ZOR – Methods and Models of Operations Research (1991) 35:437–472

Algorithms for Stochastic Games – A Survey[1]

By T.E.S. Raghavan[2] and J.A. Filar[3]

Abstract: We consider finite state, finite action, stochastic games over an infinite time horizon. We survey algorithms for the computation of minimax optimal stationary strategies in the zerosum case, and of Nash equilibria in stationary strategies in the nonzerosum case. We also survey those theoretical results that pave the way towards future development of algorithms.

Zusammenfassung: In dieser Arbeit werden unendlichstufige stochastische Spiele mit endlichen Zustands- und Aktionenräumen untersucht. Es wird ein Überblick gegeben über Algorithmen zur Berechnung von optimalen stationären Minimax-Strategien in Nullsummen-Spielen und von stationären Nash-Gleichgewichtsstrategien in Nicht-Nullsummen-Spielen. Einige theoretische Ergebnisse werden vorgestellt, die für die weitere Entwicklung von Algorithmen nützlich sind.

Introduction

The subject of Stochastic Games was initiated in 1953 in a fundamental paper of Shapley [49]. These games are dynamic, stochastic models of noncooperative competitive behavior. They include as special cases the static noncooperative games, the repeated games with complete information, and the Markovian Decision Processes (MDP's). To the extent that the subject of MDP's or "Discrete Dynamic Programming" evolved almost independently in the 1960's and 1970's, Shapley's 1953 paper was literally ahead of its time.

[1] This paper is based on the invited lectures given by the authors at the 12th Symposium for Operations Research in Passau, 1987. We are indebted to M. Abbad, Evangelista Fe, F. Thuijsman and O.J. Vrieze for valuable comments and discussion. Any remaining errors of either misinterpretation or of omission are the authors' alone.

[2] T.E.S. Raghavan, Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, IL 60680. Supported in part by the NSF under the grant # DMS-82601403.

[3] Jerzy A. Filar, Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21288. Supported in part by the AFOSR and the NSF under the grant # ECS-8704954.

0340–9422/91/6/437–472 $2.50 © 1991 Physica-Verlag, Heidelberg

Stochastic Games now constitute a significant topic within the discipline of Game Theory, as demonstrated by the high level of research activity in this area. As early as 1977, a survey paper by Parthasarathy and Stern [40] contained more than 150 references. Since then the level of research activity has, if anything, increased. Unavoidably, perhaps, the subject of Stochastic Games began fragmenting into quite a number of research directions that were being developed by particular groups of researchers. Unfortunately, many of the results of these researches are dispersed over a large number of scientific journals (spanning quite a few disciplines), conference proceedings, doctoral dissertations, and research memoranda. Consequently, it is difficult, especially for the uninitiated, to develop a global view of even the major trends in the evolution of the subject.

The objective of this survey is to report on what, in our opinion, constitutes one of the major trends today, namely, the Algorithms for Solving Stochastic Games. With the maturing of the existence theory (e.g., see Mertens and Neyman [32]) we believe that the above is the natural trend, paralleling the development of many branches of Applied Mathematics. Furthermore, it is the trend that should be of most interest to the Operations Researchers since its results may enable them to use Stochastic Games as modeling tools. The interest in the latter is indicated by the relatively recent emergence of quite a variety of "academic applications" of Stochastic Games that range from the models of super-power arms race and military combat, through some models of fisheries, to models of inspection processes, and even some models of sporting competition (e.g., see [65], [9], [54], [15] and [66]).

In order to keep the size of this survey manageable, and also to increase its potential appeal to the Operations Research/Mathematical Programming community, we tried to follow the guidelines below:

1. Only the finite state/finite action Stochastic Games with complete information were considered.
2. Only algorithms for solutions in the class of the so-called "stationary strategies" were considered.
3. The term "algorithm" was interpreted broadly enough to permit us to mention characterization results that may only ultimately lead to efficient algorithms.

While the decisions to include or exclude any given result were (of necessity?) made subjectively, we hope that the bibliography is comprehensive enough to enable the reader to at least begin tracing the contributions of most of the researchers active in the algorithmic aspects of Stochastic Games.

1 Preliminaries

The celebrated Minimax Theorem of von Neumann for matrix games asserts the following [37]: *Given a real $m \times n$ matrix $A = (a_{ij})$, there exists a pair of probability vectors* $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_m^*)$ and $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_n^*)$ such that for a unique constant n

$$\sum_i a_{ij} x_i^* \geqslant v \geqslant \sum_j a_{ij} y_j^* \quad \text{for all} \quad i,j \ . \tag{1}$$

The interpretation is that if action j were to be secretly chosen by player II and action i were to be secretly chosen by player I resulting in a payment of a_{ij} to player I by player II, then the strategy of choosing i with probability x_i^*, $i = 1,2,\dots,m$ guarantees an expected income of v to player I. Similarly, the strategy of choosing j with probability y_j^*, $j = 1,2,\dots,n$ guarantees player II an expected loss not exceeding v. Thus under repeated play of such a *matrix game* (with I and II using strategies $\mathbf{x}^*$ and $\mathbf{y}^*$ respectively) the expected reward to player I is v. Here $v := \text{val}\,[A]$ is called *the value* of the matrix game. The vectors $\mathbf{x}^*$ and $\mathbf{y}^*$ are called *optimal strategies*. The above game is called *zerosum* since what one player gets, the other player loses, that is, when $a_{ij} < 0$ it results in a payment of $-a_{ij}$ by player I to player II.

