1,K_2)J_{mn} \leq K_2 \leq K_1+d(K_1,K_2)J_{mn}$, we derive by (b) and (c) of lemma A.1.7:

$$\mathrm{Val}(K_1)-d(K_1,K_2)=\mathrm{Val}(K_1-d(K_1,K_2)J_{mn}) \leq \mathrm{Val}(K_2)$$

$$\leq \mathrm{Val}(K_1+d(K_1,K_2)J_{mn}) = \mathrm{Val}(K_1)+d(K_1,K_2).$$

This proves the lemma.

□

For a matrix game $[K] \in M_{mn}$ it is well-known that the optimal action spaces $O_1(K)$ and $O_2(K)$ are polytopes (i.e. convex hulls of finite sets). This fact is stated in the following theorem, due to Shapley & Snow (1950).

A.1.9. THEOREM. *Let* $[K] \in M_{mn}$, $x^* \in \mathcal{O}_1(K)$ *and* $y^* \in \mathcal{O}_2(K)$. *Then the following two assertions are equivalent.*

(i) x^* *is an extreme point of* $\mathcal{O}_1(K)$ *and* y^* *is an extreme point of* $\mathcal{O}_2(K)$.

(ii) *There exists a square* k,k*-submatrix* K^* *of* K, *such that*

$$1_k^T.\mathrm{adj}(K^*).1_k \neq 0, \qquad \mathrm{Val}(K) = \frac{\det(K^*)}{1_k^T.\mathrm{adj}(K^*).1_k}$$

$$x^{*k} = \frac{1_k^T.\mathrm{adj}(K^*)}{1_k^T.\mathrm{adj}(K^*).1_k} \qquad y^{*k} = \frac{\mathrm{adj}(K^*).1_k}{1_k^T.\mathrm{adj}(K^*).1_k}$$

(*Here* x^{*k} *and* y^{*k} *are the vectors obtained from* x^* *and* y^* *by removing the coordinates, which play no role in* K^* *and which are zero*).

The statement in the next theorem is called the dimension relation for matrix games and is due to Bohnenblust, Karlin & Shapley (1950). It shows the way in which the sets $\mathcal{O}_1(K)$ and $\mathcal{O}_2(K)$ are topologically related. Let $A=\mathbb{N}_m$ and $B=\mathbb{N}_n$. Let the polytope $E_1 \subset \mathcal{P}(A)$ and define $\mathrm{Car}(E_1)$ as $\mathrm{Car}(E_1) := \{i \in A \mid$ there exists an $x \in E_1$ with $x(i)>0\}$. $\mathrm{Car}(E_1)$ is named the carrier of E_1. By $\mathrm{Unf}(E_1)$ we denote the number of unnatural faces of E_1, i.e. faces which are not entirely contained in the relative boundary of $\mathcal{P}(A)$. Similarly we define $\mathrm{Car}(E_2)$ and $\mathrm{Unf}(E_2)$ for a polytope $E_2 \subset \mathcal{P}(B)$. For a set $E \subset \mathbb{R}^k$ we denote by $\dim(E)$ the dimension of E.

A.1.10. DEFINITION. *Let* $A=\mathbb{N}_m$ *and* $B=\mathbb{N}_n$. *A pair of polytopes* $(E_1,E_2) \in \mathcal{P}(A) \times \mathcal{P}(B)$ *is said to possess the* (m,n)-BKS *property if*

(a) $|\mathrm{Car}(E_1)| - \dim(E_1) = |\mathrm{Car}(E_2)| - \dim(E_2)$.

(b) $|\mathrm{Car}(E_1)| + \mathrm{Unf}(E_2) \leq m$ and $|\mathrm{Car}(E_2)| + \mathrm{Unf}(E_1) \leq n$.

A.1.11. THEOREM. *Let* $A=\mathbb{N}_m$ *and* $B=\mathbb{N}_n$. *For a pair of polytopes* $(E_1,E_2) \in \mathcal{P}(A_1) \times \mathcal{P}(B)$ *there exists a matrix game* $[K] \in M_{mn}$ *such that* $\mathcal{O}_1(K) = E_1$ *an* $\mathcal{O}_2(K) = E_2$ *if and only if* (E_1,E_2) *possesses the* (m,n)-BKS *property.*

A.1.12. COROLLARY. *Let* $A=\mathbb{N}_m$ *and* $B=\mathbb{N}_n$. *Let* $v \in \mathbb{R}$ *and let the pair of polytopes* $(E_1,E_2) \in \mathcal{P}(A) \times \mathcal{P}(B)$ *have the* (m,n)-BKS *property. Then, for each* $\varepsilon>0$, *there exists a matrix game* $[K] \in M_{mn}$ *with* $\mathrm{Val}(K)=v$, $\mathcal{O}_1(K)=E_1$, $\mathcal{O}_2(K)=E_2$ *and such that* $|K(i,j)-v|<\varepsilon$ *for all* $(i,j) \in A \times B$.

