A CONTRIBUTION TO THE THE THEORY OF STOCHASTIC GAMES

Sagnik Sinha

THESIS

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

to

Indian Statistical Institute,
New Delhi, India.

June,1989

[41] Stern, M.A(1975): On stochastic games with limiting average pay off, Ph.D. thesis, Univ. of Illinois, Chicago.

[42] Takahashi, M(1964): Equilibrium points of stochastic, noncooperative n-person games, J. Sci. Hiroshima Univ. series A, 28, 95-99.

[43] Thuijsman, F. and Vrieze, O.J(1986): On equilibria in repeated games with absorbing state, technical report, Dept. of Math., Catholic University, Nijmegen, Netherland.

[44] Tijs, S.H(1975): Semi-infinite and infinite matrix games and bimatrix games, Ph.D. Thesis.

[45] Tijs, S.H. and Vrieze, O.J(1979): Characterizing properties of the value function of stochastic games, Report BW 110/79, Math. Centre, Amsterdam, Netherland.

[46] Vrieze, O.J(1983): Stochastic games with finite state and action spaces, Ph.D. thesis, Math. Centre, Amsterdam, Netherland.

[47] Wald, A(1950): Statistical decision functions, New York, Wiley and Sons Inc.

[48] Weyl, H(1950): Elementary proof of a minimax theorem due to von Neumann, Contributions to the theory of games, I, 19-25.

[49] Whitt, W(1980): Representation and approximation of noncooperative sequential games, SIAM Journal on Control and Optimization, 18, 33-48.

Preface

This thesis concentrates on noncooperative stochastic games. Both zero sum and nonzero sum stochastic games have been analysed. The main results are the following:

1. For a zero sum two person finite stochastic game, where some states are switching control states, and the remaining are states with additive reward and transition structure, the value and a pair of stationary optimal strategies exist under discounted and undiscounted payoff criteria. Further, if the data describing such games are rational numbers, the value and a pair of stationary optimal strategies can be found whose entries are rational numbers. In other words, such games have the ordered field property under discounted and undiscounted payoff criteria.

2. For a zero sum two person semi-infinite stopping stochastic game, with finitely many states, finite actions for one player and countably infinite actions for the other player in each state, the value exists and the player with finite actions has stationary optimal strategies. However, the other player (with countable actions), has near optimal stationary strategies. One can charcterize the value vector as a maximal solution of the Shapley equation of this game. Optimal stationary strategies for the player with finite actions can be charterized from the Shapley matrices. A near optimal stationary strategy for the player with countable actions can be constructed by considering different parts of the Shapley matrices in different states.

3. For an N-person discounted stochastic game with compact state space and finite action spaces, there exist stationary equilibrium strategies when the reward function is continuous and bounded and the transition function is state independent and is dominated by a fixed nonatomic measure.

Acknowledgements

First of all, I want to thank my supervisor Professor T. Parthasarathy. I am indebted to him for introducing the field of stochastic games to me. Without his help and constant encouragement, this work might not have been complete.

There are a number of people who helped me directly or indirectly in the preparation of this thesis. Among them, in particular, I want to thank Professor S.H.Tijs, Catholic University at Nijmegen; Professor O.J.Vrieze and F.Thuijsman, University of Limburg; Professor T.E.S.Raghavan, University of Illinois at Chicago; Professor J.A.Filar, University of Maryland, Baltimore and Professor A.Nowak, University of Wroclaw, for helping me through various stimulating discussions. I want to thank Professor Tijs and Frank Thuijsman for their kind hospitality during my stay at Nijmegen.

I thank Mr.B.Ganeshan who learnt a word-processor package practically from scratch to type my thesis properly.

I also acknowledge National Council of Educational Research and Training and National Board of Higher Mathematics for the financial support they provided for this work.

I want to thank an unknown referee for detecting a mistake in an earlier version of this thesis.

At last, I want to thank my wife and family who provided the necessary inspiration and moral support to finish this work.

Sagnik Sinha
New Delhi.

Contents

Chapter 1

Introduction

1.1 Theory of Games — A Related Coverage:

A mathematical theory of *Games of Strategy* was born in several stages between 1928 and 1941. John von Neumann is known as its father. The culmination of the pioneering work of von Neumann and Morgenstern was the publication of the *Theory of Games and Economic Behavior* [Ref.28] in 1944. It is said of the book '.... posterity may regard this book as one of the major scientific achievements of the first half of the twentieth century'. Emphasizing a new approach to competetive behaviour through a mathematical reduction to suitable games of strategy, this giant work laid bare a host of problems in the mathematical theory of games.

A *game* is simply a set of descriptive rules. A *play* of the game includes every particular instance in which the game is played from beginning to end. The participants are the ***players***. A game consists of a sequence of *moves* of the players, while a play comprises a sequence of choices made by them. The decisive step in the mathematical treatment of games is the normalization achieved by introduction of pure strategies (actions). A ***pure strategy*** is a plan formulated by a player prior to the start of a play, which covers all of the possible decisions which he may face during any play permitted by the rules of the game. The expected course of a play is thereby completely determined by the selection of a pure strategy by each player in ignorance of that chosen by any other player. von Neumann first considered games with a finite number of pure strategies *i.e.* finite games. The theory was later extended to games with an infinite number of pure strategies . Both the notions will be useful in the sequel .

1.2 (Finite) Zero Sum Two Person Games:

Such games are played by two players and what one player wins, the other loses. A zero sum two person game with a finite number of pure strategies can be described by an $m \times n$ matrix $A = ((a_{ij}))$, where a_{ij}(or $a(i,j)$) is the payment to player **I** by player **II** if **I** chooses his i^{th} pure strategy and **II** chooses his j^{th} pure strategy. However, simplest games like matching pennies show that a player is worse off if he always uses the same pure strategy. Instead, during repeated plays of the game, he can be better off by mixing his various pure strategies randomly, but with chosen relative frequencies. Subsequently, a passage from such a 'sample' of plays of the game to the underlying 'population' reveals the notion of a mixed strategy (mixed actions)as a probability distribution over the set of pure strategies. *Mixed strategies* for I and II will be denoted by x and y respectively, where $x = (x_1, x_2.....x_m)$, $y = (y_1, y_2....y_n)$; $x_i \geq 0$ for all i and $y_j \geq 0$ for all j with $\sum_i x_i = 1 = \sum_j y_j$. When I plays a mixed strategy x and II plays a mixed strategy y, the expected payment by II to I is the value of a bilinear function ϕ at the point (x, y). The real valued function ϕ is defined on the cartesian product of X, the set of mixed strategies available to I, and Y, the same for II and $\phi(x, y) = \sum_i \sum_j x_i a_{ij} y_j = x^t A y$. All discussions of zero sum two person games start from the *Minimax theorem* [Ref.26] which asserts:

$$\max_X \min_Y \phi(x, y) = \min_Y \max_X \phi(x, y)$$

. The unique minimax value of ϕ is called the value of the game A and will be denoted by $v(A)$. Mixed strategies x^* and y^* such that $\phi(x^*, y) \geq \phi(x, y^*)$ for all mixed strategies x and y are called *optimal* or *good* strategies for player I and player II respectively .

This study deals with the following three major consequences of minimax theorem viewed in the light of *Stochastic games*, a class of recursive games which have moves as plays of other games .

1.2.1 Ordered field property in Matrix Games:

In terms of the preceding discussion, the minimax theorem can be stated as:

Any matrix game A has a value v and non-empty sets of optimal mixed strategies $x^* = (x_1^*, x_2^* \ldots, x_m^*)$ and $y^* = (y_1^*, y_2^* \ldots, y_n^*)$ characterised by the following linear inequalities:

$$\sum_i x_i^* a_{ij} \geq v \text{ for all } j; \; \sum_j a_{ij} y_j^* \leq v \text{ for all } i$$

and x_i^*'s are nonnegative numbers adding to unity and the same is true for the y_j^*'s. Thus, in the above form the minimax theorem is nothing but a problem in systems in linear inequalities. Noting the above fact plus the intimate relation between linear inequalities and the geometry of convex polyhedra, Herman Weyl(1950) [Ref.48] was the first to prove the minimax theorem in a completely algebraic way: The value and a pair of optimal strategies of any matrix game can be found using only finitely many elementary operations (i.e addition, subtraction, multiplication and division by a nonzero number) and the elementary decision to judge whether a given number is nonnegative or not. Thus if the data describing the matrix game are from an ordered field (like the field of rational numbers), the value and a pair of optimals can be found without going out of that field. Nowadays this property of a zero sum two person finite game is referred as *ordered field property* in matrix games.

1.2.2 Semi-Infinite Matrix Games:

The fundamental minimax theorem of von Neumann explains the optimal strategic behaviour of the players in a zero sum two person finite game. In such a game, if the set of pure strategies of one player ceases to be finite, suitable modifications are needed to obtain a value concept. Abraham Wald [Ref.47] first considered such games where the action space of one player (say player **II**) is infinite and in particular, countably infinite, while the other player has finitely many pure strategies. Such games are referred as semi-infinite matrix games after Stef Tijs [Ref.44], who did a systemic study of such games. To obtain a minimax theorem for such games, one need to consider the extension from pure strategies to mixed strategies of the players, particularly of player **II**. Unless this is done properly, it is meaningless to talk about the value because the expected pay off may not be well defined. Tijs considered three different extensions and one of them, called compact mixed extension (c-mixed extension) which restricts player **II** within the class of mixed strategies with finite support, will be useful in the sequel.

Tijs proved a minimax theorem in the class of semi-infinite matrix games where player **II** (with countably many pure strategies) uses mixed strategies with finite support only. The theorem is proved without any assumption on the boundedness of the pay off. As a result, the value of such games can be a real number or $-\infty$, player **I** has an optimal strategy and player **II** has near optimal strategies. The result can be stated as:

Let $A_{m\times\infty}$ be a semi-infinite matrix game. Suppose that player **I** is free to use any mixed strategy but player **II** can use mixed strategies only from the set $Y^c = \{(y_1, y_2, \ldots)$: there exists a natural number n such that $y_k = 0$ for all $k > n$; $y_k \geq 0$ for all $k; \sum_{k=1}^{\infty} y_k = 1\}$. Then there exists : (a) a $v \in \mathbf{R} \bigcup \{-\infty\}$, (b) a mixed strategy $\bar{x}$ for player **I**. (c) for each $u > v$ (in case $v = -\infty, u$ can be any real number), a mixed strategy $\bar{y} \in Y^c$ for player II, satisfying

$$\sum_{i=1}^{m} \bar{x}_i a_{ij} \geq v \text{ for all } j,$$

$$\sum_{j=1}^{\infty} a_{ij} \bar{y}_j \leq u \text{ for all } i$$

1.2.3 Noncooperative games:

For finite zero sum two person games, the minimax theorem of von Neumann explains the strategic optimal behaviour of the players and guarantees the existence of a (unique) value for such games. In a zero sum two person game, the interests of the two players are mutually opposite, because what a player wins, the other loses. There are certain other classes of two person games (e.g. constant sum games) which can be transformed into a zero sum game without changing the optimal strategic behaviour of the players. However, there are games involving possibly more than two players where they do not antagonise one another directly as in zero sum two person games. Such games are called N-person nonzero sum games. Minimax theorems fail to analyse the strategic behaviour in such game situations. A class of nonzero sum games is known as noncooperative games. In a noncooperative game, players cannot cooperate in the sense that no side payments, threats or promises are allowed. As in a two person zero sum game, players have to choose their strategies independently of one another.

Formally, a general N-person finite noncooperative game can be described by a $2N$ tuple $\{(A_i, K_i) : i = 1, 2, \ldots, N\}$. N players play the game and for each i, A_i is a finite set of pure strategies available for the i^{th} player. K_i is a real valued function, defined on the cartesian product of the A_i's and is called the reward function or the pay off function for player i. A play of the game involves (i) a simultaneous (and hence independent) choice of a pure strategy by each player, and (ii) if $(a_1, a_2, \ldots, a_N)$ is a choice made by the players, player i gets a reward $K_i(a_1, a_2, \ldots, a_N)$. Each player wants to maximise his pay off through strategic choice. For such games, minimax theorems fail to answer the question of optimal behaviours of the players.

In 1950, Nash [Ref.25] introduced the notion of equilibrium in the class of mixed strategies in a noncooperative game. He showed that there exists an equilibrium point which is a set of mixed strategies, one for each player, such that no player can be better off by deviating unilaterally from this point. Mathematically, this guarantees the existence of a mixed strategy tuple $\{\alpha_i^* : i = 1, 2, \ldots N\}$, α_i^* being a mixed strategy of player i (i.e. α_i^* is a probability distribution over A_i), such that

$$K_i(\alpha_1^*, \alpha_2^*, \ldots \alpha_{i-1}^*, \alpha_i, \alpha_{i+1}^* \ldots \alpha_N^*)$$
$$\leq K_i(\alpha_1^*, \alpha_2^*, \ldots \alpha_{i-1}^*, \alpha_i^*, \alpha_{i+1}^*, \ldots, \alpha_N^*),$$

for any mixed strategy α_i for player i, for each $i = 1, 2, \ldots N$. Such equilibrium points are called Nash equilibrium points or equilibrium in the sense of Nash and have wide applications in the literature as a suitable solution concept.

1.3 Stochostic Games:

History: In 1951, a group of mathematicians at the RAND corporation initiated the study of a class of *multimove games*: infinite games which are built up out of a set of ***components*** (now known as ***states***) which are games, or game like structures, themselves of finite length. One of such components (states) is prescribed as ***start***, and the outcome of each of the components (states) is an instruction to play another component (state) of the set of components (state space), possibly together with a numerical pay off. Thus the entire infinite game is recursively defined by the set of rules of the components (states).

The aforesaid group of mathematicians concentrated on repeated play of ***simultaneous move games*** . In their models, each component is already a game in normalized form, each player being uninformed of the others' present choice but completely informed of all choices made in components (states) played previously and hence knowing in which component (state) he is playing at the present moment of choice. Shapley's [Ref.38] *Stochastic Games* (which are now called *Stopping Stochastic Games*) evolved as a class of the above mentioned games which have ***moves as plays of other games***.

Motivation: The notion of mixed actions in a matrix game incorporates intrinsically the repeated plays of the same game with a fixed relative frequency allotted for each pure actions available. In 1953, Shapley [Ref.38] introduced the idea of not playing the same matrix game everyday (i.e. in every repitition, now called a ***stage*** of the game), but playing one of a finite collection of matrix games, with a motion among this collection from stage to stage governed by the current matrix game and the actions chosen there. To be more precise, there are z matrices $M^1, M^2, \ldots M^z$, matrix M^s has size $m_s \times n_s$ and entry (i,j) of matrix M^s is given as: $\boxed{\begin{array}{c} r(s,i,j) \\ \hline p(s,i,j) \end{array}}$, where $r(s,i,j) \in \mathbf{R}$ is a payment to player **I** and $p(s,i,j) = (p(1 \mid s,i,j), p(2 \mid s,i,j) \ldots, p(z \mid s,i,j)) \in \Delta_0^z = \{x \in \mathbf{R}^z : x \geq 0, \sum_{s=1}^{z} x_s < 1\}$. The interpretation is: if in state s player **I** chooses $i \in A^s = \{1, 2, \ldots, m_s\}$ and player **II** chooses $j \in B^s = \{1, 2, \ldots, n_s\}$ then player **II** pays $r(s,i,j)$ to player **I** and on the next stage the play moves from state s to state t with probability $p(t \mid s,i,j)$, and the play stops with probability $1 - \sum_{t=1}^{z} p(t \mid s,i,j) > 0$.

Thus, these *stopping stochastic games* were born quite naturally from matrix games . In a stochastic game, the players consider *strategies*: infinite plans that tell a player at each stage, at each state and for every history of the play, what mixed action to be chosen. Strategies for which the prescribed mixed actions do not depend on the histories are called *Markov strategies*; Markov strategies for which the prescribed mixed actions do not depend on the stages are called *Stationary strategies.* Stationary strategies are simply z-tuples of mixed actions, one mixed action for each state. It is obvious that stationary strategies are most easy to handle and therefore Shapley, who is the initiator of this field, naturally concentrated only on stationary strategies. In the literature of stochastic games, the notion of nonstationary strategies was introduced by Everett [Ref.8] and Gillette [Ref.13] in 1957.

An initial state $s \in S$, together with a pair of strategies (π_1, π_2) for players **I** and **II** respectively, determine a stochastic process over the set of states $S = 1, 2, \ldots\ldots, z$, and hence for all stages $\tau \in \mathbf{N}$, an expected direct payoff

$R_{s\pi_1\pi_2}(\tau)$ by player **II** to player **I** is determined. The players evaluate this stream of expected payoffs $(R_{s\pi_1\pi_2}(1), R_{s\pi_1\pi_2}(2), \ldots)$ to be worth the reward $v(s, \pi_1, \pi_2) = \sum_{\tau=1}^{\infty} R_{s\pi_1\pi_2}(\tau)$.

By Shapley [Ref.38] and Blackwell [Ref.5] , it is known that there exists a unique $v = (v(s) : s \in S) \in \mathbf{R}^s$, a stationary strategy π^* for player **I** and a stationary strategy σ^* for player **II** such that for all strategies π and σ for players **I** and **II** respectively , the following holds:

$$v(s, \pi, \sigma^*) \leq v(s) \leq v(s, \pi^*, \sigma).$$

v is called the value of the stochastic game, π^* is called an optimal stationary strategy for player **I** and similarly σ^* for player **II**.