One can ask the following question: what happens if the players play not just one matrix game, but different matrix games, one at each stage with a movement among these games depending on which entry was selected at the previous stage of the game? Games with just such a conceptual structure are called *Stochastic Games* and were introduced by Shapley [49]. Let $A^1, A^2, \dots, A^N$ be real matrices known to the two players. By *state s* we mean the matrix game A^s. Players start in say, state s. They play the matrix game A^s. Immediately after, player I receives the payoff from player II and the game moves to A^k with probability $q(t \mid s, i_s, j_s)$ that depends on the choices i_s, j_s by players I and II in state s. At the next stage they play A^k and so on. The probability of transitions, known to both players, are assumed to be Markovian, in the sense that the movement among games is determined only by the immediate past and not by the entire history. The aim of player I is to maximize his gain. The aim of player II is to minimize his loss. Of course, a matrix game is a very special case of this game where $A^1 = A^2 = \dots = A^n$. Since the stochastic game never ends, the overall payoff needs to be specified. We shall emphasize two particular payoff criteria that are commonly considered in the literature. In the *discounted payoff*, with $0 \leq \beta < 1$ one takes as payoff

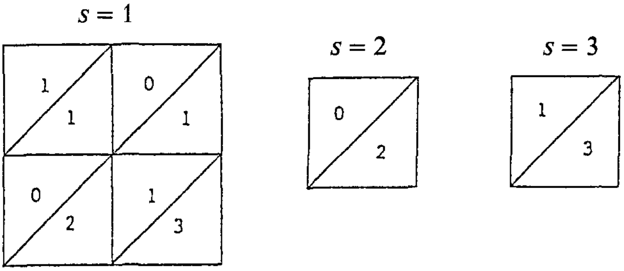
$$\sum_{n=1}^{\infty} \beta^{n-1} r(s_n, i_n, j_n) \tag{2}$$

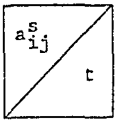
where $r(s_n, i_n, j_n) = a_{i_n j_n}^{(s_n)}$ = payoff on the n-th stage, the matrix game A^{s_n} is played and row i_n, and column j_n are chosen there. Under the above criterion, the current rewards are more important than the future prospects. The second payoff criterion is the so-called *limiting average payoff* (also called the *undiscounted payoff*) defined by

$$\lim_{T\to\infty} \inf \left[\frac{1}{T} \sum_{n=1}^{T} r(s_n, i_n, j_n) \right] . \tag{3}$$

Here one is concerned about long run average reward per play. We shall now introduce some further important notions through the following example.

Example 1: The "Big Match" (see [24], [5]):



The notation  should be interpreted as follows: if in state s the players I and II choose the i-th row and the j-th column of A^s respectively, then II pays I the amoung a_{ij}^{s} and at the next stage the players will be obliged to play the matrix game A^t. We assume, that the undiscounted payoff criterion is used, and that the game starts in state 1. The game remains in state 1 as long as player I chooses row 1. The first time player I chooses row 2, the game moves to state 2 or 3 depending on the column choice of player II. Once state 2 is reached the game stays in that state forever with a payoff of 0 all the time. Similarly, once the games reaches state 3, it stays in that state forever with a constant payoff of 1. Players in such a game may benefit from remembering opponents' past behavior. Here, player I has the option to "terminate" the game by choosing the second row. However, the consequences of such an action are so permanent that it seems advisable to carefully monitor player II's actions before player I decides to terminate the game.

The above suggest the following definition of a general strategy: each player observes at any time $t = n$, the history consisting of states and actions, $h_n := (s_1, i_1, j_1, s_2, i_2, j_2, \ldots, s_{n-1}, i_{n-1}, j_{n-1}, s_n)$. His strategy when playing A^{s_n} selects actions according to a probability distribution that depends on this complete history. Such strategies are called *behavior strategies.* For the class of games we are dealing with, Aumann [1] showed that these behavior strategies are adequate to guard against any strategy of the opponent. A much simpler class of strategies are the so-called *stationary strategies,* where the players play in a "memoryless" way as follows: for each matrix game A^s, the players select a probability distribution on the rows (or columns) of A^s, and every time A^s is reached the rows (or columns) are chosen according to that specific probability distribution. These stationary strategies are adequate for certain classes of stochastic games. Simpler still are strategies that select for each matrix game A^s a particular row (or column) to be played whenever state s is reached. These are called *pure stationary strategies,* and will be seen to be adequate for some very special classes of stochastic games.