PROOF. In view of theorem A.1.11 there is a game $[\hat{K}]$ with solution v, E_1, E_2. Then for τ large enough the matrix game $[K]$ defined as $K(i,j)=v+\tau^{-1}(\hat{K}(i,j)-v)$ has the desired property.

□

A.2. MARKOV DECISION PROBLEMS.

In this section we give some well-known properties of Markov decision problems.

A.2.1. DEFINITION. *A finite (stationary) Markov decision situation is an ordered quadruple* $< S, \{A_s | s \in S\}, r, p >$, *where the finite set* S *is the state space, the finite set* A_s *the action set in state* s, r *is the reward function and* p *the transition map.*

The meanings of the parameters of the Markov decision situation are the same as in definition 2.1.1 of chapter I for a stochastic game situation. Also one should think of a Markov decision situation as a dynamic system which may be in certain states. At discrete points in time the course of the system can be influenced by selecting an action from a set which depends on the current state. This action results in an immediate reward and determines the next state according to a chance experiment. This chance experiment only depends on the current state and the action subsequently chosen. We assume decision epochs $\tau=0,1,2,\ldots$.

A Markov decision situation can be regarded as a stochastic game situation with only one player.

Strategies for Markov decision situations are defined in an analogous way as for stochastic game situations (definition 2.2.2). The different types of strategy spaces are denoted by ST, SMST, MST, SST and PSST respectively.

Also for Markov decision problems one can differentiate between a number of optimality criteria, each of them specifying its own manner of evaluating the stream of immediate (expected) rewards. In our definition of a Markov decision problem we have implicitly assumed that one wishes to maximize the evaluation function over the set of strategies.

A.2.2. DEFINITION. *A discounted Markov decision problem with interest rate* $\alpha \in (0,\infty)$, *is a Markov decision situation for which the stream of payoffs is evaluated as*

$$V_{s\mu} := \sum_{\tau=0}^{\infty} \left(\frac{1}{1+\alpha}\right)^{\tau} . V_{s\mu}^{\tau}.$$

Note that, in definition A.2.2, $V_{s\mu}^{\tau}$ equals the expected payoff at decision epoch τ for initial state s and strategy μ. So $V_{s\mu}$ is the total discounted expected payoff for starting state s, strategy μ and discount factor $(1+\alpha)^{-1}$.

A.2.3. DEFINITION. *An average reward Markov decision problem is a Markov decision situation for which the stream of payoffs is evaluated as*

$$W_{s\mu} := \liminf_{k\to\infty} \frac{1}{k+1} \sum_{\tau=0}^{k} V_{s\mu}^{\tau}.$$

In definition A.2.3 $V_{s\mu}^{\tau}$ has the same meaning as in definition A.2.2. So $W_{s\mu}$ equals the average expected reward per unit time for starting state s and strategy μ.

Like in stochastic games obviously $V_{s\mu}$ and $W_{s\mu}$ exist for each s and μ.

A.2.4. DEFINITION. *Let* $G_{s\mu}$ *be the evaluation function for a Markov decision problem. A strategy* μ_{ε} *is said to be* ε*-optimal, given* $\varepsilon \geq 0$, *if for each* $s \in S$:

$$G_{s\mu_{\varepsilon}} \geq \sup_{\mu \in ST} G_{s\mu} - \varepsilon .$$

Zero-optimal strategies are called optimal.

Markov decision problems are extensively studied in the literature. See for example Blackwell (1962, 1965), Derman (1970), Hordijk (1974) and Federgruen (1978).

We now quote a number of results of Markov decision theory, which are referred to in this monograph. The way in which we use these theorems is as follows: fix for the two-person zerosum stochastic game a stationary strategy for one player; then the other player faces a Markov decision problem; apply the results of the Markov decision theory to this problem

and then return to the stochastic game.
In the following theorems $r(s,\rho_s)$ and $p(t|s,\rho_s)$ with $\rho_s \in \mathcal{P}(A_s)$ are defined as $r(s,\rho_s) := \Sigma_{i \in A_s} r(s,i).\rho_s(i)$ and $p(t|s,\rho_s) := \Sigma_{i \in A_s} p(t|s,i).\rho_s(i)$.