Further, Shapley [Ref.38, see also 45] showed that: v is the value of the stochastic game, π^* is an optimal stationary strategy for player **I** and σ^* an optimal stationary strategy for player **II** if and only if for all $s \in S$, one has that $v(s) = val\, G^s(v)$ and that $\pi^*(s)$ is an optimal mixed action for player **I** in $G^s(v)$ and that $\sigma^*(s)$ is an optimal mixed action for player **II** in $G^s(v)$ where $G^s(v) = ((\, r(s,i,j) + \sum_{t=1}^{s} p(t \mid s,i,j)v(t)\,))$. Hence the value and optimal stationary strategies can be found by solving the set of equations $x(s) = val\, G^s(x)$ for $s \in S$, $x \in \mathbf{R}^s$. This set of equations is known as the *Shapley equation* and can be shortened to $x = T(x)$, $x \in \mathbf{R}^s$, where T is the function from $\mathbf{R}^s$ to itself defined by $T(x)_s = val\, G^s(v)$. The equation $x = T(x)$ has a unique solution in $\mathbf{R}^s$ since T is a continuous contraction map.

Since 1953, the theory of stochastic games has been extended in various directions. Nonstopping stochastic games were introduced and have been studied under discounted and limiting average pay off criteria. A nonstopping stochastic game is different from a Shapley's stopping stochastic game in the sense that it has zero stop probabilities everywhere in the game, i.e. for $s \in S, i \in \{1, 2, \ldots, m_s\}, j \in \{1, 2, \ldots n_s\}, \sum_{t=1}^{s} p(t \mid s, i, j) = 1$. Under *$\beta-$discounted payoff criterion*, if strategies π_1 and π_2 are chosen by the players and the game starts at the s^{th} state, I gets a reward $v_\beta(s, \pi_1, \pi_2)$ $= \sum_{\tau=1}^{\infty} \beta^{\tau-1} R_{s\pi_1\pi_2}(\tau)$, where $\beta \in [0, 1)$ is a discount factor. Under ***limiting average (or undiscounted) payoff criterion*** introduced by Gillette [Ref.13] the above reward is $v(s, \pi_1, \pi_2) = \liminf_{N\to\infty}(1/N) \sum_{\tau=1}^{N} R_{s\pi_1\pi_2}(\tau)$. The question of the existence of value of a nonstopping stochastic game considered under these payoff criteria have been settled:

(1) By Shapley [Ref.38] and Blackwell [Ref.5], it is known that the value v_β and a pair of stationary optimal strategies exist in a β-discounted stochastic game for any discount factor $\beta \in [0, 1)$.

(2) Mertens and Neyman [Ref.23] have shown that an undiscounted stochastic game has a value v , but optimal strategies need not exist. However, ϵ-optimal strategies exist, for both the players, i.e., given $\epsilon > 0$, there is a pair of strategies $(\pi_\epsilon, \sigma_\epsilon)$ such that for all strategies π for player **I** and σ for player **II**, we have $v(s, \pi_\epsilon, \sigma) \geq v(s) - \epsilon$ and $v(s, \pi, \sigma_\epsilon) \leq v(s) + \epsilon$.

Following Nash's noncooperative games, nonzerosum noncooperative stochastic games (where each component is a non-cooperative game in normal form) have been studied under β-discounted and limiting average pay-off criterion. Results similar to zero sum stochastic games have been obtained in the β-discounted case [Ref.12, 37, 40, 42]. In the limiting average case, still work is going on and Thuijsman and Vrieze have proved

[Ref.43] the existence of a nonstationary equilibrium in the particular case of repeated noncooperative games with absorbing states.

1.4 About this study

This study is divided into four chapters including the present one, which is the introduction. Chapter 2 is attributed to a detail study of ordered field property in stochastic games. The known literature is reviewed and then two extension theorems have been derived. Certain other classes are also studied in the light of ordered field property and this chapter then ends with some remarks.

In Chapter 3, Semi infinite stopping stochastic games have been introduced, motivated and the existence of the value and optimal strategies have been settled. The Shapley equation is studied in a great detail and characterization of optimal and *near* optimal strategies for player I and II have been achieved. Lastly, some remarks have been made regarding possible field of research in this line.

Chapter 4, the last chapter, is devoted to Noncooperative stochastic games with uncountable state space considered under β-discounted pay off criterion. The existence of an equilibrium for such games have been recently settled by Mertens and Parthasarathy [Ref.24], but they could not prove the existence of stationary equilibrium strategies in the general case. In this chapter, the issue of existence of stationary equilibrium strategies is settled in the affirmative for a particular class of such games which have *state independent transition structure* (i.e. when $p(s,i,j)$, the transition probabilities are independent of s, the states of the game) and the transition is dominated by a fixed nonatomic probability measure.

To make each chapter self content, necessary definitions and notions have been restated in appropriate places.

Chapter 2

Ordered Field Property in Stochastic Games

2.1 Stochastic Games: A Related Coverage:

Stochastic games were introduced by Shapley [Ref.38]. Suppose that a simple matrix game is repeatedly played according to probabilities that depend on the players' choices, and the game ends with probability one after a finite number of stages . Then the payment (payoff) to player I can be formulated as a ratio of two bilinear forms. The minimax theorem for such games was established by von Neumann himself [Ref.27] and later, a simple proof was given by Loomis [Ref.20]. Shapley extended this idea in defining a stochastic game which stops with probability one after a finite number of stages. (See the introduction for a detailed study). Such games are now called *stopping stochastic games.* In the present literature , a *stochastic game* in normal form (such games are referred as nonstopping stochastic games in Chapter 1) refers to the following two person zero sum recursive simultaneous move games.

Definition: A stochastic game is defined by a collection of five objects $< S, \{A^s : s \in S\}, \{B^s : s \in S\}, p, r >$, where S is a finite set of *states* of the game. Two players I and II play the game and for each $s \in S, A^s$ (B^s respectively) is a finite set of m^s(n^s respectively) pure strategies of player I (player II respectively). For each $s \in S$, $i \in A^s$ and $j \in B^s$, $p(s, i, j)$ is a probability distribution on the state space S, and it is called the *transition law* of the game. $r(s, i, j)$ is a real valued function on the cartesian product of $S, \{A^s : s \in S\}$and$\{B^s : s \in S\}$, and it is called the *immediate reward function* of the game. The game is played over the infinite future as follows:

1. On the first stage, the game starts at a state $s_1 \in S$.

2. Both players look at s_1 and choose simultaneously (and hence independently of each other) a pure action each, say i_1 from A^{s_1} and j_1 from B^{s_1}.

3. Consequently, (a) player I gets an immediate reward $r(s_1, i_1, j_1)$ from player II. (b) On the next day (stage), the state of the system, s_2, is determined by the transition law $p(s_1, i_1, j_1)$.

4. Once s_2 is determined, the entire process starting from 2 above, with s_1 replaced by s_2, is repeated over and over again. Thus the game proceeds over the infinite time.

5. Both the players have infinite memory, which makes a stochastic game, a game with perfect recall [Ref.18].

Strategy Spaces of the players:

A *strategy* (or *behavioural strategy*) π of player I is a sequence $\{\pi_n\}_{n=1}^{\infty}$, where π_n specifies the action to be chosen by player I on the n^{th} day by associating with each history h_n of the system on the n^{th} day (where $h_n = (s_1, i_1, j_1, \ldots, s_{n-1}, i_{n-1}, j_{n-1}, s_n)$, and (s_k, i_k, j_k) are respectively the state and actions of the players on k^{th} day) a probability distribution $\pi_n(\cdot \mid h_n)$ on A^{s_n}.

A strategy $\pi = \{\pi_n\}_{n=1}^{\infty}$ for player I is called *Markov*, if each π_n is a function of s_n only. π is called *stationary* if there is a map f from the state space S to the set of probabilities on $\{A^s : s \in S\}$ such that $\pi_n = f$ for all n. Thus a Markov strategy π is stationary if each π_n is independent of the time n. It is evident that stationary strategies are most easy to work with.

Strategies, Markov strategies and Stationary strategies for player II are defined similarly.

Pay off Criterion:

In the present chapter two different pay off criteria will be considered. They are: 1. *β-discounted pay off* and 2. *limiting average* or *undiscounted pay off*:

1.*β-discounted pay off*: The notion of β-discounted pay off is borrowed from the field of dynamic programming. β-discounted pay off to player I over infinite future is calculated as $\sum_{n=1}^{\infty} \beta^{n-1} r(s_n, i_n, j_n)$ where $\beta \in [0,1)$ is a discount factor and s_n, i_n, j_n are respectively the state, action of player I and that of player II on the n^{th} day. Given a pair of strategies (π, σ) of the two players, the above quantity defines a real valued random variable on the underlying sample space (paths) of the system and the β-discounted expected pay off is defined as

$$v_\beta(s, \pi, \sigma) = E_{\pi,\sigma}[\sum_{n=1}^{\infty} \beta^{n-1} r(s_n, i_n, j_n)]$$

where $s = s_1$, and the expectation on the right hand side is taken over the probability measure induced by the strategy pair (π, σ) on the sample paths of the game.

2.*Limiting average pay off* (or *undiscounted pay off*):Gillette [Ref.13] introduced the so called limiting average (or undiscounted) pay off criterion. According to this notion, the total pay off to player I is calculated as $\liminf_{n\to\infty}(1/n)\sum_{k=1}^{n} r(s_k, i_k, j_k)$ and for a strategy pair (π, σ), the limiting average expected pay off to player I is given by $v(s, \pi, \sigma) = E_{\pi,\sigma} \liminf_{n\to\infty}(1/n)\sum_{k=1}^{n} r(s_k, i_k, j_k)$ where $E_{\pi,\sigma}$, s, s_k, i_k, j_k are as in the definition of the β-discounted pay off above.

Value and ϵ-optimal strategies:

If for each ϵ positive, there exists a pair of strategies $(\pi_\epsilon, \sigma_\epsilon)$ for the players and a vector $u = \{u(s) : s \in S\}$ such that, if s is the intitial state and player I (player II respectively) uses the strategy π_ϵ (σ_ϵ respectively), then his expected income is at least $u(s)-\epsilon$ (his average loss is at most $u(s)+\epsilon$ respectively) irrespective of the choice of strategies made by player II (player I respectively), then u is called the *value vector* of the stochastic game $< S, \{A^s : s \in S\}, \{B^s : s \in S\}, p, r >$ and π_ϵ and σ_ϵ are called *ϵ-optimal strategies* for player I and player II respectively. *0-optimal strategies* (if ϵ

can be taken to be zero above) are simply called *optimal strategies.*

Note that the notion of the value and ϵ-optimal strategies, as defined above, crucially depends on the pay off criterion being used to calculate the income and loss of the players.

Known results:

1. The existence of the value and a pair of optimal stationary strategies for the two players for a β-discounted stochastic game follows from Shapley [Ref.38] and Blackwell [Ref.5]. This amounts to the existence of a vector $v_\beta = \{v_\beta(s) : s \in S\}$ and a pair of stationary strategies (f_β^*, g_β^*) for each $\beta \in [0,1)$, such that $v_\beta(s, f_\beta^*, \sigma) \geq v_\beta(s) \geq v_\beta(s, \pi, g_\beta^*)$ for all choices of π and σ. Further, if for each $s \in S$, the matrix game G_β^s is considered, where G_β^s is an $m_s \times n_s$ matrix with $(i,j)^{th}$ entry $G_\beta^s(i,j) = r(s,i,j) + \beta \sum_{t \in S} p(t \mid s,i,j) v_\beta(t)$ then $v_\beta(s) = val\, G_\beta^s$ for all $s \in S$. In vector notation, $v_\beta = val\, G_\beta$.

The set of optimal strategies of a player in the β-discounted stochastic game is precisely the cartesian product of set of optimal strategies of that player in the matrix games G_β^s. The set of equations $v_\beta = val\, G_\beta$ is extremely important in finding the value and optimal strategies in a β-discounted stochastic game. This is known as *Shapley equation or optimality equation* in the literature [Ref.45].

2. Mertens and Neyman [Ref.23] proved the existence of the value and a pair of ϵ-optimal strategies (not necessarily stationary) for the players for an undiscounted stochastic game. From Mertens and Neyman, it is also known that limit as β tends to 1 of the normalised β-discounted values is the undiscounted value in a zero sum two person stochastic game. Mathe-

matically, $(1-\beta)v_\beta \to v$ componentwise.

2.2 Ordered field property in Stochastic Games

In the early 1970's, the study of stochastic games was directed towards undiscounted stochastic games. The existence of a value or ϵ-optimal strategies were unknown at that time. Some results were known about particular cases like *the big match* [Ref.7] (which is a particular example of *repeated games with absorbing states* [Ref.17]). At this time, two different schools of thought were working in this field. In the University of Illinois at Chicago Circle, Parthasarathy and Raghavan were working on various structured stochastic games. In Harvard University, Truman Bewley and Elon Kohlberg were working in the general undiscounted stochastic games which have stationary optimals for both the players. Shapley, in the seminal paper *Stochastic games*, commented that a stochastic game with rational coefficients does not necessarily have a rational value, so the *ordered field property* (see Chapter 1 for a detail) does not hold in the class of stochastic games. Bewley and Kohlberg [Ref.4] identified the field of real Puiseux series as a particular ordered field where this property holds. These works culminated in a series of papers [Ref.2, 3, 4] published from Harvard University. They were able to give a necessary condition and a (separate) sufficient condition for the existence of value and a pair of stationary optimal strategies in an undiscounted stochastic game.

The other school, led by Parthasarathy and Raghavan, approached the problem through structured stochastic games. Parthasarathy's graduate student Martin Stern [Ref.41] proved that if the transition probabilities in a stochastic game are always independent the choices of a *particular* player (such games are called ***one player controlled stochastic games***) then the undiscounted stochastic game has a value and a pair of stationary optimal strategies for the players. Parthasarathy and Raghavan directed this field of research towards finding (preferably in finitely many steps) the value and a pair of stationary optimal strategies for various structured stochastic games. They proved that one player controlled stochastic games (β-discounted and undiscounted) have the Archemedian ordered field property. In the β-discounted case, they exhibited a finite step method of finding the value and a pair of stationary optimals for such games. (for a detailed coverage of various methods of finding the value and optimal strategies in stochastic games, see Vrieze [Ref.46], Filar and Raghavan [Ref.11]). Using a finite step method, Raghavan's graduate student Jerzy Filar [Ref.9] first proved the above result in the undiscounted case. By this time, the Dutch school of game theorists have joined this field and Vrieze [Ref.46], a student of Stef Tijs, solved the one player controlled undiscounted stochastic games using a one step method (i.e. just by solving a single linear programming problem). In 1981, Filar [Ref.10] generalised this class of games to *Switching control* (SC) stochastic games, where the property of one player controlled transition remained the same but now the controller of transition (player I or player II) may change from state to state. Filar proved the existence of value and a pair of optimal stationary strategies for such (β-discounted and undiscounted) stochastic games and with it, he showed that one can find them without going out of any Archemedian ordered field. Filar [Ref.9] also extended the *repeated games with absorbing states* of Kohlberg [Ref.17] to a class called *single loop stochastic game which one player can terminate* which has value in the undiscounted case. All these efforts culminated in the general existence theorem for the value and ϵ-optimal strategies for undiscounted stochastic games. Mertens and Neyman [Ref.23] being the first to give a proof of it.

The issue of ordered field property and various methods in solving stochastic games was not yet solved. For a switching control stochastic game, Vrieze again [Ref.46] proved the ordered field property for the β-discounted case using a finite step method. Finally, the undiscounted case was solved in finitely many steps by Filar, Raghavan, Tijs and Vrieze [Ref.46]. In the present literature, there are two more classes of stochastic games which obey the ordered field property. They are called (i)*stochastic games with additive reward and additive transition* ($ARAT$) [Ref.35] and (ii)*stochastic games with separable reward and state independent transition* ($SeR-SIT$) [Ref.34]. The ordered field property holds for these two classes in the β-discounted as well as the undiscounted case. For the $ARAT$ class, finite step methods exist for finding the value and a pair of stationary optimals. The $SeR-SIT$ class is a simple class in this regard. A $SeR-SIT$ game is nothing but a matrix game which has been endowed with a stochastic game structure and so by solving a single matrix game, one can get the value vector and a pair of state indepedent stationary optimal strategies under β- discounted as well as undiscounted pay off criteria.

It is interesting to note that Bewley and Kohlberg's results [Ref.4] can be used to prove the existence of value and a pair of ϵ-optimal stationary strategies in all the above there general classes (viz SC, $ARAT$ and $SeR-SIT$) of stochastic games. One can also prove that the value and a pair of ϵ-optimal stationary strategies lie in the same ordered field containing the data of the game. However, it is not clear whether stationary optimals exist or how to find them actually. The history of ordered field property in the literature of stochastic games shows that whenever ordered property is proved for a class of stochastic games, sooner or later a finite step method came into being to prove the same result. The same tune is maintained in this chapter.