2 Iterative Algorithms for Discounted Zerosum Games

In his fundamental paper, Shapley [49] showed that the β-discounted stochastic games[4] can be played optimally using stationary strategies. That is, for each starting state s (or equivalently matrix A^s to be played first) player I by using an optimal stationary strategy f^0 can guarantee the expected discounted payoff of $v(s)$, no matter what strategy the opponent adopts. Namely, the expected discounted reward $\phi_\beta(f^0, g)(s)$ starting at state s is at least $v(s)$ against any strategy g (stationary, or otherwise) of player II. Similarly for an optimal stationary g^0, $\phi_\beta(f, g^0)(s)$ is at most $v(s)$ against all strategies f of player I. Shapley's proof contained an algorithm to approximately compute the value and optimal stationary strategies. This method is outlined below.

Let (f, g) be any pair of stationary strategies. Thus $f = (\mathbf{f}(1), \mathbf{f}(2), \ldots, \mathbf{f}(N))$ where $\mathbf{f}(s)$ is itself a mixed strategy on the rows of A^s for each s. Similarly, $g = (\mathbf{g}(1), \mathbf{g}(2), \ldots, \mathbf{g}(N))$ where $\mathbf{g}(s)$ is a mixed strategy on the columns of A^s. We will denote by $r(s, i, j)$ the immediate reward $a_{ij}^{(s)}$ at state s when i,j are the choices of the two players. Let the expected current payoff be defined by

$$r(f,g)(s) := \sum_i \sum_j a_{ij}^{(s)} f_i(s) g_j(s) = \sum_i \sum_j r(s,i,j) f_i(s) g_j(s) \ . \tag{1}$$

[4] Actually, Shapley considered games with positive stop probabilities in every instance, however, their analysis is the same as those of discounted games, and it is the latter class that was studied subsequently.

Also let

$$q(t|s,f,g) := \sum_i \sum_j q(t|s,i,j) f_i(s) g_j(s) \ , \tag{2}$$

and $Q(f,g)$ be an $N \times N$ transition matrix with (s,t)-th entry $q(t|s,f,g)$. The matrix $Q(f,g)$ is the "law of motion" induced by the stationary strategies f, g. Consider the map $T_{f,g}: \mathbb{R}^N \to \mathbb{R}^N$ defined in vector notation by

$$T_{f,g}: \mathbf{u} \to r(f,g) + \beta Q(f,g) \mathbf{u} \ . \tag{3}$$

That is, $(T_{f,g}\mathbf{u})(s) = r(f,g)(s) + \beta \sum_t q(t|s,i,j) u(t)$, where the argument (s) denotes the s-th entry of the corresponding vector. Clearly, $T_{f,g}$ is a contraction map. Let $\mathbf{u}^*$ be the fixed point of $T_{f,g}$. By iterating the preceding equation we have $\mathbf{u}^* = T_{f,g}\mathbf{u}^* = \phi_\beta(f,g)$, *the vector of β-discounted expected payoffs under f, g*. For any $\mathbf{u} \in \mathbb{R}^N$ define the *Shapley matrix game* $A^s(\mathbf{u}) := [a_{ij}^{(s)} + \beta \sum_t q(t|s,i,j) u(t)]$ for every s. Similarly to the above, the map T defined by $(T\mathbf{u})(s) := \text{val}\,[A^s(\mathbf{u})]$ for every s is a contraction. Hence if $\mathbf{v}$ is the fixed point of T, and if $f^*(s)$, $g^*(s)$ are optimal for $A^s(\mathbf{v})$ for every s, then $\mathbf{v} = T\mathbf{v} = \phi_\beta(f^*, g^*)$, and it follows that

$$\phi_\beta(f^*, g)(s) \geq v(s) \geq \phi_\beta(f, g^*)(s) \quad \text{for all stationary} \quad f,g \ . \tag{4}$$

With f^* fixed, the problem is a "Markov Decision Process" (MDP), for short[5] and equation (4) implies that $\phi_\beta(f^*, g)(s) \geq v(s)$ for all strategies g of player II. Similarly, $\phi_\beta(f, g^*)(s) \leq v(s)$ for all strategies f of player I. Therefore f^* and g^* are optimal stationary strategies for players I and II respectively.

An attractive feature of the above procedure due to Shapey [49] is the built-in algorithm stated below.

Algorithm 1: [Shapley [49]].

Step 1: Start with any approximation for the true value $v(s)$ of the stochastic game, say $v^1(s)$, for every state s.

[5] A Markov Decision Process can be regarded as a stochastic game in which one player is a "dummy" with only one action to choose in every state. For these processes it is known (e.g., see Derman [10]) that there exist optimal pure stationary strategies in both the discounted and the undiscounted models.

Step 2: Define recursively, for each state s,

$$v^n(s) = \text{val}\,[A^s(\mathbf{v}^{n-1})] \;. \tag{5}$$

It can be easily shown that the above sequence of approximations converges to $v(s)$, the unique fixed point of the *Shapley equations* (one for each s)

$$v(s) = \text{val}\,[a_{ij}^{(s)} + \beta \sum_t q(t|s,i,j)\, v(t)] \;. \tag{6}$$

Note that the above equation is analogous to the "optimality equation" of dynamic programming.