A.2.5. THEOREM. *For an infinite horizon discounted Markov decision problem, the vector* $V^* \in \mathbb{R}^z$ *defined as* $V^*_s := \sup_\mu V_{s\mu}$ *for each* $s \in S$, *is the unique solution of the following set of functional equations in the variable* $x \in \mathbb{R}^z$:

$$x_s = \max_{i \in A_s} \{r(s,i) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,i).x_t\} \quad , \ s \in S.$$

A stationary strategy $\rho^* = (\rho^*_1, \rho^*_2, \ldots, \rho^*_z)$ *is optimal if and only if for each* $s \in S$:

$$V^*_s = r(s,\rho^*_s) + \frac{1}{1+\alpha} \sum_{t=1}^{z} p(t|s,\rho^*_s).V^*_t.$$

Also an optimal pure stationary strategy exists.

A proof of this theorem can be found in Blackwell (1965) and in Derman (1970).

A.2.6. THEOREM. *For an infinite horizon average reward Markov decision problem let the vector* $W^* \in \mathbb{R}^z$ *be defined as* $W^*_s := \sup_{\mu \in ST} W_{s\mu}$, *for each* $s \in S$. *Consider the following set of functional equations in the variables* $x,y \in \mathbb{R}^z$:

(A.2.1) $$x_s = \max_{i \in A_s} \sum_{t=1}^{z} p(t|s,i).x_t \quad , \ s \in S$$

and

(A.2.2) $$x_s + y_s = \max_{i \in E_s} \{r(s,i) + \sum_{t=1}^{z} p(t|s,i).y_t\} \quad , \ s \in S,$$

where $E_s := \{i \in A_s \mid x_s = \sum_{t=1}^{z} p(t|s,i).x_t\}$.

Then this set of equations is solvable and for each solution (x^*,y^*) *we have* $x^*=W^*$. *Furthermore a stationary strategy* $\rho^*=(\rho_1^*,\rho_2^*,\ldots,\rho_z^*)$ *is optimal if and only if the following holds*: (i) $\rho_s(i)=0$ *for* $i\notin E_s$ (*hence* $W_s^*=\sum_{t=1}^{z} p(t|s,\rho_s^*).W_t^*$) *and* (ii) *for any solution* (W^*,y^*) *of the equations* (A.2.1) *and* (A.2.2) *it holds that for each state* $s\in S$ *which is recurrent with respect to* ρ^* *we have* $W_s^*+y_s^*=r(s,\rho_s^*)+\sum_{t=1}^{z} p(t|s,\rho_s^*).y_t^*$. *Also an optimal pure stationary strategy exists.*

A proof of this theorem can be found in Schweitzer & Federgruen (1978).

A.2.7. REMARK. If we have to do with a minimizing Markov decision problem, then in the theorems A.2.5 and A.2.6 "max" must be replaced by "min".

We conclude this section with a theorem which is used in chapter III of this monograph.

A.2.8. THEOREM. *For an infinite horizon average reward Markov decision problem* MD, *let the optimal value be* W^*. *Let* $\overline{MD}$ *be the Markov decision problem which only differs from* MD *by the immediate rewards:* $\bar{r}(s,i)=r(s,i)-W_s^*$. *Then the average reward problem* $\overline{MD}$ *has optimal value* 0_z.

PROOF. For the problem MD, let (W^*,y^*) be a solution to (A.2.1) and (A.2.2). As 0_z trivially obeys (A.2.1) it follows immediately that $(0_z,y^*)$ is a solution to (A.2.1) and (A.2.2) for problem $\overline{MD}$. Then, by theorem A.2.6, 0_z is the optimal value of $\overline{MD}$.

□

A.3. RECENT LITERATURE ON STRUCTURED STOCHASTIC GAMES

The trend to analyse stochastic games with additional structure on the game parameters (rewards and transitions) has been continued during the last years. One reason is that for the general case computational procedures are complex, accentuated by the fact that even in the discounted case the value may be irrational while all the data are rational (cf. remark 4.2.5). Another reason is that structured stochastic games are often more suitable for practical applications.

Independently Sobel (1981) and Parthasarathy, Tijs & Vrieze (1984) considered SER-SIT stochastic games, i.e. games with separable reward structure and state independent transition structure. To be more specific, the reward function r is of the form $r(s,i,j) = c(s) + a(i,j)$ and the transition map p is of the form $p(t|s,i,j) = q(t|i,j)$. The action spaces are assumed to be the same for each state. Hence the rewards are built up by a term depending on the actual state and a term depending on the chosen actions, while the transitions only depend on the actions. In Sobel (1981) it is outlined how such models can be used in inventory problems and in Parthasarathy, e.a. (1984) an application to air pollution problems is presented.

For discounted zero-sum SER-SIT games the following results hold. The value V^* of the discounted game equals

$$V_s^* = c(s) + \alpha(1+\alpha)^{-1} \operatorname{Val}_{A\times B} [a(.,.)+(1+\alpha)^{-1} \Sigma_{t=1}^{z} q(t|.,.)c(t)].$$

Both players have optimal myopic stationary strategies. By a myopic strategy we mean a strategy that in each state prescribes the same (mixed) action. For a SER-SIT game an optimal myopic stationary strategy can be composed by an optimal (mixed) action of that player for the matrix game $[a(.,.)+ (1+\alpha)^{-1} \Sigma_{t=1}^{z} q(t|.,.)\, c(t)]$. The above facts can be derived in a straight way from theorem 4.2.4.