In this chapter, the statement "a stochastic game has the ordered field property under a certain payoff criterion" means that if the data describing the stochastic game (*i.e.* entries of reward and transition functions and under discounting, β) comes from an Archemedian ordered field, then the value and a pair of stationary optimal strategies exist and have entries in that Archemedian ordered field under the payoff considered.

2.3 About This Chapter

This chapter concentrates on the generalisation of the above three classes of stochastic games with ordered field property, namely, switching control stochastic games, stochastic games with additive reward and additive transition and stochastic games with state independent transition and separable reward. To form an extended class, one need to mix the above three classes in an efficient manner. For this purpose, the notion of ***structured states*** is introduced:

Structured states in a stochastic game:

Let $< S, \{A^s : s \in S\}, \{B^s : s \in S\}, p, r >$ be a stochastic game:

1. A state $s \in S$ is called an *ARAT state* if

   (i) there exist vectors $r_1(s) = (r_1(s,1), r_1(s,2), \ldots r_1(s, m_s)) \in \mathbf{R}^{m_s}$ and $r_2(s) = (r_2(s,1), r_2(s,2), \ldots r_2(s, n_s)) \in \mathbf{R}^{n_s}$ such that the reward function r has two additive components r_1 and r_2. Mathematically, $r(s,i,j) = r_1(s,i) + r_2(s,j)$ for all $i \in A^s, j \in B^s$.

   (ii) there exist nonnegative vectors $p_1(s) = (p_1(s,1), p_1(s,2), \ldots p_1(s, m_s))$ and $p_2(s) = (p_2(s,1), p_2(s,2), \ldots p_2(s, n_s))$ such that the transition probability p has two additive subprobability components p_1 and p_2 i.e $p(s,i,j) = p_1(s,i) + p_2(s,j)$ for all $i \in A^s$, $j \in B^s$.

2. A state $s \in S$ is called a *SC state* if the transition law $p(s,i,j)$ in that state depends only on i, or on j but not on both.

3. A finite collection of states $\tilde{S} = \{s_1, s_2, \ldots s_k\} \subset S$, is called *a collection of SeR-SIT states* if (i) and (ii) are satisfied :

   (i) All states in $\tilde{S}$ have the separable reward structure which means (a) both the players have the same set of pure actions available in all states of $\tilde{S}$. Mathematically, $m_s = m$, $n_s = n$ for all $s \in \tilde{S}$, where m and n are natural numbers and (b) there exists an $m \times n$ matrix A and a vector $c = \{c(s) : s \in \tilde{S}\} \in \mathbf{R}^k$ such that $r(s,i,j) = c(s) + a(i,j)$for all$s \in \tilde{S}, i \in A^s, j \in B^s$.

   (ii) All states in $\tilde{S}$ have the same *SIT* structure i.e. $p(s,i,j)$ does not depend on the state s for all $s \in \tilde{S}$ i.e. for each pair (i,j) , there exists a probability distribution $p(i, j)$ on S such that $p(s, i, j) = p(i, j)$ for all $s \in S$. Note that the notion of a class of $SeR - SIT$ states is well defined because both players have the same of pure actions available in all states belonging to that class.

All the existing classes of stochastic games with ordered field property can be defined directly from the above definitions. A switching control stochastic game is a stochastic game where each state is an SC state. Similarly, a game is a stochastic game with additive reward and additive transtion if each state is an $ARAT$ state. For defining a $SeR - SIT$ stochastic game, the state space S itself must be a collection of $SeR - SIT$ states. It may be interesting to note here that the SC stucture and the $ARAT$ structure are structures on particular states. But $SeR - SIT$ structure is a structure imposed on a *class of states* and not on any state in particular.

In this chapter, mixtures of various structured states and structured class of states have been studied extensively. It is easy to see that in a stochastic game, if the state space S is a disjoint union of three class of states, say S_1, S_2 and S_3, such that (i) S_1 is a class of $ARAT$ states (i.e. all states in S_1 are $ARAT$), S_2 is an SC class of states and S_3 is a collection of $SeR - SIT$ states, and (ii) starting from any class S_l, the game *cannot* go to any other class S_k with positive probability (i.e. $p(s_k \mid s_l, i, j) = 0$ for all $l \neq k \in \{1, 2, 3\}$, for all $s_k \in S_k, S_l \in S_l$,) then the game, defined this way, has the ordered field property. Since S_1, S_2 and S_3 are noncommunicating disjoint classes, essentially three different games are considered jointly through the above definition. Since each of the three classes have ordered field property, this big game also has the property under both β-discounted and undiscounted pay off criteria.

In what follows, it is shown that the above mentioned structured states (viz. SC and $ARAT$) can be mixed without going out of the ordered field on which the data of this mixed game is defined. But a class of $SeR-SIT$ states cannot be combined with any structured states if one is willing to solve them in an ordered field. In section 2.4, mixture of two different $SeR-SIT$ classes are studied. An example is analysed to show that if the class property of $SeR-SIT$ is not maintained, the ordered field property does not hold in the β-discounted as well as in the undiscounted case. Section 2.5 answers in the negative the issue of ordered field property for the most general $SeR-SIT/SC/ARAT$ mixture stochastic game under β-discounted pay off criterion. As for the undiscounted case, a version of the *big match* is exhibited to show that such a general mixture need not have optimal strategies for the two players. Hence it is meaningless to talk about ordered field property under such a general set up. In section 2.6, a general mixture of SC and $ARAT$ states has been studied and a nonconstructive proof has been given to show that such stochastic games have the ordered field property under the β-discounted as well as the undiscounted payoff criterion. In section 2.7, some remarks have been made.

2.4 Mixture of two $SeR - SIT$ classes:

Example 2.4.1.

Consider the following stochastic game with four states.

state 1

0 (1,0,0,0)	0 (0,0,1/2,1/2)
0 (0,1,0,0)	0 (1,0,0,0)

state 2

-1 (1,0,0,0)	-1 (0,0,1/2,1/2)
-1 (0,1,0,0)	-1 (1,0,0,0)

state 3

-1 (1,0,0,0)	-1 (0,0,1/2,1/2)
0 (0,1,0,0)	0 (1,0,0,0)

state 4

0 (1,0,0,0)	0 (0,0,1/2,1/2)
1 (0,1,0,0)	1 (1,0,0,0)

A cell $\begin{array}{|c|}\hline r \\ (p_1, p_2, p_3, p_4) \\ \hline\end{array}$ corresponds to an immediate pay off r to player I by player II and p_i is the probability with which the state i in the next stage is played if this cell is choosen today. This game is a stochastic game with state independent transition. Also this is a mixture of two $SeR - SIT$ classes $S_1 = \{1,2\}$ and $S_2 = \{3,4\}$. Consider this game under β-discounted pay off. For the value vector $v_\beta = (\ v_\beta(1), v_\beta(2), v_\beta(3), v_\beta(4)\ )$, from the Shapley equation one gets the following relations:

$$v_\beta(1) = \beta val \begin{pmatrix} v_\beta(1) & \frac{1}{2}(v_\beta(3)+v_\beta(4)) \\ v_\beta(2) & v_\beta(1) \end{pmatrix}, \qquad v_\beta(2) = -1 + v_\beta(1);$$

$$v_\beta(4) = val \begin{pmatrix} \beta v_\beta(1) & \frac{\beta}{2}(v_\beta(3)+v_\beta(4)) \\ 1+\beta v_\beta(2) & 1+\beta v_\beta(1) \end{pmatrix}, \quad v_\beta(3) = -1 + v_\beta(4).$$

Noting the fact $v_\beta(2) < v_\beta(1)$, it is easy to see that the matrix game related to $v_\beta(4)$ above has a saddle point in the south west corner and so $v_\beta(4) = 1 + \beta v_\beta(2)$ Using all the available informations, the Shapley equation for state 1 takes the form:

$$1 + v_\beta(2) = \beta val \begin{pmatrix} 1+v_\beta(2) & \frac{1}{2}+\beta v_\beta(2) \\ v_\beta(2) & 1+v_\beta(2) \end{pmatrix}$$

The matrix game here cannot have a saddle point. So, it is a completely mixed game i.e. a game where the players use both the pure available actions with nonzero probabilities to play optimally. By Kaplansky [Ref.15], the value of this game is

$$\beta \frac{1 + 2v_\beta(2) + (v_\beta(2))^2 - (1/2)v_\beta(2) - \beta(v_\beta(2))^2}{2 + v_\beta(2) - (1/2) - \beta v_\beta(2)}$$

and by solving the resulting equation, $v_\beta(2) = (1-\beta)^{-1}(\frac{-5}{4} + \frac{1}{2}\sqrt{\frac{1}{4} + 4\beta})$. So, for $\beta = 0.75$, $v_{.75}(2) = -5 + \sqrt{13}$. Thus, a discounted stochastic game with two different $SeR - SIT$ classes do not satisfy the ordered field property. The undiscounted version of the above game has value $(-5 + \sqrt{17})/4$ in all the four states. and hence it does not have ordered field property also.

2.5 Mixture of SC and $ARAT$ states with a $SeR-SIT$ class of states

Example 2.5.1 : Discounted case

Consider the following stochastic game with four states.

state 1

3 $(\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$	0 $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$
0 $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$	1 $(\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$

state 2

3 $(\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$	0 $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$
0 $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$	1 $(\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$

state 3

0 (1,0,0,0)

state 4

0 (0,0,0,1)

The notations are same as in the previous section. It is easy to note that states 1 and 2 constitute a $SeR-SIT$ class. State 3 may be taken as an $ARAT$ state and state 4 similarly is an SC state. Consider this game under discounting with a discount factor $\beta = 1/2$. For brevity, let the value vector $v_{1/2}$ be denoted by $v = (v_1, v_2, v_3, v_4)$. Clearly $v_1 = v_2, v_3 = (1/2)v_1, v_4 = 0$. The Shapley equation of this game gives:

$$v_1 = val \begin{pmatrix} 3 + (1/4)v_1 & (1/4)v_1 + (1/8)v_3 \\ (1/4)v_1 + (1/8)v_3 & 1 + (1/4)v_1 \end{pmatrix}$$

$$= val \begin{pmatrix} 3 + (1/4)v_1 & (5/16)v_1 \\ (5/16)v_1 & 1 + (1/4)v_1 \end{pmatrix}$$

$$= \frac{3 + (1/16)v_1^2 + v_1 - (25/256)v_1^2}{4 - (1/8)v_1};$$

whence $v_1 = v_{1/2}(1) = \frac{16}{23}(24 - 13\sqrt{3})$ which is irrational while the data of the game is from the rational field. Hence, a general $SeR-SIT/SC/ARAT$ mixture does not have the ordered field property under discounting.

Example 2.5.2 : Undiscounted Case:

Consider the following stochastic game with four states.

state 1

1 (1/2,1/2,0,0)	0 (1/2,1/2,0,0)
0 (0,0,1,0)	1 (0,0,0,1)

state 2

1 (1/2,1/2,0,0)	0 (1/2,1/2,0,0)
0 (0,0,1,0)	1 (0,0,0,1)

state 3

0 (0,0,1,0)

state 4

1 (0,0,0,1)

A close look reveals that the above game is the famous *big match*, where the actual playable (nonabsorbing) state is split up into two states 1 and 2 to create a $SeR - SIT$ class containing two states. As before, state 3 is an $ARAT$ state and state 4 is an SC state. Thus, this is a general $SeR - SIT/SC/ARAT$ mixture and as in the big match, player I cannot have any optimal strategy in the undiscounted version of this game. To observe this, note that:

(i) A strategy for any player tells him what to do in states 1 and 2 at each stage given the history of the game upto that stage, because in states 3 and 4, each player has only one option;

(ii) The value of the undiscounted game is $(1/2, 1/2, 0, 1)$. [redacted].

Let π be any strategy for player **I**. π has to be one of the following four types of strategies π_1, π_2, π_3 and π_4:

π_1: Never choose the second row in state 1 or in state 2 with a positive probability.

π_2: Choose the second row for the first time in the $(m_1+1)^{th}$ stage, in state 1, with a positive probability ϵ_1, where m_1 is some nonnegative integer.

π_3: Choose the second row for the first time in the $(m_2+1)^{th}$ stage, in state 2, with a positive probability ϵ_2, where m_2 is some nonnegative integer.

π_4: Choose the second row for the first time in the $(m+1)^{th}$ stage, both in states 1 and 2 with a positive probability ϵ where m is some nonnegative integer.

Against each type π_i, player **II** has a strategy σ_i which brings down his losses below the value. These four strategies and the pay off to player **I** if the pair (π_i, σ_i) is chosen, is described below:

σ_1: Never choose the first column in state 1 or in state 2 with a positive probability. If (π_1, σ_1) is played, the expected undiscounted pay off to **I** is zero in states 1 and 2. Hence π_1 cannot be optimal.

σ_2: For the first m_1 stages, choose the second column in both states 1 and 2. In the $(m_1+1)^{th}$ stage, choose the i^{th} column in the i^{th} state.

$i \in \{1, 2\}$. From stage $(m_1 + 2)$ onwards, choose each column with equal probability in both states 1 and 2. If (π_2, σ_2) is played, the expected undiscounted pay off to player I is $(1/2) - (\epsilon_1/4)$ in states 1 and 2. Hence π_2 cannot be optimal for I.

σ_3: Choose the second column in both states 1 and 2 for the first m_2 stages. In the $(m_2 + 1)^{th}$ stage, choose column 1 in state 2 and column 2 in state 1. From stage $(m_2 + 2)$ onwards, choose each column with equal probability in both states 1 and 2. If (π_3, σ_3) is played, the expected undiscounted pay off to player I is $(1/2) - (\epsilon_2/4)$ in states 1 and 2 . Hence π_3 cannot be optimal for player I.

σ_4: Choose the second column in both states 1 and 2 for the first m stages. In the $(m+1)^{th}$ stage, choose the first column in both states 1 and 2 and in the rest of the game, play column 1 and 2 with equal chances in both states. If (π_4, σ_4) is played, the expected pay off to I becomes $(1/2) - (\epsilon/2)$. Hence π_4 cannot be optimal for player I.

Thus no strategy can be optimal for player I in this game. Hence, the question of the ordered field property does not arise in the general mixture of $SeR - SIT/SC/ARAT$ under the undiscounted pay off criterion.

2.6 Mixture of structured states: SC and $ARAT$

As mentioned earlier, the structures switching control (SC) and additive reward and additive transition ($ARAT$) are properties of a state in a stochastic game. Unlike the $SeR-SIT$ class, it is easy to mix a collection of $ARAT$ states with another collection of the same without disturbing the $ARAT$ structure (i.e. the resultant bigger collection of states retains the $ARAT$ structure). Same is true for the SC states also. Hence as far as the ordered field property is concerned, mixing a collection of structured states with another of the same kind do not violate this property. In this section, mixture of a collection of SC states with that of $ARAT$ states have been studied and it is shown that mixed games resulting from such a combination have the ordered field property.

Definition:

A stochastic game $< S, \{A^s :\in S\}, \{B^s : s \in S\}, p, r >$ is called a $SC/ARAT$ mixture if the state space S is the union of three disjoint subjects S_1, S_2 and S_3 such that (i) all states in S_i are player i controlled states for $i \in \{1,2\}$ and (ii) all states in S_3 are $ARAT$ states.

More specifically, if $S = S_1 \bigcup S_2 \bigcup S_3$; S_i's are disjoint such that

(a) $p(s,i,j)$ is independent of j for all $s \in S_1$ and is denoted by $p(s,i)$;

(b) $p(s,i,j)$ is independent of i for all $s \in S_2$ and is denoted by $p(s,j)$;

and (c) If $s \in S_3$, $r(s,i,j) = r_1(s,i) + r_2(s,j)$, for all i,j , for some r_1 and r_2 and $p(s,i,j) = p_1(s,i) + p_2(s,j)$ for all i,j for some nonnegative vectors p_1 and p_2.

To prove the ordered field property for a $SC/ARAT$ mixture we need some preliminary lemmata which are described below.

Lemma 2.6.1: Let $r(t) = p(t)/q(t)$ be a rational function defined on $t \in [\beta_0, 1)$; If $r(t)$ is a rational number whenever t is rational then $r(t) = p^*(t)/q^*(t)$ where p^* and q^* are polynomials with rational coefficients.

Proof: The proof is omitted here. Interested readers can see Parthasarathy and Raghavan [Ref.32] for a proof.

Lemma 2.6.2: Let $f(t)$ be a rational function bounded in some interval $[\beta_0, 1)$. If $f(t)$ is rational whenever t is rational then $\lim_{t\to 1-} f(t)$ is a rational number.

Proof: omitted. See [Ref.32]

Lemma 2.6.3: Let $f(t)$ be a continuous function on $(0,1)$. Let $u_j(t)$, $j = 1, 2 \ldots k$ be rational functions in t such that for each t, $f(t)$ coincides with one of these $u_j(t)$. Then there exists some $\beta_0 \in (0,1)$ and a $j_0 \in \{1, 2, \ldots k\}$ such that $f(t) = u_{j_0}(t)$ for all $t \in [\beta_0, 1)$.

Proof: omitted. See [Ref.32].

