CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
Telephone 31 - 20 592 9333, telex 12571 (mactr nl),
telefax 31 - 20 592 4199

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

CWI Tract 82

Optimality and equilibria in stochastic games

F. Thuijsman

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1990 Mathematics Subject Classification: 90D15, 90D10, 90D05, 90C47, 90C50, 93E05.
ISBN 90 6196 406 7
NUGI-code: 811

Preface

Being my PhD advisors, Stef Tijs and Koos Vrieze guided me on a journey into the world of stochastic games. In this monograph, which differs from my thesis only at some minor points, I present some of the mysteries we encountered on this trip and how they were dissolved. For helping me to find my way, I wish to express my gratitude to Koos and Stef.

The research that led to this monograph was financially supported by the Netherlands Organization for Scientific Research NWO via the Stichting Mathematisch Centrum, for which I am very thankful. I am also very grateful to Yvonne Samseer, who accurately typed this work, and to the CWI Publications Department for publishing this book.

Frank Thuijsman
Maastricht, January 1991

Contents

Chapter 1

Preliminaries

1.1 Introduction and summary

This monograph deals with two-person stochastic games with finite state and action spaces. The theory of stochastic games started by a paper of Shapley [1953]: 'Stochastic games'. In this fundamental article Shapley combined the dynamic programming model of Bellman [1952, 1957] with the matrix games considered by Von Neumann [1928] and Von Neumann & Morgenstern [1944].

In the dynamic programming model problems of the following type are considered. At a discrete number of stages in time, a person has to choose one of finitely many actions; that choice will determine an immediate payoff as well as a probability vector according to which a new state is appointed, where an action has to be chosen at the next stage. There are finitely many states, each with its own finite action space. The person faces the problem to decide which action choices give the highest income. Here the income is determined by discounting, averaging or, in some special cases, by simply adding all immediate payoffs.

In matrix games two persons, usually called players, face an $m\times n$-matrix with real entries. Simultaneously and independently player 1 has to choose a row and player 2 has to choose a column. The real number in the chosen entry is the amount player 2 has to pay to player 1. Of course, the assumption is made that player 1 wants to maximize the expected payoff and, at the same time, player 2 wants to minimize the expected payoff.

Shapley [1953] combined the features of dynamic programming with those of a matrix game. Thus a stochastic game can be seen as a finite collection of matrix games, one to be played at each stage, where the motion among the matrix games depends at each stage on the current state and on the actions chosen. The collection of stages is assumed to be $\mathbb{N}=\{1,2,3,...\}$. The stochastic game is a non-cooperative game, meaning that the players are not allowed to make binding agreements. These stochastic games as examined by Shapley [1953] are zero-sum stochastic games, i.e. one player is paying the other player and the gain of one player is the loss of the other player. In zero-sum games the two players have strictly opposite interests. The question in such games is whether there is a certain amount which player 1 can guarantee to receive (in expectation) regardless of the choices of player 2, while player 2 has a strategy such that he will not need to pay more than that amount (in expectation), regardless of the choices of player 1. Whenever it exists, this unique amount is called the value of the game and the strategies used by the players to guarantee this value are called optimal strategies. If the players can only achieve near-optimality we speak of ϵ-optimal strategies.

If it is not required that one player is paying the other player, then the game is called a general-sum stochastic game. For such a game the states no longer correspond with matrix games but with bimatrix games: in each entry of the matrix there are two real numbers, the first indicating the payoff to row-player 1, the second indicating the payoff to column-player 2. Now the players need no longer have strictly opposite interests and hence the notions 'value' and 'optimality' lose their meaning. In general-sum stochastic games the usual solution concept is that of (ϵ-)equilibria. An (ϵ-)equilibrium is a pair of strategies with the property that neither player can gain (more than ϵ ($\epsilon>0$)) by unilateral deviation. This concept of equilibrium was introduced by Nash [1951] for bimatrix games, and it is therefore known as Nash-equilibrium. Fink [1964] combined Shapley's (zero-sum) stochastic game model with Nash's (general-sum) solution concept to examine general-sum stochastic games.

In this monograph we shall deal with the general-sum stochastic game model as well as with the zero-sum stochastic game model. The existence of ϵ-equilibria, or of the value and ϵ-optimal strategies, may depend on the initial state. It should be clear that any of the states in a stochastic game can function as the starting state. Just as in dynamic programming it is often useful to consider the problems for the different initial states simultaneously. Thus the value of a stochastic game is in fact a value-vector, where coordinates correspond with the starting states. Likewise an ϵ-equilibrium is a pair of strategies which is an ϵ-equilibrium for all initial states. A further remark to be made is that both the zero-sum and general-sum solution concept depend on the criterion that is used to evaluate the incomes of the players. As in dynamic programming models this criterion can lead to the discounted incomes, the limiting average incomes or the total incomes, where the latter may be quite meaningless if the stochastic game has no specific properties.

In his stochastic game model Shapley required that in each state, for any pair of actions chosen, there is a strictly positive probability that the play terminates. Hence Shapley could derive his results with respect to the total income, since any play would ever terminate with probability 1. If, in such a terminating stochastic game all stopping probabilities are equal to each other, then examining total incomes in such a game, is equivalent to examining discounted incomes in a related non-terminating stochastic game (cf. Shapley [1953]). Gillette [1957] was the first to examine limiting average incomes in (non-terminating) stochastic games.

In this monograph we consider non-terminating stochastic games. We will deal with all three evaluation criteria. The emphasis however is on stochastic games with respect to the limiting average reward criterion since they have turned out to be quite hard to solve and since the existence of limiting average ϵ-equilibria can be seen as the major open problem in stochastic game theory nowadays. For stochastic games fine solutions are known to exist with respect to the β-discounted reward criterion, whereas with respect to the total reward criterion similar problems as for the limiting average reward criterion occur.

More precisely: for the discounted reward criterion solutions exist in terms of stationary strategies, i.e. strategies for which the action choices of the players only depend on the state that is currently visited (cf. Shapley [1953] and Fink [1964]). For the limiting average reward criterion this need not be; the players may have to take into account which actions their opponent has used in the past. This was illustrated by an example in Gillette [1957] which has been solved by Blackwell & Ferguson [1968] using history dependent strategies (cf. example 1.7.4 below). For the total reward criterion history dependent strategies may be required as well (cf. section 5.4). In the example of Gillette [1957], which became known as 'the big match', player 1 has no history independent limiting average ϵ-optimal strategies. So the solution by Blackwell & Ferguson [1968] of this big match clearly showed that for the limiting average reward criterion history dependent strategies are really indispensable. Unfortunately history dependent strategies have a rather complex structure and often lead to computational difficulties. Stationary strategies can be seen as the most simple strategies in stochastic games. Any pair of stationary strategies is related with a Markov process on the set of states. This implies, as will be clear in the sequel of this chapter, that for stationary strategies rewards can be computed rather straightforwardly. Hence from the point of view of computations, and hence of potential applications of stochastic games, the class of stationary strategies is particularly interesting. Therefore in literature, as in this monograph, a lot of attention is given to stationary strategies.

Now, knowing that for the limiting average reward criterion, as well as for the total reward criterion, solutions (ϵ-optimal strategies/ϵ-equilibria) may fail to exist if the players are restricted to stationary strategies, it is of special interest to find out what characterizes stochastic games which do have stationary solutions. For the limiting average criterion such a characterization, by means of a system of equations, is presented in chapter 5, due to Vrieze [1987-a]. We present a similar characterization for the existence of stationary total optimal strategies in chapter 5. In chapter 6, which is based on Filar et al. [1991], we completely characterize the existence of stationary solutions by means of global optima of suitably constructed non-linear mathematical programs. This is done for each of the three evaluation criteria and for zero-sum as well as for general-sum stochastic games. Previously characterizations for the existence of stationary solutions have also been reported in Sobel [1971], Bewley & Kohlberg [1978], Filar & Schultz [1986] and Schultz [1987]. So far these characterizations are formulated for zero-sum or general-sum stochastic games with finite state and action spaces without some specific extra structure. Besides, several classes of stochastic games, i.e. stochastic games with a special condition on the payoff and/or transition structure, have been examined for which stationary solutions exist. We mention: unichain/irreducible stochastic games (cf. Gillette [1957], Hoffman & Karp [1966], Rogers [1969], Sobel [1971], Federgruen [1978]); recursive games (cf. Everett [1957], Orkin [1972], Thuijsman & Vrieze [1990-b]); single controller stochastic games (cf. Stern [1975], Parthasarathy & Raghavan [1981], Hordijk & Kallenberg [1981-b], Filar [1984, 1986], Filar & Raghavan [1984], Vrieze [1987-a]); stochastic games with perfect

information (cf. Gillette [1957], Liggett & Lippman [1969]); switching control stochastic games (cf. Filar [1981-b], Filar & Schultz [1987], Vrieze [1987-a], Vrieze et al. [1983]); stochastic games with state independent transitions and separable rewards (cf. Sobel [1981], Parthasarathy et al. [1984]); stochastic games with additive rewards and additive transitions (cf. Raghavan et al. [1985], Filar & Schultz [1987]).
For many of these classes there are algorithms known to solve such games. For a survey on algorithms we refer to Raghavan & Filar [1989].

Although apparently for many classes of stochastic games stationary solutions exist, in general they do not, as was illustrated by the big match. Kohlberg [1974] extended the work of Blackwell & Ferguson [1968] by showing that for any zero-sum repeated game with absorbing states (cf. section 4.4) the limiting average value exists. Based on techniques of these papers and using results of Bewley & Kohlberg [1976] on asymptotic properties of discounted solutions for zero-sum stochastic games, Mertens & Neyman [1981] derived that for any zero-sum stochastic game the limiting average value exists.

However, as mentioned before, history dependent strategies will be needed to achieve ϵ-optimality. Since we would prefer stationary solutions, it is fortunate to know that in any zero-sum stochastic game there is, for each player, a non-empty set of initial states for which this player has a stationary limiting average optimal strategy. A first proof for this result is given by Tijs & Vrieze [1986]. In chapter 2 of this monograph we present a new, and more elementary, proof for this result. Besides we give a sufficient condition for player 1 to have stationary limiting average ϵ-optimal strategies for all initial states with maximal or minimal limiting average value.

For the general-sum case a similar result is presented in chapter 2: there is always a non-empty set of initial states for which an 'almost-stationary' limiting average ϵ-equilibrium exists. In chapter 3 this result is extended by formulating sufficient conditions for the existence of an 'almost-stationary' limiting average ϵ-equilibrium in any general-sum stochastic game. The existence of limiting average ϵ-equilibria is one of the major remaining problems in stochastic game theory.

The history of general-sum stochastic games started with Fink [1964], who proved the existence of stationary β-discounted equilibria. Other proofs for this result have been given by Takahashi [1964], Rogers [1969] and Sobel [1971]. The existence of stationary limiting average equilibria has been shown for several classes of stochastic games, most of those mentioned above. Inspired by Sorin [1986] (cf. example 1.8.6 in this chapter), the existence of (history dependent) limiting average ϵ-equilibria for general-sum repeated games with absorbing states was shown by Vrieze & Thuijsman [1989] using Kohlberg [1974]. In chapter 4 we give a slightly modified proof for this result.

Although stochastic games with just one state, better known as 'repeated games', are part of the model we discuss in this monograph, the theory on such games developed in a rather specific direction. Therefore we do not discuss repeated games in particular in this monograph. For surveys on repeated games we refer to Aumann [1981], Mertens [1986] and Sorin [1988].

Closing this brief introduction to stochastic game theory, we wish to refer to the surveys on stochastic games by Parthasarathy & Stern [1977], Raghavan & Filar [1989], Vrieze [1987-b] and Thuijsman [1987].

We now describe the set up of this monograph.

In the remainder of this chapter we give formal definitions of the stochastic game model with its solution concepts. Furthermore we formulate the major historic results in this field, in view of the topics in this monograph, and we derive several preliminary lemmas and discuss some examples.

In chapter 2 we show that for any general-sum stochastic game there is a non-empty set of initial states for which there exists an almost stationary limiting average ϵ-equilibrium, i.e. a limiting average ϵ-equilibrium consisting of stationary strategies amplified with threat-strategies. For zero-sum stochastic games we give an elementary proof for the existence of easy initial states for each player, i.e. starting states for which this player has a stationary limiting average optimal strategy. Tijs & Vrieze [1986] already proved this result, but our proof is significantly simpler. For the set of initial states with maximal or minimal limiting average value, we give a sufficient condition for each player to have stationary limiting average ϵ-optimal strategies. We also show that there may be states which are neither (ϵ-)easy for player 1 nor for player 2.

In chapter 3 we extend the general-sum results of chapter 2 to formulate sufficient conditions for the existence of a limiting average ϵ-equilibrium (for all starting states). These sufficient conditions are formulated in terms of properties of an arbitrary sequence of stationary β-discounted equilibria, which without loss of generality can be assumed to converge for β going to 1.

In chapter 4 we show that our results of chapters 2 and 3 imply the existence of limiting average (ϵ)-equilibria for several subclasses: unichain stochastic games (which includes irreducible stochastic games), stochastic games with state independent transitions (SIT), repeated games with absorbing states.

Chapter 5 is devoted to zero-sum stochastic games with the total reward criterion. We show that the total value may fail to exist, even on the condition that the limiting average value is 0. On the stronger condition of limiting average value 0 and both players possessing stationary limiting average optimal strategies, history dependent behavior strategies may still be indispensable for the players to achieve total ϵ-optimality. This is illustrated by an example: 'the bad match'. We give characterizations for the existence of stationary total optimal strategies (as well as for stationary β-discounted optimal and stationary limiting average optimal strategies). We relate this total reward criterion with the β-discounted and the limiting average reward criterion.

Chapter 6 deals with mathematical programs connected to stochastic games. With respect to all three evaluation criteria non-linear programs are given that completely characterize the existence of stationary equilibria / (ϵ-)optimal strategies. Our characterization with respect to the total reward criterion is restricted by the assumption that the limiting average reward is 0 for all pairs of stationary strategies.

1.2 The stochastic game model

1.2.1 Definition

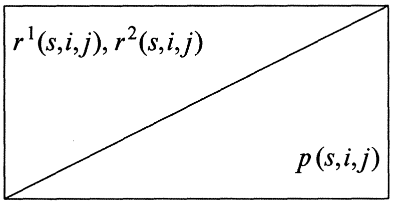
A stochastic game Γ *is a 6-tuple* $<S, \{A_s : s \in S\}, \{B_s : s \in S\}, r^1, r^2, p>$, *where:*
$S := \{1,2,\ldots,z\}$, $z \in \mathbb{N}$, *is the set of states, or state space;*
$A_s := \{1,2,\ldots,m_s\}$, $m_s \in \mathbb{N}$, *is the action space of player 1 in state* $s \in S$;
$B_s := \{1,2,\ldots,n_s\}$, $n_s \in \mathbb{N}$, *is the action space of player 2 in state* $s \in S$;
$r^k : \bigcup_{s \in S} \{s\} \times A_s \times B_s \to \mathbb{R}$ *is the payoff function for player* $k \in \{1,2\}$;
$p : \bigcup_{s \in S} \{s\} \times A_s \times B_s \to \Delta^z$ *is the transition probability map, with*

$$p(s,i,j) = (p(1|s,i,j), p(2|s,i,j), \ldots, p(z|s,i,j)).$$

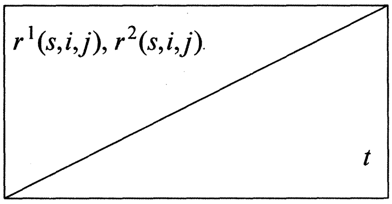
Here $\Delta^n := \{a \in \mathbb{R}^n : a \geqslant 0, \sum_{i=1}^{n} a_i = 1\}$, *for any* $n \in \mathbb{N}$.

1.2.2 Notation

In the examples in this monograph stochastic games will be given as a collection of matrices $\{matrix(1), matrix(2), \ldots, matrix(z)\}$, *where matrix(s) has size* $m_s \times n_s$ *and entry* (i,j) *of matrix(s) is given as*



or, in case for some $t \in S$ *we have* $p(t|s,i,j) = 1$, *as*



A play of the stochastic game , a 'round' of the game, develops in the following way. At each stage $n \in \mathbb{N}$ play is in precisely one of the states in S. Play starts at stage 1 in some state $s_1 \in S$, the initial state. If at stage $n \in \mathbb{N}$ play is in state $s_n \in S$, then simultaneously and independently, without making binding agreements, player 1 has to choose some $i_n \in A_{s_n}$ and player 2 has to choose some $j_n \in B_{s_n}$. Once these choices are made, player 1 receives $r^1(s_n,i_n,j_n)$, player 2 receives $r^2(s_n,i_n,j_n)$ and next play moves with probability $p(s_{n+1}|s_n,i_n,j_n)$ to state $s_{n+1} \in S$, where the players choose actions at stage $n+1$.

The players are allowed to randomize over their actions, i.e. in state s player 1 (for example) can use some 'mixed action' $x=(x(1),x(2),\ldots,x(m_s))\in\Delta^{m_s}$ which is to be interpreted as choosing 'pure action' $i\in A_s$ with probability $x(i)$. At each stage $n\in\mathbb{N}$ both players know Γ as well as the 'history' $h_n=(s_1,i_1,j_1,s_2,i_2,j_2,\ldots,s_{n-1},i_{n-1},j_{n-1},s_n)$ but neither player knows the mixed actions his opponent has used in the past. Each of the players is interested in maximizing his individual income, which is some kind of evaluation of the payoffs over all stages. Both players are assumed to use the same evaluation criterion (cf. 1.4 below).

Here we already make two remarks:
First, notice that once a play is started, it never stops, although some stopping-like things may happen as we will see in the sequel. Second, it should be observed that with any stochastic game situation there are in fact z games to be considered, one for each starting state. It is often useful to treat these z games simultaneously.

1.3 Strategies

Any plan a player uses to play a stochastic game, is called a strategy. So a strategy tells a player at all stages, in any state and for any history, what mixed action to use. Within the set of all these strategies one can discern several classes with different complexities. The most simple class of strategies is that of the stationary strategies. A player who uses a stationary strategy has fixed a mixed action for each state, which he uses at any stage the play is in that state, no matter what history preceded.

1.3.1 Definition

A stationary strategy for player 1 is given as an element $x\in X:=\bigtimes_{s=1}^{z}\Delta^{m_s}$.
A stationary strategy for player 2 is given as an element $y\in Y:=\bigtimes_{s=1}^{z}\Delta^{n_s}$.

These stationary strategies are of fundamental importance in the analysis of stochastic games.

A class of slightly more complex strategies, is that of Markov strategies. A player who uses a Markov strategy, has fixed a mixed action for each state and stage, to be used at that stage regardless of the history that preceded.

1.3.2 Definition

A Markov strategy for player 1 is given as a function $f:\mathbb{N}\rightarrow X$.
A Markov strategy for player 2 is given as a function $g:\mathbb{N}\rightarrow Y$.
The class of Markov strategies for player 1 (2) is denoted by F (G).

Observe that stationary strategies are stage independent Markov strategies.

The most complex strategies to be considered in this monograph, are behavior strategies. A player who uses a behavior strategy will consider the history of the play, in any state and at any stage, to decide what mixed action is to be used. Since this type of strategies is the most general to be considered (cf. Aumann [1964]), we will often leave out the word behavior.

1.3.3 Definition

For $n \in \mathbb{N}$ let $h_n := (s_1, i_1, j_1, s_2, i_2, j_2, \dots, s_{n-1}, i_{n-1}, j_{n-1}, s_n)$ be the history up to stage n, i.e. h_n is the sequence of states and actions that occurred up to appearance in some state s_n at stage n.

Let $H_n := \{(s_1, i_1, j_1, s_2, i_2, j_2, \dots, s_{n-1}, i_{n-1}, j_{n-1}, s_n) : s_k \in S,\ i_k \in A_{s_k}, j_k \in\ B_{s_k}\}$ be the set of histories up to stage n.

A (behavior) strategy for player 1 is given as a function π: $\bigcup_{n=1}^{\infty} H_n \to X$.

A (behavior) strategy for player 2 is given as a function σ: $\bigcup_{n=1}^{\infty} H_n \to Y$.

The class of (behavior) strategies for player 1 (2) is denoted by Π (Σ).

So we have $X \subset F \subset \Pi$ and $Y \subset G \subset \Sigma$. Although at each stage the current state is part of the history up to that stage, we say that Markov strategies are history independent strategies.

In the above definitions for strategies the players use mixed actions. The sets of mixed actions contain pure actions, i.e. choosing some row or column with probability 1. Therefore we can also define pure strategies.

1.3.4 Definition

A pure strategy is a strategy for which, for all states, stages and histories, pure actions are used. The set of pure strategies for player 1 is denoted by Π^p*; his set of pure Markov strategies (pure stationary strategies) is denoted by* F^p (X^p). *For player 2 the notations* Σ^p*,* G^p *and* Y^p *have a similar meaning.*

1.4 Evaluation criteria

As in other game theoretic models the assumption in stochastic games is, that each player wants to maximize his individual income. However, a play of a stochastic game never ends and payoffs occur at all stages. Therefore the players should use some kind of criterion to evaluate those sequences of payoffs in order to decide what strategy they prefer to use. More precisely, each player wants to be able to compare the expected income for several pairs of strategies in order to choose a good strategy. In this monograph we look at the expected income because of the stochastic element caused by the transition probabilities and by the use of mixed actions.

1.4.1 Definition

Let $(\pi,\sigma)\in\Pi\times\Sigma$ be given and let $s\in S$ be the initial state. Define $R^k(n)$ as the random variable representing the payoff at stage n to player k. Define $E_{s\pi\sigma}[R^k(n)]$ as the expected payoff at stage n to player k conditional on s,π,σ.

The above definition is possible because a play starting in state s, with the players using π and σ, leads to a well-defined stochastic process on the set of states. For: at stage 1 both π and σ prescribe some mixed action to be used in state s; hence an expected payoff at stage 1 is well-defined for both players, just as are the transitions to the next state. In any new state, at stage 2, strategies π and σ again prescribe mixed actions to be used, which determines an expected payoff for stage 2, etc.

In this monograph we consider three evaluation criteria: the β-discounted reward criterion, the limiting average reward criterion and the total reward criterion. Thus a reward to a player for a pair of strategies and an initial state is the evaluated worth to this player of a corresponding sequence of expected payoffs over the stages. We use the word 'reward' for an income for some pair of strategies for a whole play, while 'payoff' is always something for just one stage of a play.

A lot of literature in stochastic game theory is on the β-discounted reward criterion. For the β-discounted reward criterion stochastic games turn out to have very fine properties and the results for the β-discounted reward criterion are of fundamental importance for deriving results on the other two criteria. For stochastic games the β-discounted reward criterion is first mentioned as a remark in Shapley [1953].

1.4.2 Definition

Let $\beta\in[0,1)$. The β-discounted reward to player k for initial state s under $(\pi,\sigma)\in\Pi\times\Sigma$ is given by

$$\gamma_\beta^k(s,\pi,\sigma):=(1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}\,E_{s\pi\sigma}[R^k(n)].$$

We also use $\gamma_\beta^k(\pi,\sigma):=(\gamma_\beta^k(1,\pi,\sigma),\ \gamma_\beta^k(2,\pi,\sigma),\dots,\ \gamma_\beta^k(z,\pi,\sigma))$.

In this definition the factor $(1-\beta)$ is used to normalize the β-discounted rewards, because in the sequel we want to relate β-discounted rewards with limiting average rewards.

Observe that, by the finiteness of the state and action spaces, $E_{s\pi\sigma}[R^k(n)]\in[-M,M]$ for all s,π,σ,k,n, where $M:=\max\ \{|r^k(s,i,j)|:k\in\{1,2\},\ i\in A_s, j\in B_s, s\in S\}$. Hence $\gamma_\beta^k(s,\pi,\sigma)\in[-M,M]$ for all s,π,σ,k,β. Discounting with a factor $\beta\in(0,1)$ reflects an interest rate $(1-\beta)/\beta$, because an amount $\beta^{n-1}\alpha$ at stage 1 grows to an amount α at stage n under this interest rate. Discounting with factor β can also be interpreted as having at each stage probability $1-\beta$ that the play stops and probability β that play continues.

A second important evaluation criterion is the limiting average reward criterion introduced by Gillette [1957]. Most of the results in this monograph are on stochastic games with respect to this criterion.

1.4.3 Definition
The limiting average reward to player k for initial state s under $(\pi,\sigma)\in\Pi\times\Sigma$ is given by

$$\gamma^k(s,\pi,\sigma) := \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} E_{s\pi\sigma}\,[R^k(n)].$$

We also use $\gamma^k(\pi,\sigma) := (\gamma^k(1,\pi,\sigma), \gamma^k(2,\pi,\sigma),\ldots, \gamma^k(z,\pi,\sigma))$.

In this definition we use 'lim inf' because 'lim' may fail to exist. The 'lim inf' can be interpreted as a pessimistic view of player k: in the long run his average income will be at least 'lim inf'. We could also have chosen 'lim sup' or some Banach limit in the above definition. Of course one can find strategies in a stochastic game such that the limiting average reward for those strategies is different for 'lim inf' and 'lim sup'. However, for stationary strategies 'lim inf' and 'lim sup' lead to the same average reward.

The third evaluation criterion to be considered in this monograph is that of total rewards, introduced according to the following definition in Thuijsman & Vrieze [1987] and Vrieze & Thuijsman [1987].

1.4.4 Definition
The total reward to player k for initial state s under $(\pi,\sigma)\in\Pi\times\Sigma$ is given by

$$\gamma_T^k(s,\pi,\sigma) := \liminf_{N\to\infty} \frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m} E_{s\pi\sigma}\,[R^k(n)].$$

We also use $\gamma_T^k(\pi,\sigma) := (\gamma_T^k(1,\pi,\sigma), \gamma_T^k(2,\pi,\sigma),\ldots, \gamma_T^k(z,\pi,\sigma))$.

The use of 'lim inf' in this definition will be clear.

Notice that if $\sum_{n=1}^{\infty} E_{s\pi\sigma}[R^k(n)]$ exists, then it necessarily equals $\gamma_T^k(s,\pi,\sigma)$. For a general stochastic game however, the total rewards will often be $-\infty$ or $+\infty$. The total reward criterion is of particular interest in stochastic games for which the limiting average reward is 0 for all, or certain, pairs of stationary strategies. In chapter 5 we discuss this total reward criterion in detail and we examine relations among the three above evaluation criteria.

Observe that the above definitions are all based on the expected payoffs at the stages. This is possible because the triple (s,π,σ) determines for each history h_n, $n\geq 2$, a probability of occurrence $Prob^n_{s\pi\sigma}(h_n)$.

However, by the Kolmogorov extension theorem (cf. Kolmogorov [1933]) this sequence of probability measures $Prob^1_{s\pi\sigma}(.)$, $Prob^2_{s\pi\sigma}(.)$,... can be extended to a probability measure $Prob^{\infty}_{s\pi\sigma}(.)$ on the set of infinite histories, i.e. on the set consisting of sequences $(s_1,i_1,j_1,s_2,i_2,j_2,\ldots.)$. Therefore, instead of the above definitions, we could have used alternative criteria defined by

$$\tilde{\gamma}^k_\beta(s,\pi,\sigma) := E_{s\pi\sigma}\,[(1-\beta)\sum_{n=1}^{\infty} \beta^{n-1}\, R^k(n)];$$

$$\tilde{\gamma}^k(s,\pi,\sigma) := E_{s\pi\sigma}\,[\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} R^k(n)];$$

$$\tilde{\gamma}^k_T(s,\pi,\sigma) := E_{s\pi\sigma}\,[\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N}\sum_{n=1}^{m} R^k(n)].$$

For the β-discounted reward criterion it holds that $\gamma^k_\beta(s,\pi,\sigma) = \tilde{\gamma}^k_\beta(s,\pi,\sigma)$ for all s,π,σ; hence also for the solution concepts we will use (cf. section 1.7 and 1.8) $\gamma^k_\beta(.)$ and $\tilde{\gamma}^k_\beta(.)$ will give the same results. For the limiting average reward criterion $\gamma^k(s,\pi,\sigma)$ not necessarily equals $\tilde{\gamma}^k(s,\pi,\sigma)$; however for stationary strategies $\gamma^k(.)$ and $\tilde{\gamma}^k(.)$ give the same reward. For the total reward criterion $\gamma^k_T(s,x,y)$ is not necessarily equal to $\tilde{\gamma}^k_T(s,x,y)$ for stationary strategies x,y; moreover, as will be pointed out in chapter 5, it is not clear whether or not $\tilde{\gamma}^k_T(.)$ makes any sense at all.

1.5 Rewards for stationary strategies

As is mentioned above, stationary strategies are the least complex strategies. This is reflected in the fact that for stationary strategies there are fine expressions for the rewards. In this section we introduce those expressions and we give some elementary results needed in the sequel of this monograph.

1.5.1 Definition

For a pair of stationary strategies $(x,y) \in X \times Y$ we define:

a) $Car^z(x) := \times_{s=1}^{z} Car(x_s)$ *with* $Car(x_s) := \{i \in A_s : x_s(i) > 0\}$, *the carrier of* x *and* x_s *respectively.* $Car^z(y)$ *and* $Car(y_s)$ *are defined similarly.*

b) $r^k(x,y) := (r^k(1,x_1,y_1), r^k(2,x_2,y_2), \ldots, r^k(z,x_z,y_z))$, *with* $r^k(s,x_s,y_s) := \sum_{i=1}^{m_s}\sum_{j=1}^{n_s} x_s(i) r^k(s,i,j) y_s(j)$ *being the direct expected payoff to player* k *in state* s.

c) $P(x,y)$ *is the transition matrix of size* $z \times z$. *Entry* (s,t) *of* $P(x,y)$ *is* $p(t|s,x_s,y_s) := \sum_{i=1}^{m_s}\sum_{j=1}^{n_s} x_s(i) p(t|s,i,j) y_s(j)$, *which is the probability of a direct transition from* s *to* t *if in state* s *the players use* x_s *and* y_s.
$P(x,y)_s$ *denotes row* s *of* $P(x,y)$.

d) $Q(x,y) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P^n(x,y)$, *where* $P^n(x,y)$ *denotes the* n*-fold matrix product of* $P(x,y)$ *with itself.* $Q(x,y)_s$ *denotes row* s *of* $Q(x,y)$.

Observe that $P(x,y)$, for each $(x,y) \in X \times Y$, determines a stochastic process, or Markov chain, on the state space. It is obvious that the (strategy dependent) ergodic structure of such a chain has its impact on the rewards.

1.5.2 Lemma

Let $(x,y)\in X\times Y$ and let I denote the $z\times z$ identity matrix.

a) *Entry (s,t) of $P^{n-1}(x,y)$ equals the probability that at stage n play is in state t if the players use (x,y) and the initial state is s. Here $P^0(x,y)=I$.*

b) $E_{sxy}\,[R^k(n)] = P^{n-1}(x,y)_s\, r^k(x,y)$.

c) *Entry (s,t) of $Q(x,y)$ is the expected average number of visits to state t if play starts in state s and the players use (x,y).*

d) *$Q(x,y)_s$ equals the unique stationary distribution of the Markov chain which starts in s and is related with (x,y).*

e) $Q(x,y)\,P(x,y)= Q(x,y)$.

f) *$(I-\beta P(x,y))$ and $(I-P(x,y)+Q(x,y))$ are non-singular matrices for all $\beta\in[0,1)$. Hence $(I-\beta P(x,y)+Q(x,y))$ is non-singular for β close to 1.*

g) $Q(x,y)= \lim\limits_{\beta\uparrow 1}(1-\beta)\,(I-\beta P(x,y))^{-1}$.

Proof:

(a) - (d) follow directly from the definitions; (e) - (g) can be found in Kemeny & Snell [1960] or in Blackwell [1962]. ■

Observe that (c) and (d) of the above lemma imply that, if s and t are in the same ergodic set of the Markov chain related with $P(x,y)$, then $Q(x,y)_s = Q(x,y)_t$ and entry (s,t) of $Q(x,y)$ is strictly positive. For $s,t\in S$ with t transient with respect to $P(x,y)$, entry (s,t) of $Q(x,y)$ equals 0.

1.5.3 Lemma

Let $(x,y)\in X\times Y$ and $\beta\in[0,1)$. Then the following statements hold.

a) $\gamma^k_\beta(x,y)= (1-\beta)\sum\limits_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y)\,r^k(x,y)$.

b) $\gamma^k_\beta(x,y)= (1-\beta)\,(I-\beta P(x,y))^{-1}\,r^k(x,y)$.

c) *$\gamma^k_\beta(x,y)$ is the unique $\alpha^k\in\mathbb{R}^z$ satisfying $\alpha^k= (1-\beta)r^k(x,y) + \beta P(x,y)\alpha^k$.*

Proof:

By definition $\gamma^k_\beta(s,x,y)= (1-\beta)\sum\limits_{n=1}^{\infty}\beta^{n-1}E_{sxy}\,[R^k(n)]$ for all $s\in S$. By lemma 1.5.2 (b) this implies $\gamma^k_\beta(x,y)= (1-\beta)\sum\limits_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y)\; r^k(x,y)$. Since $(1-\beta)\sum\limits_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y)= (1-\beta)(I-\beta P(x,y))^{-1}$ for any stochastic matrix $P(x,y)$, we have $\gamma^k_\beta(x,y)= (1-\beta)(I-\beta P(x,y))^{-1}r^k(x,y)$. This implies $(I-\beta P(x,y))\gamma^k_\beta(x,y)= (1-\beta)r^k(x,y)$ and hence $\gamma^k_\beta(x,y)$ is a solution of $\alpha^k= (1-\beta)r^k(x,y) + \beta P(x,y)\alpha^k$. By the non-singularity of $(I-\beta P(x,y))$ this solution is unique. ■

1.5.4 Lemma

Let $(x,y)\in X\times Y$, $\beta\in[0,1)$ and $\alpha\in\mathbb{R}^z$.

a) *If $\alpha\leqslant(1-\beta)r^k(x,y)+\beta\,P(x,y)\alpha$, then $\alpha\leqslant\gamma^k_\beta(x,y)$.*

b) *If $\alpha\lneqq(1-\beta)r^k(x,y)+\beta\,P(x,y)\alpha$, then $\alpha\lneqq\gamma^k_\beta(x,y)$.*

c) *Similar statements hold when reversing the inequality signs.*

Proof:

If $\alpha\lneqq(1-\beta)r^k(x,y)+\beta P(x,y)\alpha$, then $(I-\beta P(x,y))\alpha\lneqq(1-\beta)r^k(x,y)$. Since $(I-\beta P(x,y))^{-1}$ is non-negative and each column has at least one positive entry, it follows that $\alpha\lneqq(1-\beta)\,(I-\beta P(x,y))^{-1}r^k(x,y)=\gamma^k_\beta(x,y)$. ■

1.5.5 Lemma

Let $(x,y)\in X\times Y$. Then the following statements hold.

a) $\gamma^k(x,y)=\lim\limits_{N\to\infty}\dfrac{1}{N}\sum\limits_{n=1}^{N}P^{n-1}(x,y)\,r^k(x,y)$.

b) $\gamma^k(x,y)=Q(x,y)\,r^k(x,y)$.

c) $\gamma^k(x,y)=\alpha^k$ *for any pair* $(\alpha^k,\delta^k)\in\mathbb{R}^z\times\mathbb{R}^z$ *satisfying* $\alpha^k=P(x,y)\alpha^k$ *and* $\alpha^k+\delta^k=r^k(x,y)+P(x,y)\delta^k$.

d) $\gamma^k(x,y)=\lim\limits_{\beta\uparrow 1}\gamma^k_\beta(x,y)$.

Proof:

By definition $\gamma^k(s,x,y)=\liminf\limits_{N\to\infty}\dfrac{1}{N}\sum\limits_{n=1}^{N}E_{sxy}\,[R^k(n)]$ for all $s\in S$. Lemma 1.5.2(b) implies $\gamma^k(x,y)=\liminf\limits_{N\to\infty}\dfrac{1}{N}\sum\limits_{n=1}^{N}P^{n-1}(x,y)r^k(x,y)$. It is well-known (cf. Kemeny & Snell [1960]) that $\lim\limits_{N\to\infty}\dfrac{1}{N}\sum\limits_{n=1}^{N}P^{n-1}(x,y)$ exists, and equals $Q(x,y)$. Hence (a) and (b) hold.

If $(\alpha^k,\delta^k)\in\mathbb{R}^z\times\mathbb{R}^z$ and $\alpha^k=P(x,y)\alpha^k$ as well as $\alpha^k+\delta^k=r^k(x,y)+P(x,y)\delta^k$, then multiplying the second equation with $Q(x,y)$, using lemma 1.5.2 (e), gives $Q(x,y)\alpha^k+Q(x,y)\delta^k=Q(x,y)r^k(x,y)+Q(x,y)\delta^k$. Hence $Q(x,y)\alpha^k=Q(x,y)r^k(x,y)=\gamma^k(x,y)$ by (b). Furthermore $\alpha^k=P(x,y)\alpha^k$ implies $\alpha^k=Q(x,y)\alpha^k$, so we have $\alpha^k=\gamma^k(x,y)$. Finally (d) follows directly from lemma 1.5.2 (g) and from lemma 1.5.3 (b). ■

1.5.6 Lemma.

Let $(x,y)\in X\times Y$ and let $\alpha,\delta\in\mathbb{R}^z$.

a) *If $\alpha\leqslant P(x,y)\alpha$ and $\alpha+\delta\leqslant r^k(x,y)+P(x,y)\delta$, then $\alpha\leqslant\gamma^k(x,y)$.*

b) *A similar statement holds, when reversing the inequality signs.*

Proof:

$\alpha\leqslant P(x,y)\alpha$ implies $\alpha\leqslant Q(x,y)\alpha$ and likewise $\alpha+\delta\leqslant r^k(x,y)+P(x,y)\delta$ implies $Q(x,y)\alpha\leqslant Q(x,y)r^k(x,y)$. Hence $\alpha\leqslant Q(x,y)r^k(x,y)=\gamma^k(x,y)$. ■

The condition $\alpha+\delta\leqslant r^k(x,y)+P(x,y)\delta$ in lemma 1.5.6 can be weakened to:

$\alpha_s + \delta_s \leqslant r^k(s,x_s,y_s) + \sum_{t \in S} p(t|s,x_s,y_s)\delta_t$ for all states s that are recurrent with respect to (the Markov chain related with) $P(x,y)$.

This is possible because, by the fact that for all $s \in S$ and any transient state t entry (s,t) of $Q(x,y)$ equals 0, we could still derive $Q(x,y)\alpha \leqslant Q(x,y)r^k(x,y)$.