Remark 2.1: While near-optimal stationary strategies can be derived from the above scheme when $\mathbf{v}^n$ is sufficiently close to $\mathbf{v}$, it should be noted that Shapley's algorithm does not utilize the information contained in the optimal strategies of $A^s(\mathbf{v}^n)$'s at each iteration.

The literature on Stochastic Games now contains a number of algorithms that attempt to improve on the preceding basic scheme of Shapley's. We outline some of the better known among these.

Algorithm 2: [Hoffman and Karp [26]].

Step 1: Set $\mathbf{v}^0(s) = 0$ for each state s and $\tau = 0$.

Step 2: Find an optimal strategy for player II in the matrix games $A^s(\mathbf{v}^\tau)$ for each state s. Let $g_{\tau+1}$ be one such optimal strategy.

Step 3: Solve the MDP problem $\mathbf{v}^{\tau+1} = \max_f \phi_\beta(f, g_{\tau+1})$.

Step 4: Put $\tau := \tau + 1$ and return to step 2.

It can be shown that $\mathbf{v}^\tau \to \mathbf{v}$, the value vector of the stochastic game, as $\tau \to \infty$. Note that this algorithm iterates in both the value space and the strategy space.

Pollatschek and Avi-Itzhak [42] proposed another algorithm that is based on utilizing both the approximate value vectors and the associated optimal strategies of the matrix games constructed at the intermediate steps.

Algorithm 3: [Pollatschek and Avi-Itzhak [42]].

Step 1: Select an arbitrary initial approximation $\mathbf{v}^0 = (v^0(1), \dots, v^0(N))$ to the value vector.

Step 2: At iteration τ, $\mathbf{v}^\tau$ is known. Solve the N matrix games $A^s(\mathbf{v}^\tau)$ for optimal strategies $\mathbf{f}^\tau(s)$, $\mathbf{g}^\tau(s)$ for players I and II.

Step 3: Set $\mathbf{f}^\tau = (\mathbf{f}^\tau(1), \dots, \mathbf{f}^\tau(N))$ and $\mathbf{g}^\tau = (\mathbf{g}^\tau(1), \dots, \mathbf{g}^\tau(N))$. Compute $\mathbf{v}^{\tau+1} = [I - \beta Q(f^\tau, g^\tau)]^{-1} \mathbf{r}(f^\tau, g^\tau)$.

Step 4: Set $\tau := \tau + 1$ and return to step 2.

Remark 2.2: Pollatschek and Avi-Itzhak [42] proved that the above algorithm converges under the rather restrictive condition $\max_s \{ \sum_t [\max_{i,j} q(t|s,i,j) - \min_{i,j} q(t|s,i,j)] \} \le \frac{1-\beta}{\beta}$, however, they used it successfully even on problems that did not satisfy it.

Despite the preceding remark, the algorithm of Pollatschek and Avi-Itzhak has special importance because of its relationship with the classical Newton-Raphson procedure. This relationship is sketched below.

Define the operator $(T\mathbf{v})(s) := \operatorname{val} A^s(\mathbf{v})$, and note that it is Lipschitz continuous. Let $\phi_s(\mathbf{v}) = (T\mathbf{v})(s)$. In case $\phi_s(\mathbf{v})$ has a partial derivative at $\mathbf{v}$, we can show that $\frac{\partial \phi_s(v)}{\partial v(t)} = \beta q(t|s,f,g)$ where $\mathbf{f}(s)$ and $\mathbf{g}(s)$ are optimal strategies for the matrix games $A^s(\mathbf{v})$. The problem of finding the value vector $\mathbf{v}^*$ of the stochastic game is the same as solving the nonlinear equations.

$$\psi_s(\mathbf{v}) = \phi_s(\mathbf{v}) - \mathbf{v}(s) = 0 \ , \quad \text{for all} \quad s \ . \tag{7}$$

If $\frac{\partial^2 \phi_s}{\partial v(t)^2}$ exists and is bounded in a neighborhood of $\mathbf{v}^*$, then we can use Newton-Raphson procedure for solving the system $\psi_s(\mathbf{v}) = 0$.

Remark 2.3: The precise conditions under which the above interesting algorithm converges are still unknown. Some attempts have been made to prove the convergence of the algorithm under less restrictive assumptions; for instance, by Rao, Chandrasekaran and Nair [44]. A counter-example by Van der Wal [64] shows that these authors have a gap in their proof. In a recent paper, Filar and Tolwinski [22] have shown that a variant of the above algorithm always converges. Their "Modified Newton's Method" uses a variable size Newton's iteration to ensure descent at every major iteration, which together with the uniqueness of the value vector guarantees convergence.

Another algorithm that deserves a mention is due to Van der Wal [64] and can be viewed as an extension of Algorithm 2 by Hoffman and Karp [26].

Algorithm 4: [Van der Wal [64]].

Step 1: Choose arbitrary $\mathbf{v}^0 = (v^0(1), \ldots, v^0(N))$, and an integer $m > 0$.