Lemma 2.6.4: Let $A = ((a_{ij} + a_i))_{m\times n}$ be a nonsingular matrix with $a_{ij} > 0$ for all i, j. Let $x^t A = \alpha \cdot 1$ have a nonnegative solution x with $\sum_{i=1}^{m} x_i = 1$. Then the matrix $\bar{A} = (a_{ij})$ is also nonsingular and $x^t \bar{A} = \theta \cdot 1$ for some θ. Analogous result holds if A is of the form $((a_{ij} + b_j))$.

Proof: Omitted. See [Ref.32].

Lemma 2.6.5:

In a matrix game, the set of extreme optimal strategies for any player is a nonempty finite set.

(In a matrix game $A = ((a_{ij}))_{m\times n}$ with value v not equal to zero, a pair of optimal strategies (x, y) is a pair of extreme optimal strategies of A if and only if A has a nonsingular submatrix $\dot{A}$ for which $\dot{x}^t \cdot \dot{A} = v \cdot 1$, $\dot{A} \cdot \dot{y} = v \cdot 1$, where $\dot{x}$ and $\dot{y}$ are obtained from x and y respectively by deleting entries corresponding to rows and columns of A which are not present in $\dot{A}$.)

Proof: Omitted. This is a direct corollary of Snow-Shapley theorem.The reader may find it in the book of [Ref.48].

Lemma 2.6.6:

Let s be an *ARAT* state in a β-discounted stochastic game. Then the stochastic game has a pair of extreme optimal strategies with pure s^{th} components. In other words, the *ARAT* states have pure extreme optimal stationary strategy components for the two players in a β-discounted stochastic game.

Proof: Consider the Shapley equation of an *ARAT* state.

$$\begin{aligned}
v_\beta(s) &= val\,((\,r_1(s,i) + r_2(s,j) + \beta \textstyle\sum_{t\in S}[p_1(t \mid s,i) + \\
&\qquad\qquad\qquad\qquad p_2(t \mid s,j)]v_\beta(t)\,)) \\
&= val\,((\,\{r_1(s,i) + \beta \textstyle\sum_{t\in S} p_1(t \mid s,i)v_\beta(t)\} + \\
&\qquad\qquad\qquad\qquad \{r_2(s,j) + \beta \textstyle\sum_{t\in S} p_2(t \mid s,j)v_\beta(t)\}\,)) \\
&= \min_\mu \max_\lambda [\{r_1(s,\lambda) + \beta \textstyle\sum_{t\in S} p_1(t \mid s,\lambda)v_\beta(t)\} + \\
&\qquad\qquad\qquad\qquad \{r_2(s,\mu) + \beta \textstyle\sum_{t\in S} p_2(t \mid s,\mu)v_\beta(t)\}] \\
&= \min_\mu [r_1(s,i_0) + \beta \textstyle\sum_{t\in S} p_1(t \mid s,i_0)v_\beta(t) + \\
&\qquad\qquad\qquad\qquad \{r_2(s,\mu) + \beta \textstyle\sum_{t\in S} p_2(t \mid s,\mu)v_\beta(t)\}], \\
&\qquad\qquad \text{where } i_0 \text{ is pure and maximises the first part,} \\
&= r_1(s,i_0) + \beta \textstyle\sum_{t\in S} p_1(t \mid s,i_0)v_\beta(t) + r_2(s,j_0) + \beta \textstyle\sum_{t\in S} p_2(t \mid s,j_0)v_\beta(t), \\
&\qquad\qquad \text{where } j_0 \text{ is pure and minimises the second part.}
\end{aligned}$$

Hence the proof of lemma 2.6.6.

Lemma 2.6.7 Let s be a player I controlled state in a β-discounted stochastic game (which means $p(s,i,j)$ is independent of j and hence can be denoted by $p(s,i)$). Let R^s be the $m_s \times n_s$ matrix $((\,r(s,i,j)\,))$. Consider the s^{th} component $f^\beta(s)$ of an extreme optimal stationary strategy f^β for player I in the β-discounted game. There exists a nonsingular submatrix $\dot{R}^s$ of R^s and a real number $\theta(s)$ such that $\dot{f}^\beta(s)$ is the unique solution of the system of equations $\xi^t \cdot \dot{R}^s = \theta(s) \cdot 1^t$; where ξ is a probability vector. Here $\dot{f}^\beta(s)$ is obtained from $f^\beta(s)$ by deleting the entries associated with those rows of R^s which are not in $\dot{R}^s$.

Analogous result holds for an extreme optimal strategy for player II if s is a player II controlled state (i.e. if $p(s,i,j) = p(s,j)$ for all i).

Proof: By Tijs and Vrieze [Ref.45], the s^{th} component of the set of extreme optimal stationary strategy pairs in a β-discounted stochastic game is precisely the set of extreme optimal strategy pairs of the matrix game G_s^β where G_s^β is the s^{th} Shapley matrix with entries $(\, r(s,i,j) + \beta \sum_{t \in S} p(t \mid s,i) v_\beta(t)\,)$. By lemma 2.6.5 applied to G_β^s, for every pair $(f^\beta(s), g^\beta(s))$ of extreme optimal strategies, there exists a nonsingular submatrix $\dot{G}_\beta^s$ of G_β^s such that $\dot{f}^\beta(s) \cdot \dot{G}_\beta^s = v_\beta(s) \cdot 1^t$ and $\dot{G}_\beta^s \cdot \dot{g}^\beta(s) = v_\beta(s) \cdot 1$, where $\dot{f}^\beta(s)$ and $\dot{g}^\beta(s)$ are obtained from $f^\beta(s)$ and $g^\beta(s)$ after appropriate truncation of zero entries which corresponds to rows and columns of G_β^s which are not present in $\dot{G}_\beta^s$. By lemma 2.6.4, applied to $\dot{G}_\beta^s$ which is of the form $((\, a_{ij} + a_i\,))$, the corresponding submatrix $\dot{R}^s$ of R^s is also nonsingular and satisfies $\dot{f}^\beta(s) \cdot \dot{R}^s = \theta(s) \cdot 1$, for some $\theta(s)$.

Since $\dot{f}^\beta(s)$ is a probability vector, its entries and the value of $\theta(s)$ depend only on the choice of the submatrix $\dot{R}^s$ (for different β, however, $\dot{R}^s$ possibly varies among finitely many nonsingular submatrices of R^s).

Thus the proof of lemma 2.6.7 is complete. From lemma 2.6.7 and its analogue for a player II controlled state, one can conclude that for $s \in S_1 \bigcup S_2$, in a β-discounted $SC/ARAT$ mixture, there is a choice (depending on β) of a nonsingular submatrix $\dot{R}^s$ of the reward matrix R^s which determine the following:

(a) a real number $\theta(s)$ for every $s \in S_1 \bigcup S_2$, where $\theta(s) = (det\dot{R}^s)/(\sum_i \sum_j \dot{R}^s_{ij})$, where *det* denotes the determinant and $\dot{R}^s_{ij}$ is the $(i,j)^{th}$ cofactor of $\dot{R}^s$.

(b) an extreme optimal stationary strategy component for player I $\dot{f}(s) = (\dot{f}_i(s))$ for $s \in S_1$ where $\dot{f}_i(s) = (\sum_j \dot{R}^s_{ij}) / (\sum_i \sum_j \dot{R}^s_{ij})$.

(c) an extreme optimal stationary strategy component for player II $\dot{g}(s) = (\dot{g}_j(s))$ for $s \in S_2$ where $\dot{g}_j(s) = \sum_i \dot{R}^s_{ij} / \sum_i \sum_j \dot{R}^s_{ij}$.

From lemma 2.6.6 it is obvious that if s is an *ARAT* state (i.e. $s \in S_3$) in a β-discounted *SC/ARAT* mixture, there is an entry (depending on β) of R^s which define a pair of stationary pure optimal strategy component of the two players in that state.

Now for $s \in S_1 \bigcup S_2$, each R^s has only finitely many, (say k_s) nonsingular submatrices. For $s \in S_3$ each R^s has only finitely many (say k_s), entries (one by one submatrices). Let for $s \in S = S_1 \bigcup S_2 \bigcup S_3$, $R^s_1, R^s_2, \ldots R^s_{k_s}$ be an ordering of such submatrices (nonsingular if $s \in S_1 \bigcup S_2$, one by one if $s \in S_3$). Consider all permutations of the form $(k(1), k(2), \ldots k(z))$ where $k(s) \in \{1, 2, \ldots k_s\}$ for each $s \in S$. There are $\mu = \prod_{s=1}^{z} k_s$ such permutations and they can be ordered by $\alpha_1, \alpha_2, \ldots \alpha_\mu$. Then we get μ vectors $\theta_l = (\theta_{l(1)}(1), \theta_{l(2)}(2), \ldots \theta_{l(z)}(z))$ for $l = 1, 2 \ldots \mu$ where (a) l corresponds to $\alpha_l = (R^1_{l(1)}, R^2_{l(2)} \ldots R^s_{l(s)})$, $R^s_{l(s)}$ being a submatrix of R^s of the type described above depending on s. (b) for $s \in S_3, \theta_{l(s)}(s)$ is an entry of R^s chosen by the permutation α_l and for $s \in S_1 \bigcup S_2, \theta_{l(s)}(s)$ is as $\theta(s)$

above.

For each $l \in \{1, 2, \ldots \mu\}$, the permutation α_l also determine, apart from θ_l, (c) player I's strategy components $f^l(s)$ in S_1. (d) Player II's strategy components $g^l(s)$ in S_2 and (e) both players' strategy components in S_3 which are pure.

Fixing player I's strategy in S_2 and player II's strategy in S_1 as arbitary, define a strategy pair (f^l, g^l) in the $SC/ARAT$ mixture for each l.

Explicitly,

$$f^l(s) = \begin{cases} \theta_{l(s)}(s) \cdot (R^s_{l(s)})^{-1} & \text{if } s \in S_1, \\ \text{fixed and arbitary,} & \text{if } s \in S_2, \\ e_{i^l(s)}(s), & \text{if } s \in S_3. \end{cases}$$

$$g^l(s) = \begin{cases} \text{fixed and arbitrary,} & \text{if } s \in S_1, \\ \theta_{l(s)}(s) \cdot (R^s_{l(s)})^{-1} & \text{if } s \in S_2, \\ e_{j^l(s)}(s) & \text{if } s \in S_3. \end{cases}$$

Note: $f^l(s)$ and $g^l(s)$ are defined here by submatrices of R^s. So, finitely many zero entries should be added in appropriate places to make them a strategy in the original game. In the bulk of this section, we shall not distinguish between them. For $s \in S_3$, $(i^l(s), j^l(s))$ is the row and column number of R^s, chosen by the permutation α_l and e_k is the k^{th} basis vector with k^{th} coordinate 1 and rest zeros.

Thus μ strategy pairs $(f^l, g^l), l \in \{1, 2 \ldots \mu\}$ are formed. Let P_l be the probability transition matrix corresponding to the strategic choice (f^l, g^l)

in the $SC/ARAT$ mixture i.e.

$$P_l(s,t) = \sum_i \sum_j p(t \mid s,i,j) f_i^l(s) g_j^l(s) \text{ for } s,t \in S.$$

The matrix $I - \beta P_l$ is nonsingular (since $\beta \in [0,1)$). Hence for each l, define a rational function of β on $[0,1)$ by u_β^l where $u_\beta^l = (I - \beta P_l)^{-1} \cdot \theta_l, l \in \{1, 2 \ldots \mu\}$, I being the identity matrix.

Theorem 2.6.8

For $\beta \in [0,1)$, the value function v_β of the β-discounted $SC/ARAT$ mixture coincides with one of u_β^l. If the data (i.e. the entries of p and r) of the game is from the field of rational number (or in general, from any ordered Archemedian field) and β is also rational, then the value v_β and a pair of β-discounted stationary optimal strategies have entries in that field. In other words, the β-discounted $SC/ARAT$ mixture has the ordered field property.

Proof:

By lemma 2.6.5, 2.6.6, and 2.6.7, given $\beta \in [0,1)$, there exists an $\alpha_l = (R_{l(1)}^1, R_{l(2)}^2, \ldots R_{l(s)}^s)$ (l depending on β) and a pair of stationary extreme optimal strategies (f^l, g^l) satisfying the above equations.

Now if $s \in S_1$,

$$\begin{aligned} \theta_{l(s)}(s) &= \textstyle\sum_i f_i^l(s) r(s,i,j) \\ &= v_\beta(s) - \beta \textstyle\sum_{t \in S} (\sum_i f_i^l(s) p(t \mid s,i)) v_\beta(t) \\ &= v_\beta(s) - \beta \textstyle\sum_{t \in S} p(t \mid s, f^l) v_\beta(t). \end{aligned}$$

Similarly if $s \in S_2$,

$$\theta_{l(s)}(s) = v_\beta(s) - \beta \textstyle\sum_{t \in S} p(t \mid s, g^l) v_\beta(t).$$

and if $s \in S_3$,

$$\begin{aligned}\theta_{l(s)}(s) &= v_\beta(s) - \beta \sum_{t\in S}(p_1(t \mid s, i^l(s)) + p_2(t \mid s, j^l(s))v_\beta(t) \\ &= v_\beta(s) - \beta \sum_{t\in S} p(t \mid s, f^l, g^l)v_\beta(t).\end{aligned}$$

In matrix notation, $\theta_l = (I - \beta P_l) \cdot v_\beta$, or, $v_\beta = (I - \beta P_l)^{-1} \cdot \theta_l$.

Note that $\{f^l(s) : s \in S_1\}$ and $\{g^l(s) : s \in S_2\}$ have entries in the rational field since each $R^s_{l(s)}$ has that property. $\{f^l(s), g^l(s) : s \in S_3\}$ is a pure strategy pair and hence has rational entries. Further precisely these entries determine the entries of the nonsingular matrix $I - \beta P_l$. Hence $(I - \beta P_l)^{-1}$ has entries from the rational field and θ_l (which is derived from α_l by elementary row operation) has the same property. Hence v_β has rational entries.

To complete the proof it remains to show that $\{f^l(s) : s \in S_2\}$ and $\{g^l(s) : s \in S_1\}$ can be chosen from the field of rational numbers. By lemma 2.6.5 and the Shapley equation,

$$g^l(s) = (G^s_{\beta,l(s)})^{-1} \cdot 1v_\beta(s) \text{ for } s \in S_1,$$

where $G^s_{\beta,l(s)}$ is the nonsingular submatrix of the Shapley matrix G^s_β in the s^{th} state corresponding to the submatrix $R^s_{l(s)}$ of R^s. Since entries on the right hand side of the above equation are from the rational field $\{g^l(s) : s \in S_1\}$ has rational components. Similarly $\{f^l(s) : s \in S_2\}$ has rational components and hence the proof of theorem 2.6.8.follows.

Lemma 2.6.9: Consider the β-discounted $SC/ARAT$ mixture. There exists $\beta_0 \in [0, 1)$ such that

(i) $v_\beta(s)$ is a rational function of β for all $\beta \in [\beta_0, 1)$.

(ii) If $s \in S_1$, player **I** has a stationary strategy component which is optimal for all $\beta \in [\beta_0, 1)$.

(iii) If $s \in S_2$, player **II** has a stationary strategy component which is optimal for all $\beta \in [\beta_0, 1)$.

(iv) If $s \in S_3$, each player has a stationary strategy component which is pure and optimal for all $\beta \in [\beta_0, 1)$.

Proof: From lemma 2.6.8, it follows that v_β, being continuous in β, is equal to one of μ rational functions u^l_β for each $\beta \in [0, 1)$. So lemma 2.6.3 guarantees the existence of a $\beta_0 \in [0, 1)$ and an $l_0 \in \{1, 2 \ldots \mu\}$ such that $v_\beta = u^{l_0}_\beta$ for all $\beta \in [\beta_0, 1)$. This proves (i). Proofs of (ii), (iii) and (iv) follow from lemma 2.6.8 directly.

Henceforth, $R^s_{l_0}, f^{l_0}(s), g^{l_0}(s), i^{l_0}(s)\, j^{l_0}(s)$, $\theta_{l_0(s)}(s)$ will be denoted by $R^s_0, f^0(s), g^0(s), i^0(s), j^0(s)$, $\theta_0(s)$ respectively. The Shapley submatrix of G^s_β corresponding to R^s_0 will be denoted by $G^s_{\beta,0}$.

From the concluding part of theorem 2.6.8, $f^\beta(s) = v_\beta(s)\, 1^t \cdot (G^s_{\beta,0})^{-1}$ for $s \in S_2$, and $g^\beta(s) = v_\beta(s)(G^s_{\beta,0})^{-1} \cdot 1$ for $s \in S_1$. Since $v_\beta(s)$ is a rational function of β, $f^\beta(s)$ for $s \in S_2$ and $g^\beta(s)$ for $s \in S_1$ are also so. Since $f^\beta(s)$ is a probability vector $\tilde{f}(s) = \lim_{\beta \to 1-} f^\beta(s)$ for $s \in S_2$ exists and is again a probability vector. Similarly $\tilde{g}(s) = \lim_{\beta \to 1-} g^\beta(s)$ for $s \in S_1$ is a probability vector.