1.5.7 Lemma

Let $(x,y) \in X \times Y$ and assume $\gamma^k(x,y)=0$ for $k=1,2$. Then the following statements hold.

a) $\gamma_T^k(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)r^k(x,y)$.

b) $\gamma_T^k(x,y) = (I - P(x,y) + Q(x,y))^{-1} r^k(x,y)$.

c) $\gamma_T^k(x,y) = \alpha^k$ *for any pair* $(\alpha^k,\delta^k) \in \mathbb{R}^z \times \mathbb{R}^z$ *satisfying* $\alpha^k = r^k(x,y) + P(x,y)\alpha^k$ *and* $\alpha^k + \delta^k = P(x,y)\delta^k$.

d) $\gamma_T^k(x,y) = \lim_{\beta \uparrow 1} (1-\beta)^{-1}\, \gamma_\beta^k(x,y)$.

Proof:

By definition $\gamma_T^k(s,x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{sxy}[R^k(n)]$ for all $s \in S$. Hence $\gamma_T^k(x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)r^k(x,y)$. Recall that by lemma 1.5.5 (b) we have $0 = \gamma^k(x,y) = Q(x,y)r^k(x,y)$. Observe that for each $N \in \mathbb{N}$:

$$\begin{aligned}
&(I-P(x,y)+Q(x,y))\Big(\frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m} P^{n-1}(x,y)r^k(x,y)\Big) \\
&\quad = \frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m} [P^{n-1}(x,y)r^k(x,y) - P^n(x,y)r^k(x,y)] \\
&\quad = \frac{1}{N}\sum_{m=1}^{N} [r^k(x,y) - P^m(x,y)r^k(x,y)] \\
&\quad = r^k(x,y) - \frac{1}{N}\sum_{m=1}^{N} P^m(x,y)r^k(x,y).
\end{aligned}$$

Again using $\gamma^k(x,y)=0$ and using lemma 1.5.2 (f) for the non-singularity of $(I - P(x,y) + Q(x,y))$, we derive that $\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)\,r^k(x,y)$ exists and equals $(I - P(x,y) + Q(x,y))^{-1} r^k(x,y)$, which proves (a) and (b).

As for (c), suppose that α^k and δ^k solve $\alpha^k = r^k(x,y) + P(x,y)\alpha^k$ and $\alpha^k + \delta^k = P(x,y)\delta^k$. Multiplying both sides of the last equation with $Q(x,y)$ gives $Q(x,y)\alpha^k = 0$ by lemma 1.5.2 (e). Combining this with the first equation gives that $\alpha^k - P(x,y)\alpha^k + Q(x,y)\alpha^k = r^k(x,y)$. Finally the non-singularity of $(I - P(x,y) + Q(x,y))$ implies $\alpha^k = (I - P(x,y) + Q(x,y))^{-1}\ r^k(x,y)$ and hence we have $\alpha^k = \gamma_T^k(x,y)$ by (b).

In order to show (d) notice that $Q(x,y)\gamma_\beta^k(x,y) = 0$ because $Q(x,y)\gamma_\beta^k(x,y) = Q(x,y)[(1-\beta)\sum_{n=1}^{\infty} \beta^{n-1} P^{n-1}(x,y)r^k(x,y)] = (1-\beta)\sum_{n=1}^{\infty} \beta^{n-1}\ Q(x,y)r^k(x,y) = 0$

by $\gamma^k(x,y)=0$. Hence by lemma 1.5.3 (c) we have $\gamma^k_\beta(x,y)=(1-\beta)r^k(x,y)+\beta P(x,y)\gamma^k_\beta(x,y)-Q(x,y)\gamma^k_\beta(x,y)$. By the non-singularity of $I-\beta P(x,y)+Q(x,y)$ we have $\gamma^k_\beta(x,y)=(1-\beta)\ (I-\beta P(x,y)+\ Q(x,y))^{-1}\ r^k(x,y)$. Since $I-P(x,y)+Q(x,y)$ is also non-singular we get $\lim_{\beta\uparrow 1}\ (1-\beta)^{-1}\gamma^k_\beta(x,y)=(I-P(x,y)+\ Q(x,y))^{-1}r^k(x,y)=\ \gamma^k_T(x,y)$ by (b). ■

1.5.8 Lemma

Let $(x,y)\in X\times Y$, $\alpha,\delta\in\mathbb{R}^z$ and assume $\gamma^k(x,y)=0$.

a) *If $\alpha\leqslant r^k(x,y)+P(x,y)\alpha$ and $\alpha+\delta\leqslant P(x,y)\delta$, then $\alpha\leqslant\gamma^k_T(x,y)$.*

b) *A similar statement holds, when one reverses all inequality signs.*

Proof:

From $\alpha+\delta\leqslant P(x,y)\delta$ we derive $Q(x,y)\alpha+Q(x,y)\delta\leqslant Q(x,y)\delta$ and hence $Q(x,y)\alpha\leqslant 0$. From $\alpha-P(x,y)\alpha\leqslant r^k(x,y)$ we derive $P^{n-1}(x,y)\alpha-\ P^n(x,y)\alpha\leqslant P^{n-1}(x,y)r^k(x,y)$ for all $n\in\mathbb{N}$. This implies that for all $m\in\mathbb{N}$ we have:

$$\alpha-P(x,y)^m\alpha=\sum_{n=1}^{m}(P^{n-1}(x,y)\alpha-P^n(x,y)\alpha)\leqslant\sum_{n=1}^{m}P^{n-1}(x,y)\,r^k(x,y).$$

Hence $\alpha-\frac{1}{N}\sum_{m=1}^{N}P^m(x,y)\alpha\leqslant\frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m}P^{n-1}(x,y)r^k(x,y)$ for all $N\in\mathbb{N}$. Letting N tend to infinity and using $Q(x,y)\alpha\leqslant 0$ we obtain $\alpha\leqslant\gamma^k_T(x,y)$. ■

1.6 Playing against a fixed stationary strategy

In any stochastic game both players want to maximize their individual rewards. Since they cannot make binding agreements they do not know what strategy there opponent is going to use. Nevertheless each player should hope that the strategy he chose is a best reply against the strategy of his opponent, otherwise a better strategy could have been used. Therefore it is of interest to examine what happens if the opponent fixes a strategy. For the objectives in this monograph it is sufficient to consider what happens if the opponent fixes a stationary strategy.

1.6.1 Definition

Let $y\in Y$ and $\beta\in[0,1)$.

A β-discounted best reply for player 1 against y is a strategy $\pi^*\in\Pi$ for which $\gamma^1_\beta(\pi^*,y)\geqslant\gamma^1_\beta(\pi,y)$ for all $\pi\in\Pi$. Limiting average best reply and total best reply are defined analogously.

The next lemma follows from Hordijk et al. [1983] and from Blackwell [1962] and we will often use it for our analysis of stochastic games.

1.6.2 Lemma

Let $y\in Y$ and $\beta\in[0,1)$.

There exists a pure stationary strategy $x^*\in X^P$ such that $\gamma_\beta^1(x^*,y)\geqslant\gamma_\beta^1(\pi,y)$ for all $\pi\in\Pi$. Similarly, there exists a pure stationary limiting average best reply for player 1 against y.

Hordijk et al. [1983] show that player 1 cannot do better in the stochastic game against y than to play optimal in the related Markov decision process, which we call MDP(y).

1.6.3 Definition

A Markov decision process is a stochastic game where one player has only one action available in all states. For a stationary strategy y in a stochastic game Γ, the Markov decision process MDP(y) is the stochastic game Γ^* with $S^*:=S$, $A_s^*:=A_s$, $B_s^*:=\{1\}$, $r^*(s,i,1):=r(s,i,y_s)$, $p^*(t|s,i,1):=p(t|s,i,y_s)$ for all $i\in A_s^*$, $s\in S^*$.

For Markov decision processes Blackwell [1962] has shown the existence of pure stationary optimal strategies for the β-discounted reward criterion as well as for the limiting average reward criterion. Combining this with the result of Hordijk et al. [1983] gives lemma 1.6.2.

1.6.4 Lemma

Let $y\in Y$ and $\beta\in[0,1)$.

Let $x^*\in X$ be a stationary β-discounted best reply against y. Then:

a) $\gamma_\beta^1(x^*,y)=(1-\beta)r^1(x^*,y)+\beta P(x^*,y)\gamma_\beta^1(x^*,y)$

$\qquad\qquad\geqslant(1-\beta)r^1(x,y)+\beta P(x,y)\gamma_\beta^1(x^*,y)$ *for all* $x\in X$.

b) $\gamma_\beta^1(x^*,y)=(1-\beta)r^1(x^P,y)+\beta P(x^P,y)\gamma_\beta^1(x^*,y)$ *for all* $x^P\in X^P$ *with* $Car^z(x^P)\subset Car^z(x^*)$.

c) $\gamma_\beta^1(x^*,y)=\gamma_\beta^1(\tilde{x},y)$ *for all* $\tilde{x}\in X$ *with* $Car^z(\tilde{x})\subset Car^z(x^*)$.

Proof:

The equality sign in (a) follows from lemma 1.5.3 (c).

The inequality sign in (a) follows from the fact that x^* is a β-discounted best reply against y (cf. lemma 1.5.4).

From (a) it follows that for each $s\in S$:

$$\gamma_\beta^1(s,x^*,y)=(1-\beta)r^1(s,x_s^*,y_s)+\beta\sum_{t=1}^{z}p(t|s,x_s^*,y_s)\,\gamma_\beta^1(t,x^*,y)$$

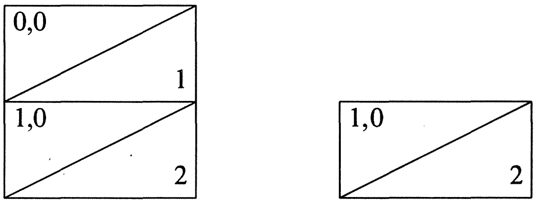
$$=\sum_{i\in A_s}x_s^*(i)[(1-\beta)r^1(s,i,y_s)+\beta\sum_{t=1}^{z}p(t|s,i,y_s)\,\gamma_\beta^1(t,x^*,y)]$$

$$\leqslant\sum_{i\in A_s}x_s^*(i)\,\gamma_\beta^1(s,x^*,y)=\gamma_\beta^1(s,x^*,y).$$

Hence $\gamma_\beta^1(s,x^*,y) = (1-\beta)r_s^1(s,i,y_s) + \beta \sum_{t=1}^{z} p(t|s,i,y_s)\ \gamma_\beta^1(t,x^*,y)$ for all $i \in Car(x_s^*)$, which proves (b).
Now (c) follows from (a), (b) and from lemma 1.5.3 (c). ■

An analogue of lemma 1.6.4 (c) does not hold for the limiting average case. Consider for instance the following example, where player 2 has only one (trivial) strategy: y. We show that a stationary strategy within the carrier of a stationary limiting average best reply against y, does not need to be a limiting average best reply against y itself.

1.6.5 EXAMPLE



State 1 State 2

Stationary strategies for player 1 in the above game, are fully determined by the mixed action which player 1 uses in state 1. So $X=\{(p,1-p):p\in[0,1]\}$. It is easy to see that $(½,½)$ is a stationary limiting average best reply against y, giving player 1 limiting average reward (1,1). Although the pure stationary strategy (1,0) is clearly contained in the carrier of $(½,½)$, it is not a limiting average best reply against y, for $\gamma^1((1,0),y) = (0,1)$.

1.7 ZERO-SUM STOCHASTIC GAMES

Shapley [1953] started the theory of stochastic games. In his model the payoffs to player 1 are the losses of player 2, i.e. $r^2(s,i,j) = -r^1(s,i,j)$ for all s,i and j. A stochastic game with this property is called a zero-sum stochastic game. Since in a zero-sum stochastic game the players have strictly opposite interests, player 1, who wants to maximize his reward, can expect that player 2 wants to minimize that same reward. We assume that player 1 is interested in maximizing his guaranteed expected reward, i.e. player 1 would prefer to use a strategy π^* such that $\inf_\sigma \gamma_\beta^1(\pi^*,\sigma) \geqslant \inf_\sigma \gamma_\beta^1(\pi,\sigma)$ for all $\pi\in\Pi$, in the β-discounted case (or similarly for the other criteria). So player 1 is interested in $\sup_\pi \inf_\sigma \gamma_\beta^1(\pi,\sigma)$. Likewise we assume that player 2 wants to minimize the reward to player 1 and is interested in $\inf_\sigma \sup_\pi \gamma_\beta^1(\pi,\sigma)$, the coordinatewise minimal level for which player 2 can guarantee that the reward to player 1 will not be greater (up to some $\epsilon>0$). It is easy to see that 'sup inf' $\leqslant$ 'inf sup' because: $\inf_\sigma \gamma_\beta^1(\pi^*,\sigma) \leqslant$

$\gamma_\beta^1(\pi^*,\sigma^*)$ for all π^*,σ^*, implies that $\sup\limits_\pi \inf\limits_\sigma \gamma_\beta^1(\pi,\sigma) \leqslant \sup\limits_\pi \gamma_\beta^1(\pi,\sigma^*)$ for all σ^* and hence $\sup\limits_\pi \inf\limits_\sigma \gamma_\beta^1(\pi,\sigma) \leqslant \inf\limits_\sigma \sup\limits_\pi \gamma_\beta^1(\pi,\sigma)$. If 'sup inf' = 'inf sup' then we call this number the value of the stochastic game.

1.7.1 Definition

a) *If for a zero-sum stochastic game there exists, for $\beta\in[0,1)$, a $v_\beta^1 \in \mathbb{R}^z$ such that $\sup\limits_\pi \inf\limits_\sigma \gamma_\beta^1(\pi,\sigma) = v_\beta^1 = \inf\limits_\sigma \sup\limits_\pi \gamma_\beta^1(\pi,\sigma)$, then v_β^1 is called the β-discounted value of the stochastic game.*

b) *If the value is v_β^1, then player 1 has, for each $\epsilon>0$, a strategy π_ϵ such that $\gamma_\beta^1(\pi_\epsilon,\sigma) \geqslant v_\beta^1 - \epsilon 1_z$ for all $\sigma\in\Sigma$. Such a strategy π_ϵ is called a β-discounted ϵ-optimal strategy for player 1. A β-discounted optimal strategy for player 1 is a strategy π^* for which $\gamma_\beta^1(\pi^*,\sigma) \geqslant v_\beta^1$ for all σ.*
A similar definition holds for β-discounted (ϵ-)optimal strategies of player 2.

c) *For the limiting average reward case and the total reward case (ϵ-)optimal strategies and value, v^1 resp. v_T^1, are defined analogously.*

In zero-sum stochastic games the players at each stage face a kind of matrix game. Therefore it is not surprising that the following theorem by Von Neumann [1928], presented here without proof, is very valuable for stochastic games.

1.7.2 Theorem

For any real matrix $A=[a_{ij}]_{i=1,j=1}^{m\ \ n}$ there exist $x^*\in\Delta^m$ and $y^*\in\Delta^n$ such that for all $x\in\Delta^m$ and $y\in\Delta^n$: $x^*Ay \geqslant x^*Ay^* \geqslant xAy^*$.
The mixed actions x^* and y^* are called optimal mixed actions, for player 1 and player 2 respectively, in the matrix game A. The number x^*Ay^* is called the value of A, denoted by val(A) or by val$[a_{ij}]$. This value of A is uniquely determined.

The importance of this theorem for stochastic games already occurs in the seminal paper on stochastic games by Shapley [1953], who examined stopping stochastic games. For stochastic games with the β-discounted reward criterion Shapley's results imply the following:

1.7.3 Theorem

For any $\beta\in[0,1)$ and any zero-sum stochastic game:

a) *The β-discounted value v_β^1 exists and both players have stationary β-discounted optimal strategies.*

b) *v_β^1 is the unique solution $\alpha\in\mathbb{R}^z$ of the 'Shapley-equation':*

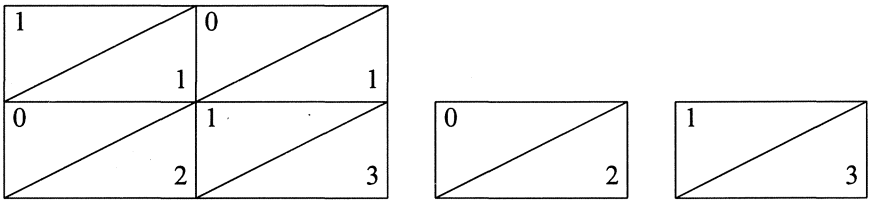
$$\alpha_s = val[(1-\beta)r^1(s,i,j) + \beta \sum_{t=1}^{z} p(t|s,i,j)\alpha_t]_{i=1,j=1}^{m_s\ \ n_s} = : val(A_\beta^{1s}(\alpha)),\quad s\in S.$$

c) *A stationary strategy x^* (y^*) for player 1 (2) is β-discounted optimal if x_s^* (y_s^*) is an optimal mixed action for player 1 (2) in the matrix game $A_\beta^{1s}(v_\beta^1)$ for each $s\in S$.*

It is well known that (c) of the above theorem is also valid when we replace 'if' by 'if and only if'. This follows for example from the results of Vrieze & Tijs [1980], who showed that for each player the set of stationary β-discounted optimal strategies is the Cartesian product of the sets of optimal mixed actions in the matrix games $A_\beta^{1s}(v_\beta^1)$, $s \in S$.

Gillette [1957] introduced the limiting average reward criterion for stochastic games. With respect to this criterion stochastic games turned out to have a more difficult nature than for the β-discounted reward criterion. Gillette [1957] gave the following example for which it was not clear for several years, whether or not it had a limiting average value.

1.7.4 Example *(the big match)*



State 1 State 2 State 3

In this zero-sum stochastic game, only player 1's payoffs are given (cf. 1.2.2). Of course state 1 is the interesting initial state in this stochastic game and for both players strategies are determined by the mixed actions used in state 1. The remarks below illustrate the beauty of this big match, which was solved by Blackwell & Ferguson [1968].

a) With respect to Markov strategies (cf. 1.3.2) one finds that:
$\sup_{f\in F} \inf_{g\in G} \gamma^1(1,f,g) = 0$, whereas $\inf_{g\in G} \sup_{f\in F} \gamma^1(1,f,g) = ½$.
Hence the limiting average value would not exist if the players were restricted to Markov strategies.

b) Allowing all strategies one finds that:

$$\sup_{\pi\in\Pi} \inf_{\sigma\in\Sigma} \gamma^1(1,\pi,\sigma) = \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi} \gamma^1(1,\pi,\sigma) = ½.$$

So the limiting average value $v^1 = (½,0,1)$.

c) For player 2 the stationary strategy determined by using the mixed action $(½,½)$ in state 1, is limiting average optimal. Player 1 has no limiting average optimal strategy and only history dependent limiting average ϵ-optimal strategies.

If we consider the above stochastic game with respect to the β-discounted reward criterion then we find by solving the Shapley-equation (cf. 1.7.3), that $v_\beta^1 = (½,0,1)$ $(=v^1)$ for all $\beta \in [0,1)$, while the unique stationary β-discounted optimal strategies are $x^\beta = (1/(2-\beta),(1-\beta)/(2-\beta))$ and $y^\beta = (½,½)$ for player 1 and player 2 respectively.

It should be observed that, with respect to $P(x^{\beta},y^{\beta})$ state 1 is transient, whereas state 1 is recurrent with respect to $P(x^1,y^1)$, with $(x^1,y^1) := \lim_{\beta\uparrow 1}(x^{\beta},y^{\beta})$.

Bewley & Kohlberg [1976] made a thorough study of asymptotic properties of v_{β}^{1} and of stationary β-discounted optimal strategies x^{β},y^{β} as β tends to 1. Using Tarski's principle on real closed fields (cf. Tarski [1951]) they derived the following remarkable theorem, which we give without proof.

1.7.5 Theorem

For any zero-sum stochastic game situation there exist $N\in\mathbb{N}$, $\{\alpha_n\in\mathbb{R}^{z}: n\in\mathbb{N}_0\}$, $\{x_n\in \mathop{\times}\limits_{s=1}^{z}\mathbb{R}^{m_s}: n\in\mathbb{N}_0\}$, $\{y_n\in \mathop{\times}\limits_{s=1}^{z}\mathbb{R}^{n_s}: n\in\mathbb{N}_0\}$ *such that for all* β *close to 1:*

a) $v_{\beta}^{1} = \sum\limits_{n=0}^{\infty}\alpha_n(1-\beta)^{n/N}$ *is the* β*-discounted value;*

b) $x^{\beta} = \sum\limits_{n=0}^{\infty}x_n(1-\beta)^{n/N}$ *is a stationary* β*-discounted optimal strategy for player 1;*

c) $y^{\beta} = \sum\limits_{n=0}^{\infty}y_n(1-\beta)^{n/N}$ *is a stationary* β*-discounted optimal strategy for player 2.*

Two remarks should directly be made about this theorem. First of all it follows that $\lim_{\beta\uparrow 1}v_{\beta}^{1}$ exists and equals α_0. Second, it follows that $\lim_{\beta\uparrow 1}x^{\beta}$ exists and equals x_0, which is therefore a stationary strategy.

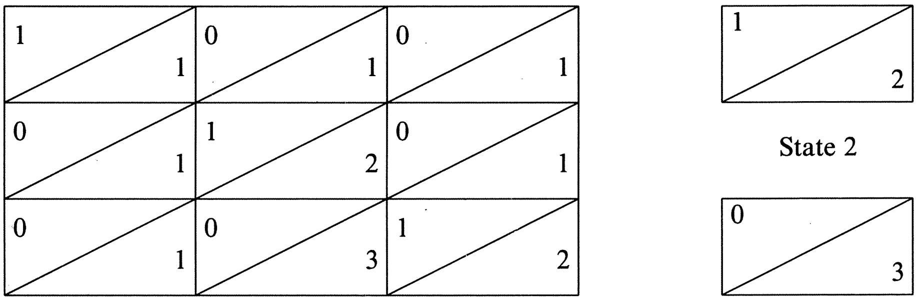
As an illustration of the above theorem observe that for the big match, example 1.7.4, we have $v_{\beta}^{1} = v^{1}$ for all $\beta\in[0,1)$ and for player 1 the unique stationary optimal strategies are given by $x^{\beta} = (1/(2-\beta),\ (1-\beta)/(2-\beta))$. Hence we have:

$$x^{\beta} = \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}-1\\1\end{bmatrix}(1-\beta) + \begin{bmatrix}1\\-1\end{bmatrix}(1-\beta)^2 + \begin{bmatrix}-1\\1\end{bmatrix}(1-\beta)^3 + \begin{bmatrix}1\\-1\end{bmatrix}(1-\beta)^4 + ...$$

For player 2 we have $y^{\beta} = \begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\end{bmatrix}$ for all β.

To give another illustration we examine one more example, where payoffs are again given only for player 1.

1.7.6 EXAMPLE



State 2

State 1

State 3

For this example, solving the Shapley-equation (cf. 1.7.3) leads to the β-discounted value v_β^1 and to the unique stationary optimal strategies x^β and y^β given, for initial state 1, by:

$$v_\beta^1 = \frac{1-(1-\beta)^{1/3}}{\beta} = 1-(1-\beta)^{1/3}+(1-\beta)-(1-\beta)^{4/3}+(1-\beta)^2-(1-\beta)^{7/3}+.....$$

$$x^\beta = \begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}-1\\1\\0\end{bmatrix}(1-\beta)^{1/3} + \begin{bmatrix}0\\-1\\1\end{bmatrix}(1-\beta)^{2/3} + \begin{bmatrix}1\\0\\-1\end{bmatrix}(1-\beta) + \begin{bmatrix}-1\\1\\0\end{bmatrix}(1-\beta)^{4/3} + \begin{bmatrix}0\\-1\\1\end{bmatrix}(1-\beta)^{5/3} +.....$$

$$y^\beta = \begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}-1\\0\\1\end{bmatrix}(1-\beta)^{1/3} + \begin{bmatrix}0\\1\\-1\end{bmatrix}(1-\beta)^{2/3} + \begin{bmatrix}1\\-1\\0\end{bmatrix}(1-\beta) + \begin{bmatrix}-1\\0\\1\end{bmatrix}(1-\beta)^{4/3} + \begin{bmatrix}0\\1\\-1\end{bmatrix}(1-\beta)^{5/3} +.....$$

The work of Blackwell & Ferguson [1968] on the big match was generalized by Kohlberg [1974] for zero-sum repeated games with absorbing states (cf. section 4.4). These papers, together with the above result of Bewley & Kohlberg [1976] for the asymptotic properties of the β-discounted solutions, were important for the derivation of the following result by Mertens & Neyman [1981].

1.7.7 THEOREM

For any zero-sum stochastic game the limiting average value v^1 exists, and it is related to the β-discounted values by $v^1 = \lim_{\beta \uparrow 1} v_\beta^1$.

Observe that this theorem implies that $v^1 = \alpha_0$, the leading term of the power series, for β close to 1, in theorem 1.7.5. In chapter 5 on stochastic games with respect to the total reward criterion we will see that if the total value v_T^1 exists in $\mathbb{R}^z$ and if both players have stationary total optimal strategies, then $\alpha_0, \alpha_1, ..., \alpha_{N-1}$ are all equal to 0 and $v_T^1 = \alpha_N = \lim_{\beta \uparrow 1} (1-\beta)^{-1} v_\beta^1$. However, for stochastic games with respect to the total reward criterion little is known and in general the total value v_T^1 will not exist. Even on the condition that the limiting average value is 0 (for all starting states), the total value is not necessarily finite. For a further discussion on stochastic games with respect to the total reward criterion we refer to chapter 5, where it is also shown that, like for the limiting average case, history dependent strategies may be indispensable to achieve total ϵ-optimality.

1.8 General-sum stochastic games

Fink [1964] started the study of general-sum stochastic games, as they are defined in definition 1.2.1. Since in a general-sum stochastic game the players not necessarily have strictly opposite interests, the solution concepts 'value' and 'optimality' become meaningless. For non-zero-sum stochastic games an alternative solution concept is required. Nash [1951] showed the applicability of the concept 'equilibrium' for bimatrix games by proving the following theorem.

1.8.1 Theorem
Let A^1 and A^2 be real $m \times n$-matrices.
Then there exist $x^* \in \Delta^m$ and $y^* \in \Delta^n$ such that:

$$x^* A^1 y^* \geqslant x A^1 y^* \text{ for all } x \in \Delta^m \text{ and}$$

$$x^* A^2 y^* \geqslant x^* A^2 y \text{ for all } y \in \Delta^n.$$

The pair (x^*, y^*) is called a (Nash-)equilibrium for the bimatrix game (A^1, A^2).

So this theorem guarantees the existence of equilibria for bimatrix games. Fink [1964] extended the definition of equilibrium to stochastic games. Here we give a more general definition.

1.8.2 Definition
Let $\epsilon > 0$. A pair of strategies $(\pi_\epsilon, \sigma_\epsilon) \in \Pi \times \Sigma$ is called a β-discounted ϵ-equilibrium for initial state s if:

$$\gamma_\beta^1(s, \pi_\epsilon, \sigma_\epsilon) \geqslant \gamma_\beta^1(s, \pi, \sigma_\epsilon) - \epsilon \text{ for all } \pi \in \Pi \text{ and}$$

$$\gamma_\beta^2(s, \pi_\epsilon, \sigma_\epsilon) \geqslant \gamma_\beta^2(s, \pi_\epsilon, \sigma) - \epsilon \text{ for all } \sigma \in \Sigma.$$

If $(\pi_\epsilon, \sigma_\epsilon)$ is a β-discounted ϵ-equilibrium for all $s \in S$, then $(\pi_\epsilon, \sigma_\epsilon)$ is called a β-discounted ϵ-equilibrium. If ϵ can be taken 0 in the inequalities, then we speak of an equilibrium. Similar definitions hold for limiting average ϵ-equilibrium and for total ϵ-equilibrium.

The idea behind the concept 'ϵ-equilibrium' is the following. Once the players have, somehow, come to use a pair of strategies $(\pi_\epsilon, \sigma_\epsilon)$, which is an ϵ-equilibrium, then neither player 1 nor player 2 can gain more that ϵ by unilateraly deviating from his strategy. So, for small ϵ, each player will remain playing his equilibrium strategy. Hence, an ϵ-equilibrium is 'self-enforcing'.

Shapley [1953] connected the β-discounted value and optimality in zero-sum stochastic games with the value and optimal mixed actions of related matrix games. Fink [1964] derived a similar result for general-sum stochastic games.

1.8.3 THEOREM

Let $\beta \in [0,1)$.

a) *For any stochastic game there exists a stationary β-discounted equilibrium.*

b) *A pair of stationary strategies (x^*,y^*) is a β-discounted equilibrium if for each $s \in S$, the pair of mixed actions (x_s^*,y_s^*) is an equilibrium in the bimatrix game $(A_\beta^{1s}(\gamma_\beta^1(x^*,y^*)), A_\beta^{2s}(\gamma_\beta^2(x^*,y^*)))$, where (cf. 1.7.3)*

$$A_\beta^{ks}(\gamma_\beta^k(x^*,y^*)) = [(1-\beta)r^k(s,i,j) + \beta \sum_{t=1}^{z} p(t|s,i,j)\gamma_\beta^k(t,x^*,y^*)]_{i=1,j=1}^{m_s \;\; n_s}$$

Other proofs for this theorem have been given by Takahashi [1964], Rogers [1969] and Sobel [1971].

In the previous section we have seen that for zero-sum stochastic games the existence of the limiting average value and of limiting average ϵ-optimal strategies remained a problem until 1980. The existence of limiting average ϵ-equilibria is even a tougher problem, for it is still open. One should observe that the existence of ϵ-equilibria $(\pi_\epsilon,\sigma_\epsilon)$, for all $\epsilon>0$, in a zero-sum stochastic game implies that the value exists and that the strategies π_ϵ and σ_ϵ are ϵ-optimal for the respective players.

If in a general-sum stochastic game an ϵ-equilibrium $(\pi_\epsilon,\sigma_\epsilon)$ exists, then player 1 would have a reward which is at least the value, possibly up to ϵ, of the zero-sum stochastic game obtained by assuming that the payoffs to player 1 have to be paid by player 2. This is due to the fact that, given a δ-optimal strategy π_δ for player 1 in that zero-sum game, we can derive $\gamma^1(\pi_\epsilon,\sigma_\epsilon) \geqslant \gamma^1(\pi_\delta,\sigma_\epsilon)-\epsilon \geqslant v^1-\delta-\epsilon$. Letting δ tend to 0 gives the result:

1.8.4 REMARK

If $(\pi_\epsilon,\sigma_\epsilon)$ is a limiting average ϵ-equilibrium, then $\gamma^k(\pi_\epsilon,\sigma_\epsilon) \geqslant v^k-\epsilon$, where v^k is the limiting average value of the zero-sum stochastic game obtained by assuming that the payoffs to player k have to be paid by player $(3-k)$: 'the k-zero-sum stochastic game'. Of course a similar statement holds for the β-discounted reward case and for the total reward case.

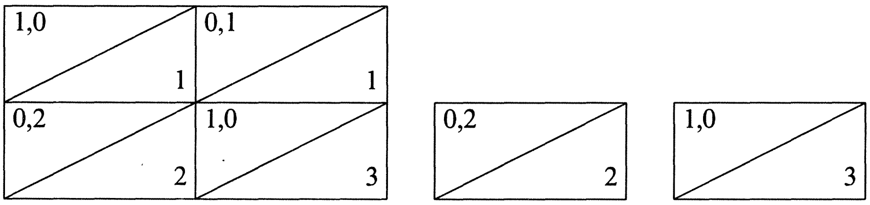
Another interesting fact concerning these k-zero-sum stochastic games is that player $(3-k)$ has for each $\epsilon>0$ a strategy to keep player k's reward below $v^k+\epsilon$. This allows player $(3-k)$ to punish player k if such is required, and therefore player $(3-k)$ can threaten to punish player k if player k deviates from a certain strategy. Punishment arguments to establish equilibria are quite common in the theory of repeated games (cf. Aumann [1981]). In the next chapter this will also become meaningful for general stochastic games.

1.8.5 DEFINITION

For a general-sum stochastic game (v^1,v^2) is called the limiting average threat-point. A retaliation strategy π_ϵ^r for player 1 is a strategy for which $\gamma^2(\pi_\epsilon^r,\sigma) \leqslant v^2+\epsilon$ for all σ. Similarly we have a retaliation strategy σ_ϵ^r for player 2.

For zero-sum stochastic games we have $v^k = \lim_{\beta\uparrow 1} v^k_\beta$ (cf. 1.7.7). For general-sum stochastic games we can take (by theorem 1.8.3), for each $\beta\in[0,1)$, a stationary β-discounted equilibrium (x^β,y^β) and, without loss of generality, we can assume that $\lim_{\beta\uparrow 1}(x^\beta,y^\beta)$ and $\lim_{\beta\uparrow 1}\gamma^k_\beta(x^\beta,y^\beta)$ exist and are equal to (x^1,y^1) and V^k respectively (cf. section 2.2). Now one could hope that V^k is related with a limiting average ϵ-equilibrium and one could even think that (x^1,y^1) may be a limiting average ϵ-equilibrium. Unfortunately this will not be true in general, as is illustrated by the next example which has been examined by Sorin [1986].

1.8.6 Example



State 1 State 2 State 3

The 1-zero-sum stochastic game of 1.8.6 is exactly 1.7.4, the big match. The 2-zero-sum stochastic game is also a kind of big match.
The unique stationary β-discounted equilibria (x^β,y^β) for this example are given by (the mixed actions in state 1):

$$(x^\beta,y^\beta) = ((2/(3-\beta),(1-\beta)/(3-\beta)),(½,½)).$$

So we have $\gamma^1_\beta(1,x^\beta,y^\beta) = ½ = v^1(1)$ and $\gamma^2_\beta(1,x^\beta,y^\beta) = ⅔ = v^2(1)$ for all $\beta\in[0,1)$. Then we find that for $(x^1,y^1) = \lim_{\beta\uparrow 1}(x^\beta,y^\beta) = ((1,0),(½,½))$ the limiting average rewards are:

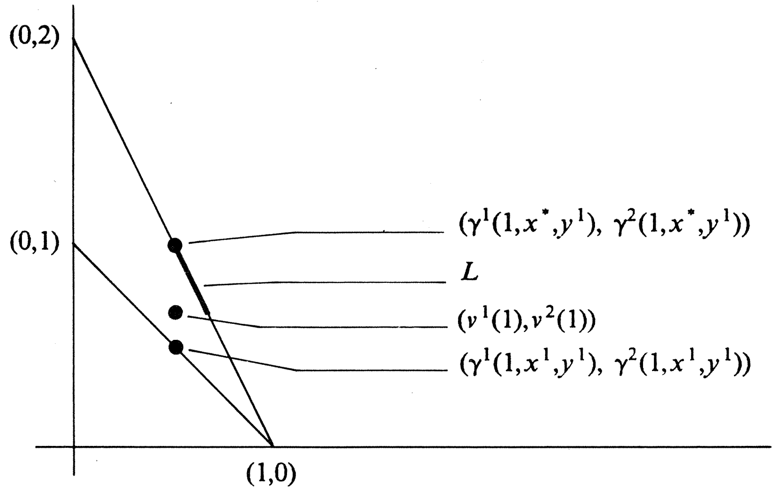
$$\gamma^1(1,x^1,y^1) = ½ = v^1(1) \text{ and } \gamma^2(1,x^1,y^1) = ½ < ⅔ = v^2(1).$$

It is obvious that (x^1,y^1) is no limiting average ϵ-equilibrium, because against x^1 player 2 could improve by playing (0,1). Sorin [1986] shows that for this example the limiting average rewards corresponding with limiting average ϵ-equilibria for state 1 are all in the convex hull L of $\{(½,1),(⅔,⅔)\}$. So $L = \{(a,2-2a): a\in[½,⅔]\}$.
It is important to observe that for $x^* = (0,1)$ we find:

$$\gamma^1(1,x^\beta,y^1) = \gamma^1(1,x^*,y^1) = ½ = v^1(1) \text{ and}$$

$$\gamma^2(1,x^\beta,y^1) = \gamma^2(1,x^*,y^1) = 1 > ⅔ = v^2(1).$$

The importance of this observation will become clear in chapter 3 (cf. example 3.2.4). In graph we have the following situation:


(0,2)
(0,1)
(1,0)
$(\gamma^1(1,x^*,y^1),\ \gamma^2(1,x^*,y^1))$
L
$(v^1(1),v^2(1))$
$(\gamma^1(1,x^1,y^1),\ \gamma^2(1,x^1,y^1))$

In this graph the area in the triangle ((1,0), (0,1), (0,2)) is the set of feasible rewards for this stochastic game, i.e. those rewards that can occur for some pair of strategies. The feasible rewards which are larger than the threat-point are called individually rational. The set L consists of the feasible, individually rational Pareto optimal rewards, where the Pareto optimal rewards are those rewards that cannot be improved simultaneously for both players.

Although the above example suggests a gap between general-sum β-discounted solutions and limiting average solutions, we exhibit in the next chapters that limiting average ϵ-equilibria may, under some condition, be derived from sequences of stationary β-discounted equilibria.

Chapter 2

Particular initial states in stochastic games

2.1 INTRODUCTION

In chapter 1 we have seen that for any zero-sum stochastic game the β-discounted value exists and that both players have stationary β-discounted optimal strategies; in the general-sum case there exist stationary β-discounted equilibria. We have also seen in chapter 1 that for any zero-sum stochastic game the limiting average value exists, which however does not guarantee the existence of optimal strategies or stationary ϵ-optimal strategies; in the general-sum case the existence of limiting average ϵ-equilibria is still an open problem. As discussed in chapter 1, for the limiting average reward criterion more complex strategies are required to play ϵ-optimal or to form an ϵ-equilibrium. There may be however, starting states for which a solution exists in terms of stationary strategies. Consider for instance the initial states 2 and 3 of the big match (example 1.7.4). This idea, of examining particular starting states, first occurred in Tijs & Vrieze [1986]. They showed that for each player there are, in any zero-sum stochastic game, 'easy initial states', i.e. starting states for which this player has a stationary limiting average optimal strategy. Their proof however is rather technical. In this chapter we give an alternative and straightforward proof for their theorem in section 2.4. There we also examine other initial states where both players can achieve ϵ-optimality, with respect to the limiting average reward criterion, by using stationary strategies.

In section 2.3 we examine special initial states in general-sum stochastic games, and we show that for certain 'strong initial states' limiting average ϵ-equilibria exist which consist of 'almost stationary strategies'. The latter are stationary strategies which are amplified with some threat to prevent profitable deviations of the opponent. So if both players stick to their ϵ-equilibrium strategies, then with probability close to 1 they will use stationary strategies throughout the whole game.

In section 2.2 we derive some basic results, which are of fundamental importance for the chapters 3 and 4 as well. Our techniques are based on properties of sequences of stationary strategies $\{(x^\beta,y^\beta):\beta\in[0,1)\}$, converging for β tending to 1.

The results of this chapter have been derived from Thuijsman & Vrieze [1990-a, 1991].