Step 2: At iteration τ, $\mathbf{v}^\tau$ is known. Solve for an optimal strategy $\mathbf{g}^\tau(s)$ of player II for the matrix game $A^s(\mathbf{v}^\tau)$, for every s. This yields a stationary strategy $g^\tau = (\mathbf{g}^\tau(1), \ldots, \mathbf{g}^\tau(N))$ in the stochastic game.

Step 3: Use a method of successive approximations to the m-th iterate to estimate the value vector of the MDP problem: $\max_f \phi_\beta(f, g^\tau)$. Let $\mathbf{v}^m$ be such an estimate.

Step 4: Set $\tau := \tau + 1$, $\mathbf{v}^{\tau+1} = \mathbf{v}^m$ and return to step 2.

Remark 2.4: Van der Wal [64] showed that in the above algorithm $\mathbf{v}^\tau$ converges to $\mathbf{v}$, the value vector of the stochastic game, and that for τ sufficiently large the optimal strategies of matrix games $A^s(\mathbf{v}^\tau)$ yield near-optimal stationary strategies of the stochastic game. It should also be noted that Algorithm 2 can be regarded as the case $m = \infty$ of Algorithm 4. Furthermore, for $m = 1$ Algorithm 4 reduces to Algorithm 1.

The natural question which now arises is: Which of the Algorithms 1 – 4 is "best" in practice for deriving solutions to Stochastic Games?

An interesting empirical study to answer the above question was recently carried out by Breton [6]. Some of its findings are reported in Breton et al. [7]. In these works, the performance of Algorithms 1 – 4 was tested on randomly generated problems with up to 15 states and 15 actions per player per state. The main conclusion based on this empirical study was that the algorithm of Pollatschek and Avi-Itzhak, whenever it converged, was consistently much faster than the others.

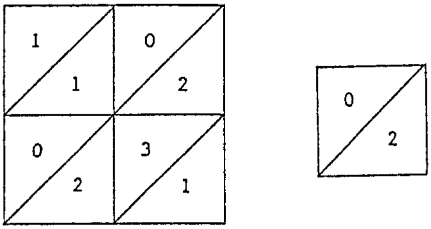
The final algorithm that we mention in this section, due to Tijs and Vrieze [61], extends the attractive notions of "fictitious play", developed by Brown [8] and Robinson [45] to the discounted zerosum stochastic games. Even for matrix games this algorithm is known to be inefficient, and it is known not to be extendable to bimatrix games [50].

3 Finite Algorithms for Structured, Discounted, 0-Sum Games

In general, solutions to stochastic games lack an important algebraic property, which suggests that efficiently solving these games is essentially more difficult than solving matrix games. This is illustrated by the following example.

Example 2: [Parthasarathy and Raghavan [39]].

Let there be two states: 1 and 2 with payoffs and transitions defined by



Here in state 1 if row 1 and column 2 are selected, the game moves to state 2, and so on. State 2 is assumed to be absorbing. The discounted stochastic game (with discount factor $\beta = 1/2$) has value $v(1) = v(1) = \frac{1}{3}[-4+2\sqrt{13}]$ with the unique optimal stationary strategy $\left(\frac{2\sqrt{13}-4}{\sqrt{13}+1}, \frac{5-\sqrt{13}}{\sqrt{13}+1}\right)$ for both players in state 1.

Remark 3.1: The above example shows that while all the data defining the stochastic game namely, the rewards, the discount factor, and the transition probabilities are rational, the value vector has irrational entries. Thus the data and the solution are not in the same ordered Archimedean field. From now on, we shall refer to this phenomenon as the lack of *ordered field property.*[6] It essentially eliminates the possibility of solving stochastic games by performing only finitely many arithmetic operations. Note that since linear programs solve a general matrix game, and since an optimal basis of that program can be found via finitely many pivots of the simplex method, matrix games do possess the ordered field property.

One line of research that has evolved from the preceding considerations is focussed on identifying those natural classes of stochastic games for which the ordered field property holds, and on developing algorithms for their solution. Five such classes have now been treated in the literature:

[6] This difficulty was anticipated in the original paper of Shapley [49].

(i) Stochastic Games with Perfect Information; (Gillette [24]).
(ii) The Single-Controller Stochastic Games; (Stern [55]).
(iii) The Switching-Controller Stochastic Games; (Filar [14]).
(iv) The Separable Reward – State Independent Transition Stochastic Games, or SER-SIT for short; (Sobel [53]).
(v) The Additive Reward – Additive Transition Stochastic Games, or AR-AT for short; (Raghavan et al. [43]).

The above classes will be defined precisely in the sequel; however, for each of these "structured" classes (i) – (v) the following result is established in [39], [14], [41], [43], respectively.

Theorem 3.1: (*The Ordered Field Property*): Let the game belong to one of the classes (i) – (v), and let all the data (rewards, transition probabilities, and the discount factor) lie in the same ordered Archimedean field. Then the value vector, and at least one set of optimal stationary strategies have all entries in the same field.

Open Problem: Characterize the class of stochastic games possessing the ordered field property. Note that the above theorem gives only sufficient conditions.