For undiscounted zerosum SER-SIT games similar results hold. The value of the undiscounted game equals

$$\operatorname{Val}_{A\times B} [a(.,.)+ \Sigma_{t=1}^{z} q(t|.,.)c(t)].1_z$$

(so state independent) and optimal myopic stationary strategies can be

composed from optimal actions of the players in the matrix game $[a(.,.)+ \Sigma_{t=1}^{z} q(t|.,.)c(t)]$. These facts follow at once from lemma 8.1.3 since $w := c$ and $g^* := \underset{A\times B}{Val}\,[a(.,.)+ \Sigma_{t=1}^{z} q(t|.,.)c(t)]$ satisfy the equations in (ii) of this lemma.

In both Sobel (1981) and Parthasarathy, e.a. (1984) also nonzerosum SER-SIT games are analysed, resulting in myopic equilibrium points.

Dirven & Vrieze (1986) too studied myopic equilibrium points. They showed how stochastic games can be used in analysing advertisement models. Their state variable corresponds to the number of customers allied to firm 1 (player 1), while the rest of the market is attracted to firm 2 (player 2). Dirven and Vrieze gave economically interpretable conditions, sufficient for the discounted payoffs to be linear in the state variable for each pair of myopic stationary strategies. These conditions are (i) $r(s,i,j) = f(i,j).s + g(i,j)$, for player $\ell = 1,2$ and (ii) $E(s,i,j) := \Sigma_{t=0}^{z} p(t|s,i,j)(t-s) = h(i,j).s+k(i,j)$. Different properties of the functions f_ℓ, g_ℓ, h and k appear to correspond to different market behaviour on advertisement budgets (the actions). They showed that if $f_\ell(.,.)$ and $h(.,.)$ are independent of i and j, then discounted myopic equilibrium points exist.

Raghavan, Tijs & Vrieze (1985) treated stochastic games with additive reward and transition structure, i.e. games for which $r(s,i,j) = r_1(s,i)$ and $p(t|s,i,j=p_1(t|s,i) + p_2(t|s,j)$. The reward structure can readily be interpreted. Concerning the transition structure, let $q_1(s,i) := \Sigma_{t=1}^{z} p(t|s,i)$ and $q_2(s,j) := \Sigma_{t=1}^{z} p(t|s,j)$. Clearly $q_1(s,i) + q_2(s,j) = 1$. Then it can be seen that the additive transition structure can be explained as: when cell (i,j) turns up in state s then with probability $q_1(s,i)$ player 1 governs the transitions according to the probability vector $(p_1(1|s,i), p_1(2|s,i), \ldots, p_1(z|s,i))q_1^{-1}(s,i)$ and with probability $q_2(s,j)$ player 2 governs the transitions according to the probability vector $(p_2(1|s,j), p_2(2|s,j), \ldots, p_2(z|s,i))q_2^{-1}(s,j)$.

For AR-AT games, when inserting the structure of the game into the discounted optimality equation, it can be seen that the matrix game $[G_{S\alpha}(V^*)]$ (cf. section 4.2) can be decomposed in a part only depending on i and a part only depending on j. Hence both players have optimal pure stationary strategies.

When regarding undiscounted games as the limit of discounted games with α tending to 0, then by the above result (and the finiteness of the action sets and the state space), there is a sequence of α's tending to 0 for which

for both players the same pure stationary strategy is optimal for each α of this sequence. It is well-known that uniformly discount optimal stationary strategies are optimal for the undiscounted version (cf. lemma 7.2.5 and Bewley & Kohlberg (1978)). Thus for AR-AT games both players possess optimal pure stationary strategies for the average evaluation criterion.

For the nonzerosum case, neither for the discounted nor for the undiscounted version equilibrium points of pure stationary strategies need to exist for AR-AT games, as examples in Raghavan, e.a. (1985) show.

The models so far discussed satisfy the orderfield property (cf. section 9.2), i.e. that the solution (value and optimal stationary strategies respectively equilibrium points) lie in the same Archimeadian field as the data of the problem. Only for models with this orderfield property a relative simple algorithm (like linear programming) might be expected and indeed for each of the above models this appears.

A further model for which the orderfield property holds is considered in Vrieze, Tijs, Parthasarathy & Dirven (1985). They analysed a two-person nonzerosum stochastic game with two states and in both states two actions of the players. Furthermore the rewards are governed by one player, say player 1, i.e. $r_\ell(s,i,j) = r_\ell(s,i)$, $\ell = 1,2$ for both states and all actions. They showed that for the extreme points of the set of stationary equilibrium points the orderfield property holds.