2.2 Limit properties for sequences of strategy pairs

By the expression

'let $\{(x^\beta,y^\beta)\in X\times Y:\beta\in[0,1)\}$ be a (converging) sequence with $\lim\limits_{\beta\uparrow 1}(x^\beta,y^\beta)=(x^1,y^1)$'

we mean:

'for some sequence $\{\beta_n\in[0,1):n\in\mathbb{N}\}$, with $\lim\limits_{n\to\infty}\beta_n=1$, it holds that $\lim\limits_{n\to\infty}(x^{\beta_n},y^{\beta_n})=(x^1,y^1)\in X\times Y$.'

Observe that $X\times Y$ is compact, which implies that any sequence in $X\times Y$ has a converging subsequence.

2.2.1 Definition

Let $\{(x^\beta,y^\beta)\in X\times Y:\beta\in[0,1)\}$ be a sequence with $\lim\limits_{\beta\uparrow 1}(x^\beta,y^\beta)=(x^1,y^1)\in X\times Y$. Without loss of generality we may assume that $Car^z(x^\beta)$ and $Car^z(y^\beta)$ are independent of $\beta<1$. By compactness arguments we can also assume that the following limits exist and we can define:

a) $V^k:=\lim\limits_{\beta\uparrow 1}\gamma^k_\beta(x^\beta,y^\beta)\in[-M,M]^z$, *for* $k=1,2$.

b) $Z^1:=\lim\limits_{\beta\uparrow 1}Z^\beta$, *where* $Z^\beta:=(1-\beta)\,(I-\beta P(x^\beta,y^\beta))^{-1}$.

c) *T is the set of states which are transient with respect to (x^1,y^1).*
$S^1,S^2,\dots,S^H$ are the ergodic sets with respect to (x^1,y^1).
For each $h\in\{1,2,\dots,H\}$ let $\tilde{q}^h\in\Delta^{|S^h|}$ be the unique stationary distribution of $P(x^1,y^1)^h$, the restriction of $P(x^1,y^1)$ to S^h; let $q^h\in\Delta^z$ be the related stationary distribution (for initial states in S^h) of $P(x^1,y^1)$ on S, i.e. $q^h_s=\tilde{q}^h_s$ for $s\in S^h$ and $q^h_s=0$ for $s\notin S^h$.

Observe that for all $\beta\in[0,1)$ all row sums of Z^β are equal to 1 and all entries of Z^β are non-negative. Hence $Z^\beta,\beta\in[0,1)$, and Z^1 are stochastic matrices.
In this section $\{(x^\beta,y^\beta):\beta\in[0,1)\}$ is a sequence as in definition 2.2.1.

2.2.2 Remark

Let $Q(x^1,y^1)^h$ denote the restriction of $Q(x^1,y^1)$ to S^h and let $s\in S^h$. Then the s-th row of $Q(x^1,y^1)^h$ equals $\tilde{q}^h$ and is strictly positive. Furthermore the s-th row of $Q(x^1,y^1)$ equals q^h.

By ordening the states, the matrices $P(x^1,y^1)$ and $Q(x^1,y^1)$ will have the following shape:

$$P(x^1,y^1)=\begin{bmatrix} P(x^1,y^1)^1 & 0 & \dots & 0 & 0 \\ 0 & P(x^1,y^1)^2 & & & \\ & & & & \\ & & & 0 & \\ 0 & 0 & & P(x^1,y^1)^H & 0 \\ P(x^1,y^1)^{T1} & P(x^1,x^1)^{T2} & \dots & P(x^1,y^1)^{TH} & P(x^1,y^1)^T \end{bmatrix}$$

$$Q(x^1,y^1)= \begin{bmatrix} Q(x^1,y^1)^1 & 0 & \dots & 0 & 0 \\ 0 & Q(x^1,y^1)^2 & & & \\ & & & 0 & \\ 0 & 0 & & Q(x^1,y^1)^H & 0 \\ Q(x^1,y^1)^{T1} & Q(x^1,x^1)^{T2} & \dots & Q(x^1,y^1)^{TH} & 0 \end{bmatrix}$$

Here $P(x^1,y^1)^{Th}$ and $Q(x^1,y^1)^{Th}$ are restrictions of $P(x^1,y^1)$, resp. $Q(x^1,y^1)$, to rows in T and columns in S^h; similarly $P(x^1,y^1)^T$ is the restriction of $P(x^1,y^1)$ to rows and columns in T.
It is well-known (cf. Kemeny & Snell [1960]) that for $h \in \{1,2,...,H\}$:

$$Q(x^1,y^1)^{Th}= (I^T - P(x^1,y^1)^T)^{-1} P(x^1,y^1)^{Th}\, Q(x^1,y^1)^h.$$

Observe that $(I^T - P(x^1,y^1)^T)^{-1}\ P(x^1,y^1)^{Th}$ has $|T|$ rows and $|S^h|$ columns. For $s \in T$ and $t \in S^h$ entry (s,t) of $(I^T - P(x^1,y^1)^T)^{-1}\ P(x^1,y^1)^{Th}$ gives the probability that a stochastic process which starts in s will ever enter the ergodic set S^h through state t.

2.2.3 Lemma
a) $\gamma^k(s,x^1,y^1)= \gamma^k(t,x^1,y^1)=: \gamma^{kh}(x^1,y^1)$ *for* $s,t \in S^h$, $h \in \{1,2,...,H\}$ *and for* $k=1,2$.
b) $V^k= P(x^1,y^1)V^k$ *for* $k=1,2$.
c) $V^k_s= V^k_t =: V^{kh}$ *for* $s,t \in S^h$, $h \in \{1,2,...,H\}$ *and for* $k=1,2$.

Proof:
By lemma 1.5.5 it holds that $\gamma^k(s,x^1,y^1)= Q(x^1,y^1)_s\, r^k(x^1,y^1)$. Now (a) follows from remark 2.2.2.
(b) follows from $\gamma^k_\beta(x^\beta,y^\beta)= (1-\beta)r^k(x^\beta,y^\beta) + \beta P(x^\beta,y^\beta)\gamma^k_\beta(x^\beta,y^\beta)$ (cf. lemma 1.5.3 (c)). Taking limits and applying definition 2.2.1 gives the result. Now (b) implies that $V^k= Q(x^1,y^1)V^k$, which by remark 2.2.2 gives (c). ■

Although its proof is rather simple, the next lemma turns out to be of great importance in the sequel.

2.2.4 Lemma
$Z^1\ P(x^1,y^1) = Z^1$.

Proof:
By definition 2.2.1 we have $Z^\beta(I - \beta P(x^\beta,y^\beta)) = (1-\beta)I$ for all $\beta \in [0,1)$.
Taking limits for β going to 1 completes the proof. ■

The strength of this lemma becomes clear in the lemmas below.

2.2.5 Lemma

Let $s \in S$ and let Z_s^1 be the s-th row of Z^1.

a) *There exists $\mu_s = (\mu_s^1, \mu_s^2, \ldots, \mu_s^H) \in \Delta^H$ such that $Z_s^1 = \sum_{h=1}^{H} \mu_s^h q^h$.*

b) *If for the Markov chain related with $P(x^\beta, y^\beta)$ the probability of ever reaching S^h is 0 when starting in $s \notin S^h$, then $\mu_s^h = 0$.*

Proof:

a) By lemma 2.2.4 it holds that Z_s^1 is a stationary distribution of the Markov chain related with $P(x^1, y^1)$. Since the set of all stationary distributions for $P(x^1, y^1)$ is the convex hull of $\{q^1, q^2, \ldots, q^H\}$, there is $\mu_s \in \Delta^H$ as desired.

b) If under (x^β, y^β) the set S^h cannot be reached when starting in s, then it follows that entry (s,t) of $P^n(x^\beta, y^\beta)$ is 0 for all $n \in \mathbb{N}$ and all $t \in S^h$.
Hence for all $t \in S^h$ we have that entry (s,t) of
$Z^\beta = (1-\beta) \sum_{n=1}^{\infty} \beta^{n-1} P^{n-1}(x^\beta, y^\beta)$ is 0. But then entry (s,t) of Z^1 is also 0 for all $t \in S^h$, which implies that $\mu_s^h = 0$. ∎

The next lemma says: 'the limit of discounted rewards equals a convex combination of the limiting average rewards for the limit strategies.'

2.2.6 Lemma

Let $s \in S$ and let $\mu_s \in \Delta^H$ be as in lemma 2.2.5.
Then $V_s^k = \sum_{h=1}^{H} \mu_s^h \gamma^{kh}(x^1, y^1)$ for $k = 1,2$.

Proof:

Let $\langle , \rangle$ denote the inner product.
By definition 2.2.1, remark 2.2.2 and lemmas 2.2.3 and 2.2.5 we have:

$$V_s^k = \lim_{\beta \uparrow 1} \gamma_\beta^k(s, x^\beta, y^\beta) = \lim_{\beta \uparrow 1} \langle Z_s^\beta, r^k(x^\beta, y^\beta) \rangle = \langle Z_s^1, r^k(x^1, y^1) \rangle$$
$$= \sum_{h=1}^{H} \mu_s^h \langle q^h, r^k(x^1, y^1) \rangle = \sum_{h=1}^{H} \mu_s^h \gamma^{kh}(x^1, y^1). \qquad ∎$$

2.2.7 Corollary

There exist $h^1, h^2 \in \{1, 2, \ldots, H\}$ such that:
$\gamma^{1h^1}(x^1, y^1) \geq \max_{s \in S} V_s^1$ *and* $\gamma^{2h^2}(x^1, y^1) \geq \max_{s \in S} V_s^2$.

2.3 Strong initial states in the general-sum case

We show that, in any stochastic game, there are some starting states for which there exists an almost stationary limiting average ϵ-equilibrium.

2.3.1 Definition

An almost stationary limiting average ϵ-equilibrium for initial state s is a pair of strategies (π^*,σ^*) such that (π^*,σ^*) is a limiting average ϵ-equilibrium for initial state s and (π^*,σ^*) consists of stationary strategies (x^*,y^*) and retaliation strategies (π^r,σ^r). Player 1 uses x^* unless he detects a deviation of player 2 from σ^*, in which case he immediately turns to using π^r. For player 2 strategy σ^* is similar.

A strong initial state is an initial state for which there exists an almost stationary limiting average ϵ-equilibrium, for all $\epsilon>0$.

Since a strategy may consist of mixed actions for all stages, the phrase 'unless he detects a deviation of player 2 from σ^*' should be interpreted as: 'unless player 1 knows that the probability of player 2 playing σ^* is close to 0.'

2.3.2 Remark

In this section let $\{(x^\beta,y^\beta):\beta\in[0,1)\}$ be a sequence of stationary β-discounted equilibria with $\lim_{\beta\uparrow 1}(x^\beta,y^\beta)=(x^1,y^1)$ and which furthermore suits definition 2.2.1.

In addition to the results developed in the previous section for such a sequence $\{(x^\beta,y^\beta):\beta\in[0,1)\}$, the fact that we are dealing with stationary β-discounted equilibria allows us to conclude the following.

2.3.3 Lemma

a) *For each $\tilde{x}\in X$ with $Car^z(\tilde{x})\subset Car^z(x^1_\beta)$ and for all $x\in X$:*

$$V^1=P(x^1,y^1)V^1=P(\tilde{x},y^1)V^1\geqslant P(x,y^1)V^1.$$

b) *For each $\tilde{y}\in Y$ with $Car^z(\tilde{y})\subset Car^z(y^1_\beta)$ and for all $y\in Y$:*

$$V^2=P(x^1,y^1)V^2=P(x^1,\tilde{y})V^2\geqslant P(x^1,y)V^2.$$

c) $V^k\geqslant v^k$ *for* $k=1,2$.

Proof:

By lemma 1.6.4 we have for all $\beta\in[0,1)$:

$$\begin{aligned}\gamma^1_\beta(x^\beta,y^\beta)&=(1-\beta)r^1(x^\beta,y^\beta)+\beta P(x^\beta,y^\beta)\gamma^1_\beta(x^\beta,y^\beta)\\&=(1-\beta)r^1(\tilde{x},y^\beta)+\beta P(\tilde{x},y^\beta)\gamma^1_\beta(x^\beta,y^\beta)\\&\geqslant(1-\beta)r^1(x,y^\beta)+\beta P(x,y^\beta)\gamma^1_\beta(x^\beta,y^\beta).\end{aligned}$$

Taking limits for β to 1 proves (a). The proof of (b) is similar. By remark 1.8.4 we have $\gamma^k_\beta(x^\beta,y^\beta)\geqslant v^k_\beta$ for all $\beta\in[0,1)$, hence (c) follows by taking limits (cf. definition 2.2.1 and theorem 1.7.7). ■

Lemma 2.3.3 implies that, if $\gamma^k(x^1,y^1)\geqslant V^k$ for both players and if each of them can check whether or not his opponent is actually using y^1 or x^1, then one could construct a limiting average ϵ-equilibrium. This is possible by using a threat to retaliate, giving less than $v^k+\epsilon$ in case of a detected deviation. This is worked out more precisely in the next lemma.

2.3.4 Lemma

If for $h\in\{1,2,...,H\}$ it holds that $\gamma^{kh}(x^1,y^1)\geqslant V^{kh}$ for $k=1$ as well as for $k=2$, then a limiting average ϵ-equilibrium for initial states in S^h can be made by supplementing x^1 and y^1 with suitable retaliation threats.

Proof:

Let $\epsilon>0$ and let $h\in\{1,2,...,H\}$ be such that $\gamma^{kh}(x^1,y^1)\geqslant V^{kh}$ for $k=1$ and for $k=2$. We divide the proof in three parts: in part 1 we show that each player can detect deviations of his opponent with probability close to 1; in part 2 we show that each player can retaliate if he detects a deviation; in part 3 we show that (x^1,y^1) supplemented with retaliation threats is a limiting average ϵ-equilibrium on S^h.

Part 1: *Player 1 can detect deviations of player 2 with probability close to 1.*

Suppose player 1 uses x^1.

It is clear that if player 2 at some stage chooses an action outside $Car^z(y^1)$, then player 1 immediately knows that player 2 is not using y^1.

As long as player 2 chooses actions within $Car^z(y^1)$, the play will remain within S^h and player 1 can count the number of times that player 2 chooses action j in state $s\in S^h$ for all j and s. Hence at each stage $n\in\mathbb{N}$ player 1 knows the action frequency $y_s^{(n)}(j)$ of action j in state s. If player 2 really uses y^1, then $y_s^{(n)}(j)$ should converge to $y_s^1(j)$ as n goes to infinity.

Let $Y_s^{(n)}(j)$ be the random variable which denotes the action frequency of action j in state s. So $y_s^{(n)}(j)$ is a realization of $Y_s^{(n)}(j)$. It is well-known (cf. Billingsley [1979]) that for every $\alpha,\delta>0$ there exists $N_{\alpha\delta}\in\mathbb{N}$ such that:

$$\text{Prob}_{x^1,y^1}\,\{\|Y_s^{(n)}-y_s^1\|>\alpha \text{ for any } s\in S^h \text{ and any } n\geqslant N_{\alpha\delta}\}<\delta.$$

If for all $n\geqslant N_{\alpha\delta}$ and all $s\in S^h$ it holds that $\|y_s^{(n)}-y_s^1\|\leqslant\alpha$ then, by continuity arguments, the limiting average reward to player k is at most $\gamma^{kh}(x^1,y^1)+\alpha K$ and at least $\gamma^{kh}(x^1,y^1)-\alpha K$ for some constant $K\in\mathbb{N}$. So if player 1 does not detect a deviation of player 2, then the limiting average reward to player 2 is at most $\gamma^{2h}(x^1,y^1)+\alpha K$.

Part 2: *Player 1 can retaliate if he detects a deviation of player 2.*

Suppose at some stage n player 1 detects a deviation of player 2, i.e. player 2 chooses j at that stage in state s and either $j\notin Car(y_s^1)$ or $n\geqslant N_{\alpha\delta}$ and $\|y_s^{(n)}-y_s^1\|>\alpha$.

If, from stage $n+1$ on, player 1 now uses a retaliation strategy $\pi_{\epsilon/2}^r$ (cf. definition 1.8.5) then the limiting average reward to player 2 will be at most $\sum_{t=1}^{z} p(t|s,x_s^1,j)\,(v_t^2+\epsilon/2)$.

By lemma 2.3.3 we have:

$$\sum_{t=1}^{z} p(t|s,x_s^1,j)(v_t^2+\epsilon/2) \leqslant \sum_{t=1}^{z} p(t|s,x_s^1,j)(V_t^2+\epsilon/2) \leqslant V_s^2+\epsilon/2.$$

Since $\gamma^{2h}(x^1,y^1) \geqslant V^{2h} = V_s^2$ (cf. lemma 2.2.3), we conclude that the limiting average reward to player 2 will be at most $\gamma^{2h}(x^1,y^1)+\epsilon/2$.

PART 3: (x^1,y^1) *can be supplemented with retaliation threats to become a limiting average* ϵ*-equilibrium for all starting states in* S^h.
Let $\alpha \in (0,\epsilon/4K)$ and take $\delta>0$ such that

$$(1-\delta)^2(\gamma^{kh}(x^1,y^1)-\alpha K) - (1-(1-\delta)^2)M \geqslant \gamma^{kh}(x^1,y^1)-\epsilon/2 \text{ for } k=1,2.$$

Now player 1 can try to keep player 2 from deviating from y^1 by using the almost stationary strategy π_ϵ^* defined by:

a) use x^1 unless:
   i) player 2 chooses an action outside $Car^z(y^1)$, or
   ii) for some $n \geqslant N_{\alpha\delta}$ and some $s \in S^h$: $\|y_s^{(n)} - y_s^1\| > \alpha$.

b) if (i) or (ii) occurs, start retaliation by using $\pi_{\epsilon/2}^r$ from that stage on.

For player 2 the almost stationary strategy σ_ϵ^* is defined analogously.

From these definitions it follows that:

$$\gamma^{kh}(\pi_\epsilon^*,\sigma_\epsilon^*) \geqslant (1-\delta)^2(\gamma^{kh}(x^1,y^1)-\alpha K)-(1-(1-\delta)^2)M \geqslant \gamma^{kh}(x^1,y^1)-\epsilon/2.$$

From parts 1 and 2 we conclude that for all $\sigma \in \Sigma$:

$$\gamma^{2h}(\pi_\epsilon^*,\sigma) \leqslant \gamma^{2h}(x^1,y^1) + \epsilon/2.$$

Similarly one can derive that for all $\pi \in \Pi$:

$$\gamma^{1h}(\pi,\sigma_\epsilon^*) \leqslant \gamma^{1h}(x^1,y^1) + \epsilon/2.$$

Hence $(\pi_\epsilon^*,\sigma_\epsilon^*)$ is an almost stationary limiting average ϵ-equilibrium for all starting states in S^h. ■

2.3.5 THEOREM
For any general-sum stochastic game there exist strong initial states.

It is clear that this theorem follows directly from lemma 2.3.6 below, which tells that the condition of lemma 2.3.4 is automatically fulfilled for some h.

2.3.6 LEMMA
There exists $h^* \in \{1,2,...,H\}$ *such that:*

$$\gamma^{1h^*}(x^1,y^1) \geqslant \max_h V^{1h} \geqslant V^{1h^*} \text{ and } \gamma^{2h^*}(x^1,y^1) \geqslant V^{2h^*}.$$

PROOF:
For non-empty $E \subset \{1,2,...,H\}$ let $S^E := \bigcup_{h \in E} S^h$.
Let $E_1 := \{h \in \{1,2,...,H\}: \gamma^{1h}(x^1,y^1) \geqslant \max_s V_s^1\}$. Then by corollary 2.2.7 we have that $E_1 \neq \varnothing$.

Now let T_1 be the set of transient states for which plays started there lead to S^{E_1} under (x^1,y^1) with probability 1. So $T_1 := \{s \in T$: entry (s,t) of $Q(x^1,y^1)$ is 0 for all $t \notin S^{E_1}\}$.
Now take a stationary strategy x^* for player 1 such that for $\beta \in [0,1)$:
$x_s^* = x_s^1$ for all $s \notin S^{E_1} \cup T_1$ and $Car(x_s^*) = Car(x_s^\beta)$ for all $s \in S^{E_1} \cup T_1$.
Remember that $Car(x_s^\beta)$ is independent of $\beta \in [0,1)$ (cf. definition 2.2.1).
Observe that each ergodic set with respect to (x^*,y^1) is either one of the sets S^h with $h \notin E_1$ or it is a subset of $T_1 \cup S^{E_1}$.
In order to prove that for some $h^* \in E_1$ it holds that $\gamma^{2h^*}(x^1,y^1) \geqslant V^{2h^*}$ we make use of the following observation: There are E_2 and T_2 with $\varnothing \neq E_2 \subset E_1$ and $T_2 \subset T_1$ such that $T_2 \cup S^{E_2}$ is an ergodic set with respect to (x^*,y^1).

> To show the correctness of this statement, suppose that it is not true. Then the ergodic sets with respect to (x^*,y^1) are necessarily the sets S^h with $h \notin E_1$, and $\varnothing \neq \{1,2,...,H\} \setminus E_1$. But then, using lemma 1.6.4, lemma 2.2.3 and analogues of lemmas 2.2.5 and 2.2.6 we conclude that for each $s \in S$ there is $\mu_s \in \Delta^H$ such that:
>
> $$V_s^1 = \lim_{\beta \uparrow 1} \gamma_\beta^1(s,x^\beta,y^\beta) = \lim_{\beta \uparrow 1} \gamma_\beta^1(s,x^*,y^\beta) = \sum_{h \notin E_1} \mu_s^h \gamma^{1h}(x^*,y^1)$$
> $$= \sum_{h \notin E_1} \mu_s^h \gamma^{1h}(x^1,y^1) < \max_{t \in S} V_t^1.$$
>
> Since this is clearly a contradiction, our statement is correct.

So there are $\varnothing \neq E_2 \subset E_1$ and $T_2 \subset T_1$ as desired. By definition of x^* and by the fact that $Car^z(x^\beta)$ is independent of $\beta \in [0,1)$ it follows that $T_2 \cup S^{E_2}$ is an ergodic set with respect to (x^β,y^1) for all $\beta \in [0,1)$. Once more applying lemma 1.6.4, lemma 2.2.3 and analogues of lemmas 2.2.5 and 2.2.6 we obtain that for each $s \in S^{E_2}$ there is $\mu_s \in \Delta^H$ such that:

$$V_s^2 = \lim_{\beta \uparrow 1} \gamma_\beta^2(s,x^\beta,y^\beta) = \lim_{\beta \uparrow 1} \gamma_\beta^2(s,x^\beta,y^1) = \sum_{h \in E_2} \mu_s^h \gamma^{2h}(x^1,y^1) \leqslant \max_{h \in E_2} \gamma^{2h}(x^1,y^1).$$

Hence there is $h^* \in E_2$ and $s \in S^{h^*}$ such that $\gamma^{2h^*}(x^1,y^1) \geqslant V_s^2 = V^{2h^*}$. ■

2.3.7 REMARK
Observe that for the almost stationary limiting average ϵ-equilibria, as constructed for some S^h in the proof of lemma 2.3.4, the property holds that a play started in S^h will remain in S^h with probability close to 1.

2.3.8 Example

$1,-1$ 1	$0,0$ 1
$0,0$ 2	$1,-1$ 3

State 1

$0,0$ 2

State 2

$1,-1$ 3

State 3

Notice that this is again the big match (cf. example 1.7.4). For this stochastic game the unique stationary β-discounted equilibria are given by $(x^{\beta},y^{\beta})=((1/(2-\beta),(1-\beta)/(2-\beta)),(\frac{1}{2},\frac{1}{2}))$ (for starting state 1). It is clear that $(x^1,y^1)=((1,0),(\frac{1}{2},\frac{1}{2}))$ is not a limiting average ϵ-equilibrium for starting state 1. However, $\gamma^1(1,x^1,y^1)=\frac{1}{2}=V_1^1=v_1^1$ and $\gamma^2(1,x^1,y^1)=-\frac{1}{2}=V_1^2=v_1^2$. So by lemma 2.3.4 the pair of strategies (x^1,y^1) can be supplemented with retaliation threats to become a limiting average ϵ-equilibrium (player 1 has to check whether or not player 2's action-frequencies for state 1 are close to $(\frac{1}{2},\frac{1}{2})$ in the long run; player 1 cannot gain by deviating against y^1).

So state 1 is a strong initial state for this stochastic game. It is also clear that state 2 and state 3 are strong initial states. Hence we have an almost stationary limiting average ϵ-equilibrium for this stochastic game. Moreover for some starting states (2 and 3) we even have a stationary limiting average equilibrium. The next example shows that in general there need not be initial states for which there is a stationary limiting average equilibrium.

2.3.9 Example

$2,-2$ 1	$-2,2$ 1
$-1,1$ 2	$1,-1$ 3

State 1

$-1,1$ 2	$-2,2$ 2
$-2,2$ 2	$0,0$ 1

State 2

$1,-1$ 3	$2,-2$ 3
$2,-2$ 3	$0,0$ 1

State 3

For this stochastic game $v^1=(0,-1,1)$ and $v^2=(0,1,-1)$. Since this stochastic game is in fact a zero-sum stochastic game, any limiting average reward for an equilibrium for initial state s should be equal to the threat-point (v_s^1,v_s^2).
For this stochastic game there is no stationary limiting average equilibrium for any of the initial states. We discuss the initial states one by one.

a) Suppose that (x,y) is a stationary limiting average equilibrium for initial state 1. If $x_1=(1,0)$ then $y_1=(0,1)$ since y is a best reply against x for player 2. But then $\gamma^1(1,x,y)=-2<v_1^1$, contradiction.
If $x_1\neq(1,0)$ then $y_1=(1,0)$ and $\gamma^2(1,x,y)\geqslant 1$ since y is a best reply against x. Hence $\gamma^1(1,x,y)\leqslant -1<v_1^1$, contradiction.

b) Suppose that (x,y) is a stationary limiting average equilibrium for initial state 2. If $y_2=(1,0)$ then $x_2=(1,0)$ and $\gamma^2(2,x,y)=1$ since x is a best reply against y for player 1. But then y is no best reply against x for player 2, contradiction.
If $y_2\neq(1,0)$ then player 1 can achieve a limiting average reward at least 1 by playing (0,1) in state 2, playing (1,0) in state 3 and by playing in state 1 the action (1,0) if $y_1(1)\geqslant 3/4$ or (0,1) if $y_1(1)<3/4$. Hence $\gamma^2(2,x,y)\leqslant -1<v_2^2$, contradiction.

c) Suppose that (x,y) is a stationary limiting average equilibrium for initial state 3. If $x_3=(1,0)$, then $y_3=(1,0)$ and $\gamma^1(3,x,y)=1$ since y is a best reply against x for player 2. But then x is no best reply for player 1 against y, for by playing (0,1) against y in starting state 3 player 1 could get limiting average reward 2, contradiction.
If $x_3\neq(1,0)$ then player 2 can get limiting average reward at least 0 by playing (1,0) in state 2, (0,1) in state 3 and $(1/2,1/2)$ in state 1. Hence $\gamma^1(3,x,y)\leqslant 0<v_3^1$, contradiction.

However, although for none of the initial states there is a stationary limiting average equilibrium, an almost stationary limiting average ϵ-equilibrium exists (for all initial states).
This follows from lemma 2.3.4, because it can be verified that for each $\beta\in[0,1)$ the pair (x^β,y^β) defined below is a stationary β-discounted equilibrium and $(x^1,y^1)=\lim_{\beta\uparrow 1}(x^\beta,y^\beta)$ satisfies the condition of lemma 2.3.4.
For $\beta\in[0,1)$ define (x^β,y^β) by:

$$x_1^\beta=(\frac{3-\beta-\sqrt{9-8\beta}}{\beta},\frac{-3+2\beta+\sqrt{9-8\beta}}{\beta}),$$

$$x_2^\beta=(\frac{3-\sqrt{9-8\beta}}{2\beta},\frac{-3+2\beta+\sqrt{9-8\beta}}{2\beta}),$$

$$x_3^\beta=(\frac{3-\sqrt{9-8\beta}}{2\beta},\frac{-3+2\beta+\sqrt{9-8\beta}}{2\beta}),$$

$$y_1^\beta=(1/2,1/2),$$

$$y_2^\beta=(\frac{3-\sqrt{9-8\beta}}{2\beta},\frac{-3+2\beta+\sqrt{9-8\beta}}{2\beta}),$$

$$y_3^\beta = \left(\frac{3-\sqrt{9-8\beta}}{2\beta}, \frac{-3+2\beta+\sqrt{9-8\beta}}{2\beta}\right).$$

It follows that $x^1 = ((1,0),(1,0),(1,0))$ and $y^1 = ((½,½),(1,0),(1,0))$. Furthermore we derive that:

$$\gamma_\beta^1(x^\beta,y^\beta) = \left(0, \frac{3-4\beta-\sqrt{9-8\beta}}{2\beta}, \frac{-3+4\beta+\sqrt{9-8\beta}}{2\beta}\right),$$

$$\gamma_\beta^2(x^\beta,y^\beta) = \left(0, \frac{-3+4\beta+\sqrt{9-8\beta}}{2\beta}, \frac{3-4\beta-\sqrt{9-8\beta}}{2\beta}\right),$$

$$\gamma^1(x^1,y^1) = (0,-1,1) = \lim_{\beta\uparrow 1} \gamma_\beta^1(x^\beta,y^\beta) = V^1,$$

$$\gamma^2(x^1,y^1) = (0,1,-1) = \lim_{\beta\uparrow 1} \gamma_\beta^2(x^\beta,y^\beta) = V^2.$$

Since the states 1, 2 and 3 are each ergodic sets with respect to (x^1,y^1), we can apply lemma 2.3.4 for each state.

2.4 (ϵ-)Easy initial states in the zero-sum case

In this section we show that for each player there are easy states in any zero-sum stochastic game. For all states with minimal limiting average value, player 1 has a stationary limiting average ϵ-optimal strategy. For the states with maximal limiting average value we give a sufficient condition for player 1 to have a stationary limiting average ϵ-optimal strategy. Similar results hold for player 2.

2.4.1 Definition

A state s is called an (ϵ-)easy initial state for player k if player k has a stationary limiting average (ϵ-)optimal strategy for the game starting in s.

It is clear that the set of ϵ-easy initial states for player k contains the set of easy initial states for this player. However, there need not be states which are easy for both players, whereas all states may be ϵ-easy for both players. Hence the set of ϵ-easy states for a player is generally larger than the set of easy states for this same player. The following theorem is due to Tijs & Vrieze [1986].

2.4.2 Theorem

For any zero-sum stochastic game each player has at least one easy initial state.

The proof presented by Tijs & Vrieze [1986] for this theorem is based on the result of Bewley & Kohlberg [1976] who showed that there are solutions for the β-discounted zero-sum case which can be written as power series in fractional powers of $(1-\beta)$ (cf. theorem 1.7.5). We give a new proof for theorem 2.4.2 based on the results derived in section 2.2.

2.4.3 Remark

In this section $\{x^\beta : \beta \in [0,1)\}$ *is a sequence of stationary* β*-discounted optimal strategies for player 1, which converges to* x^1. *Likewise we have* $\{y^\beta : \beta \in [0,1)\}$ *for player 2 converging to* y^1.

Proof of theorem 2.4.2:

We prove the existence of easy initial states for player 1. Let y^* be a stationary limiting average best reply against x^1 (cf. remark 2.4.3 and lemma 1.6.2).

Let $Z^\beta, V^k, S^1, \ldots, S^H$ etc. be as in definition 2.2.1 for the sequence $\{(x^\beta, y^*) : \beta \in [0,1)\}$.

By corollary 2.2.7 there is $h^1 \in \{1,2,\ldots,H\}$ with $\gamma^{1h^1}(x^1,y^*) \geq \max_{s \in S} V_s^1$. For all $s \in S$ we also have, using theorem 1.7.7, that:

$$V_s^1 = \lim_{\beta \uparrow 1} \gamma_\beta^1(s, x^\beta, y^*) \geq \lim_{\beta \uparrow 1} v_\beta^1(s) = v_s^1.$$

Hence for all $s \in S^{h^1}$ and all $\sigma \in \Sigma$ we have:

$$\gamma^1(s, x^1, \sigma) \geq \gamma^1(s, x^1, y^*) \geq \max_{t \in S} V_t^1 \geq \max_{t \in S} v_t^1.$$

We conclude that x^1 is limiting average optimal for all $s \in S^{h^1}$. ■

Observe that by the above proof it follows that any limit of stationary β-discounted strategies (β tending to 1) is limiting average optimal for some starting states. Moreover, for player 1 we found that among those easy initial states there are states for which the limiting average value is maximal. Similarly we conclude that for player 2 the strategy y^1 is limiting average optimal for some initial states for which the limiting average value is minimal.

The converse is not true: if for a state the limiting average value is maximal (or minimal), then this does not imply that player 1 (2) has a stationary limiting average optimal strategy. This is demonstrated by the next example.

2.4.4 Example

0 1	1 1
1 1	0 2

State 1

0 2	−1 2
−1 2	0 1

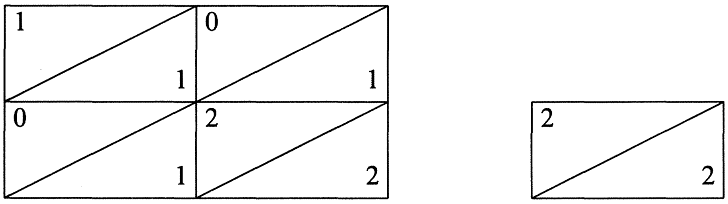
State 2

Payoffs are again those to player 1 to be paid by player 2. It is easy to verify that for this stochastic game the limiting average value $v^1 = (0,0)$. For player 1 (2) a stationary limiting average ϵ-optimal strategy is $x^\epsilon = ((1,0),(1-\epsilon,\epsilon))$ (resp. $y^\epsilon = ((1-\epsilon,\epsilon),(1,0))$).

It is clear that for initial state 2 (1) player 1 (resp. 2) has no stationary limiting average optimal strategy. So, although the value is maximal as well as minimal for both starting states, each player has a stationary limiting average optimal strategy for just one of them.

As we have just remarked, there are easy states for player 1 among the states with maximal limiting average value. The next example illustrates that there need not be easy initial states for player 1 among the states with minimal limiting average value.

2.4.5 EXAMPLE



State 1 State 2

Payoffs are again those to player 1 to be paid by player 2.
For this stochastic game the limiting average value v^1 equals (1,2). For player 1 a stationary limiting average ϵ-optimal strategy is given by $x^\epsilon = (1-\epsilon,\epsilon)$, the mixed action to be used in state 1. For player 2 a stationary limiting average optimal strategy is $y^* = (1,0)$.
It is easy to see that for state 1, the state with minimal limiting average value, player 1 has no stationary limiting average optimal strategy.
Nevertheless, states for which the limiting average value is maximal or minimal are special, as is illustrated in the following two theorems.

2.4.6 THEOREM
Let $v^{\min} := \min\limits_{s\in S} v_s^1$ *and let* $S^{\min} := \{s \in S : v_s^1 = v^{\min}\}$.
All states in $S^{\min}$ *are* ϵ*-easy for player 1.*

PROOF:
Let $y \in Y$ be arbitrary.
Using that $\gamma_\beta^1(x^\beta,y) \geqslant v_\beta^1$ (cf. remark 2.4.3) and using theorem 1.7.3 we have:

$$(1-\beta)r^1(x^\beta,y) + \beta P(x^\beta,y)v_\beta^1 \geqslant v_\beta^1 \text{ for all } \beta\in[0,1).$$

Multiplying this inequality with $Q(x^\beta,y)$ gives that:

$$Q(x^\beta,y)r^1(x^\beta,y) \geqslant Q(x^\beta,y)v_\beta^1 \text{ for all } \beta\in[0,1).$$

Hence for β such that $\|v_\beta^1 - v^1\| < \epsilon$ (cf. theorem 1.7.7) we have, by lemma 1.5.5:

$$\gamma^1(x^\beta,y) = Q(x^\beta,y)r^1(x^\beta,y) \geqslant Q(x^\beta,y)v_\beta^1$$
$$\geqslant Q(x^\beta,y)v^1 - \epsilon 1_z \geqslant (v^{\min} - \epsilon)1_z.$$

Thus x^β, with β such that $\|v_\beta^1 - v^1\| < \epsilon$, is a stationary limiting average ϵ-optimal strategy for player 1 for all initial states in $S^{\min}$. ■

Recall that by theorem 1.7.5 (Bewley & Kohlberg [1976]) we may assume that there are $N \in \mathbb{N}$, $x_0 \in \underset{s=1}{\overset{z}{\times}} \Delta^{m_s}$, $x_1, x_2, \ldots \in \underset{s=1}{\overset{z}{\times}} \mathbb{R}^{m_s}$, such that:

$$x^\beta = \sum_{n=0}^{\infty} x_n (1-\beta)^{n/N} \text{ for all } \beta \text{ close to } 1.$$

Since $x^\beta \in X$, it holds that: $x^1 = x_0$; $\sum_{i=1}^{m_s} x_{ns}(i) = 0$ for all $n \geqslant 1$ and for all $s \in S$; if $x_{0s}(i) = x_{1s}(i) = \ldots = x_{n-1s}(i) = 0$ for $s \in S$ and $n \geqslant 1$ then $x_{ns}(i) \geqslant 0$; $\sum_{n=0}^{l} x_n (1-\beta)^{n/N} \in X$ for each $l \in \mathbb{N}$. We use these facts to examine limiting average ϵ-optimality for player 1 in states with maximal limiting average value.

2.4.7 Definition

Let $v^{\max} := \max_{s \in S} v_s^1$ and let $S^{\max} := \{s \in S : v_s^1 = v^{\max}\}$.
Let $S^* := \{s \in S^{\max} : x^1$ is limiting average optimal for intial state $s\}$.
Define $\bar{x}^\beta \in X$ by $\bar{x}_s^\beta := x_s^1$ for $s \in S^*$ and $\bar{x}_s^\beta := \sum_{n=0}^{N-1} x_n (1-\beta)^{n/N}$ for $s \in S \setminus S^*$.
Let $\bar{y} \in Y$ be a stationary limiting average best reply against $\bar{x}^\beta$, for all β sufficiently close to 1.
Define $S^{**} := S^* \cup \{E \subset S^{\max} \setminus S^* : E$ ergodic with respect to $(\bar{x}^\beta, \bar{y})\}$.
Define $A := S^{\max} \setminus S^{**}$.