(i) *Stochastic Games with Perfect Information*

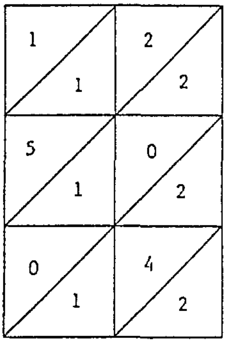
These are stochastic games in which in every state the action space of one of the players is a singleton. Even though this class is known to possess pure stationary optimal strategies for the two players it has not been studied from an algorithmic point of view.

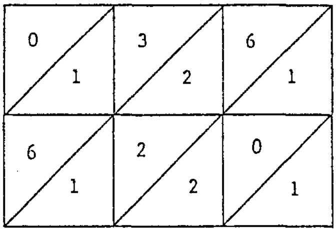
Open Problem: Find an efficient finite step algorithm for this class of games.

(ii) *The Single-Controller Stochastic Games*

In the case where player II is the "single-controller" this means that $q(s'|s,i,j) \equiv q(s'|s,j)$ for all i,j,s,s'. In this simple class we shall illustrate the ideas behind Theorem 3.1, and a linear programming algorithm with the help of the following easy example.

Example 3:





Let us write down the Shapley equations (6) of Section 2 for this example. We have

$$v(1) = \text{val}\begin{bmatrix} 1+\beta v(1) & 2+\beta v(2) \\ 5+\beta v(1) & 0+\beta v(2) \\ 0+\beta v(1) & 4+\beta v(2) \end{bmatrix} ,$$

$$v(2) = \text{val}\begin{bmatrix} 0+\beta v(1) & 3+\beta v(2) & 6+\beta v(1) \\ 6+\beta v(1) & 2+\beta v(2) & 0+\beta v(1) \end{bmatrix}$$

Let (x_1,x_2,x_3) be optimal for player I in the first matrix game above. Also let (ξ_1,ξ_2) be optimal for player I in the second matrix game above. These vectors must satisfy the following system of constraints

$$\begin{aligned}
&x_1 + 5x_2 + 0x_3 && +\beta v(1)[x_1+x_2+x_3] \geq v(1) \\
&2x_1 + 0x_2 + 4x_3 && +\beta v(2)[x_1+x_2+x_3] \geq v(1) \\
&0\xi_1 + 6\xi_2 && +\beta v(1)[\xi_1+\xi_2] \quad \geq v(2) \\
&3\xi_1 + 2\xi_2 && +\beta v(2)[\xi_1+\xi_2] \quad \geq v(2) \\
&6\xi_1 + 0\xi_2 && +\beta v(1)[\xi_1+\xi_2] \quad \geq v(2) \\
&x_1,x_2,x_3,\xi_1,\xi_2 \geq 0 && x_1+x_2+x_3 = 1 \ , \quad \xi_1+\xi_2 = 1 \ .
\end{aligned}$$

Since $x_1+x_2+x_3 = 1$ and $\xi_1+\xi_2 = 1$ we have a system of linear constraints in x_1, $x_2, x_3, v(1), v(2), \xi_1, \xi_2$. It now seems natural, and is not hard to check, that this stochastic game can be solved by linear programming using the objective function

$$\text{maximize } (v(1)+v(2))$$

subject to the constraints above. For $\beta=\frac{1}{2}$, the solution is $v(1)=4.673$, $v(2)=5.084$, $(x_1,x_2,x_3)=(0, 0.467, 0.533)$, $(\xi_1,\xi_2)=0.542, 0.458)$. Using this we can also determine an optimal stationary strategy for player II (from the above matrix) as $(y_1,y_2)=(0.446, 0.554)$ in state 1 and $(\eta_1,\eta_2,\eta_3)=(0.856, 0.144, 0)$ in state 2.

Algorithm 5: [For player II-controlled game].

Step 1: Solve the linear program

$$\text{maximize } \sum_s v(s) \; .$$

Subject to

$$\begin{array}{ll} \mathbf{f}(s)A^s(\mathbf{v})\mathbf{e}_j \geq v(s) & \text{for all } \; s,j \\ \mathbf{f}(s)\cdot\mathbf{1}=1 & \text{for all } \; s \\ \mathbf{f}(s)\geq\mathbf{0} & \text{for all } \; s \; , \end{array}$$

where $\mathbf{e}_j$ is the jth-vector of the unit basis of appropriate dimension (for each s), and $\mathbf{1}$ is the vector with unity in every component, and is of appropriate dimension. Let $\mathbf{f}^0=(\mathbf{f}^0(1),\ldots,\mathbf{f}^0(N))$ and $\mathbf{v}^0$ be an optimal solution of this program.

Step 2: Compute an optimal strategy $\mathbf{g}^0(s)$ for player II in matrix game $A^s(\mathbf{v}^0)$ for each s, and set $g^0=(\mathbf{g}^0(1),\ldots,\mathbf{g}^0(N))$.

It was shown in [39] that the vector $\mathbf{v}^0$ is the value vector of the single controller game, and (f^0,g^0) is a pair of optimal stationary strategies.