Define a stationary strategy pair $(\bar{f}, \bar{g})$ in the $SC/ARAT$ mixture as follows:

$$\bar{f}(s) = \begin{cases} f^0(s) & \text{if } s \in S_1 \\ \tilde{f}(s) & \text{if } s \in S_2 \\ e_{i^0(s)}(s) & \text{if } s \in S_3 \end{cases}$$

$$\bar{g}(s) = \begin{cases} \tilde{g}(s) & \text{if } s \in S_1 \\ g^0(s) & \text{if } s \in S_2 \\ e_{j^*(s)}(s) & \text{if } s \in S_3 \end{cases}$$

In the sequel, we shall show that $(\bar{f}, \bar{g})$ is an optimal pair of strategies in the undiscounted $SC/ARAT$ mixture.

Lemma 2.6.10 Let $P(f,g)$ be the probability transition matrix associated with the stationary strategy pair (f,g) in the $SC/ARAT$ mixture. Then we have:

(a) $P(f^\beta, g^\beta) = P(\bar{f}, \bar{g})$ for all $\beta \in [\beta_0, 1)$; here (f^β, g^β) is an extreme optimal stationary strategy pair in the β-discounted game.

(b) $P(f, g^\beta) = P(f, \bar{g})$ for all $\beta \in [\beta_0, 1)$ for any stationary f.

(c) $r(s, f^\beta, g^\beta) = \theta(s) = r(s, \bar{f}, \bar{g})$ for all $\beta \in [\beta_0, 1)$.

(d) $\lim_{\beta \to 1-}(1-\beta)v_\beta = v(\bar{f}, \bar{g}) = P^*(\bar{f}, \bar{g}) \cdot r(\bar{f}, \bar{g})$; where P^* is the cesaro limit of P.

(e) $\sum_{n=0}^{\infty} \beta^n [P^n(f, \bar{g}) - P^*(f, \bar{g})] \to [I - P(f, \bar{g}) - P^*(f, \bar{g})]^{-1} - P^*(f, \bar{g})$ as $\beta \to 1-$.

(f) $r(f, g^\beta) \to r(f, \bar{g})$ as $\beta \to 1-$.

(g) $(1-\beta) \sum_{n=0}^{\infty} \beta^n P^n(f, \bar{g}) \cdot [r(f, g^\beta) - r(f, \bar{g})] \to 0$ as $\beta \to 1-$.

(h) $(1-\beta) v_\beta(f, g^\beta) \to v(f, \bar{g})$ as $\beta \to 1-$ for any stationary f.

(i) $v(\bar{f}, \bar{g}) \geq v(f, \bar{g})$ for any stationary f.

(j) $v(\bar{f}, \bar{g})$ is the value vector in the undiscounted $SC/ARAT$ mixture. $(\bar{f}, \bar{g})$ is a pair of stationary optimal strategies in this game.

Proof:

(a) Let $\beta \geq \beta_0$ and fix $t \in S$. If $s \in S_1$, then $P(f^\beta, g^\beta)(s,t) = p(t \mid s, f^\beta) =$

$p(t \mid s, f^0(s)) = p(t \mid s, \bar{f}) = P(\bar{f}, \bar{g})(s,t)$ by lemma 2.6.9 and lemma 2.6.7.

If $s \in S_2$, the proof is similar.

For $s \in S_3, P(f^\beta, g^\beta)(s,t) = p(t \mid s, f^\beta, g^\beta) = p(t \mid s, i^0(s), j^0(s)) = P(\bar{f}, \bar{g})(s,t)$ by lemma 2.6.9 and lemma 2.6.6.

Thus (a) is proved.

(b) Let $\beta \geq \beta_0$ and fix $t \in S$ as before. Let f be a stationary strategy for player I.

If $s \in S_1, P(f, g^\beta)(s,t) = p(t \mid s, f(s)) = P(f, \bar{g})(s,t)$.

If $s \in S_2$, $P(f, g^\beta)(s,t) = p(t \mid s, g^\beta(s)) = p(t \mid s, g^0(s)) = P(f, \bar{g})(s,t)$.

If $s \in S_3, P(f, g^\beta)(s,t) = p(t \mid s, f(s), j^0(s)) = P(f, \bar{g})(s,t)$. The proof of (b) is thus completed.

(c) For $\beta \geq \beta_0, s \in S_1, r(s, f^\beta, g^\beta) = (f^\beta(s))^t \cdot R_0^s \cdot (g^\beta(s)) =$
$\theta_0(s) \cdot 1^t \cdot (g^\beta(s)) = \theta_0(s) = r(s, \bar{f}, \bar{g})$ by definition of $\theta_0(s)$.

For $s \in S_2$, the proof is similar.

For $s \in S_3, r(s, f^\beta, g^\beta) = r(s, i_0(s), j_0(s)) = r(s, \bar{f}, \bar{g}) = \theta_0(s)$.

Hence (c) is proved.

(d) Without loss of generality, consider only $\beta \geq \beta_0$.

$$
\begin{aligned}
(1-\beta)v_\beta &= (1-\beta)v_\beta(f^\beta, g^\beta) \\
&= (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n P^n(f^\beta, g^\beta) \cdot r(f^\beta, g^\beta) \\
&= (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n P^n(\bar{f}, \bar{g}) \cdot \theta_0 \text{ by (a) and (c)above.} \\
&\longrightarrow P^*(\bar{f}, \bar{g}) \cdot \theta_0 \text{ by Blackwell [Ref.5].} \\
&= P^*(\bar{f}, \bar{g}) \cdot r(\bar{f}, \bar{g}) \text{ by (c).} \\
&= v(\bar{f}, \bar{g}).
\end{aligned}
$$

So, (d) is proved.

(e) If player I fixes his strategy as f (which is stationary) the problem reduces to a dynamic programming problem and by Blackwell [Ref.5], lemma 1.(d) Page 721, the proof of (e) is completed.

(f) As $\beta \to 1-, g^\beta \to \bar{g}$ and $r(f, g^\beta)$ is a continuous function of β. Hence $r(f, g^\beta) \to r(f, \bar{g})$ and thus (f) is proved.

(g) follows from (f) directly as other entries are bounded.

(h)

$$\begin{aligned}
&(1-\beta)v_\beta(f, g^\beta) \\
&= (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n P^n(f, g^\beta) \cdot r(f, g^\beta) \\
&= (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n [P^n(f, g^\beta) - P^n(f, \bar{g})] \cdot r(f, g^\beta) \\
&\quad + (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n [P^n(f, \bar{g}) - P^*(f, \bar{g})] \cdot r(f, g^\beta) \\
&\quad + (1-\beta)\textstyle\sum_{n=0}^{\infty} \beta^n P^*(f, \bar{g}) \cdot r(f, g^\beta).
\end{aligned}$$

Taking limit as $\beta \to 1-$, first two terms of the last expression tends to zero by (b) and (g) while the last term tends to $P^*(f, \bar{g}) \cdot r(f, \bar{g}) = v(f, \bar{g})$ by (f). Hence the proof of (h) is completed.

(i)

$$(1-\beta)v_\beta = (1-\beta)v_\beta(f^\beta, g^\beta) \geq (1-\beta)v_\beta(f, g^\beta),$$

Taking limits on both sides, as $\beta \to 1-$ and using (d) and (h), $v(\bar{f}, \bar{g}) \geq v(f, \bar{g})$ for any stationary f. Thus (i) is proved.

(j) As shown in (i), it can also be proved similarly that:

$$v(\bar{f}, g) \geq v(\bar{f}, \bar{g}) \text{ for any stationary } g.$$

Now let player I uses $\bar{f}$ in the undiscounted $SC/ARAT$ mixture. Let π_2 be an arbitrary strategy of player II. Then by Blackwell [Ref.5], there exists a stationary strategy g (depending on π_2) for player II such that

$v(\bar{f}, \pi_2) \geq v(\bar{f}, g)$. Similarly, if II uses $\bar{g}$, every strategy π_1 for player I has a stationary improvement say f. Thus, (i) and its analogue jointly with these facts imply that for all π_1 and π_2, $v(\bar{f}, \pi_2) \geq v(\bar{f}, \bar{g}) \geq v(\pi_1, \bar{g})$ which proves (j).

Note: From Mertens and Neyman [Ref.23], (d) directly implies $v(\bar{f}, \bar{g})$ is the undiscounted value. But the author chose to use the arguments of Blackwell here.

Theorem 2.6.11:

If the data of the $SC/ARAT$ mixture is taken from the field of rational numbers (or in general, from an Archamedian ordered field), $v(\bar{f}, \bar{g})$ and $(\bar{f}, \bar{g})$ will have rational entries. In other words, the undiscounted $SC/ARAT$ mixture has the ordered field property.

Proof:

The undiscounted value $v = \lim_{\beta \to 1-}(1-\beta)v_\beta$ by lemma 2.6.10(d), But $(1-\beta)v_\beta$ is a (component wise) rational function of β which takes on rational values if β is rational. Hence lemma 2.6.2 guarantees rationality of the components of v. To show that $\bar{f}$ has rational entries, it is enough to prove for $\{\bar{f}(s) : s \in S_2\}$ because $\{\bar{f}(s) : s \in S_1 \bigcup S_3\}$ have rational entries (by theorem 2.6.8 and lemma 2.6.9).

By definition, $\bar{f}(s) = \lim_{\beta \to 1-} f^\beta(s)$ if $s \in S_2$. Note that $f^\beta(s)$ is a (entrywise) rational function of β for $\beta \geq \beta_0$ and each entry is a rational number if β is rational. So lemma 2.6.2 guarantees the rationality of $\{\bar{f}(s) : s \in S_2\}$ The other part for player II follows similarly.

This concludes the proof of theorem 2.6.11.

2.7 Concluding Remarks:

1. The main results of this chapter (theorems 2.6.8. and 2.6.11.) are existential and directly raise the question that whether a finite step algorithm can be developed in a $SC/ARAT$ mixture. It is strongly believed that both the discounted and the undiscounted version of such games can be solved in finitely many steps by modifying existing algorithms for solving switching control games. But an efficient solution method is yet to be found out. It will be nice to have an algorithm which solves such games in finitely many steps.

2. Using results from Bewley and Kohlberg [Ref.3], it is possible to prove that in a $SC/ARAT$ mixture stochastic game, the undiscounted value and a pair of stationary optimal strategies lie in the same ordered field where the data of the game lies. But we chose to follow the approach of Parthasarathy and Raghavan [Ref.32].

Chapter 3

Semi-Infinite Stochastic Games

3.1 Introduction

In this chapter, we consider Shapley's stopping stochastic games [Ref.38]. There are z matrices $M^1, M^2, \cdots, M^z$, matrix M^s is of the size $m_s \times n_s$ and entry (i,j) of matrix M^s is given as $((\, r(s,i,j)/p(s,i,j)\,))$; where $r(s,i,j) \in \mathbf{R}$ is a payoff to player I and $p(s,i,j) = (p(1 \mid s,i,j), p(2 \mid s,i,j), \ldots, p(z \mid s,i,j)) \in \Delta_0^z = \{x \in \mathbf{R}^z : x \geq 0, \sum_{s=1}^{z} x_s < 1\}$. The interpretation is: if in state s, player I chooses $i \in A_s = \{1, 2, \ldots, m_s\}$ and player II chooses $j \in B_s = \{1, 2, \ldots, n_s\}$ then player II pays $r(s,i,j)$ to player I and on the next stage, the play moves from state (matrix) s to state t with probability $p(t \mid s,i,j)$ and the play stops with probability $1 - \sum_{i=1}^{z} p(t \mid s,i,j) > 0$.

In a stochastic game, the players consider strategies: infinite plans that tell a player at each stage, at each state and for every history of the play, what mixed action to choose at that stage in that state. Strategies for which the prescribed mixed actions do not depend on the histories are called Markov strategies; Markov strategies for which the prescribed mixed actions do not depend on the stages are called stationary strategies. Stationary strategies are simply z-tuples of mixed actions, one mixed action for each state. It is obvious that stationary strategies are most easy to work with.

An initial state $s \in S$, together with a pair of strategies (π_1, π_2) for players **I** and **II** respectively, determine a stochastic process over the set of states $S = \{1, 2, \ldots, z\}$ and hence for all stages τ, an expected direct payoff $R_{s\pi_1,\pi_2}(\tau)$ by player **II** to player **I** is determined. The players evaluate this stream of expected payoffs $(R_{s\pi_1\pi_2}(1), R_{s\pi_1,\pi_2}(2), \ldots)$ to be worth the reward $v(s, \pi_1, \pi_2) = \sum_{\tau=1}^{\infty} R_{s\pi_1\pi_2}(\tau)$.

By Shapley, it is known that there exists a unique $v \in \mathbf{R}^z$, a stationary strategy ρ^* for player **I** and a stationary strategy σ^* for player **II** such that for all π_1 and π_2 the following holds:

$$v(s, \pi_1, \sigma^*) \leq v(s) \leq v(s, \rho^*, \pi_2)$$

Here v is called the value of the stochastic game, $\rho^*(\sigma^*)$ is called an optimal stationary strategy for player **I** (**II**) .

Further, Shapley showed that : v is the value of the stochastic game, ρ^* is an optimal stationary strategy for player **I** and σ^* an optimal stationary strategy for player **II** if and only if for all $s \in S$, one has that $v(s) = val\, M_s(v)$ and that $\rho^*(s)$ is an optimal mixed action for player **I** in $M_s(v)$ and that $\sigma^*(s)$ is an optimal mixed action for player **II** in $M_s(v)$, where $M_s(v) = ((r(s,i,j) + \sum_{t=1}^{z} p(t \mid s,i,j)v(t)))_{i=1}^{m_s}{}_{,}{}_{j=1}^{n_s}$. Hence the value and optimal stationary strategies can be found by solving the set of equations:

$x_s = val\, M_s(x)$ for $s \in S, x \in \mathbf{R}^s$. This set of equations is known as the Shapley equation and can be shortened to $x = T(x)$ ($x \in \mathbf{R}^s$), where T is the function from $\mathbf{R}^s$ to $\mathbf{R}^s$ defined by $(T(x))_s = val\, M_s(x)$. The equation $x = T(x)$ ($x \in \mathbf{R}^s$) has a unique solution in $\mathbf{R}^s$ since T is a continuous contraction map.

Since 1953, the theory of stochastic games has been extended in many ways. Non stopping stochastic games, i.e. stochastic games for which $p(s,i,j) \in \Delta^s$ for all s, i, j, have been examined under the β-discounted as well as the limiting average payoff criteria, two different ways of evaluating the infinite streams of expected payoffs; non-zerosum stochastic games have been studied; stochastic games have been examined with infinite state and/or action spaces. The present chapter deals with the last category. In this last category, one mostly need some boundedness conditions on the payoffs to be sure that the rewards are well defined for all pairs of strategies. In this chapter, we investigate stopping stochastic games with a finite state space where for each state, player **I** has a finite action space and player **II** has a countably infinite action space. We put no restriction on the payoffs. Player **II**'s strategy space however is restricted to the set of strategies having finite support, i.e. strategies for which there exists an $N \in \mathbf{N}$ such that for all states, stages and histories, the probability that player **II** will choose a column in $\{1, 2, ...N\}$ equals 1. This is done to assure that all direct expected payoffs are well defined.

In section 3.2, we show that these semi-infinite stochastic games have a value in $\mathbf{R}_-^s$, where $\mathbf{R}_- = \mathbf{R} \cup \{-\infty\}$, that player **I** has optimal stationary strategies and that player **II** has near optimal stationary strategies. In section 3.3, we relate solutions for these stochastic games with solutions of the Shapley equation in $\mathbf{R}_{-}^{z}$.

We finish this section with some examples to illustrate that these semi-infinite (stopping) stochastic games are the only reasonable countably infinite extension of the finite stochastic games as introduced by Shapley if one wants to avoid putting any boundedness condition on the payoffs. The results of this chapter is taken from Sinha, Thuijsman and Tijs [Ref.39].

Example 3.1.1

For a stochastic game with countable state space and finite action spaces, one need bounded payoffs to get well defined rewards. Take for example the stochastic game with state space $\mathbf{N}$, and action set 1 for both players in all states. The transition is given by $p(s+1 \mid s, 1, 1) = \frac{1}{2}$ for all $s \in \mathbf{N}$ and $p(t \mid s, 1, 1) = 0$ otherwise; the payoffs to player I are given by $r(s, 1, 1) = (-2)^{s-1}$. Consider the game starting in state 1. In that case the reward $v(s, \pi_1, \pi_2)$ is not defined if one wants to take $\sum_{\tau=1}^{\infty} E_{s\pi_1\pi_2} R(\tau)$.

Example 3.1.2

For a stochastic game with finite state space and both players having countably many actions in the same state, the value need not exist. Take for example the stochastic game with $S = \{1\}, A_1 = B_1 = \mathbf{N}$, the transition is given by $p(1 \mid 1, i, j) = \frac{1}{2}$ for all i, j and the payoffs are given by $r(1, i, j) = 1$ if $i > j$, $r(1, i, j) = 0$ if $i = j$, $r(1, i, j) = -1$ if $i < j$. It is clear that $\sup_{\pi_1} \inf_{\pi_2} v(1, \pi_1, \pi_2) = -2$, whereas $\inf_{\pi_2} \sup_{\pi_1} v(1, \pi_1, \pi_2) = 2$.