2.4.8 Theorem

a) $S^* \neq \varnothing$.

b) *$\bar{x}^\beta$ is limiting average ϵ-optimal for initial states in S^{**} for β close to 1.*

c) *If $\lim_{\beta \uparrow 1} (1-\beta)(I^A - \beta P(\bar{x}^\beta, \bar{y})^A)^{-1} = 0$, then $\bar{x}^\beta$ is limiting average ϵ-optimal for all initial states in $S^{\max}$ for β close to 1. (Here the superscript A denotes the restriction to rows and columns corresponding with states in A)*

Proof:

a) In the above we noticed that, by the proof of theorem 2.4.2, there are states in $S^{\max}$ which are easy for player 1, i.e. for which x^1 is limiting average optimal. Hence $S^* \neq \varnothing$.

b) We already noticed that:

$$(1-\beta) r^1(x^\beta, y) + \beta P(x^\beta, y) v_\beta^1 \geqslant v_\beta^1 \text{ for all } y \in Y.$$

Letting β tend to 1, this gives us that $P(x^1, y) v^1 \geqslant v^1$ for all $y \in Y$.
Hence, if player 1 uses x^1 in $S^{\max}$, then play will remain in $S^{\max}$ with probability 1, no matter what strategy is used by player 2.

From the fact that for initial states in S^* the strategy x^1 is limiting average optimal, one can conclude that, if player 1 uses x^1 for initial states in S^*, then play will remain within S^* with probability 1. Again this does not depend on the strategy used by player 2.
We define (cf. theorem 1.7.5):

$$x^\beta(N) := \sum_{n=0}^{N-1} x_n(1-\beta)^{n/N} \text{ and } \underline{x}^\beta(N) := \sum_{n=N+1}^{\infty} x_n(1-\beta)^{n/N}.$$

Notice that $\lim_{\beta\uparrow 1}(1-\beta)^{-1}\,\underline{x}^\beta(N)=0$ and that $x_s^\beta(N)=\overline{x}_s^\beta$ for $s\in S\setminus S^*$.

We show:

For any set $E\subseteq S^{\max}\setminus S^*$, such that E is ergodic with respect to $(\overline{x}^\beta,\overline{y})$, we have that $\gamma^1(s,\overline{x}^\beta,\overline{y})\geqslant v^{\max}-\epsilon$ for any initial state $s\in E$ and β close to 1.

For β close to 1 we have $v_\beta^1\leqslant(1-\beta)r^1(x^\beta,\overline{y})+\beta P(x^\beta,\overline{y})$ and hence:

$$\begin{aligned} v_\beta^1 \leqslant &(1-\beta)r^1(x^\beta(N),\overline{y})+\beta P(x^\beta(N),\overline{y})v_\beta^1\\ &+(1-\beta)^2r^1(x_N,\overline{y})+\beta(1-\beta)P(x_N,\overline{y})v_\beta^1\\ &+(1-\beta)r^1(\underline{x}^\beta(N),\overline{y})+\beta P(\underline{x}^\beta(N),\overline{y})v_\beta^1. \end{aligned}$$

Let Q_E^β denote the restriction of $Q(x^\beta(N),\overline{y})$ to rows corresponding with states in E. Hence Q_E^β has size $|E|\times z$. Multiplying the above inequality by Q_E^β yields (cf. lemma 1.5.2 (e)):

$$\begin{aligned} Q_E^\beta v_\beta^1 \leqslant &(1-\beta)Q_E^\beta r^1(x^\beta(N),\overline{y})+\beta Q_E^\beta v_\beta^1\\ &+(1-\beta)^2Q_E^\beta r^1(x_N,\overline{y})+\beta(1-\beta)Q_E^\beta P(x_N,\overline{y})v_\beta^1\\ &+(1-\beta)Q_E^\beta r^1(\underline{x}^\beta(N),\overline{y})+\beta Q_E^\beta P(\underline{x}^\beta(N),\overline{y})v_\beta^1. \end{aligned}$$

Hence:

$$\begin{aligned} Q_E^\beta v_\beta^1 \leqslant &Q_E^\beta r^1(x^\beta(N),\overline{y})\\ &+(1-\beta)Q_E^\beta r^1(x_N,\overline{y})+\beta Q_E^\beta P(x_N,\overline{y})v_\beta^1\\ &+Q_E^\beta r^1(\underline{x}^\beta(N),\overline{y})+\beta(1-\beta)^{-1}Q_E^\beta P(\underline{x}^\beta(N),\overline{y})v_\beta^1. \end{aligned}$$

It can be verified that:

$$\lim_{\beta\uparrow 1}(1-\beta)Q_E^\beta r^1(x_N,\overline{y})=0,\quad \lim_{\beta\uparrow 1}\beta Q_E^\beta P(x_N,\overline{y})v_\beta^1\leqslant 0$$

$$\lim_{\beta\uparrow 1}Q_E^\beta r^1(\underline{x}^\beta(N),\overline{y})=0,\quad \lim_{\beta\uparrow 1}\beta(1-\beta)^{-1}Q_E^\beta P(\underline{x}^\beta(N),\overline{y})v_\beta^1=0.$$

To show that $\lim_{\beta\uparrow 1}\beta Q_E^\beta P(x_N,\overline{y})v_\beta^1\leqslant 0$, take $\delta>0$. For $s\in E$ we have:

$$\sum_{t=1}^{z} p(t|s,x_{Ns},\overline{y}_s)v_\beta^1(t)=\sum_{i=1}^{m_s}\sum_{t=1}^{z} x_{Ns}(i)p(t|s,i,\overline{y}_s)v_\beta^1(t).$$

For β close to 1 we have:

$$\sum_{t=1}^{z} p(t|s,i,\overline{y}_s)v_\beta^1(t) \leqslant v^{\max} + \delta \text{ for all } i \in \{1,2,\ldots,m_s\}.$$

For i with $x_{Ns}(i)<0$, we have $i \in Car(x_s^\beta(N))$, as remarked above, and hence for those i and β close to 1:

$$\sum_{t=1}^{z} p(t|s,i,\overline{y}_s)v_\beta^1(t) = \sum_{t\in E} p(t|s,i,\overline{y}_s)v_\beta^1(t) \geqslant v^{\max} - \delta.$$

Combining these inequalities yields:

$$\sum_{i=1}^{m_s}\sum_{t=1}^{z} x_{Ns}(i)p(t|s,i,\overline{y}_s)\,v_\beta^1(t) \leqslant \sum_{i,x_{N_s}(i)<0} x_{Ns}(i)(v^{\max}-\delta) + \sum_{i,x_{N_s}(i)\geqslant 0} x_{Ns}(i)(v^{\max}+\delta)$$
$$= 2\delta \sum_{i,x_{N_s}(i)\geqslant 0} x_{Ns}(i).$$

Since $\delta>0$ was arbitrary, we conclude that $\lim\limits_{\beta\uparrow 1} \beta P(x_N,\overline{y})v_\beta^1 \leqslant 0$, and hence that $\lim\limits_{\beta\uparrow 1} \beta Q_E^\beta P(x_N,\overline{y})v_\beta^1 \leqslant 0$.

Altogether we have:

$$\begin{aligned} v^{\max}1^E = \lim_{\beta\uparrow 1} Q_E^\beta v_\beta^1 &\leqslant \lim_{\beta\uparrow 1} Q_E^\beta\, r^1(x^\beta(N),\overline{y}) \\ &\quad + \lim_{\beta\uparrow 1} (1-\beta)Q_E^\beta\, r^1(x_N,\overline{y}) + \lim_{\beta\uparrow 1} \beta Q_E^\beta P(x_N,\overline{y})v_\beta^1 \\ &\quad + \lim_{\beta\uparrow 1} Q_E^\beta\, r^1(\underline{x}^\beta(N),\overline{y}) + \lim_{\beta\uparrow 1} \beta(1-\beta)^{-1} Q_E^\beta P(\underline{x}^\beta(N),,\overline{y})v_\beta^1 \\ &\leqslant \lim_{\beta\uparrow 1} Q_E^\beta\, r^1(x^\beta(N),\overline{y}) = \lim_{\beta\uparrow 1} \gamma^1(\overline{x}^\beta,\overline{y})^E. \end{aligned}$$

Because $\overline{y}$ is a limiting average best reply against $\overline{x}^\beta$, for all β close to 1, this implies that $\overline{x}^\beta$ is a limiting average ϵ-optimal strategy for all initial states in E, for β close to 1. This result together with (a) shows that $\overline{x}^\beta$ is limiting average ϵ-optimal for all initial states in S^{**}, for β close to 1.

c) We show:

If $\lim\limits_{\beta\uparrow 1}(1-\beta)(I^A-\beta P(\overline{x}^\beta,\overline{y})^A)^{-1}=0$, *then* $\overline{x}^\beta$ *is limiting average* ϵ*-optimal for all initial states in A, for* β *close to 1.*

Let $v_\beta^{1A}, r^1(x^\beta,\overline{y})^A, 1^A$ (etc.) denote the restriction of v_β^1, $r^1(x^\beta,\overline{y})$ and 1 to coordinates in A. Let v_β^{1AC} (etc.) denote the restriction to coordinates in $A^c{:}= S\setminus A$. Also let $P(x^\beta,\overline{y})^A$ (etc.) denote the restriction of $P(x^\beta,\overline{y})$ to rows and columns corresponding with states in A; let $P(x^\beta,\overline{y})_A$ (etc.) denote restriction of $P(x^\beta,\overline{y})$ only to rows corresponding with states in A.

As above we start off with:

$$v_\beta^1 \leqslant (1-\beta)r^1(x^\beta,\overline{y}) + \beta P(x^\beta,\overline{y})v_\beta^1.$$

This time we derive:

$$v_\beta^{1A} \leqslant (1-\beta)r^1(x^\beta,\bar{y})^A + \beta P(x^\beta(N),\bar{y})^A v_\beta^{1A} + \beta P(x^\beta(N),\bar{y})^{AC} v_\beta^{1AC}$$
$$+\, \beta(1-\beta)P(x_N,\bar{y})_A v_\beta^1 + \beta P(\underline{x}^\beta(N),\bar{y})_A v_\beta^1.$$

Subtracting $\beta P(x^\beta(N),\bar{y})^A\, v_\beta^{1A}$ from both sides, multiplying both sides with $(I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1}$, which exists because $x^\beta(N) \in X$, and by taking limits we obtain:

$$v^{\max} 1^A = \lim_{\beta\uparrow 1} v_\beta^{1A} \leqslant \lim_{\beta\uparrow 1} \beta(I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1} P(x^\beta(N),\bar{y})^{AC} v_\beta^{1AC}$$
$$+ \lim_{\beta\uparrow 1}(1-\beta)(I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1}[r^1(x^\beta,\bar{y})^A + \beta P(x_N,\bar{y})_A v_\beta^1$$
$$+\beta(1-\beta)^{-1} P(\underline{x}^\beta(N),\bar{y})_A v_\beta^1].$$

Observe that each term within the square brackets is bounded uniformly in β. Hence the condition in (c) gives:

$$v^{\max} 1^A \leqslant \lim_{\beta\uparrow 1} \beta(I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1} P(x^\beta(N),\bar{y})^{AC} v_\beta^{1AC}$$
$$= \lim_{\beta\uparrow 1} (I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1} P(x^\beta(N),\bar{y})^{AC} v^{1AC}$$

Now we will use the following relation, which holds for any square matrix P such that $(I-P)^{-1}$ exists:

$$(I-\beta P)^{-1} = (I-P)^{-1} - (1-\beta)(I-\beta P)^{-1}P(I-P)^{-1}.$$

This can easily be verified by (left-) multiplying both sides with $(I-\beta P)$. Applying this for $P = P(x^\beta(N),\bar{y})^A$, using that
$P(x^\beta(N),\bar{y})^A\,(I^A - P\,(x^\beta(N),\bar{y})^A)^{-1}\, P(x^\beta(N),\bar{y})^{AC}$ is bounded and that
$\lim\limits_{\beta\uparrow 1}(1-\beta)(I^A - \beta P(x^\beta(N),\bar{y})^A)^{-1} = 0$, yields:

$$v^{\max} 1^A \leqslant \lim_{\beta\uparrow 1} (I^A - P(x^\beta(N),\bar{y})^A)^{-1} P(x^\beta(N),\bar{y})^{AC} v^{1AC}.$$

Since $v^{1AC} \leqslant v^{\max} 1^{AC}$, the inequality sign in the above inequality can be replaced by an equality sign. Next observe that entry (s,t) of matrix $(I^A - P(x^\beta(N),\bar{y})^A)^{-1} P(x^\beta(N),\bar{y})^{AC}$ denotes the total probability of ever entering A^c at state t when starting in $s \in A$. Hence the probability of entering S^{**} when starting in A is close to 1 for β sufficiently near 1.
Thus we have that $\gamma^1(s,\bar{x}^\beta,\bar{y}) \geqslant v^{\max} - \epsilon$ for each $s \in A$, for β close to 1. ■

2.4.9 Corollary
If for a zero-sum stochastic game we have that $S = S^{\min} = S^{\max}$, then both players have stationary limiting average ϵ-optimal strategies.

This result can also be found in Bewley & Kohlberg [1978] or Vrieze [1987-a].

2.4.10 Remark
Observe that each entry of $(I^A - \beta P(\bar{x}^\beta,\bar{y})^A)^{-1}$ can be written as $\sum\limits_{n=l}^{\infty} c_n(1-\beta)^{n/N}$

with $l\in\mathbb{Z}$ and $c_n\in\mathbb{R}$. Then the condition in theorem 2.4.8 (c) is fulfilled if and only if $l>-N$ for each entry. This holds for instance if all states A are transient with respect to $(x^1,\bar{y})$, since in this case $\lim_{\beta\uparrow 1}(I^A-\beta P(\bar{x}^\beta,\bar{y})^A)^{-1}$ exists.
If $P(\bar{x}^\beta,\bar{y})_s^{AC}\neq 0$ for each $s\in A$, then the condition of theorem 2.4.8 (c) automatically holds because $\|P^n(\bar{x}^\beta,\bar{y})^A\|\leq(1-c(1-\beta)^{1/N})^n$ for some constant $c\in\mathbb{R}$.

One might think by now, that maybe all states in $S^{\max}$ are always ϵ-easy for player 1. The following example however illustrates that in $S^{\max}$ there can be states that are neither easy nor ϵ-easy for player 1.

2.4.11 EXAMPLE

0 1	0 2
1 3	0 4

State 1

0 2	2 2
1 3	0 1

State 2

1 3

State 3

0 4

State 4

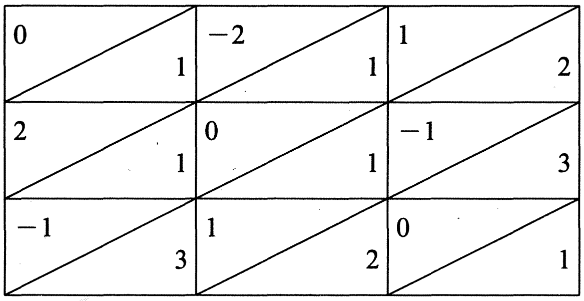
In this example $v=(1,1,1,0)$. It is not hard to verify that for any stationary strategy x player 2 has a best reply y with $\gamma^1(1,x,y)=\gamma^1(2,x,y)=0$. To see that $v=(1,1,1,0)$ examine the stationary stratey $\tilde{x}^\beta$ given by the mixed action $((1-\sqrt{1-\beta})/\beta,(-1+\beta+\sqrt{1-\beta})/\beta)$ for states 1 and 2, and find that:

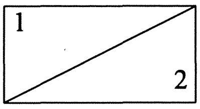
$$1 = \lim_{\beta\uparrow 1}\gamma^1_\beta(s,\tilde{x}^\beta,y)\leq \lim_{\beta\uparrow 1} v^1_\beta(s) = v^1(s)\leq 1$$

for $s=1,2$ and any pure stationary β-discounted best reply y. We do not claim that $\tilde{x}^\beta$ is β-discounted optimal, we just use it to provide a lower bound for v^β.

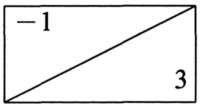
The next example shows that there may be states which are neither ϵ-easy for player 1 nor for player 2 and that for each player $S^{\min}\cup S^{\max}$ may be the set of all his ϵ-easy initial states.

2.4.12 Example





State 2



State 1

State 3

Payoffs are again those to player 1 to be paid by player 2.
The unique stationary β-discounted optimal strategies for this stochastic game are (for starting state 1) given by:

$$x^{\beta}=y^{\beta}=(1/(4-2\beta),\ 1/(4-2\beta),\ (1-\beta)/(2-\beta)).$$

For all $\beta \in [0,1)$ we have $v_{\beta}^{1}=(0,1,-1)$, hence also $v^{1}=(0,1,-1)$.
So $S^{\max}=\{2\}$ and $S^{\min}=\{3\}$.
State 1 is neither ϵ-easy for player 1 nor for player 2, since each of them faces a 'kind of big match' for starting state 1 (cf. example 1.7.4).

2.4.13 Remark

For a zero-sum stochastic game a strong initial state is not-necessarily an ϵ-easy initial state.

This remark is illustrated by the examples 1.7.4 and 2.3.8. In these examples state 1 is a strong initial state but state 1 is not ϵ-easy for player 1.

Chapter 3

Existence of limiting average ϵ-equilibria

3.1 Introduction

Since Mertens & Neyman [1981] showed that the limiting average value exists for any zero-sum stochastic game, the major remaining problem in stochastic game theory is that of existence of limiting average ϵ-equilibria for the general-sum case.

By putting extra conditions on the payoffs and/or transition structure of the stochastic game, several authors have shown that limiting average (ϵ-)equilibria exist for subclasses of stochastic games (cf. chapter 4). In this chapter we present sufficient conditions for the existence of (almost stationary) limiting average ϵ-equilibria. However, our conditions are of a more general nature. We do not put conditions on the payoff/transition structure from the start, but our conditions are formulated in terms of asymptotic properties of sequences of stationary β-discounted equilibria. Remember that stationary β-discounted equilibria exist for any general-sum stochastic game. In chapter 4 we show that our conditions are automatically fulfilled for several of the subclasses that have been examined in literature. It is not clear whether our conditions hold for any general-sum stochastic game. Nevertheless our approach shows that in general the set of strong initial states is larger than the union of ergodic sets for which '$\gamma^{kh}(x^1,y^1) \geqslant V^{kh}$ for $k=1,2$' (cf. lemma 2.3.4).

3.2 Finding more strong initial states

3.2.1 Remark

In this section let $\{(x^\beta,y^\beta):\beta\in[0,1)\}$ *be a sequence of stationary β-discounted equilibria with* $\lim_{\beta\uparrow 1}(x^\beta,y^\beta)=(x^1,y^1)$ *and which furthermore suits definition 2.2.1.*

Observe that all results in section 2.3 were derived for such a sequence, so we can use those results here. We introduce some more notations.

3.2.2 Definition

For our sequence $\{(x^\beta,y^\beta):\beta\in[0,1)\}$ *we define:*

$I := \{h\in\{1,2,...,H\}:\gamma^{1h}(x^1,y^1)\geqslant V^{1h}$ *and* $\gamma^{2h}(x^1,y^1)\geqslant V^{2h}\}$,

$I_0 := \{h\in\{1,2,...,H\}:\gamma^{1h}(x^1,y^1)= V^{1h}$ *and* $\gamma^{2h}(x^1,y^1)= V^{2h}\}$,

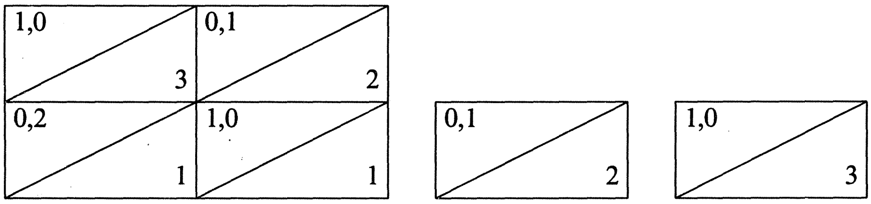
$I_1 := \{h\in\{1,2,...,H\}:\gamma^{1h}(x^1,y^1)< V^{1h}\}$,

$I_2 := \{h \in \{1,2,...,H\} : \gamma^{1h}(x^1,y^1) \geqslant V^{1h}$ *and* $\gamma^{2h}(x^1,y^1) < V^{2h}\}$,

$E := \{s \in S :$ *for initial state s limiting average* ϵ*-equilibra exist, for all* $\epsilon > 0\}$.

Observe that $I \neq \varnothing$ by lemma 2.3.6 and $I_0 \subset I \subset E$ by lemma 2.3.4, hence $E \neq \varnothing$ (cf. theorem 2.3.5). Furthermore notice that I, I_1 and I_2 have empty intersections and $S = T \cup S^I \cup S^{I_1} \cup S^{I_2}$. Recall that T is the set of states that are transient with respect to (x^1,y^1) and that $S^A = \bigcup_{h \in A} S^h$ for any $A \subset \{1,2,...,H\}$. The following example shows that T, I_1 and I_2 may all be empty, whereas I_0 does not need to be equal to I.

3.2.3 Example



State 1 State 2 State 3

For this example the unique stationary β-discounted equilibria are given by (the mixed actions in state 1): $(x^\beta,y^\beta) = (((2-2\beta)/(3-2\beta), 1/(3-2\beta)), (½,½))$ and the corresponding β-discounted rewards are given by:

$\gamma^1_\beta(1,x^\beta,y^\beta) = ½ = v^1_\beta(1)$, $\gamma^2_\beta(1,x^\beta,y^\beta) = ⅔ = v^2_\beta(1)$ for all $\beta \in [0,1)$.

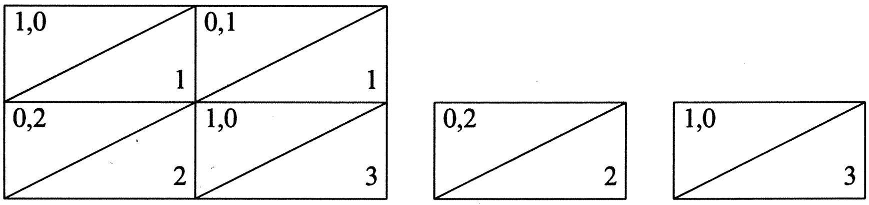
Hence $(x^1,y^1) = ((0,1),(½,½))$,

$\gamma^1(1,x^1,y^1) = ½ = V^1_1 = v^1_1$ and $\gamma^2(1,x^1,y^1) = 1 > ⅔ = V^2_1 = v^2_1$.

For (x^1,y^1) there are three ergodic classes: $S^1 = \{1\}$, $S^2 = \{2\}$, $S^3 = \{3\}$. It is easy to see that $T = \varnothing$, $I = \{1,2,3\}$, $I_0 = \{2,3\}$, $I_1 = I_2 = \varnothing$.
Since for all ergodic sets in this example the condition in lemma 2.3.4 holds, the proof of that lemma supplies a limiting average ϵ-equilibrium for this stochastic game.

It is clear that for any stochastic game for which $T \cup S^{I_1} \cup S^{I_2} = \varnothing$ for some sequence of stationary β-discounted equilibria (x^β,y^β), we can apply lemma 2.3.4 to conclude the existence of a limiting average ϵ-equilibrium (for each starting state). But how to proceed our search for a limiting average ϵ-equilibrium if $T \cup S^{I_1} \cup S^{I_2} \neq \varnothing$? We deal with these matters in this section. We start by examining one more example and then we return to the general problem of existence. The objective of the discussion on next example is to create intuitive understanding for the lemmas that follow.

3.2.4 Example



State 1 State 2 State 3

We already gave a brief discussion of this stochastic game (due to Sorin [1986]) in example 1.8.6. There however, we did not actually give a limiting average ϵ-equilibrium for this stochastic game. Recapitulate that the unique stationary β-discounted equilibria were given by:
$(x^\beta,y^\beta)=((2/(3-\beta),(1-\beta)/(3-\beta)),(½,½))$.
The corresponding β-discounted rewards for starting state 1 were given by:
$\gamma_\beta^1(1,x^\beta,y^\beta)= ½= v_\beta^1(1)$ and $\gamma_\beta^2(1,x^\beta,y^\beta)= ⅔= v_\beta^2(1)$ for all $\beta\in[0,1)$.
Hence $(x^1,y^1)=((1,0),(½,½))$, $V^1=v^1=(½,0,1)$ and $V^2=v^2=(⅔,2,0)$. So there are three ergodic sets with respect to (x^1,y^1): $S^1=\{1\}$, $S^2=\{2\}$ and $S^3=\{3\}$. It is clear that $I_0=\{2,3\}$, $T=\varnothing$, $I_1=\varnothing$ and $I_2=\{1\}$, because $\gamma^1(1,x^1,y^1)= ½\geqslant V_1^1$ and $\gamma^2(1,x^1,y^1)= ½< V_1^2$.
Now observe that state 1 is transient with respect to (x^β,y^β) for all $\beta\in[0,1)$, whereas state 1 is recurrent with respect to (x^1,y^1). This means that for all $\beta\in[0,1)$ the payoffs in states 2 and 3 may partly determine the β-discounted rewards for state 1 and hence the payoffs in states 2 and 3 may have their impact on $\lim\limits_{\beta\uparrow1}\gamma_\beta^k(1,x^\beta,y^\beta)=V_1^k$ for $k=1$ and 2. Next observe that with respect to (x^1,y^1) any play will remain in state 1 and hence $\gamma^k(1,x^1,y^1)$ is determined only by the payoffs in state 1; moreover $\gamma^k(1,x^1,y^1)=r^k(1,x^1,y^1)$.
Let us examine $\lim\limits_{\beta\uparrow1}\gamma_\beta^k(1,x^\beta,y^\beta)$, which equals $\lim\limits_{\beta\uparrow1}\gamma_\beta^k(1,x^\beta,y^1)$ since $y^\beta=y^1$ for all $\beta\in[0,1)$. By lemma 2.2.5 and lemma 2.2.6 there are $\mu_1^1,\mu_1^2,\mu_1^3\in[0,1]$ such that $\sum\limits_{h=1}^{3}\mu_1^h=1$ and $V_1^k=\sum\limits_{h=1}^{3}\mu_1^h\gamma^{kh}(x^1,y^1)$ for $k=1,2$. Furthermore those lemmas state that, if under (x^β,y^1) there are no transitions possible from state 1 to S^h, $h\in\{1,2,3\}$, then $\mu_1^h=0$. For this example we have that $\gamma^{21}(x^1,y^1)= ½< ⅔= V^{21}$. Recall that $\gamma^{21}(x^1,y^1)$ and V^{21} are respectively the worth of $\gamma^2(x^1,y^1)$ and V^2 for ergodic set $S^1=\{1\}$. Hence it follows that $\mu_1^1<1$. This implies, as will be worked out below in a more general setting:

$$\sum_{t=2}^{3}p(t|1,(0,1),y^1)>0 \text{ and } \sum_{t=1}^{3}p(t|1,(0,1),y^1)V_t^2\geqslant V_1^2=V^{21}.$$

Since $\gamma^{22}(x^1,y^1)\geqslant V^{22}$ and $\gamma^{23}(x^1,y^1)\geqslant V^{23}$ we conclude that if, against y^1, player 1 uses the pure stationary strategy (0,1) in state 1 and x^1 in S^2 and S^3, then with probability 1 a transition to $S^2\cup S^3$ will occur and hence $\gamma^2(1,x^*,y^1)\geqslant V_1^2$, where x^* is strategy (0,1) for player 1 (cf. example 1.8.6).

By lemma 1.6.4 (b) we also have that $\sum_{t=1}^{3} p(t|1,x^*,y^1)\, V_t^1 = V_1^1$, which follows by taking limits for β to 1 in

$$\gamma_\beta^1(1,x^\beta,y^1) = \gamma_\beta^1(1,x^*,y^1) = (1-\beta)r^1(1,x^*,y^1) + \beta \sum_{t=1}^{3} p(t|1,x^*,y^1)\gamma_\beta^1(t,x^*,y^1).$$

Since $\gamma^{12}(x^1,y^1) \geqslant V^{12}$ and $\gamma^{13}(x^1,y^1) \geqslant V^{13}$ we can also conclude that $\gamma^1(1,x^*,y^1) \geqslant V_1^1 = V^{11}$ (cf. example 1.8.6).
Hence we have that $\gamma^k(s,x^*,y^1) \geqslant V_s^k$ for all $s \in S$ and for $k=1,2$. So both players should be rather satisfied with these limiting average rewards. By lemma 1.5.5 and lemma 1.6.4 we even have that $\gamma^1(1,x^*,y^1) = \gamma^1(1,x^1,y^1)$ so player 1 has no profitable deviations against y^1. Unfortunately y^1 is not a best reply for player 2 against x^*. Hence (x^*,y^1) is not a limiting average equilibrium. However it should be observed that if player 1 uses the stationary strategy $x^\lambda := (1-\lambda)x^1 + \lambda x^*$ $(= (1-\lambda,\lambda)$ for this example) then, for any $\lambda \in (0,1]$ we still have that for (x^λ,y^1) a transition from state 1 to $S^2 \cup S^3$ will occur with probability 1. Moreover $\gamma^k(s,x^\lambda,y^1) = \gamma^k(s,x^*,y^1)$ for all $s \in S$, all $\lambda \in (0,1]$ and $k=1,2$.
Now we can construct an almost stationary limiting average ϵ-equilibrium.
Let $\epsilon > 0$ and let $Y^{(n)}$ and $y^{(n)}$ be as in the proof of theorem 2.3.4. Let $X^{(n)}$ be the random variable denoting the action frequencies of player 1 within $Car(x^1)$ up to stage n; let $x^{(n)}$ be a realization of $X^{(n)}$. Then, pretending that absorption does not take place, for each $\alpha > 0$ and $\delta > 0$ there is $N_{\alpha\delta} \in \mathbb{N}$ such that:

$\mathrm{Prob}_{x^\lambda,y^1}\{\|X^{(n)} - x^1\| > \alpha$ for any $n \geqslant N_{\alpha\delta}\} < \delta$ and

$\mathrm{Prob}_{x^\lambda,y^1}\{\|Y^{(n)} - y^1\| > \alpha$ for any $n \geqslant N_{\alpha\delta}\} < \delta$.

Choose $\alpha \in (0,\epsilon/4M)$ and $\delta > 0$ such that:

$(1-\delta)^4(\gamma^k(x^*,y^1) - \alpha M) - (1-(1-\delta)^4)M \geqslant \gamma^k(x^1,y^1) - \epsilon/2$ for $k=1,2$.

Choose $\lambda \in (0,\epsilon/4M)$ such that:

$$\mathrm{Prob}_{x^\lambda,y^1}\{\text{absorption before stage } N_{\alpha\delta}\} < \delta.$$

Choose $N_\lambda \in \mathbb{N}$, $N_\lambda > N_{\alpha\delta}$ such that:

$$\mathrm{Prob}_{x^\lambda,y^1}\{\text{absorption before stage } N_\lambda\} \geqslant 1-\delta.$$

Define π_ϵ^* by:

a) use x^λ unless:
   i) player 2 chooses $j \notin Car(y^1)$
   ii) $\|y^{(n)} - y^1\| > \alpha$ for some $n \geqslant N_{\alpha\delta}$
   iii) at stage N_λ play is still in the initial state
b) if (i), (ii) or (iii) occurs, then use some retaliation strategy $\pi_{\epsilon/2}^r$.

Define σ_ϵ^* analogously.
Now it can be verified that $(\pi_\epsilon^*,\sigma_\epsilon^*)$ is an almost stationary limiting average ϵ-equilibrium.

In order to generalize these ideas we introduce the following notations.

3.2.5 Definition

For a non-empty $A \subset S$, $A \neq S$, a pair of stationary strategies (x,y) and $\beta \in [0,1)$ we define:

a) *$P(x,y)^A$ is the restriction of $P(x,y)$ to rows and columns corresponding with A.*

b) *$P(x,y)^{AC}$ is the restriction of $P(x,y)$ to rows corresponding with A and columns corresponding with $A^c = S \setminus A$.*

c) *For any $\alpha \in \mathbb{R}^z$ the restriction of α to coordinates corresponding with A (A^c) is denoted as α^A (α^{AC}).*

d) *I^A is the identity matrix of size $|A| \times |A|$.*

e) *$M_A^\beta := [m_{st}^\beta]_{s \in A, t \in A} := (1-\beta)(I^A - \beta P(x^\beta,y^1)^A)^{-1}$ and*
$M_A^1 := \lim_{\beta \uparrow 1} M_A^\beta$, which limit we assume to exist without loss of generality.

f) *$N_A^\beta := [n_{st}^\beta]_{s \in A, t \in A^c} := \beta(I^A - \beta P(x^\beta,y^1)^A)^{-1} P(x^\beta,y^1)^{AC}$ and*
$N_A^1 := \lim_{\beta \uparrow 1} N_A^\beta$, which limit we may also assume to exist.

Using this definition the next lemma follows from elementary calculations. We therefore omit the proof.

3.2.6 lemma

Notations us above. Then:

a) *$m_{st}^\beta \geqslant 0$ for all $s \in A, t \in A$ and $\beta \in [0,1)$;*
$n_{st}^\beta \geqslant 0$ for all $s \in A, t \in A^c$ and $\beta \in [0,1)$.

b) *$\sum_{t \in A} m_{st}^\beta + \sum_{t \in A^c} n_{st}^\beta = 1$ for all $s \in A$ and $\beta \in [0,1)$.*

c) *$\gamma_\beta^2(x^\beta,y^1)^A = M_A^\beta r^2(x^\beta,y^1)^A + N_A^\beta \gamma_\beta^2(x^\beta,y^1)^{AC}$ for all $\beta \in [0,1)$.*

d) *$V^{2A} = M_A^1 r^2(x^1,y^1)^A + N_A^1 V^{2AC}$.*

e) *$n_{st}^\beta = \dfrac{\beta}{1-\beta} \sum_{s^* \in A} \sum_{i^*=1}^{m_{s^*}} m_{ss^*}^\beta x_{s^*}^\beta(i^*)\, p(t|s^*,i^*,y_{s^*}^1)$ for each $s \in A$, $t \in A^c$.*

3.2.7 Lemma

Notations as above.

Let $A = S^{h^1} \cup S^{h^2} \cup \ldots \cup S^{h^a}$ and let $s \in A$. Then there is $\mu_s \in \Delta^a$ such that:

$$\sum_{t \in A} m_{st}^1 r^2(t,x_t^1,y_t^1) = \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha} \gamma^{2h^\alpha}(x^1,y^1), \text{ where } \lambda_s = \sum_{t \in A} m_{st}^1 \in [0,1].$$

Proof:

By definition 3.2.5 we have that $M_A^\beta(I^A - \beta P(x^\beta,y^1)^A) = (1-\beta)I^A$ for all $\beta \in [0,1)$. Hence $M_A^1 = M_A^1 P(x^1,y^1)^A$, which implies that for each $s \in A$ the s-*th* row of M_A^1 is a multiple λ_s of a stationary distribution for $P(x^1,y^1)^A$. Remember that $q^{h^\alpha A}$ is the restriction of the unique stationary distribution q^{h^α} of $P(x^1,y^1)$ on S^{h^α} to coordinates corresponding with states in A (cf. definition 2.2.1). So we have that for some $\mu_s \in \Delta^a$:

$$\begin{aligned}\sum_{t\in A} m_{st}^{1}\, r^2(t,x_t^1,y_t^1) &= \sum_{t\in A}\lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha}\, q_t^{h^\alpha A}\, r^2(t,x_t^1,y_t^1)\\ &= \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha} \sum_{t\in A} q_t^{h^\alpha A}\, r^2(t,x_t^1,y_t^1)\\ &= \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha}\, \gamma^{2h^\alpha}(x^1,y^1)\end{aligned}$$

and

$$\sum_{t\in A} m_{st}^{1} = \sum_{t\in A} \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha}\, q_t^{h^\alpha A} = \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha} \sum_{t\in A} q_t^{h^\alpha A} = \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha} = \lambda_s \in [0,1]$$

by lemma 3.2.6 (a) and (b). ■

3.2.8 LEMMA

Notations as above.

For $s^* \in A$ *and* $i^* \in Car(x_{s^*}^{\beta})$ *with* $\sum_{t\in A^c} p(t|s^*,i^*,y_{s^*}^{1}) > 0$ *define:*

$$\lambda_s(s^*,i^*) := \lim_{\beta\uparrow 1} \frac{\beta}{1-\beta}\, m_{ss^*}^{\beta}\; x_{s^*}^{\beta}(i^*) \sum_{t\in A^c} p(t|s^*,i^*,y_{s^*}^{1}).$$

Suppose that there is at least one pair (s^*,i^*) *with this property. Then:*

$$\sum_{t\in A^c} n_{st}^{1}\, V_t^2 = \sum_{s^*,i^*} \lambda_s(s^*,i^*) \sum_{t\in A^c} \left[\frac{p(t|s^*,i^*,y_{s^*}^{1})V_t^2}{\sum_{t^*\in A^c} p(t^*|s^*,i^*,y_{s^*}^{1})}\right] \textit{ and}$$

$$\sum_{t\in A^c} n_{st}^{1} = \sum_{s^*,i^*} \lambda_s(s^*,i^*) = 1-\lambda_s$$

PROOF:

By lemma 3.2.6 (e): $n_{st}^{\beta} = \dfrac{\beta}{1-\beta} \displaystyle\sum_{s^*\in A} \sum_{i^*=1}^{m_{s^*}} m_{ss^*}^{\beta}\; x_{s^*}^{\beta}(i^*)\, p(t|s^*,i^*,y_{s^*}^{1})$.

$$\begin{aligned}\text{Hence } n_{st}^{1} &= \lim_{\beta\uparrow 1} \frac{\beta}{1-\beta} \sum_{s^*\in A} \sum_{i^*=1}^{m_{s^*}} m_{ss^*}^{\beta}\; x_{s^*}^{\beta}(i^*)\, p(t|s^*,i^*,y_{s^*}^{1})\\ &= \sum_{s^*,i^*} \lambda_s(s^*,i^*)\, \frac{p(t|s^*,i^*,y_{s^*}^{1})}{\sum_{t^*\in A^c} p(t^*|s^*,i^*,y_{s^*}^{1})}\end{aligned}$$

for $s\in A$ and $t\in A^c$, which proves this lemma (cf. lemma 3.2.6 (b) and lemma 3.2.7). ■

3.2.9 LEMMA

Let $A = \bigcup_{\alpha=1}^{a} S^{h^\alpha}$ *and* $\varnothing \neq A \neq S$. *Suppose* $V_s^2 = V_t^2 =: V_A^2$ *for all* $s,t\in A$.

If $V_A^2 > \gamma^{2h^\alpha}(x^1,y^1)$ *for* $\alpha = 1,2,\ldots,a$, *then there is* $s^*\in A$ *and* $i^* \in Car(x_{s^*}^{\beta})$ *such that:* $\sum_{t\in A^c} p(t|s^*,i^*,y_{s^*}^{1}) > 0$ *and* $\sum_{t\in S} p(t|s^*,i^*,y_{s^*}^{1})V_t^2 \geqslant V_A^2 = V_{s^*}^2$.

Furthermore: $\sum_{t\in S} p(t|s^*,i^*,y_{s^*}^{1})\, V_t^1 = V_{s^*}^1$.