Remark 3.2: Note that unlike the Algorithms 1 – 5, the above algorithm terminates in finitely many steps provided that a finite algorithm (e.g., simplex method) is used to implement steps 1 and 2. Furthermore, it is not hard to check that we could have used the dual of the linear program in step 1 to obtain an optimal stationary strategy g^0 for player II (e.g., see [39] and [59]). Of course, it is the single-controller assumption that ensures that the constraints of the program in step 1 are indeed linear.

(iii) *The Switching-Controller Stochastic Games*

In this case we suppose that the set of states is the union of two disjoint nonempty sets S_1 and S_2 such that for all t, i, j

$$q(t\,|\,s,i,j) = \begin{cases} q(t\,|\,s,i) & \text{if} \quad s \in S_1 \\ q(t\,|\,s,j) & \text{if} \quad s \in S_2 \end{cases}.$$

While the above transition structure is a natural generalization of the single-controller game, from the algorithmic point of view this class of games appears to be more difficult. The following two algorithms tackle the problem by constructing an appropriate finite sequence of single-controller games; an approach that was outlined in Filar and Raghavan [17].

We shall need the following notation: Let Γ denote the given switching-controller game. Suppose that player I fixed his strategy $\mathbf{f}(s)$ for every $s \in S_1$, then we can define a player II-controlled game $\hat{\Gamma}(f)$ with state and action spaces the same as in Γ, but whose rewards and transitions are modified in the following manner. The data in the states of S_2 remain unchanged, but for each $s \in S_1$ all of player I's actions will be identical to action 1 and will result in the rewards $\hat{r}(s,1,j) = \sum_i r(s,i,j) f_i(s)$, and in the transitions of $\hat{q}(t\,|\,s,1,J) \equiv \sum_i q(t\,|\,s,i,j) f_i(s) = q(t\,|\,s,f,j)$. We illustrate this reduction with the following simple example

Example 4: Let the data of the switching-controller game be

2 2	0 2
0 1	3 1

3 1	5 2
6 1	4 2

Clearly, $S_1 = \{1\}$ and $S_2 = \{2\}$ in this example. Suppose now that player I fixes his strategy in state 1 to $\mathbf{f}(1) = (0.2,\ 0.8)$, then the single controller game $\hat{\Gamma}(f)$ is

.4 (.8,.2)	2.4 (.8,.2)
.4 (.8,.2)	2.4 (.8,.2)

3 1	5 2
6 1	4 2

where the transitions (0.8, 0.2) in state 1 denote the probability of 0.8 of remaining in that state, and of 0.2 of moving to state 2, irrespective of the actions the two players select in state 1. Of course, player I is effectively a "dummy" in state 1 of $\hat{\Gamma}(f)$.

Algorithm 6: [Vrieze [59]]

Step 1: Set $\tau := 0$, choose an arbitrary $\mathbf{v}^0 = (v^0(1), \ldots, v^0(N))$, and find an extreme optimal strategy $\mathbf{f}^0(s)$ for player I in the matrix game $A^s(\mathbf{v}^0)$ for each $s \in S_1$.

Step 2: Set $\tau := \tau + 1$. Solve the player II-controlled game $\hat{\Gamma}(f^{\tau-1})$, and denote its value by $\mathbf{v}^\tau$.

Step 3: If $v^\tau(s) = \operatorname{val} A^s(\mathbf{v}^\tau)$ for each $s \in S$, then stop. Otherwise, find an extreme optimal strategy $\mathbf{f}^\tau(s)$ for player I in the matrix game $A^s(\mathbf{v}^\tau)$ for each $s \in S_1$ and return to step 2.

Theorem 3.2: [Vrieze [59]]. The estimates $\mathbf{v}^\tau$ are componentwise monotone decreasing in τ. Algorithm 6 terminates in finitely many iterations, and if it stops at the τ-th iteration, then $\mathbf{v}^\tau$ is the value vector of the switching-controller stochastic game.

The preceding algorithm can be perceived as being somewhat "one-sided" since at each iteration the induced single-controller game is always a player II-controlled game. Mohan and Raghavan [35] proposed an algorithm that alternates between a player I and a player II-controlled game in an attempt to exploit the dual simplex method.

More precisely, let Γ be the underlying switching-controller stochastic game, where, without loss of generality $S_1 = \{1, 2, \ldots, k\}$ are the states controlled by player I and $S_2 = \{k+1, k+2, \ldots, N\}$ are the states controlled by player II. Let P_1 be the polyhedron defined below by the linear constraints in the variables f, g, $\mathbf{v}$ a k-vector, and $\mathbf{w}$ an N-k-vector

$$\mathbf{f}(s)A^s(\mathbf{w})\mathbf{e}_j \geq v(s) \quad \text{for all } j \text{ and } s \in S_1 \ ,$$

$$\mathbf{f}(s)\cdot\mathbf{1} = 1 \quad \text{for all } s \in S_1$$

$$\mathbf{f}(s) \geq \mathbf{0} \quad \text{for all } s \in S_1 \ .$$

Similarly, let P_2 be the polyhedron defined by the linear constraints

$$
\begin{aligned}
&[\mathbf{e}_i]^T A^s(\mathbf{v})\mathbf{g}(s) \leq w(s) && \text{for all} \quad i \text{ and } s \in S_2 \ , \\
&\mathbf{g}(s)\cdot\mathbf{1} = 1 && \text{for all} \quad s \in S_2 \ , \\
&\mathbf{g}(s) \geq \mathbf{0} && \text{for all} \quad s \in S_2 \ .
\end{aligned}
$$