Example 3.1.3

For a stochastic game with finite state space for which in each state one of the players has a finite action set and the other a countable action set, the value need not exist either. Consider the stochastic game with $S = \{1,2\}; A_1 = B_2 = \mathbf{N}, A_2 = B_1 = \{1\}; p(2 \mid 1, i, j) = \frac{1}{2} = p(1 \mid 2, i, j)$ for all i and j and $p(t \mid s, i, j) = 0$ elsewhere; $r(1, i, 1) = i, r(2, 1, j) = -j$ for all i and j.

If the players are not restricted to strategies with finite support, then $\sum_{\tau=1}^{\infty} E_{s\pi_1\pi_2} R(\tau)$ need not exist. If the players are restricted to strategies with finite support, then $\sup_{\pi_1} \inf_{\pi_2} v(s, \pi_1, \pi_2) = -\infty$ whereas $\inf_{\pi_2} \sup_{\pi_1} v(s, \pi_1, \pi_2) = +\infty$ for both $s = 1$ and $s = 2$.

3.2 Semi-Infinite Games:

Tijs [Ref.44] examined semi-infinite matrix games and showed that for such games the value exists, though it may be $-\infty$ if player **II** is the player with action set $\mathbf{N}$. For player **II** one can not allow all probability distributions over $\mathbf{N}$ as mixed actions without facing problems concerning the expected direct payoffs. Tijs considered serveral restrictions on player **II**'s set of mixed actions. One of them is restricting player **II** to use only mixed actions with finite support, i.e. mixed actions in $\cup_{n \in \mathbf{N}} \Delta^n$ where we identify Δ^n with the set $\{x \in \mathbf{R}^{\infty} : x \geq 0, \sum_{i=1}^{n} x_i = 1, x_i = 0 \text{ for } i > n\}$.

In this section, we extend the work of Tijs [Ref.44] on semi-infinite matrix games, with the above restriction for player **II**, to semi-infinite stochastic games.

Preliminaries

A semi-infinite stochastic game Γ_{∞} is given by a finite set of states S, where for each $s \in S$ there is an $m_s \times \infty$ matrix M^s of which each entry

(i,j) contains $r(s,i,j) \in \mathbf{R}$, the payoff to player **I** and $p(s,i,j) \in \Delta_0^s$, a transition vector. However, we assume that $\sup_{s,i,j}\{\sum_{t\in S} p(t \mid s,i,j)\} < 1$. Play proceeds in stages as explained in section 3.1. Player **II** is restricted to strategies π_2 with finite support, i.e. for each π_2 there exists an $N \in \mathbf{N}$ such that, before stopping, with probability 1 player **II** will choose one of the first N columns at all stages, in all states, for all histories.

Related with Γ_∞, for all $n \in \mathbf{N}$ one can look at the n-truncated stochastic game Γ_n, which we get by deleting all columns j with $j > n$, for all states. By completion with zeros, strategies for player **II** in Γ_n can be identified with finite support strategies for player **II** in Γ_∞; similarly, by deleting zeros, finite support strategies for player **II** in Γ_∞ can be identified with strategies for player **II** in Γ_n for n sufficiently large. We also identify Markov strategies for player **I** in Γ_∞ with those in $\Gamma_n, n \in \mathbf{N}$.

If there exists a $v_\infty \in \mathbf{R}_-^s$ such that for all $\epsilon > 0$ and all $\delta < 0$, there are strategies $\pi^1_{\epsilon\delta}$ and $\pi^2_{\epsilon\delta}$ for the respective players for which for all $s \in S$ all π^1, π^2 (with finite support),

$$v_\infty(s, \pi^1_{\epsilon\delta}, \pi^2) \geq v_\infty(s) - \epsilon$$

$$v_\infty(s, \pi^1, \pi^2_{\epsilon\delta}) \leq \begin{cases} v_\infty(s) + \epsilon & \text{if } v_\infty(s) \in \mathbf{R} \\ \delta & \text{if } v_\infty(s) = -\infty \end{cases}$$

then v_∞ is called the value of Γ_∞, and $\pi^1_{\epsilon\delta}$ is called an $\epsilon\delta$-optimal strategy for player **I** and similarly $\pi^2_{\epsilon\delta}$ for player **II**. If one can take $\epsilon = 0$ and $\delta = -\infty$ then $\epsilon\delta$-optimal strategies are called optimal strategies.

It is known that the value v_n of Γ_n exists for all n.

Theorem 3.2.1

(i) $\lim_{n\to\infty} v_n$ exists in $\mathbf{R}_-^s$ and equals v_∞, the value of Γ_∞.

(ii) Player **I** has optimal stationary strategies in Γ_∞.

(iii) Player **II** has near optimal (i.e. $\epsilon\delta$ – *optimal* for each $\epsilon > 0, \delta < 0$) stationary strategies in Γ_∞.

Proof:

For all $n \in \mathbf{N}$, it holds that $v_n \geq v_{n+1}$ coordinatewise, since in Γ_{n+1} player **II**'s action set is larger than in Γ_n, and player **I**'s action set remains the same. As $v_1, v_2, \dots$ is nonincreasing, it converges to some $w \in \mathbf{R}_-^s$ coordinatewise. We prove that w is the value of Γ_∞.

For every $n \in \mathbf{N}$, player **I** has an optimal stationary strategy ρ^n in Γ_n. All ρ^n, $n \in \mathbf{N}$, can be seen as elements of the compact set $\Delta^{m_1} \times \Delta^{m_2} \times \cdots \times \Delta^{m_s}$. Hence, without loss of generality we can assume that $\rho^1, \rho^2, \dots$ converges to some stationary strategy ρ^*. Now let π^2 be any strategy with finite support for player **II**. Then π^2 can be seen as a strategy for player **II** in Γ_n for n sufficiently large. Fix n_0 large enough. It is well-known that playing against a fixed stationary strategy in a stopping stochastic game with finite state and action spaces, Γ_{n_0} for instance, an optimal reply can be found among the stationary strategies. Blackwell [Ref.5] proved this for the special class of β-discounted stochastic games, but his proof can be applied to stopping games as well. Hence there is a σ for player **II** in Γ_{n_0} such that $v(\rho^*, \sigma) \leq v(\rho^*, \pi^2)$, coordinatewise. We also know $v(\rho^n, \sigma) \geq v_n$ for all $n \geq n_0$. Now using the fact that $\rho \mapsto v(\rho, \sigma)$ is a continuous function on the set of stationary strategies of player **I**, and using the fact that $v_1, v_2, \dots$ converges to w, we conclude $v(\rho^*, \sigma) \geq w$, and finally $v(\rho^*, \pi^2) \geq w$. So the lower value of Γ_∞ is at least w.

Take $\epsilon > 0, \delta < 0$. Since $v_1, v_2 \ldots$ converges to w, there exists $n_1 \in \mathbb{N}$ such that $v_{n_1}(s) \leq w(s) + \epsilon$ if $w(s)$ is finite, and $v_{n_1}(s) \leq \delta$ if $w(s) = -\infty$, for all $s \in S$. Let σ^* be an optimal stationary strategy for player **II** in Γ_{n_1}. Then again applying the argument that against σ^* player **I** has a stationary strategy as best answer, we have that for all strategies π^1 for player **I** and all $s \in S$, it holds that $v(s, \pi^1, \sigma^*) \leq v_{n_1}(s) \leq w(s) + \epsilon$ if $w(s)$ is finite; $v(s, \pi^1, \sigma^*) \leq v_{n_1}(s) \leq \delta$ if $w(s) = -\infty$. So the upper value of Γ_∞ is at most w. Combination of the above arguments proves the theorem.

It is clear that player **II** need not possess an optimal stationary strategy (with finite support). A simple example for this is the stochastic game consisting of one $1 \times \infty$ state matrix for which $r(1,1,j)$ equals $-j$ and $p(1,1,j) = 0$ for all $j \in \mathbb{N}$.

In the following theorem, we give a necessary and sufficient condition for player **II** to have optimal stationary strategies. First we introduce the concept of a *critical number* for a semi-infinite stochastic game Γ_∞. The critical number c is defined by $c = min\{n \in \mathbb{N} : v_n = v_\infty\}$, where $min\emptyset = \infty$. If c is finite, then Γ_∞ is called an *essentially finite* stochastic game.

Theorem 3.2.2

A semi-infite stochastic game Γ_∞ is essentially finite if and only if player **II** has optimal stationary strategies.

Proof:

Suppose $c < \infty$. Let σ^* be an optimal stationary strategy for player **II** in Γ_c. Playing against σ^* in Γ_∞, player **I** has a stationary best answer. For all stationary ρ we have $v(\rho, \sigma^*) \leq v_c = v_\infty$, so for all strategies π^1 for player **I** in Γ_∞, $v(\pi^1, \sigma^*) \leq v_\infty$. Hence σ^* is an optimal stationary strategy for player **II** in Γ_∞.

Starting with a stationary optimal strategy $\hat{\sigma}$ for player **II** in Γ_∞ which only uses the first n_1 columns, it is clear that for all stationary strategies ρ for player **I** we have $v(\rho, \hat{\sigma}) \leq v_\infty$. Hence $v_{n_1} \leq v_\infty$. But for all $n \in \mathbf{N}$ we also had $v_n \geq v_\infty$. So $v_{n_1} = v_\infty$ and therefore $c \leq n_1 < \infty$. This concludes the proof.

For essentially finite stochastic games Γ_∞, we look at the relation between sets of optimal stationary strategies in $\Gamma_\infty, O^1_\infty$ and O^2_∞, and the sets of optimal stationary strategies in Γ_n, O^1_n and O^2_n, $n \in \mathbf{N}$. Note that for an essentially finite stochastic game, both O^1_n and O^2_n are nonempty for all $n \in \mathbf{N} \cup \{\infty\}$ by theorems 3.2.1 and 3.2.2.

Theorem 3.2.3

Let Γ_∞ be an essentially finite stochastic game. Then $O^1_\infty = \bigcap_{n \geq c} O^1_n$ and $O^2_\infty = \bigcup_{n \geq c} O^2_n$.

Proof:

By the proof of theorem 3.2.1, we know that, if some sequence of stationary strategies $\rho^1, \rho^2, \ldots$, with $\rho^n \in O^1_n$ for all $n \in \mathbf{N}$, converges to a stationary strategy ρ^*, then $\rho^* \in O^1_\infty$. So we have $O^1_\infty \supset \bigcap_{n \geq c} O^1_n$. Conversely, if $\rho^* \in O^1_\infty$, then for all stationary strategies σ for player **II** in $\Gamma_n, n \geq c$, it holds that $v(\rho^*, \sigma) \geq v_\infty = v_n$. Hence $O^1_\infty \subset \bigcap_{n \geq c} O^1_n$. Combining, we have proved the first part of the theorem.

The player **II** part follows directly from the proof of theorem 3.2.2.

3.3 Shapley Equation

As already mentioned in the introduction, for finite two person stopping stochastic games, the value and optimal stationary strategies can be found by solving the Shapley equation $x = Tx$ ($x \in \mathbf{R}^s$), where $(Tx)_s = val\, M_s(x)$ $= val\,((\, r(s,i,j) + \sum_{t=1}^{s} p(t \mid s,i,j)x_t\,))_{i=1}^{m_s}{}_{j=1}^{n_s}$ which, in the finite case, has a unique solution.

One may also consider the Shapley equation for semi-infinite stochastic games. The difference is that for these stochastic games the $M_s(x)$ will be semi-infinite matrices and the system $x = Tx$ should be solved over $\mathbf{R}_-^s$. Examining $x = Tx$ ($x \in \mathbf{R}_-^s$) one quickly notes that $(-\infty, -\infty, \ldots, -\infty)$ will often be a solution, hence unlike the finite case we may no longer have a unique solution. However, some interesting results can be derived.

Theorem 3.3.1

The value of the semi-infinite stochastic game, v_∞ is a solution of the Shapley equation.

Proof:

For all $x \in \mathbf{R}_-^s$ and all $s \in S$, let $M_s(x)$ denote the semi-infinite matrix game with entries $((r(s,i,j) + \sum_{t=1}^{s} p(t \mid s,i,j)x_t))$ in $\mathbf{R}_-$ and let $M_s^n(x)$ denote the corresponding n-truncated matrix game for $n \in \mathbf{N}$. We know by Shapley that $v_n(s) = val\, M_s^n(v_n)$ for all $s \in S$.

First, suppose $v_\infty \in \mathbf{R}^s$ and take $\epsilon > 0$.

There exists $n_o \in \mathbf{N}$ such that for all $n \geq n_o$:

(i) $\mid v_\infty(s) - v_n(s) \mid < \epsilon$ for all $s \in S$, because $\lim_{n\to\infty} v_n(s) = v_\infty(s) \in \mathbf{R}$ for all $s \in S$.

(ii) $| \, val\, M_s^n(v_n) - val\, M_s^n(v_\infty) \,| < \epsilon$ for all $s \in S$, because payoffs in $M_s^n(v_n)$ differ less than ϵ from corresponding payoffs in $M_s^n(v_\infty)$ for all $s \in S$.

(iii) $| \, val\, M_s^n(v_\infty) - val\, M_s(v_\infty) \,| < \epsilon$ for all $s \in S$, because $\lim_{k \to \infty} val\, M_s^k (v_\infty) = val\, M_s(v_\infty)$.

Combining (i), (ii) and (iii) yields $| \, v_\infty(s) - val\, M_s(v_\infty) \,| < 3\epsilon$ for all $s \in S$. Since ϵ is arbitrary, we have $v_\infty(s) = val\, M_s(v_\infty)$ for all $s \in S$.

Now, if $v_\infty \notin \mathbf{R}^z$ then, without loss of generality, there is $k \in \{0, 1, \ldots, z-1\}$ such that $v_\infty(s) = -\infty$ if and only if $s \in \{k+1, k+2, \ldots, z\}$. For all $s \in \{1, 2, \ldots, k\}$, player **I** can prevent the play to move to any state outside $\{1, 2, \ldots, k\}$, otherwise $v_\infty(s)$ would be $-\infty$ as well. So for $s \in \{1, 2, \ldots, k\}$, player **I** can restrict to rows i in $M_s(v_\infty)$ for which $p(t \mid s, i, j) = 0$ for all $t \in \{k+1, k+2, \ldots, z\}$ and all $j \in$N. Hence, restricting player **I**'s action in this way yields a stochastic game $\tilde{\Gamma}$ for which play remains in $\{1, 2, \ldots, k\}$ forever. It is obvious that the value of this new stochastic game with state space $\{1, 2, \ldots, k\}$ equals the value of the original stochastic game on these states. So the value of the new stochastic game is finite and we can conclude $v_\infty(s) = val\, M_s(v_\infty)$ for $s \in \{1, 2, \ldots, k\}$ from above. If $v_\infty(s) = -\infty$ for some $s \in S$, then this implies that player **I** cannot force a play to remain within the set of states $\{1, 2, \ldots, k\}$ with probability 1. Thus each row of $M_s(v_\infty)$ contains some entries equal to $-\infty$. Hence $v_\infty(s) = val\, M_s(v_\infty)$ for all states (with value $-\infty$, or otherwise).

Theorem 3.3.2

if $x \in \mathbf{R}^z$ is a solution to the Shapley equation, then $x = v_\infty$.

Proof:

Suppose $x \in \mathbf{R}^z$ is a solution of the Shapley equation. Then for all $s \in S$:

$x_s = val\ M_s(x) = val\ ((r(s,i,j) + \sum_{t=1}^{z} p(t \mid s,i,j)x_t))_{i=1 j=1}^{m_s \infty}$.

Let ρ_s^* be an optimal mixed action for player **I** in $M_s(x)$ for all $s \in S$. Then for all stationary strategies σ for player **II**, we have $x_s \leq r(s,\rho_s^*,\sigma_s) + \sum_{t=1}^{z} p(t \mid s,\rho_s^*,\sigma_s)x_t$, where $r(s,\rho_s^*,\sigma_s) = \sum_{i=1}^{m_s}\sum_{j=1}^{\infty} \rho_s^*(i)r(s,i,j)\sigma_s(j)$ and $p(t \mid s,\rho_s^*,\sigma_s) = \sum_{i=1}^{m_s}\sum_{j=1}^{\infty} \rho_s^*(i)p(t \mid s,i,j)\sigma_s(j)$. In vector notation, letting $r(\rho^*,\sigma) = (r(1,\rho_1^*,\sigma_1),\ldots r(z,\rho_z^*,\sigma_z))$, and $P(\rho^*,\sigma^*)$, the $z \times z$ matrix with $p(t \mid s,\rho_s^*,\sigma_s)$ in entry (s,t), we get $x \leq r(\rho^*,\sigma) + P(\rho^*,\sigma)x$.

This implies $x \leq \sum_{\tau=1}^{n} P^{\tau-1}(\rho^*,\sigma)r(\rho^*,\sigma) + P^n(\rho^*,\sigma)x$ for all $n \in \mathbf{N}$, where $P^\tau(\rho^*,\sigma)$ is the τ-fold product of $P(\rho^*,\sigma)$ and $P^o(\rho^*,\sigma)$ is the identity matrix. Since, by the stopping play assumption, $\lim_{n\to\infty} P^n(\rho^*,\sigma)x = 0$ we have $x \leq \sum_{\tau=1}^{\infty} P^{\tau-1}(\rho^*,\sigma)r(\rho^*,\sigma) = v(\rho^*,\sigma)$. Hence $x \leq v_\infty$.