PROOF:
Take $s \in A$. By lemma 3.2.6 (d): $V_s^2 = \sum_{t\in A} m_{st}^1\, r^2(t,x_t^1,y_t^1) + \sum_{t\in A^c} n_{st}^1\, V_t^2$.
Applying lemmas 3.2.7 and 3.2.8 we derive:

$$V_s^2 = \lambda_s \sum_{\alpha=1}^{a} \mu_s^{h^\alpha}\, \gamma_s^{h^\alpha}(x^1,y^1) + \sum_{s^*,i^*} \lambda_s(s^*,i^*) \sum_{t\in A^c} \frac{p(t|s^*,i^*,y_s^{1\cdot})V_t^2}{\sum_{t^*\in A^c} p(t^*|s^*,i^*,y_s^{1\cdot})}$$

for certain $\lambda_s \in [0,1]$, $\mu_s \in \Delta^a$ and $\lambda_s(s^*,i^*) \in [0,1]$, and it holds that:

$$\lambda_s + \sum_{s^*,i^*} \lambda_s(s^*,i^*) = 1.$$

Since $\gamma^{2h^\alpha}(x^1,y^1) < V_s^2$ for all $\alpha \in \{1,2,...,a\}$, we conclude that $\lambda_s < 1$ and for at least one (s^*,i^*) with $\lambda_s(s^*,i^*) > 0$ we have:

$$\sum_{t\in A^c} p(t|s^*,i^*,y^1) > 0 \quad \text{and} \quad \sum_{t\in A^c} \frac{p(t|s^*,i^*,y_s^{1\cdot})V_t^2}{\sum_{t^*\in A^c} p(t^*|s^*,i^*,y_s^{1\cdot})} \geqslant V_s^2 = V_A^2.$$

Since $V_t^2 = V_A^2$ for all $t \in A$, it follows that $\sum_{t\in S} p(t|s^*,i^*,y_s^{1\cdot})V_t^2 \geqslant V_A^2$.
By lemma 2.3.3 we also have $\sum_{t\in S} p(t|s^*,i^*,y_s^{1\cdot})V_t^1 = V_s^{1\cdot}$. ■

The following lemma can be proved analogously.

3.2.10 LEMMA
Let $A = \bigcup_{\alpha=1}^{a} S^{h^\alpha}$ *and* $\varnothing \neq A \neq S$. *Suppose* $V_s^1 = V_t^1 =: V_A^1$ *for all* $s,t \in A$.
If $V_A^1 > \gamma^{1h^\alpha}(x^1,y^1)$ *for* $\alpha = 1,2,...,a$, *then there is* $s^* \in A$ *and* $j^* \in Car(y_s^{\beta})$ *such that:* $\sum_{t\in A^c} p(t|s^*,x_s^{1\cdot},j^*) > 0$ *and* $\sum_{t\in S} p(t|s^*,x_s^{1\cdot},j^*)V_t^1 \geqslant V_A^1 = V_s^{1\cdot}$.
Furthermore: $\sum_{t\in S} p(t|s^*,x_s^{1\cdot},j^*)\, V_t^2 = V_s^{2\cdot}$.

Observe that in the above lemmas $i^* \in Car(x_s^{\beta})\setminus Car(x_s^{1\cdot})$ and $j^* \in Car(y_s^{\beta})\setminus Car(y_s^{1\cdot})$, since $\sum_{t\in A^c} p(t|s^*,x_s^{1\cdot},y_s^{1\cdot}) = 0$.

3.2.11 THEOREM
Let (x^*,y^*) *be a pair of stationary strategies with the following properties.*

a) *For each* $h \in I_1$ *there is* $s^h \in S^h$ *and* $j^h \in Car(y_{s^h}^{\beta_h})$ *with* $y_{s^h}^* = j^h$ *such that:*
$\sum_{t\notin S^h} p(t|s^h,x_{s^h}^1,j^h) > 0$,
$\sum_{t\in S} p(t|s^h,x_{s^h}^1,j^h)V_t^2 = V_{s^h}^2$ *and* $\sum_{t\in S} p(t|s^h,x_{s^h}^1,j^h)V_t^1 \geqslant V_{s^h}^1 = V^{1h}$.
Also $y_s^* = y_s^1$ *for all* $s \in S \setminus \bigcup_{h\in I_1} \{s^h\}$.

b) *For each* $h \in I_2$ *there is* $s^h \in S^h$ *and* $i^h \in Car(x_{s^h}^{\beta_h})$ *with* $x_{s^h}^* = i^h$ *such that:*
$\sum_{t\notin S^h} p(t|s^h,i^h,y_{s^h}^1) > 0$,

$\sum_{t\in S} p(t|s^h,i^h,y^1_{s^h})V^1_t = V^1_{s^h}$ *and* $\sum_{t\in S} p(t|s^h,i^h,y^1_{s^h})V^2_t \geqslant V^2_{s^h} = V^{2h}$.
Also $x^*_s = x^1_s$ *for all* $s\in S\setminus \bigcup_{h\in I_2}\{s^h\}$.

c) *Each* $s\in S\setminus S^I$ *is transient with respect to* (x^*,y^*).
Then $P(x^*,y^*)V^k\geqslant V^k$ *and* $\gamma^k(x^*,y^*)\geqslant V^k$ *for* $k=1,2$.

PROOF:
By lemma 2.3.3 and by (a) and (b) it follows that $P(x^*,y^*)V^k\geqslant V^k$, which also implies that $Q(x^*,y^*)V^k\geqslant V^k$. For $s\in S^I$ we have by definition that $\gamma^k(s,x^*,y^*)=\gamma^k(s,x^1,y^1)\geqslant V^k$.
Since all states in $S\setminus S^I$ are transient with respect to (x^*,y^*) it holds that (cf. lemma 1.5.2 (e) and lemma 1.5.5):

$$\begin{aligned}\gamma^k(x^*,y^*) &= Q(x^*,y^*)r^k(x^*,y^*) = Q(x^*,y^*)\,Q(x^*,y^*)\,r^k(x^*,y^*) = \\ &= Q(x^*,y^*)\gamma^k(x^*,y^*)\geqslant Q(x^*,y^*)V^k\geqslant V^k.\end{aligned}$$ ■

3.2.12 LEMMA
For (x^*,y^*) *as in theorem 3.2.11 and* $\lambda\in(0,1)$ *define:*
$x^\lambda_s := (1-\lambda)x^1_s + \lambda x^*_s$ *and* $y^\lambda_s := (1-\lambda)y^1_s + \lambda y^*_s$ *for all* $s\in S$.
Then the following statements hold:

a) *Each* $s\in S^I$ *is recurrent with respect to* (x^λ,y^λ) *and each* $s\in S\setminus S^I$ *is transient with respect to* (x^λ,y^λ).

b) $P(x^\lambda,y^\lambda)V^k\geqslant V^k$ *and* $\gamma^k(x^\lambda,y^\lambda)\geqslant V^k$ *for* $k=1,2$.

PROOF:
For all $s\in S\setminus \bigcup_{h\in I_1\cup I_2}\{s^h\}$ we have $(x^\lambda_s,y^\lambda_s)=(x^*_s,y^*_s)$ and hence it follows that $p(t|s,x^\lambda_s,y^\lambda_s)= p(t|s,x^*_s,y^*_s)$ for those s and for all $t\in S$.
For $s\in S^h, h\in I_1\cup I_2$, we have that $p(t|s,x^\lambda_s,y^\lambda_s)=\lambda p(t|s,x^*_s,y^*_s)$ for all $t\notin S^h$.
Since $\lambda>0$ the ergodic sets for (x^λ,y^λ) are precisely the same as those for (x^*,y^*). By lemma 2.3.3 and theorem 3.2.11 we have that $P(x^\lambda,y^\lambda)V^k\geqslant V^k$ for $k=1,2$. Using arguments as in the proof of theorem 3.2.11 we derive that $\gamma^k(x^\lambda,y^\lambda)\geqslant V^k$ for $k=1,2$. ■

Now we can formulate the main theorem of this section.

3.2.13 THEOREM
Let (x^λ,y^λ) *be as in lemma 3.2.12.*
If $p(t|s,x^\lambda_s,y^\lambda_s)=0$ *for all* $s\in T$ *and* $t\in S\setminus(S^{I_0}\cup T)$,
then (x^λ,y^λ) *can be supplemented with suitable retaliation threats to achieve an almost stationary* ϵ*-equilibrium, for* $\lambda\in(0,1)$ *sufficiently small.*

PROOF:
Let $\epsilon>0$. For each $\eta\in(0,1)$ there is $N_\eta\in\mathbb{N}$ such that, with probability at least $1-\eta$, the expected number of transitions among elements of the set

$\{S^h : h \in I_1 \cup I_2\} \cup T$ will be at most N_η with respect to (x^λ, y^λ) for any initial state $s \in S \setminus S^I$. Furthermore N_η does not depend on $\lambda \in (0,1)$.
Now choose $\eta, \delta \in (0,1)$ such that for all $s \in S \setminus S^I$:

$$(1-\eta)(1-\delta)^4 (\gamma^k(s,x^\lambda,y^\lambda)-\epsilon/4) - (1-(1-\eta)(1-\delta)^4) M \geqslant \gamma^k(s,x^\lambda,y^\lambda)-\epsilon/2.$$

Let $Y_s^{(n)}$ $(X_s^{(n)})$ be random variables denoting the action frequencies for player 2 (1) after n stages in state s, and let $y_s^{(n)}(x_s^{(n)})$ denote realizations of these.
Let $K \in \mathbb{N}$ be a constant such that for $\alpha > 0$:
if for all s and for all n sufficiently large $|y_s^{(n)} - y_s^1| \leqslant \alpha$ and $|x_s^{(n)} - x_s^1| \leqslant \alpha$, then for each ergodic set S^h the limiting average reward to player k is between $\gamma^{kh}(x^1,y^1) - \alpha K$ and $\gamma^{kh}(x^1,y^1) + \alpha K$.
Next choose $\alpha \in (0, \epsilon/4KN_\eta)$ and $N_{\alpha\delta} \in \mathbb{N}$ such that
$\text{Prob}_{x^\lambda,y^\lambda} \{\| Y_s^{(n)} - y_s^1 \| > \alpha$ for any $n \geqslant N_{\alpha\delta}$ and any $s \in S\} < \delta$ and
$\text{Prob}_{x^\lambda,y^\lambda} \{\| X_s^{(n)} - x_s^1 \| > \alpha$ for any $n \geqslant N_{\alpha\delta}$ and any $s \in S\} < \delta$.
Observe that $N_{\alpha\delta}$ is independent of the initial state.
Choose $\lambda \in (0, \epsilon/4K)$ and $N_\lambda \geqslant N_{\alpha\delta}$ such that:
$\text{Prob}_{x^\lambda,y^\lambda}$ {transition from S^h to $S \setminus S^h$ within $N_{\alpha\delta}$ stages, with $h \in I_1 \cup I_2\} < \delta$,
$\text{Prob}_{x^\lambda,y^\lambda}$ {transition from S^h to $S \setminus S^h$ within N_λ stages, with $h \in I_1 \cup I_2\}$ $\geqslant (1-\delta)^{1/N_\eta}$.
Define strategy π_ϵ^* for player 1 by:

a) use strategy x^λ unless (i), (ii) or (iii) below occurs:
   i) for some $s \in S$ and some $n \geqslant N_{\alpha\delta} : \|y_s^{(n)} - y_s^1\| > \alpha$, where $y_s^{(n)}$ is a realization of $Y_s^{(n)}$.
   ii) player 2 chooses an action outside $Car^z(y^\lambda)$.
   iii) after N_η transitions among elements of $\{S^h : h \in I_1 \cup I_2\} \cup T$, play is still in $S \setminus S^I$.
   iv) play remains in a set S^h, $h \in I_1 \cup I_2$ for more than N_λ stages.
b) if (i), (ii), (iii) or (iv) occurs, then use some retaliation strategy $\pi_{\epsilon/4}^r$, from that moment on.

For player 2 the strategy σ_ϵ^* is defined analogously.

If the players use $(\pi_\epsilon^*, \sigma_\epsilon^*)$ then with probability at most $(1-(1-\eta)(1-\delta)^4)$ some player may start using his retaliation strategy and with probability at least $(1-\eta)(1-\delta)^4$ the players remain using (x^λ, y^λ) forever.
Hence, by choice of λ, η, α and δ, we have $\gamma^k(\pi_\epsilon^*, \sigma_\epsilon^*) \geqslant \gamma^k(x^\lambda, y^\lambda) - \epsilon/2$ for $k = 1,2$.
Now suppose player 1 uses π_ϵ^* and player 2 uses some arbitrary σ.

If player 1 detects a deviation at stage n, where player 2's action was j in state s, then with probability at least $(1-\lambda)$ player 1 was using x_s^1 at that stage and with probability at most λ player 1 was using some i^h, where $s = s^h$. Hence by lemma 2.3.3, player 1's retaliation in that case gives that the limiting average reward will be at most $(1-\lambda)(V_s^2 + \epsilon/4) + \lambda M \leqslant V_s^2 + \epsilon/2 \leqslant \gamma^k(s, x^\lambda, y^\lambda) + \epsilon/2$.

If player 1 does not detect any deviation, then at each transition among elements of $\{S^h : h \in I_1 \cup I_2\} \cup T$ player 2 can gain at most $\alpha K \leqslant \epsilon/2N_\eta$. Since there are at most N_η transitions without player 1 starting retaliation, we find that the limiting average reward to player 2 will be at most $\gamma^k(x^\lambda,y^\lambda)+\epsilon/2$.

Here it should be noticed that we use the condition of the theorem to conclude that player 2 cannot deviate in T (and neither can player 1). This is due to the following argument.

If player 2 chooses an action outside $Car(y^\lambda)$ in T, then player 1 will retaliate directly according to the definition of π_ϵ^*. Retaliation is possible because by lemma 2.3.3 we have $P(x^1,y)V^2 \leqslant V^2$ for all stationary strategies y. By lemma 1.6.2 it is sufficient to consider deviations by stationary strategies.

Now suppose that the play is in T and player 2 uses a stationary strategy y with $Car(y_s) \subset Car(y_s^\lambda)$ for all $s \in T$ and $y_s = y_s^1$ for $s \in S^{I_0}$.

We denote restrictions of $r^2(\,,\,)$, $\gamma^2(\,,\,)$, $P(\,,\,)$, $Q(\,,\,)$ to coordinates, respectively rows and columns, corresponding with $T \cup S^{I_0}$ by $\overline{r}^2(\,,\,)$, $\overline{\gamma}^2(\,,\,)$, $\overline{P}(\,,\,)$, $\overline{Q}(\,,\,)$. If the condition of this theorem holds, then any play which is in T at some stage, will remain in $T \cup S^{I_0}$ forever, in case the players use (x^1,y) from that stage on. If some state $t \in T$ is recurrent with respect to (x^1,y), then player 1 using π_ϵ^* would start, with probability 1, to retaliate player 2. By lemma 2.3.3 this would give player 2 a limiting average reward of at most $V_t^2+\epsilon/2$. On the other hand, if all states in T are transient with respect to (x^1,y) then we have by lemma 2.3.3:

$$\begin{aligned}\overline{\gamma}^2(x^1,y) &= \overline{Q}(x^1,y)\overline{r}^2(x^1,y) = \overline{Q}(x^1,y)\,\overline{Q}(x^1,y)\overline{r}^2(x^1,y)\\ &= \overline{Q}(x^1,y)\,\overline{Q}(x^1,y^1)\overline{r}^2(x^1,y^1)\\ &= \overline{Q}(x^1,y)\,\overline{V}^2\\ &\leqslant \overline{V}^2.\end{aligned}$$

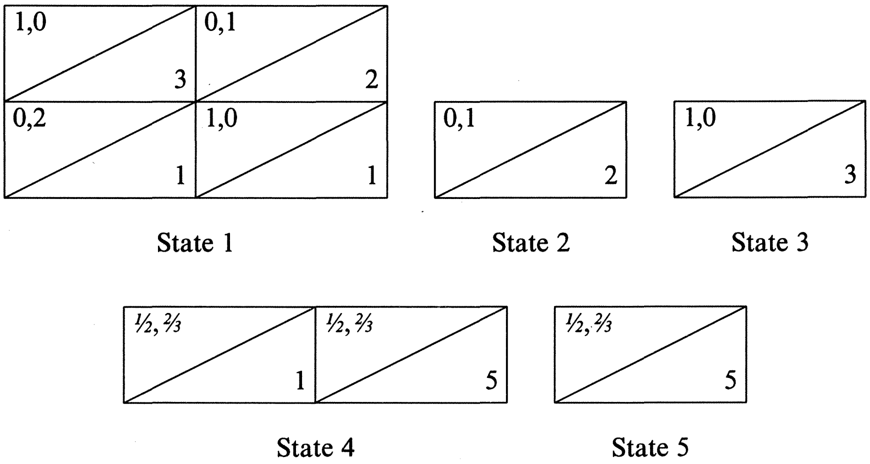
Since we know that, once the play is in S^I neither player 1 nor player 2 can gain more that ϵ by deviating from $(\pi_\epsilon^*,\sigma_\epsilon^*)$, the above implies that against π_ϵ^* player 1 has no profitable deviations while the play is in T. ■

3.2.14 Remark

If for any converging sequence of stationary β-discounted equilibria $\{(x^\beta,y^\beta): \beta \in [0,1)\}$ one can choose s^h, i^h, j^h, for $h \in I_1 \cup I_2$, such that the conditions of theorem 3.2.13 hold, then (x^λ,y^λ) can be supplemented with retaliation threats to achieve an almost stationary limiting average ϵ-equilibrium for λ sufficiently small.

Let us now consider an example to clarify the condition in theorem 3.2.13.

3.2.15 Example



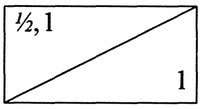
Observe that we have simply added states 4 and 5 to example 3.2.3.
A β-discounted equilibrium (x^β,y^β) is given by:
$x^\beta := (((2-2\beta)/\ (3-2\beta),\ 1/(3-2\beta)),1,1,1,1)$ and $y^\beta := ((\tfrac12,\tfrac12),1,1,(\tfrac12,\tfrac12),1)$. Then $\gamma^1_\beta(x^\beta,y^\beta) = (\tfrac12,0,1,\tfrac12,\tfrac12)$ and $\gamma^2_\beta(x^\beta,y^\beta) = (\tfrac23,1,0,\tfrac23,\tfrac23)$ for the respective initial states. Hence $V^1 = (\tfrac12,0,1,\tfrac12,\tfrac12)$ and $V^2 = (\tfrac23,1,0,\tfrac23,\tfrac23)$. Also: $\gamma^1(x^1,y^1) = (\tfrac12,0,1,\tfrac12,\tfrac12)$ and $\gamma^2(x^1,y^1) = (1,1,0,5/6,\tfrac23)$. Hence $I_1 = I_2 = \varnothing$, $S^I = \{1,2,3,5\}$, $S^{I_0} = \{2,3,5\}$, $T = \{4\}$.
Notice that $p(1|4,x^1_4,y^1_4) = \tfrac12 \neq 0$, so the condition of theorem 3.2.13 is not fulfilled. Although for initial states in $\{1,2,3,5\}$ the strategies (x^1,y^1) can be supplemented to achieve an almost stationary limiting average ϵ-equilibrium (cf. example 3.2.3), this is impossible for initial state 4, since player 2 could gain $1/6$ by using $y^* = ((\tfrac12,\tfrac12),1,1,(1,0),1)$ against x^1, for initial state 4. Player 1 cannot check in state 4 whether player 2 is using y^* or y^1.
If however we had started with $y^\beta_4 = (0,1)$, then the condition of theorem 3.2.13 would have been fulfilled. If we had started with $y^\beta_4 = (1,0)$, then the condition of theorem 3.2.13 would not have been fulfilled, but since player 2 has no profitable deviations within $Car(y^\beta_4)$ we could also in this case establish an ϵ-equilibrium by supplementing (x^1,y^1) with retaliation threats.

So if the condition of theorem 3.2.13 is not fulfilled, then this does not necessarily mean that it is impossible to achieve a limiting average ϵ-equilibrium from (x^1,y^1). All one really needs is that neither player 1 nor player 2 has profitable deviations from x^1_s, respectively y^1_s, in any $s \in T$, that cannot be detected by the opponent (with probability near 1). Hence we can make the following remarkable observation:
If after each stage both players were told what mixed actions have been used at that stage, then the condition of theorem 3.2.11 would be sufficient to achieve an almost stationary limiting average ϵ-equilibrium. Any deviation could be detected immediately and hence retaliation threats could be used for

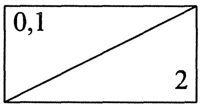
all states. Notice that for zero-sum stochastic games the limiting average value and limiting average ϵ-optimal strategies are independent of this information.

In the above construction we used just one sequence $\{(x^\beta,y^\beta)\colon \beta\in[0,1)\}$ of stationary β-discounted equilibria. One could also use an iteration argument: start with an arbitrary sequence $\{(x^\beta,y^\beta)\colon\beta\in[0,1)\}$ of stationary β-discounted equilibria; find all strong initial states that can be found by the above techniques; replace those initial states s by absorbing states s, i.e. each player has just one action in s and $p(s|s,1,1)=1$, with $r^k(s,1,1)=\gamma^k(s,x^\lambda,y^\lambda)$; in this new stochastic game again take a sequence of stationary β-discounted equilibria and try to find more strong initial states; repeat this procedure. In certain cases it may lead to an ϵ-equilibrium for all initial states. If however the stationary β-discounted equilibria are chosen arbitrary at each iterative step, then one does not necessarily find new strong initial states. We give two examples to illustrate these ideas.

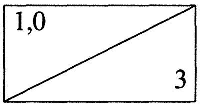
Suppose in example 3.2.15 we had indeed started with $(x^\beta,y^\beta)=((((2-2\beta)/(3-2\beta),1/(3-2\beta)),1,1,1,1),((½,½),1,1,(½,½),1))$. Then in the first step we would have found the strong initial states 1,2,3 and 5. Replacing these by absorbing states gives the following stochastic game.



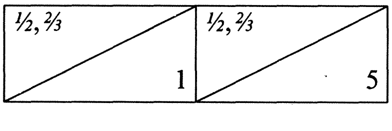
State 1



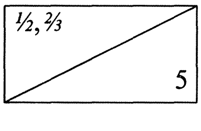
State 2



State 3



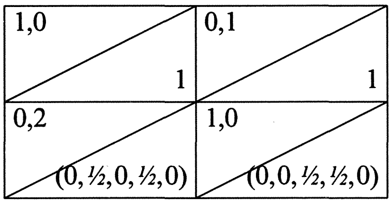
State 4



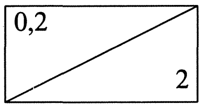
State 5

The unique stationary β-discounted equilibria for this stochastic game are given by $y_4^\beta=(1,0)$. So for this new stochastic game $S^{I_0}=\{1,2,3,5\}$ and $T=\{4\}$ and the condition of theorem 3.2.13 is fulfilled; hence we can achieve a limiting average ϵ-equilibrium for this stochastic game. This ϵ-equilibrium induces an ϵ-equilibrium for the original stochastic game.

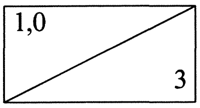
3.2.16 EXAMPLE



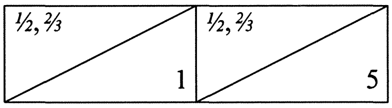
State 1



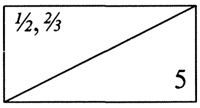
State 2



State 3



State 4



State 5

It can be verified that for this stochastic game stationary β-discounted equilibria are for instance:

$$(x^\beta,y^\beta) = ((((2-\beta)/(3-2\beta),(1-\beta)/(3-2\beta)),1,1,1,1),((½,½),1,1,(½,½),1)).$$

We find:

$$\gamma^1_\beta(x^\beta,y^\beta) = (½,0,1,½,½) \text{ and } \gamma^2_\beta(x^\beta,y^\beta) = (⅔,2,0,⅔,⅔) \text{ for all } \beta\in[0,1).$$

So $V^1 = (½,0,1,½,½)$ and $V^2 = (⅔,2,0,⅔,⅔)$.
Furthermore
$\gamma^1(x^1,y^1) = (½,0,1,½,½) = V^1$ and $\gamma^2(x^1,y^1) = (½,2,0,7/12,⅔) \lneq V^2$.
Hence $S^{I_0} = S^I = \{2,3,5\}$, $S^{I_2} = \{1\}$, $T = \{4\}$ and $I_1 = \emptyset$. Clearly the condition of theorem 3.2.13 is not fulfilled. The strong initial states we find are $\{2,3,5\}$. By the techniques in the proof of theorem 3.2.13 it is not possible to achieve an ϵ-equilibrium for initial state 1.
Iteration does not work either; replacing the strong initial states $\{2,3,5\}$ by absorbing states does not change the stochastic game situation, so in the second step we could again choose (x^β,y^β) the same as above.
Of course, by choosing $y^\beta_4 = (1,0)$ or $(0,1)$ one could establish an ϵ-equilibrium.

Chapter 4

Special classes of stochastic games

4.1 Introduction

In this chapter we discuss the impact of the results from the previous chapters on several special classes of stochastic games. Special classes of stochastic games are stochastic games with an additional property on the payoff and/or the transition structure.

The special classes we consider are: unichain stochastic games (section 4.2), stochastic games with state independent transitions (section 4.3) and repeated games with absorbing states (section 4.4).

4.2 Unichain stochastic games

4.2.1 Definition

A unichain stochastic game is a stochastic game with the property that, for any pair of stationary strategies (x,y)*, there is just one irreducible set of states.*

Unichain stochastic games were considered by Gillette [1957] and by Hoffman & Karp [1966] who proved that in the zero-sum case both players have stationary limiting average optimal strategies. Later Rogers [1969], Sobel [1971] and Federgruen [1978] independently showed that in the general-sum case there exist stationary limiting average equilibria. Those proofs are all based on continuity properties of $\gamma^k(x,y)$ on $X \times Y$ and they all use some fixed point theorem. Using results of chapter 2 we give new proofs for these facts.

4.2.2 Theorem

Let $\{(x^\beta,y^\beta):\beta\in[0,1)\}$ *be a sequence of stationary* β*-discounted equilibria in a general-sum unichain stochastic game and let* $(x^1,y^1)=\lim_{\beta\uparrow 1}(x^\beta,y^\beta)$.

Then (x^1,y^1) *is a stationary limiting average equilibrium.*

Proof:

Since we are dealing with a unichain stochastic game, there is just one ergodic set for any pair of stationary strategies (x,y). Hence each row of $Q(x,y)$ is equal to the unique stationary distribution for the related Markov chain. In view of lemma 1.5.5 (b) we conclude that $\gamma^k(s,x,y)=\gamma^k(t,x,y)$ for all $s,t\in S$ and $k=1,2$. Now suppose player 2 is using $y\in Y$ against x^1. By lemma 2.2.6 we have that $\gamma^2(s,x^1,y)=\lim_{\beta\uparrow 1}\gamma^2_\beta(s,x^\beta,y)\leqslant \lim_{\beta\uparrow 1}\gamma^2_\beta(s,x^\beta,y^\beta)=\gamma^2(s,x^1,y^1)$, for all $s \in S$. From lemma 1.6.2 it now follows that y^1 is a limiting average best reply for player 2 against x^1. Similarly it can be shown that x^1 is a limiting average best reply for player 1 against y^1. ■

The above theorem implies that for zero-sum unichain stochastic games the limiting average value exists and equals the limit of β-discounted values; moreover it follows that the limiting average value for initial states s and t is the same for all $s,t \in S$. Furthermore both players have stationary limiting average optimal strategies. This is formulated in the next theorem.

4.2.3 THEOREM

For a zero-sum unichain stochastic game the limiting average value v^1 exists and $v^1 = \lim_{\beta \uparrow 1} v^1_\beta$. Furthermore $v^1_s = v^1_t$ for all $s,t \in S$.

Let $\{x^\beta : \beta \in [0,1)\}$ $(\{y^\beta : \beta \in [0,1)\})$ be a sequence of stationary β-discounted optimal strategies for player 1 (2) and let $x^1 = \lim_{\beta \uparrow 1} x^\beta$ $(y^1 = \lim_{\beta \uparrow 1} y^\beta)$.

Then x^1 (y^1) is a stationary limiting average optimal strategy for player 1 (2).

PROOF:

It is easy to verify that for a zero-sum stochastic game a pair of stationary strategies (x,y) is an equilibrium if and only if x is optimal for player 1 and y is optimal for player 2. Since by theorem 4.2.3 the pair of strategies (x^1,y^1) is a limiting average equilibrium and since $\gamma^1(x^1,y^1) = \lim_{\beta \uparrow 1} \gamma^1_\beta(x^\beta,y^\beta) = \lim_{\beta \uparrow 1} v^1_\beta$ is independent of the initial state, the proof is complete. ■

4.3 STOCHASTIC GAMES WITH STATE INDEPENDENT TRANSITIONS

4.3.1 DEFINITION

A stochastic game with state independent transitions (SIT) is a stochastic game for which there are $m,n \in \mathbb{N}$ such that $m_s = m$ and $n_s = n$ for all $s \in S$ and for which furthermore $p(s,i,j) = p(t,i,j)$ for all $s,t \in S$ and all i,j.

A stochastic game with state independent transitions and separable rewards (SER-SIT) is a SIT stochastic game with the additional property that there are $c^k : S \to \mathbb{R}$ and $a^k : \{1,2,...,m\} \times \{1,2,...,n\} \to \mathbb{R}$, for $k = 1,2$, such that $r^k(s,i,j) = c^k(s) + a^k(i,j)$ for all s,i,j and $k = 1,2$.

An early appearance of the SER-SIT conditions can be found in Sobel [1981]. As a class of games, SER-SIT stochastic games were introduced by Parthasarathy et al. [1984]. They showed, among other results, that for this class of stochastic games: in the zero-sum case the limiting average value is independent of the initial state and both players have state independent stationary limiting average optimal strategies; in the general-sum case there exists a state independent stationary limiting average equilibrium. In this section we derive some results for SIT stochastic games, without using the SER-property.

4.3.2 THEOREM

For any zero-sum SIT stochastic game the limiting average value v^1 is independent of the initial state: $v_s^1 = v_t^1$ for all $s,t \in S$.
Furthermore both players have stationary limiting average optimal strategies.

PROOF:
Let $\{x^\beta : \beta \in [0,1)\}$ be a sequence of stationary β-discounted optimal strategies for player 1 and let $x^1 = \lim_{\beta \uparrow 1} x^\beta$.
Let y^* be a stationary limiting average best reply for player 2 against x^1.
Let S^* be the set of states s with $\gamma^1(s,x^1,y^*) = v_s^1 = \max_{t \in S} v_t^1$.
By the proof of theorem 2.4.2 we have that $S^* \neq \emptyset$.
Now observe that $p(t|s,x_s^1,y_s) = 0$ for all $s \in S^*$, $t \in S \setminus S^*$ and any $y \in Y$. This can be seen by the following argument. Suppose there were $s \in S^*$, $t \in S \setminus S^*$ and $y \in Y$ such that $p(t|s,x_s^1,y_s) > 0$; let g be the Markov strategy (definition 1.3.2) defined by using y at stage 1 and y^* at all stages $n \geqslant 2$; then it follows that $\gamma^1(s,x^1,g) < v_s^1$, which contradicts the optimality of x^1 in s.
Hence, if player 1 uses x^1 then the play will never leave the set of states S^*, once it has been reached.
Take $s^* \in S^*$ and define x^* by $x_s^* := x_s^1$ for $s \in S^*$ and $x_s^* := x_{s^*}^1$ for $s \in S \setminus S^*$.
Now, for any initial state, if player 1 uses x^*, then after 1 stage the play will be in S^* with probability 1 because we have state independent transitions. Since for $s \in S^*$ the strategies x^1 and x^* are equal, the play will remain in S^* forever. Hence we have that for any initial state s and any $y \in Y$:

$$\gamma^1(s,x^*,y) \geqslant \gamma^1(s^*,x^1,y^*) = v_{s^*}^1 = \max_{t \in S} v_t^1.$$

This implies that $v_s^1 = v_t^1$ for all $s,t \in S$ and x^* is limiting average optimal. A limiting average optimal strategy for player 2 can be derived analogously. ■

4.3.3 THEOREM

For every general-sum SIT stochastic game there exists an almost stationary limiting average ϵ-equilibrium.

PROOF:
Let $\{(x^\beta,y^\beta) : \beta \in [0,1)\}$ be a sequence of stationary β-discounted equilibria, converging to (x^1,y^1).
By lemma 2.3.6 there exists $h^* \in \{1,2,...,H\}$ (defined as in definition 2.2.1) such that $\gamma^{1h^*}(x^1,y^1) \geqslant \max_h V^{1h} \geqslant V^{1h^*}$ and $\gamma^{2h^*}(x^1,y^1) \geqslant V^{2h^*}$.
Take some $s^* \in S^{h^*}$. Then $p(t|s,x_{s^*}^1,y_{s^*}^1) = 0$ for $s \in S$, $t \in S \setminus S^{h^*}$, by the state independent transitions and by the irreducibility of S^{h^*}. Define x^* by $x_s^* := x_s^1$ for $s \in S^{h^*}$ and $x_s^* := x_{s^*}^1$ for $s \in S \setminus S^{h^*}$. Similarly define y^* by $y_s^* := y_s^1$ for $s \in S^{h^*}$ and $y_s^* := y_{s^*}^1$ for $s \in S \setminus S^{h^*}$. Then $\gamma^k(s,x^*,y^*) \geqslant V^{kh^*}$ for $k = 1$ as well as for $k = 2$ and for any initial state s, since for any initial state the play will be in S^{h^*} after at most one stage.
Now using the fact that the limiting average value is independent of the initial

state and using arguments similar to those in the proof of lemma 2.3.4, one can obtain an almost stationary limiting average ϵ-equilibrium ($\epsilon>0$). ■

4.4 Repeated games with absorbing states

4.4.1 Definition

In a stochastic game a state $s \in S$ is called an absorbing state if $p(s|s,i,j)=1$ for all $i \in \{1,2,...,m_s\}$ and all $j \in \{1,2,...,n_s\}$. A repeated game with absorbing states is a stochastic game for which all states except one are absorbing states.

In this monograph we have already seen several examples of repeated games with absorbing states: 1.6.5, 1.7.4, 1.7.6, 1.8.6, 2.3.8, 2.4.5, 2.4.12, 3.2.3, 3.2.4. The class of zero-sum repeated games with absorbing states was first examined by Kohlberg [1974], who extended the work of Blackwell & Ferguson [1968] on the big match (cf. example 1.7.4).

Kohlberg [1974] showed that the limiting average value exists for any zero-sum repeated game with absorbing states. Using the techniques of Kohlberg [1974] and inspired by Sorin [1986], example 1.8.6, Vrieze & Thuijsman [1989] showed the existence of ϵ-equilibria for general-sum repeated games with absorbing states. In their proof Vrieze & Thuijsman construct ϵ-equilibrium strategies for which it may occur that one of the players has to adjust the mixed action he uses in the non-absorbing state at all stages. Since for a repeated game with absorbing states the condition of theorem 3.2.13 is necessarily fulfilled, the next theorem follows immediately.

4.4.2 Theorem

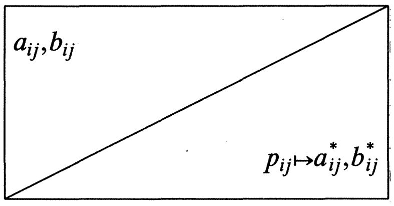
For any general-sum repeated game with absorbing states there exists an almost stationary limiting average ϵ-equilibrium for each $\epsilon>0$.

Observe that this theorem provides ϵ-equilibrium strategies which are different from those in Vrieze & Thuijsman [1989]. Observe also that for several other classes of stochastic games, e.g. unichain stochastic games, single-loop stochastic games (cf. Filar [1981-a]), the condition of theorem 3.2.13 is automatically fulfilled, which immediately provides the existence of almost stationary limiting average equilibria for such games. However, as for the unichain stochastic games, several results can be derived more explicitly for repeated games with absorbing states.

In this section we make the assumption that all absorbing states are of size 1×1. With respect to the existence of limiting average ϵ-equilibria this assumption can be made without loss of generality, since for each of the absorbing states the existence of a stationary limiting average equilibrium is obvious (cf. theorem 1.8.1). Hence each of the absorbing states can be contracted to a 1×1 absorbing state with payoffs according to some equilibrium.
We introduce the following notation.

4.4.3 Notation

A repeated game with absorbing states can be given by one $m \times n$-matrix of which entry (i,j) is



where a_{ij}, b_{ij}, a^*_{ij}, $b^*_{ij} \in \mathbb{R}$ and $p_{ij} \in [0,1]$.

The interpretation of the above notation is the following. If in the initial state, the non-absorbing one, player 1 chooses action i and player 2 chooses action j, then player 1 (2) receives a_{ij} (b_{ij}) and with probability p_{ij} a transition takes place to an absorbing state where the players get a^*_{ij}, b^*_{ij} from that stage on; with probability $1-p_{ij}$ the play remains in the initial state, at least until next stage. In this framework a stationary strategy for player 1 is simply some $x \in \Delta^m$; for player 2 stationary strategies are elements $y \in \Delta^n$.

4.4.4 Lemma

For a pair of stationary strategies (x,y) we have:

a) $$\gamma^1_\beta(x,y) = \frac{(1-\beta)\sum\limits_i \sum\limits_j x_i a_{ij} y_j + \beta \sum\limits_i \sum\limits_j x_i p_{ij} a^*_{ij} y_j}{1-\beta+\beta \sum\limits_i \sum\limits_j x_i p_{ij} y_j} \quad \textit{for } \beta \in [0,1).$$

b) $$\gamma^1(x,y) = \lim_{\beta \uparrow 1} \frac{(1-\beta)\sum\limits_i \sum\limits_j x_i a_{ij} y_j + \beta \sum\limits_i \sum\limits_j x_i p_{ij} a^*_{ij} y_j}{1-\beta+\beta \sum\limits_i \sum\limits_j x_i p_{ij} y_j}.$$

c) *Similar statements hold for player 2's rewards.*

Proof:

It is clear that (b) follows from (a) by lemma 1.5.5 (d) and that (c) needs no further comment.

The formula in (a) follows directly from (cf. lemma 1.5.3 (c)):

$$\gamma^1_\beta(x,y) = (1-\beta)\sum_i \sum_j x_i a_{ij} y_j + \beta \sum_i \sum_j x_i p_{ij} a^*_{ij} y_j + \beta \sum_i \sum_j x_i (1-p_{ij}) y_j \gamma^1_\beta(x,y). \quad \blacksquare$$

4.4.5 Definition

A pair of stationary strategies (x,y) is called absorbing if $\sum\limits_i \sum\limits_j x_i p_{ij} y_j > 0$.

The next two lemmas can be derived directly from lemma 4.4.4.