Finally we like to mention a paper of Raghavan (1984).He surveyed nearly all algorithms for as well discounted as undiscounted two-person zerosum stochastic games. Moreover he started interrelating the subclasses of games determined by the structure on the parameters. Most of these algorithms can also be found in this monograph (cf. chapter 6 and chapter 9).

DIRVEN, C.A.J.M. & O.J. VRIEZE (1986), *Advertisement models, stochastic games and myopic strategies,* to appear in Operations Research.

PARTHASARATHY, T., S.H. TIJS & O.J. VRIEZE (1984),*Stochastic games with state independent transitions and separable rewards.* In: Hammer, G. & D. Pallaschke (eds.), *Selected topics in Operations research and mathematical economics,* Springer Verlag, Berlin, 262-271.

RAGHAVAN, T.E.S. (1984), *Algorithms for stochastic games, a survey,* Dep. of Math., Statistics and Computer Science, University of Illinois at Chicago, Chicago, Illinois.

RAGHAVAN, T.E.S., S.H. TIJS & O.J. VRIEZE (1985), *On stochastic games with additive reward and transition structure,* J.O.T.A. 47, 451-464.

SOBEL, M.J. (1981), *Myopic solutions of Markov decision processes and stochastic games, Operations Research* 29, 995-1009.

VRIEZE, O.J., S.H. TIJS, T. PARTHASARATHY & C.A.J.M. DIRVEN (1985), *A class of stochastic games with the ordered field property,* submitted to J.O.T.A.

REFERENCES

AUMANN, R.J. (1964), *Mixed and Behaviour strategies in infinite extensive games*. In: Dresher, M., L.S. Shapley & A.W. Tucker (eds.), *Advances in Game Theory*, Ann. of Math. Stud. no. 52, 627-650. Princeton Univ. Press, Princeton.

BATHER, J. (1973), *Optimal decision procedures for finite Markov chains, Part II*, Adv. Appl. Prob. 5, 521-540.

BERGE, C. (1959), *Espaces Topologiques*, Dunod, Paris.

BEWLEY, T. & E. KOHLBERG (1976a), *The asymptotic theory of stochastic games*, Math. of O.R. 1, 197-208.

———— & ———— (1976b), *The asymptotic solution of a recursive equation arising in stochastic games*, Math. of O.R. 1, 321-336.

———— & ———— (1978), *On stochastic games with stationary optimal strategies*, Math. of O.R. 3, 104-125.

BLACKWELL, D. (1962), *Discrete dynamic programming*, Ann. of Math. Stat. 33, 719-726.

———— (1965), *Discounted dynamic programming*, Ann. of Math. Stat. 36, 226-235.

———— & T.S. FERGUSON (1968), *The big match*, Ann. of Math. Stat. 39, 159-163.

BOHNENBLUST, H.F., S. KARLIN & L.S. SHAPLEY (1950), *Solutions of discrete two-person games*, In: Kuhn, H.W. & A.W. Tucker (eds.), *Contributions to the theory of games, Vol. I*, Ann. of Math Stud. no. 24, 51-72, Princeton Univ. Press, Princeton.

BROWN, B. (1965), *On the iterative method of dynamic programming on a finite state space discrete time Markov process*, Ann. of Math. Stat. 36, 1279-1285.

BROWN, G.W. (1949), *Some notes on computation of game solutions*, RAND Report P-78, The RAND Corporation, Santa Monica, California.

———— (1951), *Iterative solution of games by fictitious play*, In: Koopmans, T.C. (ed.), *Activity Analysis of Production and Allocation*, 374-376, John Wiley, New York.

DANSKIN, J.M. (1954), *Fictitious play for continuous games*, Nav. Res. Log. Quart. 1, 313-320.

DENARDO, E.V. (1971), *Markov renewal programs with small interest rates*, Ann. of Math. Stat. 42, 477-496.

DERMAN, C. (1970), *Finite state Markovian decision processes*, Academic Press, New York.

———— & R. STRAUCH (1966), *A note on memoryless rules for controlling sequential processes*, Ann. of Math. Stat. 37, 276-278.

FEDERGRUEN, A. (1978), *Markovian control problems*, Ph.D. Dissertation, Math. Centre, Amsterdam.

FILAR, J.A. (1979), *Algorithms for solving some undiscounted stochastic games*, Ph. D. Dissertation, Univ. of Illinois, Chicago.

———— (1981), *Ordered field property for stochastic games when the player who controls transitions changes from state to state*, J.O.T.A. 34, 503-515.

———— & T.E.S. RAGHAVAN (1979), *An algorithm for solving an undiscounted stochastic game in which one player controls transitions*, Research Memorandum, Univ. of Illinois, Chicago.