Conversely, let σ_s^* be an ϵ-optimal mixed action for player **II** in $M_s(x)$, for all $s \in S$. Then we have, for all stationary strategies ρ for player **I**: $x \geq r(\rho,\sigma^*) + P(\rho,\sigma^*)x - \epsilon \cdot 1_z$. Iterating, $x \geq v(\rho,\sigma^*) - \epsilon(1-\alpha)^{-1} \cdot 1_z$ where $\alpha = max_{s,i} \sum_{t=1}^{z} p(t \mid s,i,\sigma_s^*) < 1$. Since this can be done for all $\epsilon > 0$, we have $x \geq v_\infty$.

Theorem 3.3.2 says that there can be at most one real solution to the Shapley equation, and if there is one, it necessarily equals the v_∞.The following theorem gives the way to find v_∞ among all solutions.

Theorem 3.3.3

If $x \in \mathbf{R}_-^z$ is a solution to the Shapley equation, then $x \leq v_\infty$.

Proof:

Let $x \in \mathbf{R}_-^z$ be a solution. For $s \in S$ with $x_s = -\infty$, it is clear that $x_s \leq v_\infty(s)$. For $s \in S$ with $x_s \in \mathbf{R}$, we have $x_s = val\,((\,r(s,i,j) + \sum_{t=1}^{z} p(t \mid$

$s, i, j)x_t))_{i=1,j=1}^{m,\infty}$. This implies that for an optimal mixed action ρ_s^* of player **I** in $M_s(x)$ we have $\sum_{t=1}^{s} p(t \mid s, \rho_s^*, j)x_t \in \mathbf{R}$ *for all* $j \in \mathbf{N}$, *and hence* $p(t \mid s, \rho_s^*, j) = 0$ for all t with $x_t = -\infty$ and all $j \in \mathbf{N}$. Without loss of generality let $k \in S$ be such that $x_t = -\infty$ for $t > k$ and $x_t \in \mathbf{R}$ for $t \leq k$. Then, for $s \in \tilde{S} = \{1, 2, \ldots, k\}$, we have that for all stationary strategies σ of player **II**: $x_s \leq r(s, \rho_s^*, \sigma_s) + \sum_{t=1}^{k} p(t \mid s, \rho_s^*, \sigma_s)x_t$. In vector notation over $\tilde{S}$, $x \leq r(\rho^*, \sigma) + P(\rho^*, \sigma)x$. Now using $x_s \in \mathbf{R}$ for all $s \in \tilde{S}$, iteration of the above inequality gives $x \leq v(\rho^*, \sigma)$. Hence we have shown that $x_s \leq v_\infty(s)$ for all $s \in \tilde{S}$.

Theorem 3.3.4

A stationary strategy ρ^* for player **I** is optimal if and only if for each $s \in S$, the mixed action ρ_s^* is optimal in the semi-infinite matrix game $M_s(v_\infty)$.

Proof:

If for each $s \in S$, the mixed action ρ_s^* is optimal in $M_s(v_\infty)$ then, by the proof of theorem 3.3.3, it follows that ρ^* is an optimal stationary strategy in the stochastic game.

Conversely, let ρ^* be an optimal stationary strategy for player **I**. Suppose, for some $s \in S$, that ρ_s^* is not optimal in $M_s(v_\infty)$, then for some $j \in \mathbf{N}$ and some $\epsilon > 0$, it holds that $r(s, \rho_s^*, j) + \sum_{t=1}^{s} p(t \mid s, \rho_s^*, j)v_\infty(t) < val\ M_s(v_\infty) - \epsilon = v_\infty(s) - \epsilon$. Hence, if for the stochastic game starting in s, player **II** initially chooses column j and from then on uses an $\epsilon\delta$-optimal stationary strategy $\pi_{\epsilon\delta}^2$ against ρ^*,then for this strategy $(j, \pi_{\epsilon\delta}^2)$ of player **II**, we have $v(s, \rho_s^*, (j, \pi_{\epsilon\delta}^2)) \leq r(s, \rho_s^*, j) + \sum_{t=1}^{s} p(t \mid s, \rho_s^*, j)(v_\infty(t) + \epsilon) < v_\infty(s)$. This contradicts the optimality of ρ^* in the stochastic game.

Hence the proof follows.

For player **II**, near-optimal stationary strategies cannot directly be found as extensions of mixed actions in the matrix games $M_s(v_\infty)$. Think for instance of the stochastic game consisting of one $1 \times \infty$ state with $r(1,1,j) = -j$ and $p(1 \mid 1,1,j) = \frac{1}{2}$ for all $j \in \mathbf{N}$. It is obvious that $v_\infty = -\infty$. Hence $M_1(v_\infty)$ is the $1 \times \infty$ matrix game with payoff $-\infty$ in all entries. Clearly, the mixed action *choose column j* is an optimal mixed action for player **II** in the matrix game $M_1(v_\infty)$, but does not give any information about near optimal strategies in the stochastic game. This is due to the fact that a state s, with $v_\infty(s) = -\infty$, is either *directly good* or *indirectly good* for player **II**. The following two lemmas illustrate this phenomenon.

Lemma 3.3.5

If $v_\infty \notin \mathbf{R}^s$ and $\delta < 0$, then there is at least one state $s \in S$, in which player **II** has a mixed action q_s, such that the expected direct payoff in state s is at most δ if player **II** uses the mixed action q_s.

Proof:

If not, then player **I** has a stationary strategy ρ such that all expected direct payoffs are at least δ and hence $v \in \mathbf{R}^s$, leading to a contradiction. This completes the proof.

Let $D \subset S$, be the set of states for which player **II** has, for all $\delta < 0$, a mixed action to keep the the expected direct payoff below δ. If $v_\infty \notin \mathbf{R}^s$, then D is non-empty by lemma 3.3.5. D is called the set of states that are directly good for player **II**.

Lemma 3.3.6

If $v_\infty \notin \mathbf{R}^s$, then $ID = \{s \in S \backslash D : v_\infty(s) = -\infty\}$ and suppose $ID \neq \emptyset$. Then player **II** has mixed actions $q_s, s \in ID$, such that for all stationary strategies σ for player **II** with $\sigma_s = q_s$ for $s \in ID$, and all stationary strategies for player **I**, any play started in some state in ID will reach the set of states D with positive probability.

Proof:

If not, then $v_\infty(s) \in \mathbf{R}$ for some $s \in ID$. Contradiction. Hence the proof.

ID may be empty and is called the set of indirectly good states for player **II**. Player **II** can construct a near optimal strategy in the following way. In directly good states (belonging to D), player **II** ensures that the expected direct payoff is low enough; in indirectly good states (belonging to ID) player **II** ensures that the transitions will lead to D; in the other states, with finite value, player **II** has to consider direct payoffs as well as transitions.

Theorem 3.3.7

A near optimal stationary strategy for player **II** can be constructed by taking mixed actions $q_s, s \in S$ which are near optimal in the matrix game $[r(s,i,j)]_{i=1j=1}^{m,\infty}$ for $s \in D$, in the matrix game $[r(s,i,j) + \sum_{t=1}^{z} p(t \mid s,i,j)\, v_\infty(t)]_{i=1j=1}^{m,\infty}$, for $s \in S \backslash (D \bigcup ID)$, and for $s \in ID(k)$ to be taken near optimal in the matrix game $[\sum_{t=1}^{z} p(t \mid s,i,j) w^k(t)]_{i=1j=1}^{m,\infty}$, where $ID(k) = \{s \in ID$: player **I** cannot force the play to move from s to D within k stages, with positive probability $\}$, and where $w^k(t) = -\infty$ for $t \in D \bigcup (\bigcup_{l=1}^{k} ID(l))$ and $w^k(t) = 0$ for other $t, k \in \{0, 1, 2, \ldots, z-1\}$.

Proof:

Follows directly from Lemma 3.3.5, Lemma 3.3.6 and part of the proof of Theorem 3.3.2.

For stochastic games with finite state and action spaces and any $x \in \mathbf{R}^z$, it holds that $\lim_{n\to\infty} T^n(x) = v_\infty$ by the facts that T is a contraction operator and v_∞ is its unique fixed point. Since for semi-infinite stochastic games the value may be $-\infty$ in some coordinates, it is not directly clear whether the above method can be used to find the value. The following theorem answers this problem.

Theorem 3.3.8

For all $x \in \mathbf{R}^s$, $\lim_{n\to\infty} T^n(x) = v_\infty$.

Proof:

$D = \{s \in S : (T(x))_s = -\infty\}, ID = \{s \in S\backslash D : (T^s(x))_s = -\infty\}$. If $S = D \bigcup ID$, we are finished, else we continue in the following way: For the remaining states $(T^s(x))_s \in \mathbf{R}$ and hence $(T^n(x))_s \in \mathbf{R}$ for all these remaining states. Let $R = S\backslash(D \bigcup ID)$. Without loss of generality suppose $R = \{1, 2 \ldots, k\}$. For $y = (y_1, y_2, \ldots, y_k, -\infty, \ldots, -\infty) \in \mathbf{R}^s_-$ with $\hat{y} = (y_1, y_2, \ldots, y_k) \in \mathbf{R}^k$, and $s \in R$ it holds that $(T(y))_s = (\hat{T}(\hat{y}))_s$, where

$$(\hat{T}(\hat{y}))_s = val[r(s,i,j) + \sum_{t=1}^{z} p(t \mid s,i,j)\hat{y}_t]_{i=1\ j=1}^{m_s(R)\infty} = val\hat{M}_s(\hat{y}),$$

where, without loss of generality, $\{1, 2, \ldots, m_s(R)\}$ is the set of rows of $M_s(y)$ for which all entries are reals, and where $\hat{M}_s(\hat{y})$ is the matrix consisting of those rows.

For $a, b \in \mathbf{R}^k, a \neq b$, it holds that $||\hat{T}(a)\delta - \hat{T}(b)|| < ||a - b||$, where $||x|| = \max_s \mid x_s \mid$. Hence $\hat{T}$ is a contraction map on $\mathbf{R}^k$ and has a unique fixed point which necessarily equals $\lim_{n\to\infty} \hat{T}^n(\hat{x})$ for all $\hat{x} \in \mathbf{R}^k$. For $s \in R$, player I's optimal stationary strategies never lead the play to $D \bigcup ID$ and hence, starting in R, the stochastic game can be seen as a stochastic game $\hat{\Gamma}$ with R as set of states. Simply restrict player I's action sets as is done above. Applying theorem 3.3.1 and 3.3.2 to $\hat{\Gamma}$, we have, $v_\infty(s) =$

$\lim_{n\to\infty}(\hat{T}^n(\hat{x}))_s = \lim_{n\to\infty}(T^n(x))_s$ for $s \in R$.

For $s \in D \bigcup ID$ and $n \geq z$: $(T^n(x))_s = -\infty = v_\infty(s)$. Hence $\lim_{n\to\infty} T^n(x) = v_\infty$.

Observe that $\lim_{n\to\infty} T^n(x)$ need not equal to v_∞ if we start with $x \in \mathbf{R}_-^z \backslash \mathbf{R}^s$. This is illustrated by the following example: take $S = \{1,2,3\}$; $A_1 = \{1,2\}, A_2 = A_3 = \{1\}, B_1 = B_2 = B_3 = \mathbf{N}$; $r(1,1,j) = r(2,1,j) = r(3,1,j) = 0$ for all $j \in \mathbf{N}$; $r(1,2,j) = 2$ for all $j \in \mathbf{N}$; $p(1 \mid 1,1,j) = p(3 \mid 1,2,j) = p(1 \mid 2,1,j) = p(3 \mid 3,1,j) = \frac{1}{2}$, $p(t \mid s,i,j) = 0$ elsewhere. Then, starting with $x = (0,-\infty,-\infty)$, it follows that $T^n(x) = (0,0,-\infty)$ for all $n \in \mathbf{N}$, whereas v = (2,1,0).

3.4 Concluding remarks:

In the literature, β-discounted stochastic games have been studied extensively. These are stochastic games that continue indefinitely, since $\sum_{t=1}^{z} p(t \mid s,i,j) = 1$ for all s,i,j. The players, however, discount future payoffs by some factor $\beta \in (0,1)$. So for a pair of strategies (π_1,π_2) and an initial state s, the β-discounted payoff to player **I** is given by $\sum_{\tau=1}^{\infty} \beta^{\tau-1} R_{s\pi_1\pi_2}(\tau)$.

Notice that a β-discounted stochastic game can be seen as a stopping stochastic game if we relate transition probabilities $p(t \mid s,i,j)$ in the β-discounted non stopping game with transition probabilities $\beta\, p(t \mid s,i,j)$ in the stopping game. Hence all the results derived in this paper for stopping stochastic games, also hold for β-discounted stochastic games.

In section 3.3.2 we made the assumption $sup_{s,i,j}\{\sum_{t\in S} p(t \mid s,i,j)\} < 1$. This condition played no role until lemma 3.3.5.

Chapter 4

Stochastic Games with Uncountably Many States

4.1 Introduction

In 1950, Nash [Ref.25] introduced the concept of equilibrium points in N-person noncooperative games. He showed that every noncooperative N-person game in normal form has an equilibrium point in mixed strategies in the sense that no single player can be unilaterally better off by changing from that point. (See Chapter 1 for a detail).

In 1964, Fink [Ref.12] and Takahashi [Ref.42] extended this idea in defining a noncooperative stochastic game. They proved that every discounted noncooperative stochastic game has a Nash equilibrium point when the state and action spaces are all finite. Takahashi also considered the case when the action spaces are compact under certain topological assumptions on the reward and transition functions. Later, Rogers [Ref.37] and Sobel [Ref.40] also considered such games independently.

The notion of uncountable state space in the literature of stochastic games was introduced by Maitra and Parthasarathy [Ref.21]. They showed the existence of the value and a pair of optimal stationary strategies in a zero sum discounted stochastic game if the state and action spaces are compact provided the reward is continuous and the transition is weakly continuous.

In 1976, Himmelberg, *et.al.* [Ref.14] introduced the notion of p-equilibrium in the class of stochastic games with uncountable state space. They showed the existence of p-equilibrium stationary strategies in a non-cooperative stochastic game when the reward and transition satisfy some separbility conditions and the state space is a Borel subset of a complete separable metric space. Parthasarathy [Ref.31] proved that in such games a stationary Nash equilibrium exists when there is a p-equilibrium stationary strategies. On the existence of Nash equilibrium. Rieder [Ref.36] and Whitt [Ref.49] approximated countable person noncooperative games with uncountable state and action spaces by simpler games. Rieder proved the existence of an equilibrium pair when the state and action spaces are Borel sets and the transition is a bounded transition measure. Whitt approximated such games with games having countable state and action spaces and he proved that the equilibria of the game with countable set up become ϵ-equilibria of the uncountable game under some conditions.

Nowak [Ref.29] also approximated a discounted noncooperative stochastic game with uncountable state and compact action spaces by a sequence of finite or countable state games and showed the existence of a stationary ϵ-equilibrium pair in the original game under certain continuity and integrability assumptions.

In 1982 [Ref.31], Parthasarathy has given an open problem which asks for a stationary equilibrium for the discounted stochastic game with compact state space and finite action spaces when the rewards are bounded measurable or continuous functions. Recently, Mertens and Parthasarathy [Ref.24] provided a partial answer to this problem by proving the existence of a nonstationary equilibrium for such games.

The present chapter concentrates on the same problem and it is shown here that if the transition is state independent and is dominated by a fixed nonatomic measure, then a β-discounted N-person noncooperative stochastic game with compact state space, finite action spaces and continuous rewards, have a stationary equilibrium strategies. For related works in this line, we refer the interested readers to [Ref.1, 22 and 30]. The material of this chapter is taken from Parthasarathy and Sinha [Ref.33].

4.2 Definitions and Assumptions

A nonzero sum N-person stochastic game is determined by $< S, A_i, r_i, p;$ $i = 1, 2, \ldots, N >$. Here S =[0,1] is the state space ; $A_i = \{1, 2, \ldots, k_i\}$ finite sets, describing actions available to player i . $r_i(s, \cdot)$ is the immediate payoff to player i , which depends on the current state s and the actions of the players and $p(\cdot \mid s, \cdot)$ is a measurable transition function defined on the set of all $(s, \cdot)$, the range being probabilities on the Borel sets of S. As a consequence of the actions chosen by the N-players, the following things take place :

(a) Each player (say i) receives a reward $r_i(s, \cdot)$.

(b) On the next stage, the system moves to a new state s' according to the transition law $p(s' \mid s, \cdot)$.

(c) The whole process is repeated from the new state s'.

The game is played over the infinite future. Every player wants to maximize his accumulated rewards and the rewards are accumulated with a discount factor $\beta \in [0, 1)$, that is, on the n^{th} day, the pay off to player i is $\beta^{n-1} r_i^n(s, \cdot)$, where r_i^n stands for the reward on the n^{th} day.