4.4.6 Lemma

a) *If* $(x,y)\in\Delta^m\times\Delta^n$ *is absorbing, then:*

$$\gamma^1(x,y)=\frac{\sum_i\sum_j x_i p_{ij} a^*_{ij} y_j}{\sum_i\sum_j x_i p_{ij} y_j} \text{ and } \gamma^2(x,y)=\frac{\sum_i\sum_j x_i p_{ij} b^*_{ij} y_j}{\sum_i\sum_j x_i p_{ij} y_j}.$$

b) *If* $(x,y)\in\Delta^m\times\Delta^n$ *is non-absorbing, then:*

$$\gamma^1(x,y)=\sum_i\sum_j x_i a_{ij} y_j \text{ and } \gamma^2(x,y)=\sum_i\sum_j x_i b_{ij} y_j.$$

4.4.7 Lemma

Let $\{(x^\beta,y^\beta)\in\Delta^m\times\Delta^n:\beta\in[0,1)\}$ *be a converging sequence with* $(x^1,y^1)=\lim_{\beta\uparrow 1}(x^\beta,y^\beta)$ *and with* $Car(x^\beta)$ *and* $Car(y^\beta)$ *independent of* $\beta\in[0,1)$.

a) *If* (x^1,y^1) *is absorbing, then* (x^β,y^β) *is absorbing for* $\beta\in[0,1)$ *and* $\gamma^k(x^1,y^1)=\lim_{\beta\uparrow 1}\gamma^k_\beta(x^\beta,y^\beta)$ *for* $k=1,2$.

b) *If* (x^β,y^β) *is non-absorbing for* $\beta\in[0,1)$, *then* (x^1,y^1) *is non-absorbing and* $\gamma^k(x^1,y^1)=\lim_{\beta\uparrow 1}\gamma^k_\beta(x^\beta,y^\beta)$ *for* $k=1,2$.

4.4.8 Remark

For the remainder of this section let $\{(x^\beta,y^\beta)\in\Delta^m\times\Delta^n:\beta\in[0,1)\}$ *be a converging sequence of stationary* β*-discounted equilibria with* $(x^1,y^1)=\lim_{\beta\uparrow 1}(x^\beta,y^\beta)$ *and with* $Car(x^\beta)$ *and* $Car(y^\beta)$ *independent of* $\beta\in[0,1)$.

Let $V^k=\lim_{\beta\uparrow 1}\gamma^k(x^\beta,y^\beta)$ *for* $k=1,2$.

By lemma 1.6.4 we have the following fact.

4.4.9 Lemma

a) *If* $x\in\Delta^m$ *and* $Car(x)\subset Car(x^\beta)$, $\beta\in[0,1)$, *then* $\gamma^1_\beta(x,y^\beta)=\gamma^1_\beta(x^1,y^\beta)=\gamma^1_\beta(x^\beta,y^\beta)$.

b) *If* $y\in\Delta^n$ *and* $Car(y)\subset Car(y^\beta)$, $\beta\in[0,1)$, *then* $\gamma^2_\beta(x^\beta,y)=\gamma^2_\beta(x^\beta,y^1)=\gamma^2_\beta(x^\beta,y^\beta)$.

4.4.10 Lemma

a) *If* $x\in\Delta^m$ *and either* (x,y^1) *is absorbing or* (x,y^β) *is non-absorbing, then* $\gamma^1(x,y^1)\leqslant V^1$.

b) *If* $y\in\Delta^n$ *and either* (x^1,y) *is absorbing or* (x^β,y) *is non-absorbing, then* $\gamma^2(x^1,y)\leqslant V^2$.

Proof:

Using lemma 4.4.7 and using that the pairs of strategies (x^β,y^β) are β-discounted equilibria, we have (a) by:

$$\gamma^1(x,y^1)=\lim_{\beta\uparrow 1}\gamma^1_\beta(x,y^\beta)\leqslant\lim_{\beta\uparrow 1}\gamma^1_\beta(x^\beta,y^\beta)=V^1. \quad\blacksquare$$

4.4.11 Lemma
If $\gamma^k(x^1,y^1) \geqslant V^k$ for $k=1,2$, then (x^1,y^1) can be supplemented with retaliation threats to obtain an almost stationary ϵ-equilibrium.

Proof:
Observe that by lemma 4.4.10 neither player 1 nor player 2 can profit by an absorbing deviation from (x^1,y^1). Hence retaliation threats are only needed to counter non-absorbing deviations. As in the proof of lemma 2.3.4, let $Y^{(n)}$ be the random variable (vector) denoting the action frequencies of player 2 in the initial state up to stage n and let $y^{(n)}$ be a realization of $Y^{(n)}$. Let $X^{(n)}$ and $x^{(n)}$ be defined similarly for the action frequencies of player 1.
Let $\epsilon>0$. Then, pretending absorption does not take place, for each $\alpha>0$ there is $\delta>0$ and $N_{\alpha\delta} \in \mathbb{N}$ such that:

$$\text{Prob}_{x^1,y^1}\{\|X^{(n)}-x^1\|>\alpha \text{ for any } n \geqslant N_{\alpha\delta}\}<\delta \text{ and}$$

$$\text{Prob}_{x^1,y^1}\{\|Y^{(n)}-y^1\|>\alpha \text{ for any } n \geqslant N_{\alpha\delta}\}<\delta.$$

Choose $\alpha \in (0,\epsilon/8M)$ and choose $\delta>0$ such that:

$(1-\delta)^2(\gamma^k(x^1,y^1)-2\alpha M)-(1-(1-\delta)^2)M \geqslant \gamma^k(x^1,y^1)-\epsilon/2$ for $k=1,2$.

Define π_ϵ^* by:
a) use x^1 unless
   i) player 2 chooses $j \notin Car(y^1)$, or
   ii) $\|y^{(n)}-y^1\|>\alpha$ for some $n \geqslant N_{\alpha\delta}$
b) if (i) or (ii) occurs, use some retaliation strategy $\pi_{\epsilon/2}^r$ from that stage on.
Define σ_ϵ^* analogously.
It can be verified that $(\pi_\epsilon^*,\sigma_\epsilon^*)$ is an almost stationary limiting average ϵ-equilibrium. ■

The next lemma follows directly from lemma 4.4.7 and lemma 4.4.9.

4.4.12 Lemma
*If $\gamma^2(x^1,y^1)<V^2$,
then (x^1,y^1) is non-absorbing whereas (x^β,y^1) is absorbing for all $\beta \in [0,1)$.*

4.4.13 Definition
If $\gamma^2(x^1,y^1)<V^2$, then we define:
a) $\tilde{x}^\beta$ *and* $\tilde{x}^{*\beta} \in \mathbb{R}^m$ *by* $(i \in \{1,2,\ldots,m\})$:

$$\tilde{x}_i^\beta := \begin{cases} x_i^\beta & \text{if } (i,y^1) \textit{ is non-absorbing} \\ 0 & \text{if } (i,y^1) \textit{ is absorbing} \end{cases}$$

$$\tilde{x}_i^{*\beta} := \begin{cases} 0 & \text{if } (i,y^1) \textit{ is non-absorbing} \\ x_i^\beta & \text{if } (i,y^1) \textit{ is absorbing} \end{cases}$$

b) $\underline{x}^\beta$ *and* $\underline{x}^{*\beta} \in \Delta^m$ *by* $(i \in \{1,2,\ldots,m\})$: $\underline{x}_i^\beta := \dfrac{\tilde{x}_i^\beta}{\sum_i \tilde{x}_i^\beta}$ *and* $\underline{x}_i^{*\beta} := \dfrac{\tilde{x}_i^{*\beta}}{\sum_i \tilde{x}_i^{*\beta}}$.

c) $x^* := \lim\limits_{\beta\uparrow 1\ -} x^{*\beta}$ *(which limit exists without loss of generality).*

d) $\mu^\beta := \dfrac{1-\beta}{1-\beta+\beta\sum\limits_i\sum\limits_j x_i^\beta p_{ij} y_j^1}$ *and* $\mu^1 := \lim\limits_{\beta\uparrow 1} \mu^\beta$ *(without loss of generality).*

Observe that by these definitions we have that $x^\beta = \tilde{x}^\beta + \tilde{x}^{*\beta}$; both $\tilde{x}^\beta \neq 0$ and $\tilde{x}^{*\beta} \neq 0$ for all β; $\lim\limits_{\beta\uparrow 1} \tilde{x}^{*\beta} = 0$; $\lim\limits_{\beta\uparrow 1} \tilde{x}^\beta = \lim\limits_{\beta\uparrow 1\ -} x^\beta = \lim\limits_{\beta\uparrow 1} x^\beta = x^1$.
It is also clear that for all β we have $\mu^\beta \in [0,1]$ and $\mu^1 \in [0,1]$.

4.4.14 LEMMA
If $\gamma^2(x^1,y^1) < V^2$ *and* μ^1 *and* x^* *are as above,*
then $V^2 = \mu^1\gamma^2(x^1,y^1) + (1-\mu^1)\gamma^2(x^*,y^1)$ *and hence* $\gamma^2(x^*,y^1) \geqslant V^2$.

PROOF:
From definition 4.4.13 observe that $\sum\limits_i\sum\limits_j x_i^\beta p_{ij} y_j^1 = \sum\limits_i\sum\limits_j \tilde{x}_i^{*\beta} p_{ij} y_j^1$ and $\sum\limits_i\sum\limits_j x_i^\beta p_{ij} b_{ij}^* y_j^1 = \sum\limits_i\sum\limits_j \tilde{x}_i^{*\beta} p_{ij} b_{ij}^* y_j^1$. Using this, definition 4.4.13, lemma 4.4.9 and lemma 4.4.4 we obtain:

$$
\begin{aligned}
V^2 &= \lim_{\beta\uparrow 1} \gamma_\beta^2(x^\beta,y^1) \\
&= \mu^1 \sum_i \sum_j x_i^1 b_{ij} y_j^1 + \lim_{\beta\uparrow 1} \frac{\beta\mu^\beta}{1-\beta} \sum_i \sum_j x_i^\beta p_{ij} b_{ij}^* y_j^1 \\
&= \mu^1\, \gamma^2(x^1,y^1) + \lim_{\beta\uparrow 1} \frac{\beta\mu^\beta}{1-\beta} \Big(\sum_i \sum_j x_i^\beta p_{ij} y_j^1\Big) \frac{\sum\limits_i\sum\limits_j \tilde{x}_i^{*\beta} p_{ij} b_{ij}^* y_j^1}{\sum\limits_i\sum\limits_j \tilde{x}_i^{*\beta} p_{ij} y_j^1} \\
&= \mu^1\, \gamma^2(x^1,y^1) + \lim_{\beta\uparrow 1} (1-\mu^\beta) \frac{\sum\limits_i\sum\limits_j \underline{x}_i^{*\beta} p_{ij} b_{ij}^* y_j^1}{\sum\limits_i\sum\limits_j \underline{x}_i^{*\beta} p_{ij} y_j^1} \\
&= \mu^1\, \gamma^2(x^1,y^1) + (1-\mu^1)\gamma^2(x^*,y^1).
\end{aligned}
$$

Since $\gamma^2(x^1,y^1) < V^2$ and $\mu^1 \in [0,1]$ we have that $\gamma^2(x^*,y^1) \geqslant V^2$. ■

4.4.15 LEMMA
If $\gamma^2(x^1,y^1) < V^2$ *and* $\epsilon > 0$, *then* (x^λ,y^1), *with* $x^\lambda := (1-\lambda)x^1 + \lambda x^*$ *and* $\lambda \in (0,1)$, *can be supplemented with retaliation threats to yield an almost stationary limiting average* ϵ*-equilibrium for* λ *sufficiently small.*

PROOF:
For each $\lambda \in (0,1)$ we have $\gamma^2(x^\lambda,y^1) = \gamma^2(x^*,y^1) \geqslant V^2$ and by lemmas 4.4.7 and 4.4.9: $\gamma^1(x^\lambda,y^1) = \gamma^1(x^*,y^1) = \lim\limits_{\beta\uparrow 1} \gamma^1(x^*,y^\beta) = \lim\limits_{\beta\uparrow 1} \gamma^1(x^\beta,y^\beta) = V^1$.

Let again $Y^{(n)}$ be the random variable denoting the action frequencies of player 2 in the initial state up to stage n and let $y^{(n)}$ be a realization of $Y^{(n)}$. Let $X^{(n)}$ be the random variable denoting the action frequencies of player 1 within $Car(x^1)$ in the initial state up to stage n. Let $x^{(n)}$ be a realization of $X^{(n)}$. Let $\epsilon>0$. Then, pretending absorption does not take place, for each $\alpha>0$ and $\delta>0$ there is $N_{\alpha\delta} \in \mathbb{N}$ such that:

$$\text{Prob}_{x^\lambda,y^1} \{\|X^{(n)}-x^1\| > \alpha \text{ for any } n \geqslant N_{\alpha\delta}\} < \delta \text{ and}$$

$$\text{Prob}_{x^\lambda,y^1} \{\|Y^{(n)}-y^1\| > \alpha \text{ for any } n \geqslant N_{\alpha\delta}\} < \delta.$$

Choose $\alpha \in (0,\epsilon/8M)$ and choose δ such that for $k=1,2$:

$$(1-\delta)^4(\gamma^k(x^*,y^1)-2\alpha M)-(1-(1-\delta)^4)M \geqslant \gamma^k(x^1,y^1)-\epsilon/2.$$

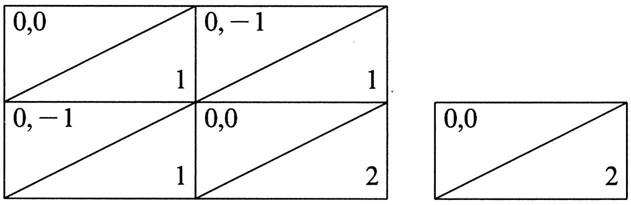
Choose $\lambda \in (0,\epsilon/8M)$ such that $\text{Prob}_{x^\lambda,y^1}\{\text{absorption before stage } N_{\alpha\delta}\} < \delta$. Choose $N_\lambda \in \mathbb{N}$, $N_\lambda > N_{\alpha\delta}$, such that
$\text{Prob}_{x^\lambda,y^1}\{$ absorption before stage $N_\lambda\} \geqslant 1-\delta$.
Define π_ϵ^* by:

a) use x^λ unless
   i) player 2 chooses $j \notin Car(y^1)$, or
   ii) $\|y^{(n)}-y^1\| > \alpha$ for some $n \geqslant N_{\alpha\delta}$
   iii) at stage N_λ play is still in the initial state
b) if (i), (ii) or (iii) occurs, then use some retaliation strategy $\pi_{\epsilon/2}^r$ (cf. 1.8.5).

Define σ_ϵ^* in a similar way.
Now $(\pi_\epsilon^*,\sigma_\epsilon^*)$ is an almost stationary limiting average ϵ-equilibrium. ■

Observe that example 3.2.4 is an illustration of the lemmas 4.4.14 and 4.4.15; in that example $\mu^1=2/3$. The next example illustrates that one may also have $\gamma^2(x^1,y^1) < V^2$ and $\mu^1=0$.

4.4.16 Example



State 1 State 2

For this example stationary β-discounted equilibria are for instance given by $(x^\beta,y^\beta)=((\frac{1-\sqrt{1-\beta}}{\beta}, \frac{-1+\beta+\sqrt{1-\beta}}{\beta}), (0,1))$, with $\beta \in [0,1)$.
Now $\gamma^2(x^1,y^1)=-1$, whereas $V^2=\lim_{\beta\uparrow 1}\gamma_\beta^2(x^\beta,y^\beta)=\lim_{\beta\uparrow 1}\frac{1-\beta-\sqrt{1-\beta}}{\beta}=0$.
Hence $x^*=(0,1)$ and $\gamma^2(x^*,y^1)=0$; $\mu^1=0$.

Examples that illustrate lemma 4.4.11 are for instance 2.3.8 and 3.2.3.

Chapter 5

The total reward criterion in zero-sum stochastic games

5.1 Introduction

In chapter 1 we briefly dealt with the total reward criterion. This criterion has been introduced in Thuijsman & Vrieze [1987] and Vrieze & Thuijsman [1987]. Based on these papers, we examine this criterion as well as its relations with the β-discounted reward criterion and with the limiting average reward criterion for the zero-sum case.

In section 5.2 we give several examples to support our choice for defining total rewards by $\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{s\pi\sigma}\,[R^1(n)]$, as is done in definition 1.4.4. Furthermore these examples lead to the conclusion that a certain property should be fulfilled if we want that the total value, whenever it exists, is finite. This property is that the limiting average value should be 0 (for all initial states) and both players should have stationary limiting average optimal strategies.

For the existence of stationary limiting average optimal strategies, several characterizations have been given (cf. Sobel [1971], Bewley & Kohlberg [1978], Filar & Schultz [1986], Schultz [1987], Vrieze [1987-a]). Several of these characterizations are by means of mathematical programs (see chapter 6). By means of equations similar to the Shapley-equation (cf. 1.7.3), existence of stationary limiting average optimal strategies has been characterized by Vrieze [1987-a].

In section 5.3 we extend Vrieze's result by characterizing existence of stationary total optimal strategies for stochastic games with above property.

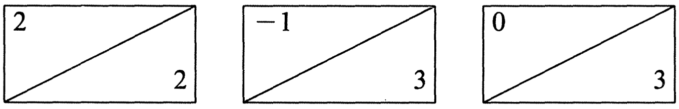
In section 5.4 we give an example to illustrate that, even with above property, history dependent strategies are indispensable for total ϵ-optimal play. This indicates that the total reward criterion and the limiting average criterion have similar features.

In section 5.5 we show that examining total rewards in a stochastic game is equivalent to examining limiting average rewards in a related stochastic game with countable state space.

5.2 The total reward criterion

For the β-discounted reward criterion the emphasis is on near-future payoffs, while for the limiting average reward criterion the emphasis is on far-future payoffs. In certain situations however, it would be more appropriate to use an intermediate criterion, where both near-future and far-future payoffs are equally important. There are several ways of defining such an intermediate criterion. Recently Krass et al. [1987] and Filar & Vrieze [1989] examined a criterion which is a 'weighted combination' of the β-discounted reward criterion and the limiting average reward criterion. In this chapter however we look at the total reward criterion, which is also an intermediate criterion. To see why this total reward criterion is interesting, think for instance of stopping stochastic games, as introduced in Shapley [1953], where for all s,i and j there is a small probability of stopping, or equivalently, a small probability of moving to an absorbing state where the payoffs are 0. In such games the limiting average reward is necessarily equal to 0 for any pair of strategies and hence the limiting average criterion does not seem to be a suitable criterion to examine the game. Instead of discounting with some factor $\beta \in [0,1)$ one would simply like to take the sum of all payoffs for any play in such a game. The same holds for the next example although it is no stopping stochastic game.

5.2.1 Example

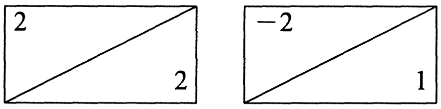


State 1 State 2 State 3

The payoffs in this example, like those in the other examples in this section, are the payoffs to player 1 to be paid by player 2.
For this example it is clear that the limiting average reward (for the unique strategies) is 0 for all initial states. However, it seems reasonable that player 1 would prefer to start in state 1, whereas player 2 would prefer initial state 2.

In the following example we can also imagine that player 1 would prefer to start in state 1 while player 2 would prefer to start in state 2. The limiting average reward is again 0 for both initial states.

5.2.2 Example



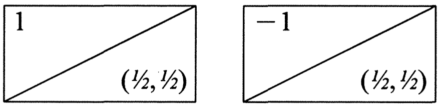
State 1 State 2

For the play starting in state 1 the sequence of partial sums of payoffs would be (2,0,2,0,2,....). Thus, on the average, player 1 owns 1, whereas it is clear that the limiting average reward is 0 for initial state 1. Similarly, for the play starting in state 2, player 1 would own -1 on the average. The β-discounted reward for initial state 1 is $2(1-\beta)\sum_{n=1}^{\infty}(-\beta)^{n-1} = 2(1-\beta)/(1+\beta)$. Leaving out normalization by $1-\beta$ gives $2/(1+\beta)$, which converges to 1 as β goes to 1.

These examples and the fact that, if $\sum_{n=1}^{\infty} E_{s\pi\sigma}[R^1(n)]$ exists in $\mathbb{R}\cup\{-\infty,+\infty\}$ then it is equal to $\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{s\pi\sigma}[R^1(n)]$, lead us to define total rewards by the latter expression (cf. definition 1.4.4). Here we use 'lim inf' since for non-stationary strategies 'lim' may fail to exist in $\mathbb{R}\cup\{-\infty,+\infty\}$. We chose 'lim inf' since, like for the limiting average criterion, this reflects a kind of 'worst case view' of player 1. However we could also have chosen 'lim sup' or any convex combination of 'lim inf' and 'lim sup'.

In lemma 1.5.7 we have shown that for any pair of stationary strategies $\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{sxy}[R^1(n)]$ is finite on condition that $\gamma^1(x,y)=0$. It does not seem to make sense to define a total reward evaluation by $E_{s\pi\sigma}$ $[\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} R^1(n)]$, since the following example, which was communicated to us by Neyman [1986], shows that this alternative definition does not express what we would like to call a total reward.

5.2.3 Example



State 1 State 2

For this example we have that $E[\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} R^1(n)]=-\infty$ for both initial states, since with probability 1, for any realization of the random walk, $\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} r^1(n)= -\infty$. Observe that $\gamma^1(x,y)=0$ for all x,y.

We now turn to some properties of the total reward criterion.

5.2.4 Theorem

For any pair of strategies $(\pi,\sigma)\in\Pi\times\Sigma$ we have:

$$\liminf_{\beta\uparrow 1} (1-\beta)^{-1}\gamma^1_\beta(\pi,\sigma) \geqslant \gamma^1_T(\pi,\sigma).$$

Proof:

Let $\{a_n\in\mathbb{R} : n\in\mathbb{N}\}$ and $C\in\mathbb{N}$ such that $|a_{n+1}-a_n|<C$ for all $n\in\mathbb{N}$. Then:

$$\liminf_{\beta\uparrow 1}(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m \geqslant \liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} a_m .$$

We now prove this inequality using that for each $\beta\in[0,1)$:

$$(1-\beta)^{-1}\sum_{m=1}^{\infty}\beta^{m-1}\,a_m=\sum_{m=1}^{\infty}\beta^{m-1}\sum_{n=1}^{m}a_n$$

(this is easy to verify, cf. page 37 in Kallenberg [1983]).
From this equation we derive

$$(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m\;=\;(1-\beta)^2\sum_{m=1}^{\infty}\beta^{m-1}\sum_{n=1}^{m}a_n\;=\;(1-\beta)^2\sum_{m=1}^{\infty}m\beta^{m-1}(\frac{1}{m}\sum_{n=1}^{m}a_n).$$

First, notice that $(1-\beta)^2\sum_{m=1}^{\infty}m\beta^{m-1}=1$. Second, if either $\liminf_{\beta\uparrow 1}(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m=\infty$ or $\liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}a_m=-\infty$, then the inequality holds. Third, notice that $|\frac{1}{N}\sum_{m=1}^{N}a_m|\leqslant CN$ for all $N\in\mathbb{N}$. If $\liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}a_m\geqslant L\in\mathbb{R}$, then let $\epsilon>0$ and $N^*\in\mathbb{N}$ such that $\frac{1}{m}\sum_{n=1}^{m}a_n\geqslant L-\epsilon$ for all $m\geqslant N^*$. Now take $\beta^*\in(0,1)$ such that $C(1-\beta)^2\sum_{m=1}^{N^*}m^2(1-\beta)^{m-1}\leqslant\epsilon$ and $(L-\epsilon)(1-\beta)^2\sum_{m=N^*+1}^{\infty}m\beta^{m-1}\geqslant L-2\epsilon$ for all $\beta\in(\beta^*,1)$. Then for all $\beta\in(\beta^*,1)$:

$$\begin{aligned}(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m\;&=\;(1-\beta)^2\sum_{m=1}^{N^*}m\beta^{m-1}(\frac{1}{m}\sum_{n=1}^{m}a_n)\\&\quad+\;(1-\beta)^2\sum_{m=N^*+1}^{\infty}m\beta^{m-1}(\frac{1}{m}\sum_{n=1}^{m}a_n)\\&\geqslant\;-C(1-\beta)^2\sum_{m=1}^{N^*}m^2\beta^{m-1}\;+\;(L-\epsilon)(1-\beta)^2\sum_{m=N^*+1}^{\infty}m\beta^{m-1}\\&\geqslant\;L-3\epsilon.\end{aligned}$$

Since this can be done for any $\epsilon>0$, and since C and L are independent of ϵ, it follows that $\liminf_{\beta\uparrow 1}(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m\geqslant L$. This completes the proof of $\liminf_{\beta\uparrow 1}(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}a_m\geqslant\liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}a_m$. From this inequality we conclude that for all $s\in S$ and all $\pi,\sigma\in\Pi\times\Sigma$:

$$\begin{aligned}\liminf_{\beta\uparrow 1}\,(1-\beta)^{-1}\,\gamma^1_\beta(s,\pi,\sigma)&=\liminf_{\beta\uparrow 1}\;\sum_{n=1}^{\infty}\beta^{n-1}\;E_{s\pi\sigma}[R^1(n)]\\&=\liminf_{\beta\uparrow 1}\,(1-\beta)\sum_{m=1}^{\infty}\beta^{m-1}\sum_{n=1}^{m}E_{s\pi\sigma}[R^1(n)]\\&\geqslant\liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m}E_{s\pi\sigma}[R^1(n)]=\gamma^1_T(s,\pi,\sigma).\end{aligned}$$ ■

5.2.5 THEOREM

For a zero-sum stochastic game with limiting average value 0 (for all initial states) let $y^* \in Y$ be a stationary limiting average optimal strategy.
Then there exists a pure stationary total best reply against y^* for player 1 and $\gamma_T^1(s,\pi,y^*) < \infty$ for all $\pi \in \Pi$ and all $s \in S$.

PROOF:

Since there are only finitely many pure stationary strategies, we can assume (by taking some subsequence) that there is a pure stationary strategy x^* such that x^* is a β-discounted best reply against y^* for all β close to 1.
By lemma 1.5.5 and by theorem 1.7.7 we have $0 \geqslant \gamma^1(x^*,y^*) = \lim_{\beta\uparrow 1} \gamma_\beta^1(x^*,y^*) \geqslant \lim_{\beta\uparrow 1} v_\beta^1 = v^1 = 0$, and hence $\gamma^1(x^*,y^*) = 0$.
Now let $\pi \in \Pi$. Then by theorem 5.2.4 and lemma 1.5.7, we have for all $s \in S$:

$$\begin{aligned}
\gamma_T^1(s,\pi,y^*) &\leqslant \liminf_{\beta\uparrow 1} (1-\beta)^{-1}\, \gamma_\beta^1(s,\pi,y^*) \\
&\leqslant \liminf_{\beta\uparrow 1} (1-\beta)^{-1}\, \gamma_\beta^1(s,x^*,y^*) \\
&= \lim_{\beta\uparrow 1} (1-\beta)^{-1}\, \gamma_\beta^1(s,x^*,y^*) \\
&= \gamma_T^1(s,x^*,y^*) \\
&= (I - P(x^*,y^*) + Q(x^*,y^*))_s^{-1} r^k(x^*,y^*) < \infty
\end{aligned}$$
■

Closely related to theorem 5.2.5 is the next theorem.

5.2.6 THEOREM

If for a zero-sum stochastic game with limiting average value 0 there are strategies $x^* \in X$ and $y^* \in Y$ which are uniform β-discounted optimal (i.e. β-discounted optimal for all β close to 1), then the total value exists in $\mathbb{R}^z$, x^* and y^* are total optimal and $v_T^1 = \lim_{\beta\uparrow 1} (1-\beta)^{-1} v_\beta^1$.

PROOF:

From lemmas 1.5.5, 1.6.2 and from theorem 1.7.7 it follows that stationary uniform β-discounted optimal strategies are also limiting average optimal. Previously this result has been shown by Bewley & Kohlberg [1978]. Hence x^* and y^* are limiting average optimal and $\gamma^1(x^*,y^*) = 0$. Now the result follows from the proof of theorem 5.2.5, since x^* is and y^* are β-discounted best replies against each other for all β close to 1. ■

5.2.7 THEOREM

If for a zero-sum stochastic game both players have stationary total optimal strategies and v_T^1 is finite, then $v_T^1 = \lim_{\beta\uparrow 1} (1-\beta)^{-1} v_\beta^1$.

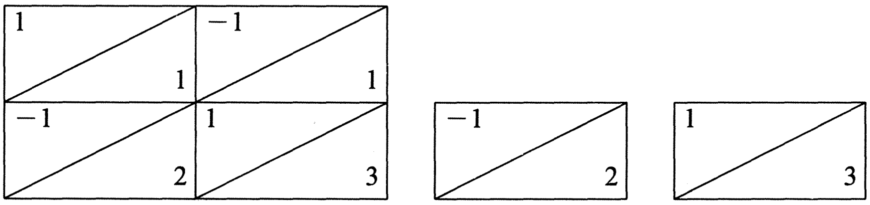
PROOF:

Let x^* be a stationary total optimal strategy for player 1. Since there are only finitely many pure stationary strategies, there is, using lemma 1.6.2, a pure stationary strategy y^p which is a β-discounted best reply for player 2 against x^* for all β close to 1.
Now let $\epsilon > 0$. Then for β close to 1 we have, using theorem 5.2.4:
$(1-\beta)^{-1}\, v_\beta^1 \geqslant (1-\beta)^{-1}\, \gamma_\beta^1(x^*,y^p) \geqslant \gamma_T^1(x^*,y^p) - \epsilon 1_z \geqslant v_T^1 - \epsilon 1_z$.
Similarly one can show that $(1-\beta)^{-1}\, v_\beta^1 \leqslant \gamma_T^1 + \epsilon 1_z$ for β close to 1. ■

Observe that the above theorems imply that, under the condions in the theorems, v_T^1 equals α_N, the coefficient for $(1-\beta)$ in the power series expansion of v_β^1 in fractional powers of $(1-\beta)$. So both the limiting average value and the total value appear in this expansion (cf. theorem 1.7.5).

It is evident that, if for a zero-sum stochastic game the limiting average value is not equal to 0 for some initial state s, then the total value $v_T^1(s)$ will either be $+\infty$ or $-\infty$. However, the existence of the total value in $\mathbb{R} \cup \{-\infty, +\infty\}$ for some initial state is not guaranteed by the limiting average value equalling 0 for that initial state. Take for instance the next example.

5.2.8 Example

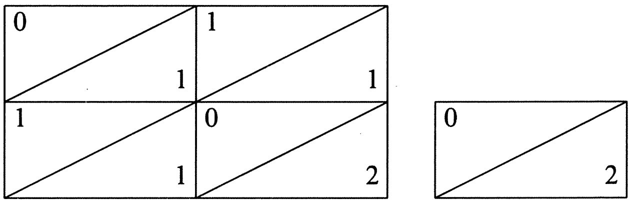


State 1 State 2 State 3

This example is essentially the big match of Blackwell & Ferguson [1968], example 1.7.4. The limiting average value for initial state 1 is 0 in this example. For this stochastic game player 1 has no limiting average optimal strategy (cf. 1.7.4. (c)). Hence for each strategy $\pi \in \Pi$ there is some $\sigma \in \Sigma$ such that $\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} E_{1\pi\sigma}[R^1(n)] < 0$ and thus for those strategies we find $\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{1\pi\sigma}\, [R^1(n)] = -\infty$. It is easy to verify that with $y^* = (½,½)$ we have that $\gamma_T^1(1,\pi,y^*) = 0$ for all $\pi \in \Pi$. Consequently, the total value does not exist for initial state 1 since we have $\sup_\pi \inf_\sigma \gamma_T^1(1,\pi,\sigma) = -\infty < 0 = \inf_\sigma \sup_\pi \gamma_T^1(1,\pi,\sigma)$.

The next example illustrates a curious phenomenon: even if the limiting average value equals 0 for all initial states and both players have limiting average optimal strategies, then the total value may still be infinite.

5.2.9 EXAMPLE



State 1 State 2

For this stochastic game the limiting average value is 0 for both initial states. For player 1 all strategies are limiting average optimal; for player 2 the stationary strategy $(1-\epsilon,\epsilon)$, with $\epsilon\in(0,1)$, is limiting average ϵ-optimal. However here, as for any other stochastic game with a state independent value, player 2 possesses a Markov strategy (cf. definition 1.3.2) that is limiting average optimal. Such a limiting average optimal Markov strategy can for instance be obtained by using:

a one-stage optimal strategy at the first stage, followed by
a two-stage optimal strategy at the next two stages, followed by
a three-stage optimal strategy at the next three stages, etc.

That such a strategy is optimal can be shown by using that $\lim\limits_{N\to\infty} \dfrac{v_s^{(N)}}{N} = v_s^1$ is independent of $s\in S$, where $v_s^{(N)}$ denotes the value of the N-stage game starting in s (cf. Mertens & Neyman [1981], Bewley & Kohlberg [1976, 1978] and Vrieze [1987-a]).

To show that the total value for initial state 1 is $+\infty$, observe that by using the stationary strategy $(1-\epsilon,\epsilon)$, with $\epsilon\in(0,1)$, player 1's total reward will be at least $1/\epsilon$ against any strategy of player 2. Hence, letting ϵ tend to 0 we find that, although player 2 can keep the limiting average reward at 0, the total reward value is $+\infty$.

It can even be verified for this example that player 1 can guarantee a total reward $+\infty$ by using the Markov strategy f^* defined by: if at stage n the play is still in state 1, then use the mixed action $(n/(n+1),1/(n+1))$, for all $n\in\mathbb{N}$. Now $\gamma_T^1(1,f^*,\sigma)=+\infty$ for any strategy $\sigma\in\Sigma$.

The above examples illustrate that only with the following property (P) we can expect the total value, if it exists, to be finite.

5.2.10 DEFINITION

We say that the stochastic game has property P if the limiting average value is 0 (for all initial states) and if, moreover, both players have stationary limiting average optimal strategies.

For a zero-sum stochastic game with property P, the total reward criterion can be seen as a refinement of the limiting average reward criterion, since it is obvious that for such a game a total (ϵ-)optimal strategy is necessarily limiting average optimal.

For the remainder of this chapter we focus on zero-sum stochastic games with property P.

5.3 Stochastic games and optimal stationary strategies

It is well-known that in any zero-sum stochastic game there exist stationary β-discounted optimal strategies (cf. theorem 1.7.3). Furthermore it is well-known that stationary limiting average optimal strategies do not always exist (cf. example 1.7.4). Vrieze [1987-a] gives the following characterization for the existence of stationary limiting average optimal strategies.

5.3.1 Theorem

For any zero-sum stochastic game both players have stationary limiting average optimal strategies if and only if there exist α, $\delta_1, \delta_2 \in \mathbb{R}^z$ such that for all $s \in S$:

a) $\alpha_s = \underset{A_s \times B_s}{val} \, [\sum_{t=1}^{z} p(t|s,i,j)\alpha_t]$

b) $\alpha_s + \delta_{1s} = \underset{O_{1s} \times B_s}{val} \, [r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{1t}]$

c) $\alpha_s + \delta_{2s} = \underset{A_s \times O_{2s}}{val} \, [r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{2t}]$.

Here $A_s = \{1,2,...,m_s\}$, $B_s = \{1,2,...,n_s\}$ and O_{1s} (resp. O_{2s}) is the set of optimal mixed actions for player 1 (resp. 2) in the matrix game $[\sum_{t=1}^{z} p(t|s,i,j)\alpha_t]_{i=1,j=1}^{m_s \ \ n_s}$.

Shapley & Snow [1950] showed that for each player the set of optimal mixed actions of a matrix game is a bounded polyhedron, with a finite number of extreme points. In view of this result $\underset{O_{1s} \times B_s}{val} \, [r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{1t}]$ is the value of a polyhedral game (cf. Wolfe [1956]), where for $(x,y) \in O_{1s} \times B_s$ the payoff equals $\sum_{i=1}^{m_s} \sum_{j=1}^{n_s} x_i(r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{1t})y_j$.

The expression $\underset{A_s \times O_{2s}}{val} \, [r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{2t}]$ should be interpreted in a similar way.

For every solution $(\alpha,\delta_1,\delta_2)$ of the equations in theorem 5.3.1 it can be verified that $\alpha = v^1$, the limiting average value of the stochastic game. Furthermore, given such a solution $(\alpha,\delta_1,\delta_2)$, a stationary limiting average optimal strategy x^* for player 1 can be constructed by letting x_s^* be an optimal mixed action for player 1 in the polyhedral game $[r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{1t}]_{O_{1s} \times B_s}$, for each $s \in S$. A stationary limiting average optimal strategy y^* for player 2 can be found in a similar way. An example in Vrieze [1987-a] shows that there may be stationary limiting average optimal strategies which cannot be found from any solution $(\alpha,\delta_1,\delta_2)$.

Observe that theorem 5.3.1 implies that a zero-sum stochastic game has property P (cf. definition 5.2.10) if and only if there exists $\delta\in\mathbb{R}^z$ such that:

$$\delta_s = \underset{A_s\times B_s}{val}\ [r^1(s,i,j)+\sum_{t=1}^{z} p(t|s,i,j)\delta_t], \text{ for each } s\in S.$$

We now give an analogue of theorem 5.3.1 for the existence of stationary total optimal strategies for stochastic games with property P.

5.3.2 THEOREM

For any zero-sum stochastic game with property P the total value exists in $\mathbb{R}^z$ and both players have stationary total optimal strategies if and only if there exist $\alpha,\delta_1,\delta_2\in\mathbb{R}^z$ and $\lambda\geqslant 0$ such that for all $s\in S$:

a) $\alpha_s = \underset{A_s\times B_s}{val}\ [r^1(s,i,j)+\sum_{t=1}^{z} p(t|s,i,j)\alpha_t]$

b) $\alpha_s+\delta_{1s} = \underset{O^*_{1s}\times B_s}{val}\ [\lambda r^1(s,i,j)+\sum_{t=1}^{z} p(t|s,i,j)\delta_{1t}]$

c) $\alpha_s+\delta_{2s} = \underset{A_s\times O^*_{2s}}{val}\ [\lambda r^1(s,i,j)+\sum_{t=1}^{z} p(t|s,i,j)\delta_{2t}]$.

Here O^*_{1s} and O^*_{2s} are the sets of optimal mixed actions for the respective players for the matrix game in (a), and the games in (b) and (c) are again polyhedral games.

PROOF:

THE 'IF'-PART:

Suppose there are $\alpha,\delta_1,\delta_2\in\mathbb{R}^z$ and $\lambda\geqslant 0$ such that for each $s\in S$ the statements (a), (b) and (c) hold. Let x^*_s be an optimal mixed action for player 1 in the polyhedral game $[\lambda r^1(s,i,j)+\sum_{t=1}^{z} p(t|s,i,j)\delta^1_t]_{O^*_{1s}\times B_s}$, for each $s\in S$ and let $y\in Y$.