———— & ———— (1980), *Two remarks concerning two undiscounted stochastic games*, Technical Report no. 329, Dep. of Math. Sciences, The John Hopkins Univ., Baltimore.

GILLETTE, D. (1957), *Stochastic games with zero stop probabilities*, In: Dresher, M., A.W. Tucker & P. Wolfe (eds.), *Contributions to the theory of games, Vol. III*, Ann. of Math. Stud. 39, Princeton Univ. Press, Princeton.

GROENEWEGEN, L. (1981), *Characterization of optimal strategies in dynamic games*, MC Tract 90, Math. Centre, Amsterdam.

———————— & J. WESSELS (1976), *On the relation between optimality and saddle-conservation in Markov games*, Eindhoven Univ. of Technology, Dep. of Math., Memorandum COSOR 76-14.

HOFFMAN, A. & R. KARP (1966), *On nonterminating stochastic games*, Man. Science 12, 359-370.

HORDIJK, A. (1974), *Dynamic programming and Markov potential theory*, Math. Centre Tract No. 51, Math. Centre, Amsterdam.

———————— & L.C.M. KALLENBERG (1979), *Linear programming and Markov decision chains*, Man. Science 25, 352-362.

———————— & ———————— (1981a), *Linear programming and Markov games I*, In: Moeschlin, O. & D. Pallaschke (eds.), *Game Theory and Mathematical Economics*, 291-305, North-Holland, Amsterdam.

———————— & ———————— (1981b), *Linear programming and Markov games II*, In: Moeschlin, O. & D. Pallaschke (eds.), *Game Theory and Mathematical Economics*, 307-320, North-Holland, Amsterdam.

———————— & H.C. TIJMS (1975), *A modified form of the iterative method of dynamic programming*, Ann. of Stat. 3, 203-208.

———————— & O.J. VRIEZE & G.L. WANROOIJ (1976), *Semi-Markov strategies in stochastic games*, Report BW 68/76, Math. Centre, Amsterdam.

———————— & ———————— & ———————— (1983), *Semi-Markov strategies in stochastic games*, Int. J. of Game Theory 12, 81-89.

HOWARD, R. (1960), *Dynamic programming and Markov processes*, John Wiley, New York.

KARLIN, S. (1959), *Mathematical methods and the theory of games, Vol. I*, Addison Wesley, London.

KEMENY, J. & J. SNELL (1961), *Finite Markov chains*, Van Nostrand, Princeton.

KOHLBERG, E. (1974), *Repeated games with absorbing states*, Ann. of Stat. 2, 724-738.

KOLMOGOROV, A. (1973), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik 2, no. 3, Springer Verlag, Berlin.

KRABS, W. (1977), *Stetige Abänderung der Daten by nichtlinearer Optimierung und ihrer Konsequenzen*, Oper. Res. Verfahren, Vol. 25, 93-113.

KUHN, H. (1953), *Extensive games and the problem of information*, In: Kuhn, H.W. & A.W. Tucker (eds.), *Contributions to the theory of games, Vol. II*, Ann. of Math. Stud. no. 28, 193-216, Princeton Univ. Press, Princeton.

KUSHNER, H.J. & S.G. CHAMBERLAIN (1969), *Finite state stochastic games: existence theorems and computational procedures*, IEEE, Trans Automatic Control AC-14, 248-255.

LIGGETT, T. & S. LIPPMAN (1969), *Stochastic games with perfect information and time average payoff*, SIAM Review 11, 604-607.

LUCE, R. & H. RAIFFA (1957), *Games and decisions*, John Wiley, New York.

MAITRA, A. & T. PARTHASARATHY (1970), *On stochastic games*, J.O.T.A. 5, 289-300.

MANGASARIAN, O.L. (1969), *Nonlinear Programming*, Mc Graw Hill Book Co., New York.

MERTENS, J.F. & A. NEYMAN (1981), *Stochastic games*, Int. J. of Game Theory 10, 53-66.

MILLS, H.D. (1956), *Marginal values of matrix games and linear programs*, In: Kuhn, H.W. & A.W. Tucker (eds.), *Linear inequalities and related systems*, Ann. of Math. Stud. no. 38, 183-193, Princeton Univ. Press, Princeton.

MIYASAWA, K. (1961), *On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game*, Research Memorandum 33, Princeton Univ. Press, Princeton.

MONASH, C.A. (1979), *Stochastic games: The minmax theorem*, Thesis submitted to the dep. of Math. of the Harvard Univ., Cambridge, Massachusetts.

PARTHASARATHY, K.R. (1967), *Probability measures on metric spaces*, Acad. Press, New York.

PARTHASARATHY, T. (1971), *Discounted and positive stochastic games*, Bull. Amer. Math. Soc. 77, 134-136.