A strategy π_i for player i is a sequence $(f_1, f_2, \ldots, f_k, \ldots)$ where f_k specifies the actions to be chosen on the k^{th} day, depending on the past history h_{k-1}. Note that f_k is a probability distribution on A_i given h_{k-1} and $f_k(E \mid h_{k-1})$ is measurable in h_{k-1} for every Borel set E. A strategy π_i is called stationary if there exists a Borel map $f : S \mapsto P_{A_i}$ such that $f_k = f$ for all k, where P_{A_i} stands for the space of probability vectors on A_i.

Let $(\pi_i : 1 \leq i \leq N)$ be an N-tuple of strategies for the N players. Associated with each initial state s, if the players play this strategy tuple, the total β-discounted expected pay off to player i is given by $I_i(\pi_1, \pi_2, \ldots, \pi_N)(s) = \sum_{n=1}^{\infty} \beta^{n-1} r_i^n(\pi_1, \pi_2, \ldots, \pi_N)(s)$. An N-tuple $(\pi_i^* : 1 \leq i \leq N)$ is an equilibrium in the sense of Nash if

$$I_i(\pi_1^*, \ldots \pi_i^*, \ldots, \pi_N^*)(s) \leq I_i(\pi_1^*, \ldots \pi_{i-1}^*, \pi_i, \pi_{i+1}^*, \ldots, \pi_N^*)(s)$$

for all $s \in S$, for all π_i and for all $i = 1, 2, \ldots, N$. The law of transition $p(ds' \mid s, \cdot)$ is called state independent (denoted by SIT) if $p(ds' \mid s_1, \cdot) = p(ds' \mid s_2, \cdot)$ for all s_1, s_2. For notational convenience, if the transition is SIT, we denote it by $p(ds' \mid \cdot)$ only.

Assumptions:

(i) r_i is bounded by some constant K for all $i = 1, 2 \ldots, N$.

(ii) r_i is continuous on $S \times A_1 \times \cdots \times A_N$ for all $i = 1, 2, \ldots N$.

(iii) There exists a fixed nonatomic measure μ such that $p(\cdot \mid \cdot) \ll \mu$ i.e. transition probabilities are absolutely continuous with respect to μ.

(iv) We shall use the expected pay off

$$I_i(\pi_1, \ldots, \pi_N)(s) = (1 - \beta) \sum_{n=1}^{\infty} \beta^{n-1} r_i^n(\pi_1, \ldots \pi_N)(s)$$

so that I_i is bounded by K for all i .

4.3 Main Theorem and Preliminaries

Main Theorem

Under assumptions (i), (ii), (iii), and (iv), a stochastic game with state independent transition (SIT) under β-discounted pay off has stationary Nash equilibrium strategies.

Preliminaries:

To prove this theorem, we need some preliminaries. Denote by F the set of all possible measurable pay off functions, each component being bounded by K . That is,

$$F = \{f : S \longrightarrow \mathbf{R}^N : f = (f_1, f_2, \dots f_N) \text{ and } | f_k(s) | \leq K \text{ for all } k = 1, 2, \dots N \text{ and } s \in S\}.$$

Given $f \in F$, $s \in [0,1]$, define an N-person non zero sum finite dummy game $G_f(s)$, where the pay off to the k^{th} player is given by

$$(1-\beta) r_k(s, i, j, \dots) + \beta \int f_k(\cdot) dp(\cdot \mid s, i, j, \dots), k = 1, 2, \dots N.$$

For every s, this finite game has a Nash equilibrium point. Let $N_f(s) = \{h(s) = (h_1(s), \dots, h_N(s)) : h_k(s)$ is a Nash equilibrum pay off to the k^{th} player in $G_f(s)$ where $k = 1, 2, ..., N\}$. Define $\bar{N}_f(s) =$ Convexhull of $N_f(s)$ for each s. Let N_f and $\bar{N}_f$ be all measurable selections from $N_f(s)$ and $\bar{N}_f(s)$ respectively. One can show that N_f and $\bar{N}_f$ are nonempty by a selection theorem in [Ref.19].

We shall endow F as well as the space of stationary strategies with weak* ($= w^*$) topology induced by the fixed nonatomic measure μ. We first prove the following basic lemma.

Lemma 4.3.1

The correspondence $\Psi : F \longrightarrow 2^F$ with $\Psi(f) = \bar{N}_f$ is closed , convex valued and $w*$-upper semicontinuous.

Proof: Clearly $\bar{N}_f$ is convex valued since $\bar{N}_f(s)$ is convex for each s .

Let $f_k \in F, f_k \xrightarrow{w^*} f, \phi_k \in \bar{N}_{f_k}$ for each k and $\phi_k \xrightarrow{w^*} \phi$. To show Ψ is upper semicontinuous, it is enough to prove that $\phi \in \bar{N}_f$. In fact, it is enough to show that $\phi(s) \in \bar{N}_f(s)$ almost everywhere, since it does not matter if we change the value of a function on a set of measure zero.

Clearly some subsequence of convex combinations of ϕ_k converges to ϕ almost everywhere. We continue to call the subsequence as ϕ_k without loss of generality. Since $f_k \longrightarrow f$ in w^*-topology, $G_{f_k}(s) \longrightarrow G_f(s)$ for every s. Hence

$$\limsup N_{f_k}(s) \subset N_f(s)$$

or

$$\limsup \bar{N}_{f_k}(s) \subset \bar{N}_f(s) \text{ for every } s.$$

This means given $\epsilon > 0$, there is a k_0 such that for all $k \geq k_0, \bar{N}_{f_k}(s)$ $\subset$ ϵ-neighbourhood of $\bar{N}_f(s)$; in other words, $\phi_k(s) \in$ ϵ-neighbourhood of $\bar{N}_f(s)$. So, $\phi(s) \in$ ϵ-neighbourhood of $\bar{N}_f(s)$. Since ϵ is arbitrary, $\phi(s) \in \bar{N}_f(s)$ almost everywhere. Thus $\phi \in \bar{N}_f$ which proves Ψ is upper semicontinuous. The same proof also shows that $\Psi(f)$ is closed for each f. This terminates the proof of the lemma.

Lemma 4.3.2

There exists a $g_0 \in F$ such that $g_0 \in \bar{N}_{g_0}$.

Proof: Observe F is compact and metrizable in the w^*- topology. Now use lemma 4.3.1. to invoke Glicksberg's fixed point theorem which yields the desired result.

Note that $g_0 \in \bar{N}_{g_0}$ almost everywhere. modify g_0 on a null set to make it measurable and $g_0 \in \bar{N}_{g_0}$ everywhere. This terminates the proof of lemma 4.3.2.

Lemma 4.3.3

There exists a g^* such that $g^* \in N_{g^*}$ under the SIT assumption.

Proof: By lemma 4.3.2, there is a $g_0 \in \bar{N}_{g_o}$. The idea is to replace g_o by a

g^* such that $g^* \in N_{g_0}$ and $N_{g^*} = N_{g_0}$.

For every s consider the game $G_{g_0}(s)$. Here the pay off to player k is $(1-\beta)r_k(s,i,j,\ldots) + \beta \int g_{ok}(\cdot)dp(\cdot \mid i,j,\ldots)$. Since $g_o(s) \in \bar{N}_{g_o}(s) =$ Convexhull of $N_{g_o}(s)$, it follows that $\int g_o(\cdot)dp(\cdot \mid i,j,\ldots) \in$ Convex hull of $\int N_{g_o} h \, d\mu$, where h is the density of $p(\cdot \mid i,j,\ldots)$ with respect to μ. Since μ is nonatomic, by Liapunov [theorem 17.1.6, page187, Ref.16], $\int N_{g_o} h \, d\mu$ is a convex set. So, there is a selection g^* from N_{g_o} such that $\int g_o(\cdot)\, dp(\cdot \mid i,j,\ldots) = \int g^*(\cdot)dp(\cdot \mid i,j,\ldots)$. Observe that finite game $G_{g_o}(s)$ does not change if we replace g_o by g^* and consequently, $N_{g^*}(s) = N_{g_o}(s)$ for every s. Hence $g^* \in N_{g^*}$ everyewhere. This terminates the proof of lemma 4.3.3.

Lemma 4.3.3. precisely means that for every s, $g^*(s)$ is an equilibrium pay off to the finite game $G_{g^*}(s)$. Next we show that g^* is an equilibrium pay off to the β-discounted stochastic game under the assumptions (i) through (iv) and SIT. We also show the existence of stationary equilibrium strategies for the N-players corresponding to the Nash payoff g^*.

Proof of the main theorem:

From lemma 4.3.3., we have a $g^* \in N_{g^*}$ everywhere. Fix the finite game $G_{g^*}(s)$ throughout the rest of the argument. Consider the following set E:

> $E = \{(s, p, \mu_1, \mu_2, \ldots, \mu_N)$: for each s, p is a feasible Nash payoff and $\mu_1, \mu_2, \ldots, \mu_N$ Nash equilibrium points that induce p corresponding to the finite game $G_{g^*}(s)\}$.

Now, we claim that E is a Borel set. To prove the claim, let X be the graph of the correspondence $s \mapsto N_{g^*}(s)$ *i.e.* $X = \{(s,p) : s \in [0,1]$ and $p \in N_{g^*}(s)\}$. It is easy to see that this correspondence is upper semicontinuous.

So, its graph, X, is a closed set. For $(s,p) \in X$, define the correspondence H, which maps each $(s,p) \mapsto H(s,p) = \{(\mu_1, \mu_2, \ldots, \mu_N) : (\mu_1, \mu_2, \ldots, \mu_N)$ is an equilibrium strategy tuple which yields a Nash pay off p at s $\}$. Clearly E is the graph of this correspondence H, which can be checked to be upper semicontinuous. Hence E is closed and in particular, Borel. The (s,p) sections of E, *i.e.* $H(s,p)$ is nonempty and compact. Hence from Kunugui-Novikoff's theorem, we can find a Borel map σ^* such that $(s, p, \sigma^*(s,p)) \in E$. In other words, $\sigma^*(s,p)$ is an equilibrium point yielding a Nash pay off p at state s. Since $g^*(s) \in N_{g^*}(s)$, $\sigma^*(s, g^*(s))$ is an equilibrium strategy tuple in $G_{g^*}(s)$. For notational convenience, let $\sigma^*(s, g(s))$ be denoted by $(\sigma_1^*(s), \sigma_2^*(s), \ldots, \sigma_N^*(s))$. Then by Blackwell [Ref.6],

$$g_k^*(s) = \sup_{\pi} I_k(\sigma_1^*, \ldots, \sigma_{k-1}^*, \pi, \sigma_{k+1}^*, \ldots, \sigma_N^*)(s).$$

where π varies over all strategies of player k in the dynamic programming problem, where all players except k use the stationary strategy σ^*. The supremum remains the same if π varies over all strategies of player k in the stochastic game [theorem 3.1, page 295, Ref.21]. Hence $\sigma^*(s, g*(s))$ is a set of stationary Nash equilibrium strategies for the stochastic game under consideration. This terminates the proof of the main theorem.

4.4 Concluding Remarks

1. The main theorem remains valid if S, the state space is a Borel subset of a complete separable metric space.
2. it is not known whether the assumption of state independent transi-

tion is redundant in the main theorem *i.e.* whether there is a set of stationary equilibrium strategies in a discounted stochastic game with uncountable state space, finite action spaces and transitions dominated by a fixed nonatomic measure.

Bibliography

[1] Amir,R(1987):On stochastic games with uncountable state and action spaces, discussion paper, Cowles Foundation, Yale Univ, U.S.A.

[2] Bewley,T. and Kohlberg,E(1976):The asymptotic theory of stochactic games, Math.O.R.,1,197-208.

[3] Bewley,T. and Kohlberg,E(1976):The asymptotic solution of a recursion equation occuring in stochastic games,Math.O.R.,1,321-336.

[4] Bewley,T. and Kohlberg,E(1978):On stochastic games with stationary optimal strategies,Math.O.R.,3,104-125.

[5] Blackwell,D(1962):Discrete dynamic programming, Ann. Math. Stat., 33, 719-726.

[6] Blackwell,D(1965):Discounted dynamic programming, Ann. Math. Stat., 36, 226-235.

[7] Blackwell,D.and Ferguson,T.S(1968):The big match, Ann. Math. stat., 39, 159-163.

[8] Everett,H(1957):Recursive games, Contributions to the theory of games III, Ann. Math. Stud., 39, 47-78.

[9] Filar,J.A(1980):Algorithms for solving some undiscounted stochastic games, Ph.D. thesis, Univ. of Illinois at Chicago.

[10] Filar,J.A(1981):Ordered field property for stochastic games when the player who controls the transition changes from state to state, J.O.T.A., 34, 505-517.

[11] Filar,J.A. and Raghavan,T.E.S(1988):Algorithms of stochastic games, mimeo.

[12] Fink, A.M(1964): Equilibrium in a stochastic n-person game, Jour. Sci. Hiroshima. Univ. Series A, 28,89-93.

[13] Gillette,D(1957): Stochastic games with zero stop probabilities, Contributions to the theory of games III, 179-187.

[14] Himmelberg, C.J., Parthasarathy, T., Raghavan, T.E.S. and Van Vleck, F.S(1976): Existence of p-equilibrium and optimal stationary strategies in stochastic games, Proc. Amer. Math. Soc, 60, 245-251.

[15] Kaplansky,I(1945): A contribution to von Neumann's theory of games,Ann. of Math. 46, 474-479.

[16] Klein, E and Thompson, A.C(1984): Theory of correspondences, including applications to mathematical economics, Wiley-Interscience Publication, John Wiley and Sons.

[17] Kohlberg, E(1974): Repeated games with absorbing states, Ann. Stat., 2, 724-738.

[18] Kuhn, H.W(1953): Extensive games and the problem of information, Contribution to the theory of games II, Ann. Math. Stud. No.28, 193-216.

[19] Kuratowski, K. and Ryll-Nardzewski(1963): A general theorem on selectors, Bull. Acad. Poland Sci. Ser. Math. Astronom Phys.,13, 397-403.

[20] Loomis, L.H(1946): On a theorem of von Neumann, Proc. Nat. Acad. Sci, 32, 213-215.

[21] Maitra, A. and Parthasarathy, T(1970): On stochastic games, J.O.T.A., 5, 289-300.

[22] Majumdar, M. and Sundaram, R(1988): Existence of stationary equilibrium in pure strategies in symmetric stochastic games with uncountable state and action spaces, Pre-print.

[23] Mertens, J.F. and Neyman, A(1981): Stochastic games, Int. J. Game Theory, 10, 53-66.

[24] Merten, J.F. and Parthasarathy, T(1987): Equilibria for discounted stochastic games, CORE discussion paper no. 8750.

[25] Nash, J.F(1950): Equilibrium points in n-person games, Proc. Nat. Acad. Sci. U.S.A, 36, 48-49.

[26] von Neumann, J(1928): Zur Theorie der Gesellschaftsspiele, Mathematische Annalen, 100, 295-320.

[27] von Neumann, J(1937): Ergebnisse eines Math. Kolloquiums, 8, 73-83.

[28] von Neumann, J. and Morgenstern, O(1944): Theory of games and economic behavior, Princeton, Princeton Univ. Press.

[29] Nowak, A(1985): Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space, J.O.T.A, 45, 591-601.

[30] Nowak, A. and Raghavan, T.E.S(1988): Existence of stationary correlated equilibria for discounted stochastic games in arbitrary state spaces,Tech. Report: 88-03, Stat. Lab, Univ. Illinois at Chicago, U.S.A.

[31] Parthasarathy, T(1982): Existence of equilibrium stationary strategies in discounted stochastic games, Sankhya; Ind. Jour. Stat, 44, 114-127.

[32] Parthasarathy, T. and Raghavan, T.E.S(1981): An ordered field property for stochastic games when one player controls transition probabilities, J.O.T.A, 33, 375-392.

[33] Parthasarathy, T. and Sinha, S(1986): Existence of stationary equilibrium strategies in non zero sum discounted stochastic games with uncountable state space and state independent transitions, pre-print, forthcoming in Int. Jour. Game Theory.

[34] Parthasarathy, T., Tijs, S.H. and Vrieze, O.J(1984): Stochastic games with state independent transition and separable rewards, In: Selected topics in operations research and mathematical economics, Springer Verlag Berlin, No.226, 262-271.

[35] Raghavan, T.E.S., Tijs, S.H. and Vrieze, O.J(1985): On stochastic games with additive reward and transition structure, J.O.T.A., 47, 451-464.

[36] Rieder, U(1979): Equilibrium plans for non zero sum markov games, Game Theory and Related Topics, Eds. Moeschlin, O. and Pallaschke, D., North-Holland, Amsterdam, Holland.

[37] Rogers, P.D(1979): Non zero sum stochastic games. Ph.D. Thesis, Report ORC 69-8, Univ. of California, Berkeley.

[38] Shapley, L(1953): Stochastic games, Proc. Nat. Acad. Sci. U.S.A., 39, 1095-1100.

[39] Sinha, S., Tijs, S.H. and Thuijsman, F(1987): Semi infinite stochastic games, report, Dept. of Math., Catholic Univ., Nijmegen, Netherland; to appear in "Stochastic Games and Related Topics", Editors: T.E.S. Raghavan, T.S. Ferguson, [redacted] T. Parthasarathy, [redacted] and O. J. Vrieze, (a Riedel publication).

[40] Sobel, M.J(1973): Continuous stochastic games, Jour. Appl. Prob, 10, 497-504.