From (a) we conclude that $\alpha\leqslant r^1(x^*,y)+P(x^*,y)\alpha$ and hence it follows that $0\leqslant Q(x^*,y)r^1(x^*,y)=\gamma^1(x^*,y)$, so x^* is limiting average optimal.

It is clear that if $\gamma^1(s,x^*,y)>0$ then $\gamma^1_T(s,x^*,y)=\infty>\alpha_s$.

So let $A\subset S$ be the set of states with $\gamma^1(s,x^*,y)=0$.

Then $p(t|s,x^*,y)=0$ for all $t\in A^c$ and $s\in A$. Let α^A, δ^A_1, $r^1(x^*,y)^A$, $P(x^*,y)^A$, $Q(x^*,y)^A$ be restrictions to states in A.

We have $0=\gamma^1(x^*,y)^A=Q(x^*,y)^A r^1(x^*,y)^A$.

From (b) we conclude that $\alpha^A+\delta^A_1\leqslant\lambda r^1(x^*,y)^A+P(x^*,y)^A\delta^A_1$ and hence it follows that $Q(x^*,y)^A\alpha^A\leqslant\lambda Q(x^*,y)^A r^1(x^*,y)^A=0$.

From (a) we have $\alpha^A\leqslant r^1(x^*,y)^A+P(x^*,y)^A\alpha^A$ which implies that:

$$\alpha^A\leqslant(\sum_{n=1}^{m}(P(x^*,y)^A)^{n-1}r^1(x^*,y)^A)+(P(x^*,y)^A)^m\alpha^A \text{ for all } m\in\mathbb{N}.$$

This implies:

$$\alpha^A \leqslant (\frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} (P(x^*,y)^A)^{n-1} r^1(x^*,y)^A) + \frac{1}{N} \sum_{m=1}^{N} (P(x^*,y)^A)^m \alpha^A$$

for all $N \in \mathbb{N}$. Letting N tend to infinity and using that $Q(x^*,y)^A \ \alpha^A \leqslant 0$ we get: $\alpha^A \leqslant \gamma_T^1(x^*,y)^A$.

By theorem 5.2.5 we derive $\inf_\sigma \ \gamma_T^1(x^*,\sigma) \geqslant \ \inf_y \gamma_T^1(x^*,y) \geqslant \alpha$ and hence $\sup_\pi \inf_\sigma \gamma_T^1(\pi,\sigma) \geqslant \alpha$.

Similarly we can show $\inf_\sigma \sup_\pi \gamma_T^1(\pi,\sigma) \leqslant \alpha$, and thus $v_T^1 = \alpha \in \mathbb{R}^z$ and both players have stationary total optimal strategies.

THE 'ONLY IF'-PART:

Suppose that the total value v_T^1 exists in $\mathbb{R}^z$ and that both players have stationary total optimal strategies $x^* \in X$ and $y^* \in Y$.

Then $v_T^1 = r^1(x^*,y^*) + P(x^*,y^*)v_T^1$ in view of lemma 1.5.7 (b), since $Q(x^*,y^*)r^1(x^*,y^*) = 0$. By the optimality of x^* and y^*:

$v_T^1 \leqslant r^1(x^*,y) + P(x^*,y)v_T^1$ for all $y \in Y$, as well as

$v_T^1 \geqslant r^1(x,y^*) + P(x,y^*)v_T^1$ for all $x \in X$.

Hence $v_T^1(s) = \underset{A_s \times B_s}{val} \ [r^1(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)v_T^1(t)]$ and x_s^* and y_s^* are optimal mixed actions in this matrix game for each $s \in S$.

Now let $y \in Y$.

Since $\gamma_T^1(x^*,y) \geqslant v_T^1 > -\infty$ it follows that $\gamma^1(x^*,y) = \ Q(x^*,y)r^1(x^*,y) \geqslant 0$. Let again A be the set of states s with $\gamma^1(s,x^*,y) = 0$, and let $B = S \setminus A$. Then $p(t|s,x_s^*,y_s) = 0$ for all $s \in A$, $t \in B$. Using similar notations as above we have (using lemma 1.5.2 and using $Q(x^*,y)^A r^1(x^*,y)^A = 0$):

$$Q(x^*,y)^A \ \gamma_T^1(x^*,y)^A = \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} Q(x^*,y)^A r^1(x^*,y)^A = 0.$$

Because $v_T^{1A} \leqslant \gamma_T^1(x^*,y)^A$ we derive $Q(x^*,y)^A v_T^{1A} \leqslant 0$, or equivalently:

$$Q(x^*,y)^A \ (\lambda r^1(x^*,y)^A - v_T^{1A}) \geqslant 0 \text{ for all } \lambda \in \mathbb{R}.$$

Let $Q(x^*,y)_B$ denote the restriction of $Q(x^*,y)$ to rows in B.

Since $Q(x^*,y)_B r^1(x^*,y) > 0$ we also have that:

$$Q(x^*,y)_B(\lambda r^1(x^*,y) - v_T^1) \geqslant 0 \text{ for } \lambda \text{ sufficiently large.}$$

Hence $Q(x^*,y)(\lambda r^1(x^*,y) - v_T^1) \geqslant 0$ for λ sufficiently large.

Then for λ sufficiently large:

$$Q(x^*,y)(\lambda r^1(x^*,y) - v_T^1) \geqslant 0 \text{ for all } y \in Y^p.$$

Likewise it can be shown that for all λ sufficiently large:

$$Q(x,y^*)(\lambda r^1(x,y^*) - v_T^1) \leqslant 0 \text{ for all } x \in X^p.$$

Hence, if for λ sufficiently large we look at the stochastic game Γ^* defined by $r^{1*}(s,i,j) := \lambda r^1(s,i,j) - v_T^1(s)$ and $p^*(t|s,i,j) := p(t|s,i,j)$ for all s,i,j, then we find that the limiting average value of Γ^* equals 0 and x^* and y^* are stationary limiting average optimal strategies in Γ^*. Now observe that the limiting average value of Γ^* is also 0 if player 1 is restricted to mixed actions in O^*_{1s} for each $s\in S$, or if player 2 is restricted to mixed actions in O^*_{2s} for each $s\in S$. Applying theorem 5.3.1 gives the existence of δ_1 and $\delta_2\in\mathbb{R}^z$ with:

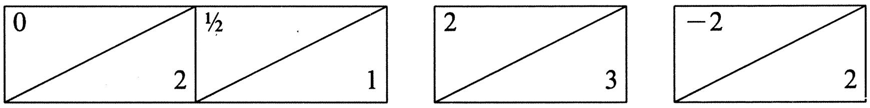
$$\delta_{1s} = \underset{O^*_{1s}\times B_s}{Val}\ [\lambda r^1(s,i,j) - v^1_T(s) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{1t}] \text{ for each } s\in S,$$

$$\delta_{2s} = \underset{A_s\times O^*_{2s}}{Val}\ [\lambda r^1(s,i,j) - v^1_T(s) + \sum_{t=1}^{z} p(t|s,i,j)\delta_{2t}] \text{ for each } s\in S.$$

Using the fact that for any matrix $[a_{ij}]$ and constant $c\in\mathbb{R}$ it holds that $val[a_{ij}+c]= c+val[a_{ij}]$, completes the proof. ■

The following example and the next remark illustrate that one cannot in general take $\lambda=0$ in the above theorem.

5.3.3 Example



State 1 State 2 State 3

For this stochastic game $v^1_T= (1,1,-1)$. Now consider the stochastic game Γ^* as defined in the proof of above theorem. It is easy to verify that the limiting average value of Γ^*, with $\lambda=0$, is not equal to 0 for state 1; in fact it equals -1. Hence for the proof of the 'only if'-part one cannot take $\lambda=0$ from the start. One has to take $\lambda\geqslant 2$.

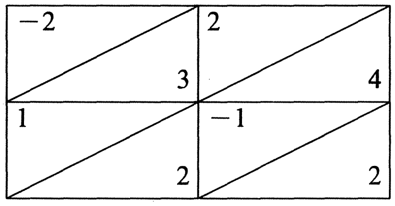
5.3.4 Remark

It should be observed that for a zero-sum stochastic game with the property that $\gamma^1(x,y)=0$ for all $(x,y)\in X\times Y$, one can take $\lambda=0$ in theorem 5.3.2.

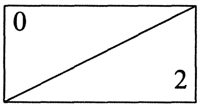
5.4 The bad match

In this section we examine a zero-sum stochastic game with property P for which the total value does exist but for which one of the players has no history independent total ϵ-optimal strategy. We have called this specific stochastic game 'the bad match' in analogy with 'the big match' of Blackwell & Ferguson [1968] (cf. example 1.7.4).

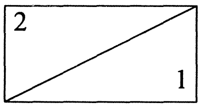
5.4.1 Example *(the bad match)*



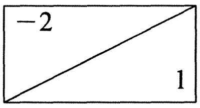
State 1



State 2



State 3



State 4

The interesting initial states are 1, 3 and 4. However, by the structure of the game it is clear that if we know how the players should play total (ϵ-)optimal for initial state 1, then the same strategies are total (ϵ-)optimal for initial states 3 and 4. Therefore we focus on state 1 as initial state.

Observe that if the play starts in state 1, then the players will only have to take (non-trivial) decisions at the odd stages. Those stages we call decision epochs and strategies are determined by the mixed actions that are to be chosen on those decision epochs in state 1. Notice that as soon as player 1 chooses action 2, then the play will move to state 2 with probability 1.

It can be verified that the limiting average reward is 0 for any pair of strategies and for all initial states. Hence property P holds for this stochastic game.

We now define history dependent strategies for player 1 that will turn out to be total ϵ-optimal for player 1 (for specific $\epsilon>0$).

5.4.2 Definition

Let $p(m)=(m+1)^{-2}$ for $m\in\{0,1,2,...\}$ and let $N\in\mathbb{N}$.

We define the history dependent strategy π^N for player 1 by:

having observed the action choices of player 2 at the first n decision epochs, say $j_1,j_2,...,j_n\in\{1,2\}$, $n\geqslant 0$, calculate the excess k_n of 2's over 1's among $\{j_1,j_2,...,j_n\}$ and choose, at decision epoch $n+1$, action 2 with probability $p(k_n+N)$.

5.4.3 Theorem

a) *The total value of the bad match exists and equals 0 (for initial state 1).*

b) *For player 2 a stationary total optimal strategy is to use the mixed action $(\frac{1}{2},\frac{1}{2})$ in state 1 at all decision epochs.*

c) *The strategy π^N is total $(N+1)^{-1}$-optimal for player 1, for all $N\in\mathbb{N}$.*

d) *Player 1 has no history independent total ϵ-optimal strategy for $\epsilon>0$*

sufficiently small.

e) *Player 1 has no total optimal strategy.*

This theorem follows directly from 5.4.4 - 5.4.15 below.

5.4.4 LEMMA

Let y^* be the stationary strategy for player 2 defined by using (½,½) in state 1 at all decision epochs. Then $\gamma^1_T(1,\pi,y^*)=0$ for each $\pi \in \Pi$.

PROOF:

Whatever actions player 1 chooses, at each stage the expected payoff will be 0 if player 2 uses y^*. Namely, in state 1 at each decision epoch the expected payoff is 0; at other stages the play is either in state 2 with payoff 0 or the play is in state 3 or state 4 with the same probability, giving also expected payoff 0. ■

The following corollary is immediate.

5.4.5 COROLLARY

$$\inf_{\sigma} \sup_{\pi} \gamma^1_T(1,\pi,\sigma) \leq 0.$$

5.4.6 LEMMA

With Markov strategies player 1 cannot guarantee a total reward larger than -1.

PROOF:

Let f be a Markov strategy for player 1. We consider two cases:

a) Suppose that the probability that player 1 will ever choose action 2 is 0. Then player 1 chooses action 1 at all decision epochs with probability 1. The stationary strategy $y^1 := (1,0)$ for player 2 leads to $\gamma^1_T(1,f,y^1) = -1$.

b) Suppose that the probability that player 1 will ever choose action 2 is $\epsilon > 0$. Then for each $\delta \in (0,\epsilon)$ there is an $N_\delta \in \mathbb{N}$ such that the probability of player 1 choosing action 2 before stage N_δ, is larger that $\epsilon - \delta$. For each $\delta \in (0,\epsilon)$ define strategy g^δ for player 2 by: at decision epochs $1,2,...,N_\delta$ choose action 2 and at all other decision epochs choose action 1. Then we have that:

$$\gamma^1_T(1,f,g^\delta) \leq (\epsilon-\delta)\cdot(-1) + \delta\cdot(1) + (1-\epsilon)\cdot(-1) = -1 + 2\delta.$$

Since player 2 can choose δ arbitrarily small, the proof is complete. ■

The next lemma says that player 1 has no limiting average optimal strategy.

5.4.7 LEMMA

For any strategy π for player 1 there is some σ such that $\gamma^1_T(1,\pi,\sigma) < 0$.

PROOF:
Let π be a strategy for player 1. If there is no sequence of action choices of player 2, such that player 1 chooses action 2 with positive probability, then it is clear that $\gamma_T^1(1,\pi,y^1) = -1$ for the stationary strategy $y^1 := (1,0)$.
So suppose there is some sequence of action choices $(j_1,j_2,\ldots,j_m)$ such that at decision epoch $m+1$ player 1 will, for the first time, choose action 2 with positive probability $\epsilon > 0$. Now define g^m for player 2 by: at decision epochs $n \leqslant m$ choose j_n with probability 1, at decision epoch $m+1$ choose action 2 with probability 1, at decision epochs $n > m+1$ use the mixed action $(\frac{1}{2},\frac{1}{2})$. It can be verified that $\gamma_T^1(1,\pi,g^m) = -\epsilon < 0$. ■

We will now show that $\gamma_T^1(1,\pi^N,\sigma) \geqslant -(N+1)^{-1}$ for all $\sigma \in \Sigma$, where π^N is the strategy as defined in definition 5.4.2.
To prove this, we fix an arbitrary strategy $\sigma \in \Sigma$ and we define several random variables which are supposed to correspond to the pair of strategies (π^N,σ):

5.4.8 DEFINITION
Let $\sigma \in \Sigma$. Suppose that the players use (π^N,σ) for some $N \in \mathbb{N}$.
Let B (Bottom) be the random variable denoting the number of decision epochs before player 1 chooses action 2.
For each $m \in \mathbb{N}$ define the event $K(m)$ by:
$K(m) := \{B \geqslant m,$ or $B < m$ and $j_{B+1} = 1\}$.
Let $P_N\{K(m)\}$ be the probability that $K(m)$ occurs.

Notice that $K(m)$ is the event that at decision epoch m player 1 either has not yet chosen action 2, or he did choose action 2 and was lucky in receiving 1. In other words we have that $K(m)$ is the event that the total reward up to decision epoch m is non-negative.

5.4.9 REMARK

$$\begin{aligned} P_N\{K(m+1)\} = {} & P_N\{B=0 \textit{ and } j_1 = 1\} \\ & + P_N\{B \geqslant m+1, \textit{ or } 1 \leqslant B < m+1 \textit{ and } j_{B+1} = 1 | j_1 = 1\} \\ & + P_N\{B \geqslant m+1, \textit{ or } 1 \leqslant B < m+1 \textit{ and } j_{B+1} = 1 | j_1 = 2\}. \end{aligned}$$

5.4.10 LEMMA
a) $P_N\{B \geqslant m+1, \textit{ or } 1 \leqslant B < m+1 \textit{ and } j_{B+1} = 1 | j_1 = 1\}$
$= (1-p(N))P_{N-1}\{K(m)\}$.
b) $P_N\{B \geqslant m+1, \textit{ or } 1 \leqslant B < m+1 \textit{ and } j_{B+1} = 1 | j_1 = 2\}$
$= (1-p(N))P_{N+1}\{K(m)\}$.

PROOF:
We only prove (a) since the proof of (b) is similar.
Notice that in the left-hand side of (a) the event $B = 0$ is excluded. Given that

$j_1=1$, the mixed action to be used according to π^N at decision epoch $n+1$, with some history $(1,j_2,j_3,\dots,j_n)$, is equal to the mixed action used according to π^{N-1} at decision epoch n with history $(j_2,j_3,\dots,j_n)$.
At decision epoch 1 player 1 using π^N chooses action 1 with probability $1-p(N)$.
Hence, given that $j_1=1$, using π^N yields the same stochastic process as initially choosing action 1 with probability $1-p(N)$ and using π^{N-1} thereafter. ■

Consequently, we have that

$$P_N\{K(m+1)\}=P_N\{B=0 \text{ and } j_1=1\}$$
$$+(1-p(N))P_{N-1}\{K(m)\}+(1-p(N))P_{N+1}\{K(m)\}.$$

The next lemma states that for all m and N the probability that the total reward up to decision epoch m is non-negative, is at least $N/2(N+1)$.

5.4.11 Lemma

$$P_N\{K(m)\}\geqslant N/2(N+1) \textit{ for all } m,\ N\in\mathbb{N}.$$

Proof:
We use induction to m.

a) Let $m=1$ and let $N\in\mathbb{N}$.
If $j_1=1$, then:
$P_N\{B\geqslant 1|j_1=1\}=1-p(N)$ and $P_N\{B<1 \text{ and } j_{B+1}=1|j_1=1\}=p(N)$.
So $P_N\{K(1)|j_1=1\}=1\geqslant N/2(N+1)$.
If $j_1=2$, then:
$P_N\{K(1)|j_1=2\}=P_N\{B\geqslant 1|j_1=2\}=1-p(N)\geqslant N/2(N+1)$.
Let q be the probability that player 2 chooses action 1 at decision epoch 1, then:

$$P_N\{K(1)\}=q\ P_N\{K(1)|j_1=1\}+(1-q)\ P_N\{K(1)|j_1=2\}$$
$$\geqslant q\ N/2(N+1)+(1-q)\ N/2(N+1)=N/2(N+1).$$

b) Suppose $P_N\{K(m)\}\geqslant N/2(N+1)$ for some $m\in\mathbb{N}$ and all $N\in\mathbb{N}$.
Then, in view of remark 5.4.9 and lemma 5.4.10, we have:

$$\begin{aligned}P_N\{K(m+1)|j_1=1\}&=P_N\{B=0 \text{ and } j_1=1|j_1=1\}\\&\quad+P_N\{B\geqslant m+1, \text{ or } 1\leqslant B<m+1\\&\qquad \text{and } j_{B+1}=1|j_1=1\}\\&=p(N)+(1-p(N))\ P_{N-1}\{K(m)\}\\&\geqslant p(N)+(1-p(N))(N-1)/2N=N/2(N+1).\end{aligned}$$

Also in view of lemma 5.4.10 we have:

$$\begin{aligned}P_N\{K(m+1)|j_1=2\}&=P_N\{B=0 \text{ and } j_1=1|j_1=2\}\\&\quad+P_N\{B\geqslant m+1, \text{ or } 1\leqslant B<m+1\\&\qquad \text{and } j_{B+1}=1|j_1=2\}\\&=0+(1-p(N))P_{N+1}\{K(m)\}\end{aligned}$$

$$\geqslant (1-p(N))(N+1)/2(N+2) = N/2(N+1).$$

Hence $P_N\{K(m+1)\} = q\,P_N\{K(m+1)|j_1=1\}$
$+ (1-q)P_N\{K(m+1)|j_1=2\} \geqslant N/2(N+1)$,
which shows the induction step. ■

The following lemma demonstrates that π^N guarantees a total reward of at least $-(N+1)^{-1}$ if the probability that player 1 will ever choose action 2, using π^N against σ, is 1.

5.4.12 Lemma

If $\lim_{m\to\infty} P_N\{B \geqslant m\} = 0$, *then* $\gamma_T^1(1,\pi^N,\sigma) \geqslant -(N+1)^{-1}$.

Proof:

Since by definition $P_N\{K(m)\} = P_N\{B \geqslant m\} + P_N\{B < m \text{ and } j_{B+1} = 1\}$, we derive from lemma 5.4.11, in view of the assumption of this lemma, that:

$$\lim_{m\to\infty} P_N\{K(m)\} = P_N\{j_{B+1} = 1\} \geqslant N/2(N+1).$$

With probability 1 player 1 will choose action 2 at some decision epoch and, since the sum of payoffs until that decision epoch equals 0, the total reward is determined by the action which player 2 chooses at that moment. So we have:

$$\begin{aligned}\gamma_T^1(1,\pi^N,\sigma) &= P_N\{j_{B+1} = 1\} - P_N\{j_{B+1} = 2\}\\ &= 2P_N\{j_{B+1} = 1\} - 1\\ &\geqslant N/(N+1) - 1 = -(N+1)^{-1}.\end{aligned}$$

■

Notice that, if for a certain n we have $k_n = -N$, then player 1 will choose action 2 with probability 1 at decision epoch $n+1$. Hence, as long as no transition to state 2 has occurred, we have $k_n \geqslant -N$.

5.4.13 Lemma

For any realization of the stochastic process associated to π^N *and* σ, *for which player 1 never chooses action 2, it holds that the corresponding total reward is at least 0.*

Proof:

Let $(r_1,r_2,\ldots)$ be the sequence of payoffs (to player 1) that occurs.

Since in this case $k_n > -N$ for all $n \in \mathbb{N}$, we have $\sum_{t=1}^{T} \sum_{n=1}^{t} r_n > -2N$ for every $T \in \mathbb{N}$. It follows that $\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{t} r_n \geqslant 0$. ■

5.4.14 Lemma

If $\lim_{m\to\infty} P_N\{B \geqslant m\} > 0$, *then* $\gamma_T^1(1,\pi^N,\sigma) \geqslant -(N+1)^{-1}$.

PROOF:
For $m \in \mathbb{N}$ let $\lambda(m) := P_N\{B < m$ and $j_{B+1} = 1\}$ and let $\mu(m) := P_N\{B < m$ and $j_{B+1} = 2\}$. Since $\{\lambda(m) : m \in \mathbb{N}\}$ and $\{\mu(m) : m \in \mathbb{N}\}$ are bounded monotone increasing sequences, we can define $\lambda := \lim_{m\to\infty} \lambda(m)$ and $\mu := \lim_{m\to\infty} \mu(m)$.
Now the probability that player 1 will ever choose action 2, equals $\lambda + \mu$ and hence $1 - \lambda - \mu$ is the probability that the play never reaches state 2. By lemma 5.4.13 and by the definitions of λ and μ we have:

$$\gamma_T^1(1,\pi^N,\sigma) \geqslant \lambda \cdot 1 + \mu \cdot (-1) + (1 - \lambda - \mu) \cdot 0 = \lambda - \mu.$$

So if we can prove that $\lambda - \mu \geqslant -(N+1)^{-1}$, then the proof is finished.

For each $m \in \mathbb{N}$ define strategy σ^m by: up to decision epoch m use σ, at all other decision epochs use the mixed action $(½,½)$. Then σ^m will give rise to sequences $(j_1,j_2,\ldots)$ for which $[k_n = -N$ for some $n \in \mathbb{N}]$ with probability 1. Hence for the strategies σ^m the condition of lemma 5.4.12 applies (where P_N now refers to (π^N,σ^m)). Hence $\gamma_T^1(1,\pi^N,\sigma^m) \geqslant -(N+1)^{-1}$ for all $m \in \mathbb{N}$.

On the other hand, with respect to (π^N,σ^m), if player 1 chooses action 2 before decision epoch m, then this contributes $\lambda(m) \cdot 1 + \mu(m) \cdot (-1)$ to $\gamma_T^1(1,\pi^N,\sigma^m)$, and choosing action 2 later contributes $(1 - \lambda(m) - \mu(m)) \cdot 0$ (cf. lemma 5.4.4).
Hence $\gamma_T^1(1,\pi^N,\sigma^m) = \lambda(m) - \mu(m) \geqslant -(N+1)^{-1}$ for all $m \in \mathbb{N}$. Taking limits for m to ∞ gives $\lambda - \mu \geqslant -(N+1)^{-1}$, which completes this proof. ■

An immediate consequence of lemma 5.4.12 and lemma 5.4.14 is the following.

5.4.15 COROLLARY

$$\sup_{\pi} \inf_{\sigma} \gamma_T^1(1,\pi,\sigma) \geqslant 0.$$

It should be remarked that the above proofs for the bad match are along the same lines as proofs by Blackwell & Ferguson [1968] for the big match.

5.5 CONCLUSIONS

The bad match illustrates that there is an analogy between the total reward criterion and the limiting average reward criterion. For both criteria history dependent strategies are indispensable for playing ϵ-optimal, which distinguishes these criteria from the β-discounted reward criterion. The relation between the total reward criterion and the limiting average reward criterion is even narrowed by the fact that with any stochastic game Γ we can relate another stochastic game Γ^* with an infinite state space, such that for all strategies the total reward in Γ equals the related limiting average reward in Γ^*. To show this, let Γ be a zero-sum stochastic game as in definition 1.2.1 and let H_n, $n \in \mathbb{N}$, be as defined in definition 1.3.3. We use asterisks to define Γ^*.
Let $S^* := \bigcup_{n=1}^{\infty} H_n$ be the state space of Γ^*.
For $s^* = h_n = (s_1,i_1,j_1,s_2,i_2,j_2,\ldots,s_{n-1},i_{n-1},j_{n-1},s_n)$ let $A_{s^*}^* := A_{s_n}$ and

$B^*_{s^*} := B_{s_n}$ be the action spaces in Γ^*.

For $s^* = h_n$ and $i^* \in A_{s^*}$, $j^* \in B_{s^*}$ let $r^{1*}(s^*,i^*,j^*) := \sum_{k=1}^{n-1} r^1(s_k,i_k,j_k) + r^1(s_n,i^*,j^*)$ and let $p^*(t^*|s^*,i^*,j^*) := p(t|s_n,i^*,j^*)$ for $t^* = (h_n,i^*,j^*,t)$ and $p^*(t^*|s^*,i^*,j^*) := 0$ for other $t^* \in S^*$.

Hence we have translated histories of Γ into states in Γ^*. Notice that in Γ^* every state s^* can be reached along precisely one history path, and there is a one-to-one relation between strategies in Γ and strategy classes in Γ^*.

Furthermore we have that at each stage $N \in \mathbb{N}$, for strategies π,σ and initial state $s^* = s$, it holds that:

$$\sum_{m=1}^{N} E_{s^*\pi\sigma}[R^{1*}(m)] = \sum_{m=1}^{N} E_{s\pi\sigma}[\sum_{n=1}^{m} R^1(n)] = \sum_{m=1}^{N}\sum_{n=1}^{m} E_{s\pi\sigma}[R^1(n)].$$

Hence $\gamma^{1*}(s^*,\pi,\sigma) = \gamma^1_T(s,\pi,\sigma)$ for all $s^* = s$, and all strategies π and σ.

Chapter 6

Stochastic games and mathematical programming

6.1 Introduction

Mangasarian & Stone [1964] proved the following theorem which relates equilibria of a bimatrix game with solutions of an associated non-linear program, with quadratic objective function and with linear constraints.

6.1.1 Theorem

For a bimatrix game (A^1,A^2) a pair of mixed actions (x^*,y^*) is an equilibrium with payoffs $(\alpha^{1*},\alpha^{2*})$ if and only if $(x^*,y^*,\alpha^{1*},\alpha^{2*})$ is a global minimum in the following non-linear program, with objective value 0.
NLP 6.1.1:

$$\begin{array}{ll} \textit{variables} & x \in \mathbb{R}^m, y \in \mathbb{R}^n, \alpha^1, \alpha^2 \in \mathbb{R} \\ \textit{minimize} & \alpha^1 - xA^1y + \alpha^2 - xA^2y \\ \textit{subject to} & \alpha^1 1_m - A^1y \geqslant 0 \textit{ and } \alpha^2 1_n - xA^2 \geqslant 0 \\ & \sum_{i=1}^{m} x_i = 1 \textit{ and } \sum_{j=1}^{n} y_j = 1 \\ & x \geqslant 0 \textit{ and } y \geqslant 0. \end{array}$$

For matrix games A, where player 1 is the maximizing player, the non-linear factors in the objective function of NLP 6.1.1 disappear (since their sum is 0) and what remains is a linear program. It is easy to verify that for matrix games we have the following result.

6.1.2 Theorem

For an $m \times n$ matrix game A the value, for player 1, is v and (x^*,y^*) are optimal mixed actions for the players, if and only if $(x^*,y^*,v,-v)$ is a global minimum in the following linear program, with objective value 0.
LP 6.1.2:

$$\begin{array}{ll} \textit{variables} & x \in \mathbb{R}^m, y \in \mathbb{R}^n, \alpha^1, \alpha^2 \in \mathbb{R} \\ \textit{minimize} & \alpha^1 + \alpha^2 \\ \textit{subject to} & \alpha^1 1_m - Ay \geqslant 0 \textit{ and } \alpha^2 1_n + xA \geqslant 0 \\ & \sum_{i=1}^{m} x_i = 1 \textit{ and } \sum_{j=1}^{n} y_j = 1 \\ & x \geqslant 0 \textit{ and } y \geqslant 0. \end{array}$$

In this chapter we will formulate analogues of the above theorems for stochastic games. We consider both the general-sum and the zero-sum case for the β-discounted reward criterion, for the limiting average reward criterion and for the total reward criterion. For the last criterion we have restricted our attention to stochastic games with the property that the limiting average reward is 0 for all pairs of stationary strategies. For the β-discounted criterion and the limiting average criterion, characterizations for stationary solutions by means of mathematical programs have been reported in Rogers [1969], Rothblum [1979], Hordijk & Kallenberg [1981], Vrieze [1981, 1983, 1987-a], Filar [1986], Filar & Schultz [1986, 1987], Schultz [1987]. However several of these characterizations are for special classes of stochastic games and/or for the zero-sum case only.

Since it is well-known that only with respect to the β-discounted reward criterion stationary optimal strategies and stationary equilibria always exist, it is of interest to know, especially for the other criteria, whether near-optimal solutions of the programs correspond with ϵ-optimal strategies or ϵ-equilibria. For the β-discounted reward criterion this is indeed the case. However for the limiting average reward criterion and for the total reward criterion this correspondence between near-optimal solutions not necessarily holds. Nevertheless, for the zero-sum case the program we formulate will for both players lead to stationary (ϵ-)optimal strategies, whenever they exist. If stationary ϵ-optimal strategies fail to exist, then our program finds 'ϵ-best' stationary strategies for both players. Here an 'ϵ-best' stationary strategy for player 1 is a strategy x_ϵ such that $\inf_{y \in Y} \gamma^1(x_\epsilon,y) + \epsilon \geqslant \sup_{x \in X} \inf_{y \in Y} \gamma^1(x,y)$.

The programs we present are based on the lemmas 1.5.3 to 1.5.8 and on lemma 1.6.2, as well as on theorems 1.7.3, 5.3.1 and 5.3.2. The results of the sections 6.2 and 6.3 can be found in Filar et al. [1991].

6.2 Programs for the β-discounted reward criterion

6.2.1 Lemma

For a general-sum stochastic game a pair of stationary strategies (x^*,y^*) is a β-discounted equilibrium with β-discounted rewards $(\alpha^{1*},\alpha^{2*})$ if and only if for all $s \in S$ and $k \in \{1,2\}$:

a) $\alpha^{k*} = (1-\beta)r^k(s,x_s^*,y_s^*) + \beta \sum_{t=1}^{z} p(t|s,x_s^*,y_s^*)\alpha_t^{k*}$

b) $\alpha_s^{1*} \geqslant (1-\beta)r^1(s,i,y_s^*) + \beta \sum_{t=1}^{z} p(t|s,i,y_s^*)\alpha_t^{1*}$ *for all* $i \in A_s$

$\alpha_s^{2*} \geqslant (1-\beta)r^2(s,x_s^*,j) + \beta \sum_{t=1}^{z} p(t|s,x_s^*,j)\alpha_t^{2*}$ *for all* $j \in B_s$.

Proof:
This lemma follows directly from lemma 1.5.3, lemma 1.6.2 and lemma 1.6.4.■

From lemma 6.2.1 we immediately obtain the next theorem.

6.2.2 THEOREM

For a general-sum stochastic game a pair of stationary strategies (x^*,y^*) is a β-discounted equilibrium with β-discounted rewards $(\alpha^{1*},\alpha^{2*})$ if and only if $(x^*,y^*,\alpha^{1*},\alpha^{2*})$ is a global minimum in the following non-linear program, with objective value 0.

NLP 6.2.2:

variables $x\in \mathop{\times}\limits_{s=1}^{z} \mathbb{R}^{m_s},\ y\in \mathop{\times}\limits_{s=1}^{z} \mathbb{R}^{n_s},\ \alpha^1,\alpha^2\in\mathbb{R}^z$

minimize $\sum\limits_{k=1}^{2}\sum\limits_{s=1}^{z}(\alpha_s^k-(1-\beta)r^k(s,x_s,y_s)-\beta\sum\limits_{t=1}^{z}p(t|s,x_s,y_s)\alpha_t^k)$

subject to

a) $\alpha_s^1\geqslant(1-\beta)r^1(s,i,y_s)+\beta\sum\limits_{t=1}^{z}p(t|s,i,y_s)\alpha_t^1$ *for all* $i\in A_s,\ s\in S$

$\alpha_s^2\geqslant(1-\beta)r^2(s,x_s,j)+\beta\sum\limits_{t=1}^{z}p(t|s,x_s,j)\alpha_t^2$ *for all* $j\in B_s,\ s\in S$

b) $\sum\limits_{i=1}^{m_s}x_s(i)=1$ *and* $\sum\limits_{j=1}^{n_s}y_s(j)=1$ *for all* $s\in S$

c) $x_s\geqslant 0$ *and* $y_s\geqslant 0$ *for all* $s\in S$.

An interesting feature of the above non-linear program is that feasible solutions with objective value near 0 are directly related with stationary β-discounted ϵ-equilibria.

6.2.3 COROLLARY

If $(x^*,y^*,\alpha^{1*},\alpha^{2*})$ is a feasible solution of NLP 6.2.2 with objective value $\delta>0$, then (x^*,y^*) is a stationary β-discounted $\delta(1-\beta)^{-1}$-equilibrium.

PROOF:

Constraints (b) and (c) give that $x^*\in X$ and $y^*\in Y$.

By the constraints (a) and by the objective value $\delta>0$ for the solution $(x^*,y^*,\alpha^{1*},\alpha^{2*})$ we have for each $s\in S$:

$$0\leqslant\alpha_s^{k*}-(1-\beta)r^k(s,x_s^*,y_s^*)-\beta\sum_{t=1}^{z}p(t|s,x_s^*,y_s^*)\alpha_t^{k*}\leqslant\delta.$$

Or, equivalently, in vector notation:

$$0\leqslant\alpha^{k*}-(1-\beta)r^k(x^*,y^*)-\beta P(x^*,y^*)\alpha^{k*}\leqslant\delta 1_z.$$

By the first inequality sign: $\alpha^{k*}\geqslant\gamma_\beta^k(x^*,y^*)$ (cf. lemma 1.5.4).

The second inequality sign gives us:

$$(I-\beta P(x^*,y^*))\alpha^{k*}\leqslant(1-\beta)r^k(x^*,y^*)+\delta 1_z,$$

and hence $\alpha^{k*}\leqslant(1-\beta)(I-\beta P(x^*,y^*))^{-1}r^k(x^*,y^*)+(I-\beta P(x^*,y^*))^{-1}\delta 1_z$

$$=\gamma_\beta^k(x^*,y^*)+\delta(1-\beta)^{-1}1_z.$$

Constraints (a) also imply:

$$\gamma_\beta^1(x,y^*) \leqslant \alpha^{1*} \text{ for all } x \in X \text{ and } \gamma_\beta^2(x^*,y) \leqslant \alpha^{2*} \text{ for all } y \in Y.$$

Combining these results proves this corollary. ■

Of course, the reduction of NLP 6.2.2 to zero-sum stochastic games is a program which finds the β-discounted value and stationary optimal strategies for both players. Rothblum [1979] proposed the following non-linear program to find the β-discounted value and a stationary β-discounted optimal strategy for player 2.

6.2.4 Theorem

For a zero-sum stochastic game with β-discounted value v_β^1 a stationary strategy $y^* \in Y$ is β-discounted optimal for player 2 if and only if (y^*, v_β^1) is a global minimum in the following non-linear program.

NLP 6.2.4:

variables $y \in \mathop{\times}\limits_{s=1}^{z} \mathbb{R}^{n_s},\ \alpha \in \mathbb{R}^z$

minimize $\sum\limits_{s=1}^{z} \alpha_s$

subject to

a) $\alpha_s \geqslant (1-\beta)r^1(s,i,y_s) + \beta \sum\limits_{t=1}^{z} p(t|s,i,y_s)\alpha_t$ *for all* $i \in A_s,\ s \in S$

b) $\sum\limits_{j=1}^{n_s} y_s(j) = 1$ *for all* $s \in S$

c) $y_s \geqslant 0$ *for all* $s \in S$.

It is obvious that a similar program can be formulated to find stationary β-discounted optimal strategies for player 1.

6.3 Programs for the limiting average reward criterion

It is well-known that stationary limiting average equilibria may fail to exist. However, below we present a non-linear program which finds a stationary limiting average equilibrium whenever one exists. Our program is based on the following lemma.

6.3.1 Lemma

A pair of stationary strategies (x^*,y^*) is a limiting average equilibrium with limiting average rewards $(\alpha^{1*},\alpha^{2*})$ if and only if there exist $\delta^{1*},\delta^{2*},\ \mu^{1*},\mu^{2*} \in \mathbb{R}^z$ with:

a) $\alpha_s^{k*} = \sum\limits_{t=1}^{z} p(t|s,x_s^*,y_s^*)\alpha_t^{k*}$ *for all* $s \in S,\ k \in \{1,2\}$

b) $\alpha_s^{k*} + \delta_s^{k*} = r^k(s,x_s^*,y_s^*) + \sum\limits_{t=1}^{z} p(t|s,x_s^*,y_s^*)\delta_t^{k*}$ *for all* $s \in S,\ k \in \{1,2\}$

c) $\alpha_s^{1*} \geqslant \sum\limits_{t=1}^{z} p(t|s,i,y_s^*)\alpha_t^{1*}$ *for all* $i \in A_s,\ s \in S$

$$\alpha_s^{2*} \geqslant \sum_{t=1}^{z} p(t|s,x_s^*,j)\alpha_t^{2*} \textit{ for all } j \in B_s,\ s \in S$$

d) $$\alpha_s^{1*} + \mu_s^{1*} \geqslant r^1(s,i,y_s^*) + \sum_{t=1}^{z} p(t|s,i,y_s^*)\mu_t^{1*} \textit{ for all } i \in A_s,\ s \in S$$

$$\alpha_s^{2*} + \mu_s^{2*} \geqslant r^2(s,x_s^*,j) + \sum_{t=1}^{z} p(t|s,x_s^*,j)\mu_t^{2*} \textit{ for all } j \in B_s,\ s \in S.$$

This lemma follows from the fact that (x^*,y^*) is a limiting average equilibrium if and only if x^* is limiting average optimal for player 1 in MDP(y^*) and y^* is limiting average optimal for player 2 in MDP(x^*) (cf. section 1.6). A stationary strategy x^* is limiting average optimal in MDP(y^*) with limiting average reward α^{1*} if and only if there exist δ^{1*} and μ^{1*} such that (cf. Blackwell [1962], Hordijk & Kallenberg [1979]):

a) $$\alpha_s^{1*} = \sum_{t=1}^{z} p(t|s,x_s^*,y_s^*)\alpha_t^{1*} \text{ for all } s \in S$$

b) $$\alpha_s^{1*} + \delta_s^{1*} = r^1(s,x_s^*,y_s^*) + \sum_{t=1}^{z} p(t|s,x_s^*,y_s^*)\delta_t^{1*} \text{ for all } s \in S$$

c) $$\alpha_s^{1*} \geqslant \sum_{t=1}^{z} p(t|s,i,y_s^*)\alpha_t^{1*} \text{ for all } i \in A_s,\ s \in S$$

d) $$\alpha_s^{1*} + \mu_s^{1*} \geqslant r^1(s,i,y_s^*) + \sum_{t=1}^{z} p(t|s,i,y_s^*)\mu_t^{1*} \text{ for all } i \in A_s,\ s \in S.$$

It is easy to verify that lemma 6.3.1 directly implies the following result.