——————— & T.E.S. RAGHAVAN (1978), *An orderfield property for stochastic games when one player controls the transitions*, Game Theory Conference, Cornell, Ithaca.

——————— & ——————— (1981), *An orderfield property for stochastic games when one player controls transition probabilities*, J.O.T.A. 33, 375-392.

POLLATSCHEK, M. & B. AVI-ITZHAK (1969), *Algorithms for stochastic games with geometrical interpretation*, Man. Science 15, 399-415.

RAO, S., R. CHANDRASEKARAN & K. NAIR (1973), *Algorithms for discounted stochastic games*, J.O.T.A. 11, 627-637.

RIEDER, U. (1979), *Equilibrium plans for non-zero-sum Markov games*, In: Moeschlin, D. & D. Pallaschke (eds.), *Game theory and related topics*, North-Holland, Amsterdam.

ROBINSON, J. (1950), *An iterative method of solving a game*, Ann. of Math. 54, 296-301.

ROGERS, P.D. (1969), *Non-zero-sum stochastic games*, Ph.D. Dissertation, Report ORC 69-8, Operations Research Centre, Univ. of California, Berkeley,

ROSENMÜLLER, J. (1971), *Über Periodizitätseigenschaften spieltheoretischer Lernprozesse*, Z. Wahrscheinlichkeitstheorie verw. Geb. 17, 259-308.

ROTHBLUM, U.G. (1978), *Solving stopping stochastic games by maximizing a linear function subject to quadratic constraints*, In: Moeschlin, O. & D. Pallaschke (eds.), *Game theory and related topics*, 103-105, North-Holland, Amsterdam.

SCHWEITZER, P.J. (1968), *Perturbation theory and finite Markov chains*, J. of Appl. Prob. 5, 401-413.

——————— & A. FEDERGRUEN (1978), *The asymptotic behaviour of undiscounted value iteration in Markov decision problems*, Math. of Oper. Res. 2, 360-382.

SHAPIRO, H.N. (1958), *Note on a computation method in the theory of games*, Comm. on Pure and Appl. Math. 11, 587-593.

SHAPLEY, L.S. (1953), *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. 39, 1095-1100.

——————— (1964), *Some topics in two-person games*, In: Dresher, M., L.S. Shapley & A.W. Tucker (eds.), *Advances in game theory*, Ann. of Math. Stud. no. 52, 1-28, Princeton Univ. Press, Princeton.

——————— & R.N. SNOW (1950), *Basic solutions of discrete games*, In: Kuhn, H.W. & A.W. Tucker (eds.), *Contributions to the theory of games, Vol. I*, Ann. of Math. Stud. no. 24, 27-35, Princeton Univ. Press, Princeton.

SOBEL, M.J. (1971), *Non-cooperative stochastic games*, Ann. of Math. Stat. 42, 1930-1935.

———— (1973), *Continuous stochastic games*, J. of Appl. Prob. 10, 597-604.

STERN, M.A. (1975), *On stochastic games with limiting average payoff*, Ph.D. Dissertation, Univ. of Illinois, Chicago.

TIJS, S.H. (1976a), *Semi-infinite programs and semi-infinite matrix games*, Report 7630, Nijmegen Catholic Univ., Dep. of Math., Nijmegen.

———— (1976b), *I: A characterization of the value of linear programs, II: A characterization of the value of zero-sum two-person games*, Report 7632, Nijmegen Catholic Univ., Dep. of Math., Nijmegen.

———— (1977), *ε-Equilibrium point theorems for two-person games*, Oper. Res. Verfahren 26, 755-766.

———— (1980), *Stochastic games with one big action space in each state*, Methods of Op. Res. 38, 161-173.

———— (1981), *Characterization of the value of zero-sum two-person games*, Nav. Res. Log. Quart. 28, 153-156.

———— & O.J. VRIEZE (1979), *Characterizing properties of the value function of stochastic games*, Report BW 110/79, Math. Centre, Amsterdam.

———— & ———— (1980), *Perturbation theory for games in normal form and stochastic games*, J.O.T.A. 30, 549-567.

———— & ———— (1981), *Characterizing properties of the value function of stochastic games*, Technical Note, J.O.T.A. 33, 145-150.

———— & ———— (1986), *On the existence of easy initial states for undiscounted stochastic games*, to appear in Math. of O.R..

VAN DEN AKKER, A.G. (1976), *Some subjects in game theory; Markov games, continuous games,* (in Dutch), Master's Thesis, Eindhoven, Univ. of Technolgy, Eindhoven.

VAN DER WAL, J. (1977), *Discounted Markov games: successive approximations and stopping times,* Int. J. of Game Theory 6, 11-22.

———————— (1981), *Stochastic dynamic programming,* MC Tract 139 Math. Centre, Amsterdam.