6.3.2 THEOREM

For a general-sum stochastic game a pair of stationary strategies (x^*,y^*) is a limiting average equilibrium with limiting average rewards $(\alpha^{1*},\alpha^{2*})$ if and only if there exist $\delta^{1*},\delta^{2*},\mu^{1*},\mu^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*},\mu^{1*},\mu^{2*})$ is a global minimum in the following non-linear program, with objective value 0.

NLP 6.3.2:

variables $x \in \times_{s=1}^{z} \mathbb{R}^{m_s},\ y \in \times_{s=1}^{z} \mathbb{R}^{n_s},\ \alpha^1,\alpha^2,\delta^1,\delta^2,\mu^1,\mu^2 \in \mathbb{R}^z$

minimize $$\sum_{k=1}^{2}\sum_{s=1}^{z} [\alpha_s^k - \sum_{t=1}^{z} p(t|s,x_s,y_s)\alpha_t^k]$$

subject to

a) $$\alpha_s^1 \geqslant \sum_{t=1}^{z} p(t|s,i,y_s)\alpha_t^1 \textit{ for all } i \in A_s,\ s \in S$$

$$\alpha_s^2 \geqslant \sum_{t=1}^{z} p(t|s,x_s,j)\alpha_t^2 \textit{ for all } j \in B_s,\ s \in S$$

b) $$\alpha_s^1 + \delta_s^1 = r^1(s,x_s,y_s) + \sum_{t=1}^{z} p(t|s,x_s,y_s)\delta_t^1 \textit{ for all } s \in S$$

$$\alpha_s^2 + \delta_s^2 = r^2(s,x_s,y_s) + \sum_{t=1}^{z} p(t|s,x_s,y_s)\delta_t^2 \textit{ for all } s \in S$$

c) $$\alpha_s^1 + \mu_s^1 \geqslant r^1(s,i,y_s) + \sum_{t=1}^{z} p(t|s,i,y_s)\mu_t^1 \textit{ for all } i \in A_s,\ s \in S$$

$$\alpha_s^2+\mu_s^2 \geqslant r^2(s,x_s,j) + \sum_{t=1}^{z} p(t|s,x_s,j)\mu_t^2 \textit{ for all } j\in B_s,\ s\in S$$

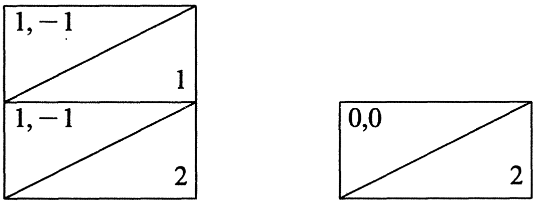
d) $\sum_{i=1}^{m_s} x_s(i)=1,\ \sum_{j=1}^{n_s} y_s(j)=1$ *for all* $s\in S$

e) $x_s\geqslant 0,\ y_s\geqslant 0$ *for all* $s\in S$.

Observe that constraints (a) of NLP 6.3.2 imply that for any feasible solution the objective value is non-negative.

For the β-discounted criterion feasible solutions with objective value near 0 in NLP 6.2.2, turned out to correspond with β-discounted ϵ-equilibria (for specific $\epsilon>0$). Unfortunately, feasible solutions with objective value near 0 in NLP 6.3.2 do not necessarily correspond with limiting average ϵ-equilibria. This is illustrated in the next example.

6.3.3 Example



State 1 State 2

Let $x=((1-\epsilon,\epsilon),1)$, $y=(1,1)$, $\alpha^1=(1,0)$, $\delta^2=(-1/\epsilon,0)$ and $\alpha^2=\delta^1=\mu^1=\mu^2=0$, where $\epsilon>0$. It is easy to verify that $(x,y,\alpha^1,\alpha^2,\delta^1,\delta^2,\mu^1,\mu^2)$ is a feasible solution of NLP 6.3.2 with objective value ϵ.
It is clear that (x,y) is not a limiting average ϵ-equilibrium for $\epsilon\in(0,1)$, because $\gamma^1(1,x,y)=0<1=\gamma^1(1,x^*,y)$ with $x^*=((1,0),1)$.

Since example 6.3.3 is a zero-sum stochastic game, it also demonstrates that the restriction of NLP 6.3.2 to zero-sum stochastic games does not yield a program for which near-optimal solutions correspond with stationary limiting average ϵ-optimal strategies.

However for zero-sum stochastic games we present a powerful non-linear program for which feasible solutions with objective value near 0 do indeed correspond with stationary limiting average ϵ-optimal strategies. The non-linear program NLP 6.3.4 (below) completely characterizes stochastic games with stationary limiting average (ϵ-)optimal strategies. For any stochastic game NLP 6.3.4 finds 'best' stationary strategies with respect to the 'distance measure' $d(\ ,\ )$ defined by:

$$d(x^*,y^*):=\sum_{s=1}^{z}[\max_x \gamma^1(s,x,y^*)-\min_y \gamma^1(s,x^*,y)].$$

Thus $d(\ ,\ )$ is a measure for the 'distance from optimality' of any pair of

stationary strategies (x^*,y^*).
Notice that for each $s \in S$ it holds that $\max_x \gamma^1(s,x,y^*) - \min_y \gamma^1(s,x^*,y)$ is non-negative and: $d(x^*,y^*) = 0$ if and only if x^* and y^* are stationary limiting average optimal strategies. If $d(x^*,y^*) > 0$, then x^* and y^* are both limiting average ϵ-optimal for some $\epsilon \in [0,d(x^*,y^*)]$.

6.3.4 THEOREM
For a zero-sum stochastic game the following results hold.
If there exist $x^* \in \times_{s=1}^{z} \mathbb{R}^{m_s}$, $y^* \in \times_{s=1}^{z} \mathbb{R}^{n_s}$, α^{1*},α^{2*}, $\delta^{1*},\delta^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a feasible solution with objective value $\epsilon(\geqslant 0)$ in NLP 6.3.4 below, then x^* and y^* are stationary limiting average ϵ-optimal strategies for the respective players.
Conversely, if x^* and y^* are stationary limiting average ϵ-optimal strategies, then there exist $\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a feasible solution with objective value $2z\epsilon$, or less, in NLP 6.3.4.
NLP 6.3.4:

variables $x \in \times_{s=1}^{z} \mathbb{R}^{m_s}$, $y \in \times_{s=1}^{z} \mathbb{R}^{n_s}$, $\alpha^1,\alpha^2,\delta^1,\delta^2 \in \mathbb{R}^z$

minimize $\sum_{s=1}^{z} (\alpha_s^1 + \alpha_s^2)$

subject to

a) $\alpha_s^1 \geqslant \sum_{t=1}^{z} p(t|s,i,y_s)\alpha_t^1$ *for all* $i \in A_s$, $s \in S$

b) $\alpha_s^1 + \delta_s^1 \geqslant r^1(s,i,y_s) + \sum_{t=1}^{z} p(t|s,i,y_s)\delta_t^1$ *for all* $i \in A_s$, $s \in S$

c) $\alpha_s^2 \geqslant \sum_{t=1}^{z} p(t|s,x_s,j)\alpha_t^2$ *for all* $j \in B_s$, $s \in S$

d) $\alpha_s^2 + \delta_s^2 \geqslant -r^1(s,x_s,j) + \sum_{t=1}^{z} p(t|s,x_s,j)\delta_t^2$ *for all* $j \in B_s$, $s \in S$

e) $\sum_{i=1}^{m_s} x_s(i) = 1$, $\sum_{j=1}^{n_s} y_s(j) = 1$ *for all* $s \in S$

f) $x_s \geqslant 0$, $y_s \geqslant 0$ *for all* $s \in S$.

PROOF:
Suppose that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is feasible in NLP 6.3.4 with objective value $\epsilon \geqslant 0$.
By constraints (e) and (f) we have $x^* \in X$ and $y^* \in Y$.
Constraints (a) and (b) imply $\alpha_s^{1*} \geqslant \gamma^1(s,x,y^*)$ for all $s \in S$, $x \in X$ by lemma 1.5.6. Similarly (c) and (d) imply $\alpha_s^{2*} \geqslant \gamma^2(s,x^*y) = -\gamma^1(s,x^*,y)$ for all $s \in S$, $y \in Y$.
Hence $\alpha_s^1 + \alpha_s^2 \geqslant \gamma^1(s,x,y^*) - \gamma^1(s,x^*,y)$ for all $x \in X$, $y \in Y$, $s \in S$.
Especially $\alpha_s^1 + \alpha_s^2 \geqslant \gamma^1(s,x^*,y^*) - \gamma^1(s,x^*,y^*) = 0$ for all $s \in S$.
Since $\sum_{s=1}^{z} (\alpha_s^1 + \alpha_s^2) = \epsilon$, we conclude that $\alpha_s^1 + \alpha_s^2 \leqslant \epsilon$ for all $s \in S$.

It follows that for all $s \in S$, $x \in X$, $y \in Y$ we have:

$$\gamma^1(s,x^*,y)+\epsilon \geqslant \gamma^1(s,x^*,y^*) \geqslant \gamma^1(s,x,y^*)-\epsilon.$$

So the strategies x^* and y^* are limiting average ϵ-optimal.

To prove the converse statement, suppose that we have stationary limiting average ϵ-optimal strategies $x^* \in X$, $y^* \in Y$. Then constraints (e) and (f) are satisfied.

By solving MDP(y^*) with respect to the limiting average reward criterion (cf. Hordijk & Kallenberg [1979]) there exist α^{1*} and δ^{1*} such that (a) and (b) are satisfied. Moreover, we have that $\alpha_s^{1*} = \max_{x \in X} \gamma^1(s,x,y^*) \leqslant \gamma^1(s,x^*,y^*)+\epsilon$ for each $s \in S$. Similarly there exist α^{2*} and δ^{2*} such that constraints (c) and (d) are satisfied and $\alpha_s^{2*} = \max_{y \in Y}(-\gamma^1(s,x^*,y)) \leqslant -\gamma^1(s,x^*,y^*)+\epsilon$.

Hence $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a feasible solution of NLP 6.3.4 and the corresponding objective value is $\sum_{s=1}^{z}(\alpha_s^{1*}+\alpha_s^{2*}) \leqslant 2z\epsilon$. ■

Theorem 6.3.4 implies that by solving NLP 6.3.4 'best' stationary strategies can be found, as can be seen from the following results.

6.3.5 Corollary

If $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a global minimum in NLP 6.3.4 with objective value $\mu \geqslant 0$, then $\mu = d(x^*,y^*) \leqslant d(x,y)$ for all $x \in X$, $y \in Y$.

Proof:

Let $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ be a global minimum in NLP 6.3.4 with objective value μ. By the constraints (a), (b), (c) and (d) it holds that:

$$\alpha_s^{1*} \geqslant \max_{x \in X} \gamma^1(s,x,y^*) \text{ for all } s \in S,$$

$$\alpha_s^{2*} \geqslant \max_{y \in Y} \gamma^2(s,x^*,y) = -\min_{y \in Y} \gamma^1(s,x^*,y) \text{ for all } s \in S.$$

Since $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a global minimum, equality must hold in all these inequalities. Otherwise, by solving MDP(x^*) and MDP(y^*) one could find variables $\alpha^1,\alpha^2,\delta^1,\delta^2$ for which equality indeed holds, and hence $\sum_{s=1}^{z}(\alpha_s^1+\alpha_s^2) < \mu$ would contradict the minimality of $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$.

We conclude that $\mu = \sum_{s=1}^{z}(\alpha_s^{1*}+\alpha_s^{2*}) = d(x^*,y^*)$.

For any pair of stationary strategies $(\tilde{x},\tilde{y})$ one can find $\tilde{\alpha}^1,\tilde{\alpha}^2,\tilde{\delta}^1,\tilde{\delta}^2$ such that $(\tilde{x},\tilde{y},\tilde{\alpha}^1,\tilde{\alpha}^2,\tilde{\delta}^1,\tilde{\delta}^2)$ is feasible in NLP 6.3.4 and such that $\tilde{\alpha}_s^1 = \max_{x \in X} \gamma^1(s,x,\tilde{y})$ and $\tilde{\alpha}_s^2 = -\min_{y \in Y} \gamma^1(s,\tilde{x},y)$ for all $s \in S$. Then $\mu \leqslant \sum_{s=1}^{z}(\tilde{\alpha}_s^1+\tilde{\alpha}_s^2) = d(\tilde{x},\tilde{y})$ by the minimality of μ. ■

6.3.6 Corollary

Suppose that the minimum in NLP 6.3.4 does not exist, but that the infimum equals $\mu (\geqslant 0)$, then for every $\epsilon>0$ there exist stationary limiting average $(\mu+\epsilon)$-optimal strategies for both players.

Proof:

For every $\epsilon>0$ there exists $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ which is feasible in NLP 6.3.4 with objective value less than $\mu+\epsilon$. From theorem 6.3.4 it follows that x^* and y^* are stationary limiting average $(\mu+\epsilon)$-optimal strategies. ■

Without proof we state.

6.3.7 Corollary

For a zero-sum stochastic game there exist stationary limiting average ϵ-optimal strategies for both players and for all $\epsilon>0$, if and only if the infimum of NLP 6.3.4 is equal to 0.

6.4 Programs for the total reward criterion

From the bad match (section 5.4) it is clear that stationary total equilibria do not necessarily exist, not even in stochastic games for which the limiting average rewards are 0 for all pairs of stationary strategies. Nevertheless we can formulate non-linear programs for which optimal solutions correspond with stationary total equilibria. We make use of the following lemma.

6.4.1 Lemma

For a stochastic game with the property that $\gamma^k(x,y)=0$ for all stationary strategies x and y, we have: a stationary strategy x^* is a total best reply against $y^* \in Y$ if and only if there exist α^*, δ^* and $\mu^* \in \mathbb{R}^z$ such that

a) $\alpha_s^* = r^1(s,x_s^*,y_s^*) + \sum_{t=1}^{z} p(t|s,x_s^*,y_s^*)\alpha_t^*$ *for all* $s \in S$

$\alpha_s^* \geqslant r^1(s,i,y_s^*) + \sum_{t=1}^{z} p(t|s,i,y_s^*)\alpha_t^*$ *for all* $i \in A_s$, $s \in S$

b) $\alpha_s^* + \delta_s^* = \sum_{t=1}^{z} p(t|s,x_s^*,y_s^*)\delta_t^*$ *for all* $s \in S$

c) $\alpha_s^* + \mu_s^* \geqslant \sum_{t=1}^{z} p(t|s,i,y_s^*)\mu_t^*$ *for all* $i \in A_s$, $s \in S$

Proof:

The 'if'-part follows directly from the lemmas 1.5.7, 1.5.8 and theorem 5.2.5. The 'only-if'-part can be shown in a similar way as the proof for the 'only-if' part of theorem 5.3.2. ■

Lemma 6.4.1 directly leads to the following theorem.

6.4.2 Theorem

For a general-sum stochastic game with $\gamma^k(x,y)=0$ for all $x \in X$, $y \in Y$, we have: a pair of stationary strategies $(x^*,y^*) \in X \times Y$ is a stationary total equilibrium with total reward $(\alpha^{1*},\alpha^{2*})$ if and only if there exist $\delta^{1*},\delta^{2*},\mu^{1*},\mu^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*},\mu^{1*},\mu^{2*})$ is a global minimum in the following non-linear program with objective value 0.

NLP 6.4.2:

variables $x \in \displaystyle\mathop{\times}_{s=1}^{z} \mathbb{R}^{m_s}, y \in \mathop{\times}_{s=1}^{z} Y^{n_s}, \alpha^1,\alpha^2,\delta^1,\delta^2,\mu^1,\mu^2 \in \mathbb{R}^z$

minimize $\displaystyle\sum_{k=1}^{2} \sum_{s=1}^{z} [\alpha_s^k - r^k(s,x_s,y_s) - \sum_{t=1}^{z} p(t|s,x_s,y_s)\alpha_t^k]$

subject to

a) $\displaystyle\alpha_s^1 \geq r^k(s,i,y_s) + \sum_{t=1}^{z} p(t|s,i,y_s)\alpha_t^1$ *for all* $i \in A_s$, $s \in S$

$\displaystyle\alpha_s^2 \geq r^k(s,x_s,j) + \sum_{t=1}^{z} p(t|s,x_s,j)\alpha_t^2$ *for all* $j \in B_s$, $s \in S$

b) $\displaystyle\alpha_s^1 + \delta_s^1 = \sum_{t=1}^{z} p(t|s,x_s,y_s)\delta_t^1$ *for all* $s \in S$

$\displaystyle\alpha_s^2 + \delta_s^2 = \sum_{t=1}^{z} p(t|s,x_s,y_s)\delta_t^2$ *for all* $s \in S$

c) $\displaystyle\alpha_s^1 + \mu_s^1 \geq \sum_{t=1}^{z} p(t|s,i,y_s)\mu_t^1$ *for all* $i \in A_s$, $s \in S$

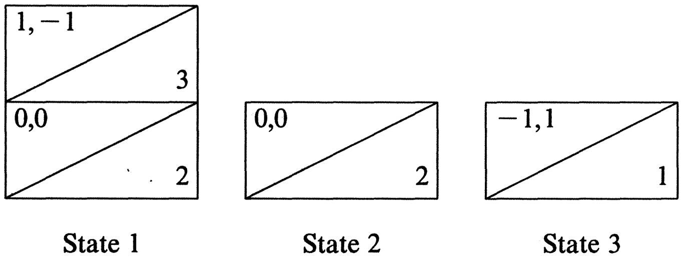
$\displaystyle\alpha_s^2 + \mu_s^2 \geq \sum_{t=1}^{z} p(t|s,x_s,j)\mu_t^2$ *for all* $j \in B_s$, $s \in S$

d) $\displaystyle\sum_{i=1}^{m_s} x_s(i) = 1, \sum_{j=1}^{n_s} y_s(j) = 1$ *for all* $s \in S$

e) $x_s \geq 0, y_s \geq 0$ *for all* $s \in S$.

Like in the previous sections we wonder whether or not near-optimal solutions of NLP 6.4.2 correspond with total ϵ-equilibria. For the total reward criterion we find that, as for the limiting average criterion, there is no such correspondence. This is illustrated by the next example.

6.4.3 Example



Let $x=((1-\epsilon,\epsilon),1,1)$, $0<\epsilon<1$, $y=(1,1,1)$, $\alpha^1=(1,0,0)$, $\delta^1=(1/\epsilon,2/\epsilon,1/\epsilon)$, $\alpha^2=(0,0,1)$, $\delta^2=(1,1/\epsilon,0)$ and let $\mu^1=\mu^2=(0,0,0)$. Furthermore let

$x^* = ((1,0),1,1)$. Then $(x,y,\alpha^1,\alpha^2,\delta^1,\delta^2,\mu^1,\mu^2)$ is a feasible solution of NLP 6.4.2 with objective value ϵ. However $\gamma_T^1(1,x,y) = 0 < ½ = \gamma_T^1(1,x^*,y)$ and hence (x,y) is no total ϵ-equilibrium for $\epsilon \in (0,½)$.

Since example 6.4.3 is a zero-sum stochastic game we conclude that the restriction of NLP 6.4.2 to zero-sum stochastic games, does not yield a program for which near-optimal solutions correspond with stationary total ϵ-optimal strategies. Nevertheless we present a non-linear program which has this property.

6.4.4 THEOREM

For a zero-sum stochastic game with $\gamma^k(x,y) = 0$ for all $x \in X$, $y \in Y$, we have:

If there exist $x^* \in \bigtimes_{s=1}^{z} \mathbb{R}^{m_s}$, $y^* \in \bigtimes_{s=1}^{z} \mathbb{R}^{n_s}$, $\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a feasible solution with objective value $\epsilon(\geqslant 0)$ in NLP 6.4.4 below, then x^* and y^* are stationary total ϵ-optimal strategies for the respective players.

Conversely, if x^* and y^* are stationary total ϵ-optimal strategies, then there exist $\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*} \in \mathbb{R}^z$ such that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a feasible solution with objective value less than $2z\epsilon$ in NLP 6.4.4.

NLP 6.4.4:

variables $x \in \bigtimes_{s=1}^{z} \mathbb{R}^{m_s}, y \in \bigtimes_{s=1}^{z} \mathbb{R}^{n_s}, \alpha^1,\alpha^2,\delta^1,\delta^2 \in \mathbb{R}^z$

minimize $\sum_{s=1}^{z} (\alpha_s^1 + \alpha_s^2)$

subject to

a) $\alpha_s^1 \geqslant r^1(s,i,y_s) + \sum_{t=1}^{z} p(t|s,i,y_s)\alpha_t^1$ *for all* $i \in A_s$, $s \in S$

$\alpha_s^2 \geqslant -r^1(s,x_s,j) + \sum_{t=1}^{z} p(t|s,x_s,j)\alpha_t^2$ *for all* $j \in B_s$, $s \in S$

b) $\alpha_s^1 + \delta_s^1 \geqslant \sum_{t=1}^{z} p(t|s,i,y_s)\delta_t^1$ *for all* $i \in A_s$, $s \in S$

$\alpha_s^2 + \delta_s^2 \geqslant \sum_{t=1}^{z} p(t|s,x_s,j)\delta_t^2$ *for all* $j \in B_s$, $s \in S$

c) $\sum_{i=1}^{m_s} x_s(i) = 1$, $\sum_{j=1}^{n_s} y_s(j) = 1$ *for all* $s \in S$

d) $x_s \geqslant 0, y_s \geqslant 0$ *for all* $s \in S$.

PROOF:

Suppose that $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is feasible in NLP 6.4.4 with objective value ϵ. By constraints (c) and (d): $x^* \in X$, $y^* \in Y$. By constraints (a) and (b) we have, using lemma 1.5.8, that:

$\gamma_T^1(s,x,y^*) \leqslant \alpha_s^{1*}$ and $-\gamma_T^1(s,x^*,y) = \gamma_T^2(s,x^*,y) \leqslant \alpha_s^{2*}$ for all $x \in X, y \in Y, s \in S$.

Hence $\gamma_T^1(s,x,y^*) - \gamma_T^1(s,x^*,y) \leqslant \alpha_s^{1*} + \alpha_s^{2*}$ for all $x \in X, y \in Y, s \in S$.

Especially $\alpha_s^{1*} + \alpha_s^{2*} \geqslant \gamma_T^1(s,x^*,y^*) - \gamma_T^1(s,x^*,y^*) = 0$ for all $s \in S$. Since $\sum_{s=1}^{z} (\alpha_s^{1*} + \alpha_s^{2*}) = \epsilon$ we conclude that $\epsilon \geqslant \alpha_s^{1*} + \alpha_s^{2*} \geqslant 0$ for all $s \in S$.

Hence for all $s \in S$, $x \in X$ and $y \in Y$ we have:

$$\gamma_T^1(s,x^*,y) + \epsilon \geqslant \gamma_T^1(s,x^*,y^*) \geqslant \gamma_T^1(s,x,y^*) - \epsilon,$$

which means that x^* and y^* are stationary total ϵ-optimal strategies.

Conversely, if x^* and y^* are stationary total ϵ-optimal strategies, then constraints (c) and (d) are satisfied. By solving MDP(x^*) there exist α^{2*} and δ^{2*} such that (a) and (b) for player 2 are satisfied (cf. lemma 6.4.1). Similarly one can find α^{1*} and δ^{1*} by solving MDP(y^*). Then $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is feasible in NLP 6.4.4. By lemma 6.4.1 we have for all $s \in S$:

$$\alpha_s^{1*} = \max_{x \in X} \gamma_T^1(s,x,y^*) \leqslant \gamma_T^1(s,x^*,y^*) + \epsilon \text{ and}$$

$$\alpha_s^{2*} = \max_{y \in Y} \gamma_T^2(s,x^*,y) \leqslant \gamma_T^2(s,x^*,y^*) + \epsilon = -\gamma_T^1(s,x^*,y^*) + \epsilon. \qquad \blacksquare$$

It should be observed that NLP 6.4.4 finds 'best' stationary strategies for the total reward criterion.
Here a 'best' pair of stationary strategies is a pair (x^*,y^*) such that
$d_T(x^*,y^*) := \sum_{s=1}^{z} [\max_{x \in X} \gamma_T^1(s,x,y^*) - \min_{y \in Y} \gamma_T^1(s,x^*,y)]$ is (near-)minimal.

Without proofs we formulate some implications of theorem 6.4.4. They can be proved in a similar way as the corresponding results for the limiting average reward criterion were proved.

6.4.5 Corollary
If $(x^*,y^*,\alpha^{1*},\alpha^{2*},\delta^{1*},\delta^{2*})$ is a global minimum in NLP 6.4.4 with objective value $\mu \geqslant 0$, then $\mu = d_T(x^*,y^*) \leqslant d_T(x,y)$ for all $x \in X$, $y \in Y$.

6.4.6 Corollary
Suppose that the minimum in NLP 6.4.4 does not exist, but that the infimum equals $\mu\,(\geqslant 0)$, then for every $\epsilon > 0$ there exist stationary total $(\mu+\epsilon)$-optimal strategies for both players.

6.4.7 Corollary
For a zero-sum stochastic game both players have stationary total ϵ-optimal strategies for all $\epsilon > 0$, if and only if the infimum of NLP 6.4.4 is equal to 0.

References

R.J. Aumann [1964]: *Mixed and behavior strategies in infinite extensive games.* In: Advances in Game Theory, Annals of Mathematical Studies 52, Princeton University Press, 627-650.

R.J. Aumann [1981]: *Survey of repeated games.* In: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern, Bibliographisches Institüt, Mannheim, 11-42.

R. Bellman [1952]: *On the theory of dynamic programming.* Proceedings of the National Academy of Sciences U.S.A. 38, 716-719.

R. Bellman [1957]: *Dynamic Programming.* Princeton University Press.

T. Bewley & E. Kohlberg [1976]: *The asymptotic theory of stochastic games.* Mathematics of Operations Research 1, 197-208.

T. Bewley & E. Kohlberg [1978]: *On stochastic games with optimal stationary strategies.* Mathematics of Operations Research 3, 104-125.

P. Billingsley [1979]: *Probability and Measure.* John Wiley & Sons, New York.

D. Blackwell [1962]: *Discrete dynamic programming.* Annals of Mathematical Statistics 33, 719-726.

D. Blackwell & T.S. Ferguson [1968]: *The big match.* Annals of Mathematical Statistics 39, 159-163.

H. Everett [1957]: *Recursive games.* In: M. Dresher, A.W. Tucker & P. Wolfe (eds.), Contributions to the Theory of Games III, Annals of Mathematical Studies 39, Princeton University Press, 47-78.

A. Federgruen [1978]: *On n-person stochastic games with denumerable state space.* Advances in Applied Probability 10, 452-471.

J.A. Filar [1981-a]: *A single-loop stochastic game which one player can terminate.* Opsearch 18, 185-203.

J.A. Filar [1981-b]: *Ordered field property for stochastic games when the player who controls transitions changes from state to state.* Journal of Optimization Theory and Applications 34, 503-515.

J.A. Filar [1984]: *On stationary equilibria of a single controller stochastic game.* Mathematical Programming 30, 313-325.

J.A. Filar [1986]: *Quadratic programming and the single controller stochastic game.* Journal of Mathematical Analysis and Applications 113, 136-147.

J.A. Filar & T.E.S. Raghavan [1984]: *A matrix game solution of the single controller stochastic game.* Mathematics of Operations Research 9, 356-362.

J.A. Filar & T.A. Schultz [1986]: *Nonlinear programming and stationary strategies in stochastic games.* Mathematical Programming 35, 243-247.

J.A. Filar & T.A. Schultz [1987]: *Bilinear programming and structured stochastic games.* Journal of Optimization Theory and Applications 53, 85-104.

J.A. Filar, T.A. Schultz, F. Thuijsman & O.J. Vrieze [1991]: *Nonlinear programming and stationary equilibria in stochastic games.* Mathematical Programming 50, 227-237.

J.A. Filar & O.J. Vrieze [1989]: *Weighted reward criteria in competitive Markov* *decision processes.* Proceedings of the 6th IFAC Symposium on Dynamic Modelling and Control of National Economies, 93-98.

A.M. Fink [1964]: *Equilibrium in a stochastic n-person game.* Journal of Science of Hiroshima University, Series A-I 28, 89-93.

D. Gillette [1957]: *Stochastic games with zero stop probabilities.* In: M. Dresher, A.W. Tucker & P. Wolfe (eds.), Contributions to the Theory of Games III, Annals of Mathematical Studies 39, Princeton University Press, 179-187.

A.J. Hoffman & R.M. Karp [1966]: *On nonterminating stochastic games.* Management Science 12, 359-370.

A. Hordijk & L.C.M. Kallenberg [1979]: *Linear programming and Markov decision chains.* Management Science 25, 352-362.

A. Hordijk & L.C.M. Kallenberg [1981]: *Linear programming and Markov games II.* In: O. Moeschlin, D. Pallaschke (eds.), Game Theory and Mathematical Economics, North Holland Publishing Company, Amsterdam, 307-320.

A. Hordijk, O.J. Vrieze & G.L. Wanrooij [1983]: *Semi-Markov strategies in stochastic games.* International Journal of Game Theory 12, 81-89.

L.C.M. Kallenberg [1983]: *Linear Programming and Finite Markovian Control Problems.* Mathematical Centre Tract 148, Centre for Mathematics and Computer Science, Amsterdam.

J. Kemeny & J. Snell [1960]: *Finite Markov Chains.* Van Nostrand, Princeton.

E. Kohlberg [1974]: *Repeated games with absorbing states.* Annals of Statistics 2, 724-738.

A. Kolmogorov [1933]: *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse der Mathematik 2, no. 3, Springer Verlag, Berlin.

D. Krass, J.A. Filar & S. Sinha [1987]: *A weighted Markov decision process.* Technical report, Department of Mathematics, University of Maryland Baltimore County, Baltimore.

T.M. Liggett & S.A. Lippman [1969]: *Stochastic games with perfect information and time average payoff.* SIAM Review 11, 604-607.

O.L. Mangasarian & H. Stone [1964]: *Two-person non-zerosum stochastic games and quadratic programming.* Journal of Mathematical Analysis and Applications 9, 348-355.

J.F. Mertens [1986]: *Repeated games.* CORE reprint 8624, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve.

J.F. Mertens & A. Neyman [1981]: *Stochastic games.* International Journal of Game Theory 10, 53-66.

J. Nash [1951]: *Non-cooperative games.* Annals of Mathematics 54, 286-295.

J. von Neumann [1928]: *Zur Theorie der Gesellschaftsspiele.* Mathematische Annalen 100, 295-320.

J. von Neumann & O. Morgenstern [1944]: *Theory of Games and Economic Behavior.* Princeton University Press.

A. Neyman [1986]: Private communication.

M. Orkin [1972]: *Recursive matrix games.* Journal of Applied Probability 9, 813-820.

T. Parthasarathy & T.E.S. Raghavan [1981]: *An orderfield property for stochastic games when one player controls transition probabilities.* Journal of Optimization Theory and Applications 33, 375-392.

T. Parthasarathy & M.A. Stern [1977]: *Markov Games, a survey.* In: E. Roxin, P. Lieu & R. Sternberg (eds.), Differential Games and Control Theory 2, Marcel Dekker, New York, 1-46.

T. Parthasarathy, S.H. Tijs & O.J. Vrieze [1984]: *Stochastic games with state independent transitions and separable rewards.* In: G. Hammer & D. Pallaschke (eds.), Selected Topics in Operations Research and Mathematical Economics, Springer Verlag, Berlin 262-271.

T.E.S. Raghavan & J.A. Filar [1989]: *Algorithms for stochastic games, a survey.* Report, Department of Mathematics, University of Illinois, Chicago.

T.E.S. Raghavan, S.H. Tijs & O.J. Vrieze [1985]: *On stochastic games with additive reward and transition structure.* Journal of Optimization Theory and Applications 47, 451-464.

P.D. Rogers [1969]: *Non-zerosum stochastic games.* Ph.D. thesis, Report ORC 69-8, Operations Research Center, University of California, Berkeley.

U.G. Rothblum [1979]: *Solving stopping stochastic games by maximizing a linear function subject to quadratic constraints.* In: O. Moeschlin & D. Pallaschke (eds.), Game Theory and Related Topics, North-Holland Publishing Company, Amsterdam, 103-105.

T.A. Schultz [1987]: *Mathematical programming and stochastic games.* Ph.D. thesis, Johns Hopkins University, Baltimore.

L.S. Shapley [1953]: *Stochastic games.* Proceedings of the National Academy of Sciences U.S.A. 39, 1095-1100.

L.S. Shapley & R.N. Snow [1950]: *Basic solutions of discrete games.* In: H.W. Kuhn & A.W. Tucker (eds.), Contributions to the Theory of Games I, Annals of Mathematical Studies 24, Princeton University Press, 27-35.

M.J. Sobel [1971]: *Noncooperative stochastic games.* Annals of Mathematical Statistics 42, 1930-1935.

M.J. Sobel [1981]: *Myopic solutions of Markov decision processes and stochastic games.* Operations Research 29, 995-1009.

S. Sorin [1986]: *Asymptotic properties of a non-zerosum stochastic game.* International Journal of Game Theory 15, 101-107.

S. Sorin [1988]: *Repeated games with complete information.* CORE discussion paper 8822, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve.

M.A. Stern [1975]: *On stochastic games with limiting average payoff.* Ph.D. thesis, University of Illinois, Chicago.

M. Takahashi [1964]: *Equilibrium points of stochastic noncooperative n-person games.* Journal of Science of Hiroshima University, Series A-I 28, 95-99.

A. Tarski [1951]: *A Decision Method for Elementary Algebra and Geometry.* Second edition, revised, University of California Press, Berkeley and Los Angeles.

F. Thuijsman [1987]: *Non-zerosum stochastic games.* In: H.J.M. Peters & O.J. Vrieze (eds.), Surveys in Game Theory and Related Topics, CWI-tract 39, Centre for Mathematics and Computer Science, Amsterdam, 133-161.

F. Thuijsman & O.J. Vrieze [1987]: *The bad match, a total reward stochastic game.* OR Spektrum 9, 93-99.

F. Thuijsman & O.J. Vrieze [1990-a]: *Stationary ϵ-optimal strategies in stochastic games.* Report M90-07, Department of Mathematics, University of Limburg, Maastricht.

F. Thuijsman & O.J. Vrieze [1990-b]: *Note on recursive games.* Report M90-11, Department of Mathematics, University of Limburg, Maastricht.

F. Thuijsman & O.J. Vrieze [1991]: *Easy initial states in stochastic games.* In: T.E.S. Raghavan, T.S. Ferguson, O.J. Vrieze & T. Parthasarathy (eds.), Stochastic Games and Related Topics, Kluwer Academic Publishers, Dordrecht, 85-100.

S.H. Tijs & O.J. Vrieze [1986]: *On the existence of easy initial states for undiscounted stochastic games.* Mathematics of Operations Research 11, 506-513.

O.J. Vrieze [1981]: *Linear programming and undiscounted stochastic games in which one player controls transitions.* OR Spektrum 3, 29-35.

O.J. Vrieze [1983]: *Discounted stochastic games and mathematical programming.* Report 8352, Department of Mathematics, Catholic University, Nijmegen.

O.J. Vrieze [1987-a]: *Stochastic Games with Finite State and Action Spaces.* CWI-tract 33, Centre for Mathematics and Computer Science, Amsterdam.

O.J. Vrieze [1987-b]: *Zero-sum stochastic games.* In: H.J.M. Peters & O.J. Vrieze (eds.), Surveys in Game Theory and Related Topics, CWI-tract 39, Centre for Mathematics and Computer Science, Amsterdam, 103-162.

O.J. Vrieze & F. Thuijsman [1987]: *Stochastic games and optimal stationary strategies, a survey.* In: W. Domschke, W. Krabs, J. Lehn & P. Spellucci (eds.), Methods of Operations Research 57, 513-529.

O.J. Vrieze & F. Thuijsman [1989]: *On equilibria in repeated games with absorbing states.* International Journal of Game Theory 18, 293-310.

O.J. Vrieze & S.H. Tijs [1980]: *Relations between the game parameters, value and optimal strategy spaces in stochastic games and construction of games with given solution.* Journal of Optimization Theory and Applications 31, 501-513.

O.J. Vrieze, S.H., Tijs, T.E.S. Raghavan & J.A. Filar [1983]: *A finite algorithm for the switching control stochastic game.* OR Spektrum 5, 15-24.

P. Wolfe [1956]: *Determinateness of polyhedral games.* In: H.W. Kuhn, A.W. Tucker (eds.), Linear Inequalities and Related Systems I, Annals of Mathematical Studies 38, Princeton University Press, Princeton, 195-198.

Author index

Subject index

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. *Preservations 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 statements.* 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. *Empirical 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 Wallmann 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 renewai 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 sequencies.* 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.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.

59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.

60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict or conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1991.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1992.

76 O.E. Flippo. *Stability, duality and decomposition in general mathematical programming.* 1991.

77 A.J. van Es. *Aspects of nonparametric density estimation.* 1991.

78 G.A.P. Kindervater. *Exercises in parallel combinatorial computing.* 1992.

79 J.J. Lodder. *Towards a symmetrical theory of generalized functions.* 1991.

80 S.A. Smulders. *Control of freeway traffic flow.* 1992.

81 P.H.M. America, J.J.M.M. Rutten. *A parallel object-oriented language: design and semantic foundations.* 1992.

82 F. Thuijsman. *Optimality and equilibria in stochastic games.* 1992.

83 R.J. Kooman. *Convergence properties of recurrence sequences.* 1992.

84 A.M. Cohen (ed.). *Computational aspects of Lie group representations and related topics. Proceedings of the 1990 Computational Algebra Seminar at CWI, Amsterdam.* 1991.