VILKAS, E.I. (1963), *Axiomatic definition of the value of a matrix game,* Theory of Prob. and its Appl. 8, 304-307.

VON NEUMANN, J. (1928), *Zur Theory der Gesellschaftspiele,* Mathematische Annalen 100, 295-320.

———————— & O. Morgenstern (1944), *Theory of games and economic behaviour,* Princeton Univ. Press, Princeton.

VRIEZE, O.J. (1976), *The stochastic non-cooperative countable-person game with countable state space and compact action spaces under the discounted payoff criterion,* Report BW 66/76, Math. Centre, Amsterdam.

———————— (1979a), *Characterization of optimal stationary strategies in undiscounted stochastic games,* Report BW 102/79, Math. Centre, Amsterdam.

———————— (1979b), *Implications of the asymptotic behaviour of* V_n *on characterizing properties of stochastic games,* Report 7929, Nijmegen Catholic Univ., Dep. of Math., Nijmegen.

———————— (1981a), *Linear programming and undiscounted stochastic games in which one player controls transitions,* Op. Res. Spektrum 3, 29-35.

———————— (1981b), *Characterization of undiscounted stochastic games with optimal stationary strategies,* Report 8111, Nijmegen Catholic Univ., Dep. of Math., Nijmegen.

———————— & S.H. TIJS (1980), *Relations between game parameters, value, and optimal strategy spaces in stochastic games and construction of games with given solution*, J.O.T.A. 31, 501-513.

———————— & ———————— (1982), *Fictitious play applied to sequences of games and discounted stochastic games*, Int. J. of Game Theory 11, 71-85.

———————— , ———————— , T.E.S. RAGHAVAN & J.A. FILAR (1983), *A finite algorithm for the switching control stochastic game*, Op. Res. Spektrum 5, 15-24.

WESSELS, J. (1977), *Markov games with unbounded rewards*. In: M. Schäl, Dynamische optimierung, Bonner Mathematische Schriften, nr 98.

WEYL, H. (1950), *Elementary proof of a minimax theorem due to Von Neumann*, In: Kuhn, H.W. & A.W. Tucker (eds.), *Contributions to the theory of games, Vol. I*, Ann. of Math. Stud. 24, 19-25, Princeton Univ. Press, Princeton.

WHITT, W. (1975), *Continuity of Markov processes and dynamic programs*, Yale Univ., New Haven, Connecticut.

———————— (1977), *Representation and approximation of non-cooperative sequential games*, Operations Research Centre, Bell Laboratories, Holmdel, New Jersey.

AUTHOR INDEX

SUBJECT INDEX

SYMBOL INDEX

(Symbols with only local significance are not included).

$Y^{\tau}_{\mu\nu}$	11
z	7
$Z^{\tau}_{\mu\nu}$	11

NOTATIONS

$\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}^m$ are the set of natural numbers, the set of real numbers and the m-fold Cartesian product of $\mathbb{R}$ respectively.

$\mathbb{N}_m := \{1,2,\ldots,m\}$ and $\mathbb{N}^- := \{-1,-2,-3,\ldots\}$.

$1_m := (1,1,\ldots,1) \in \mathbb{R}^m$ and $0_m := (0,0,\ldots,0) \in \mathbb{R}^m$.

For $x=(x_1,x_2,\ldots,x_m) \in \mathbb{R}^m$, $||x|| := \max_{k\in\mathbb{N}_m} |x_k|$

For $x,y \in \mathbb{R}^m$, $d(x,y) := ||x-y||$.

For $x \in \mathbb{R}^m$, $Car(x) := \{k \mid k \in \mathbb{N}_m,\ x_k \neq 0\}$.

For $x,y \in \mathbb{R}^m$, $x \geq y$ if and only if $x_k \geq y_k$ for each $k \in \mathbb{N}_m$ and $x > y$ if and only if $x \geq y$ and at least for one component k it holds that $x_k > y_k$.
$x \leq y$ and $x < y$ are defined analogously.

By an m×n-matrix we mean a matrix consisting of m rows and n columns.

I_{mm} denotes the m×m-matrix with each entry equal to 1.

If K is an m×n-matrix, K^T represents the transpose of K.

If K_1 and K_2 are two m×n-matrices, then
$d(K_1,K_2) := \max_{m\in\mathbb{N}_m,\, n\in\mathbb{N}_n} |K_1(m,n)-K_2(m,n)|$.
The adjoint of an m×m-matrix K is denoted by adj(K) and the determinant of K is denoted by det(K).

A vector is supposed to be a column vector. However, when no confusion arises we often write x.K instead of $x^T.K$ for $x \in \mathbb{R}^m$ and K an m×n-matrix.

For a finite set S, $|S|$ is the numer of elements of S.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.