Theorem 3.3: [Mohan and Rahavan [35]]. The switching-control stochastic game has value vector $(\mathbf{v}^*, \mathbf{w}^*)$ and optimal stationary strategies $\mathbf{f}^*$, $\mathbf{g}^*$ which form an extreme point of the polyhedron $P_1 \cap P_2$. Further $\mathbf{v}^*$ minimizes **(1.v)** on the polyhedron P_1 when the $\mathbf{w}$ coordinates are restricted to $\mathbf{w} = \mathbf{w}^*$, and $\mathbf{w}^*$ maximizes **(1.w)** on the polyhedron P_2 when the $\mathbf{v}$ coordinates are restricted to $\mathbf{v} = \mathbf{v}^*$.

Remark 3.3: Based on this theorem an algorithm is proposed in [35] which oscillates between the two linear programs mentioned earlier. Even though this procedure converges to the value vector, it need not to terminate in finitely many steps since basis vectors vary continuously. On the other hand, the worst case performance of the Algorithm 6 could conceivably grow exponentially fast with the size of the problem.

Open Problem: Find an efficient step algorithm for the switching-controller stochastic games. In particular, does there exist a polynomial algorithm? Note that the latter is the case for the single-controller stochastic games.

(iv) *The SER-SIT Stochastic Games*

In this case we assume that the *rewards are separable*, namely $r(s,i,j) = c(s) + \varrho(i,j)$ for all s,i,j, and the transitions are state independent, that is $q(t|s,i,j) \equiv q(t|i,j)$ for all s,i,j.

Note that the above is meaningful only if all the matrices A^s have the same dimensions. Now, with such a stochastic game associate the matrix game $V(\mathbf{c}) = [\varrho(i,j) + \beta \sum_t q(t|i,j)c(t)]$ of the same dimension as A^s. The SER-SIT games can be solved in "closed form" by the following simple algorithm.

Algorithm 7: [Sobel [53], and Parthasarathy et al. [41]].

Step 1: Solve the matrix game $V(\mathbf{c})$ and denote its value and a pair of optimal strategies by v and $(\mathbf{x}^0, \mathbf{y}^0)$ respectively.

Step 2: Let $\mathbf{1} = (1 \dots 1) \in \mathbb{R}^N$, and (f^0, g^0) be a pair of stationary strategies in the stochastic game such that $\mathbf{f}^0(s) \equiv \mathbf{x}^0$ and $\mathbf{g}^0(s) \equiv \mathbf{y}^0$ for all s. Also set $\mathbf{v} := \mathbf{c} + (1-\beta)^{-1} v \mathbf{1}$, and stop. The vector $\mathbf{v}$ is the value vector and (f^0, g^0) is a pair of optimal stationary strategies of the SER-SIT game.

An example in Parthasarathy et al. [41] shows that there are games possessing the state independent property (SIT) which nonetheless lack ordered field property.

(v) *The AR-AT Stochastic Games*

In this case we assume that *rewards are additive*, namely $r(s,i,j) \equiv r_1(s,i) + r_2(s,j)$ for all s, i, j, and that the *transitions are additive*, that is, $q(t \mid s,i,j) \equiv q_1(t \mid s,i) + q_2(t \mid s,j)$ for all t, s, i, j. Note that the above transition property reduces to the switching-controller property if in each state exactly one of $q_1(\cdot \mid \cdot, i)$ or $q_2(\cdot \mid \cdot, j)$ is identically zero. However, an example in Raghavan et al. [43] shows that games with the additive transition property alone lack the ordered field property that the AR-AT game possess. A large class of games with the additive transition structure also possesses the following intuitive interpretation.

Suppose we consider two stochastic games Γ_1 and Γ_2 where in Γ_1 player I controls the transitions and in Γ_2 player II controls the transitions. Let both games have identical data for the transition probabilities. Now a referee selects, for each state s, a number $0 \le \lambda(s) \le 1$. A new game Γ is constructed as follows: in any state s, once the immediate reward is paid, the referee allows the law of motion dictated by Γ_1 with probability $\lambda(s)$ and allows the law of motion dictated by Γ_2 with the probability $(1-\lambda(s))$. This induces the AT-property in the resulting game. For the AR-AT games we have the following finite algorithm.

Algorithm 8: [Raghavan et al. [43]].

Step 1: Set $\tau := 0$, $M := \min_{s,i,j} r(s,i,j)$ and $\mathbf{v}^0 = (M/(1-\beta))\mathbf{1}$ where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$.

Step 2: Determine for player I a pure stationary strategy $f^\tau = (\mathbf{f}^\tau(1), \dots, \mathbf{f}^\tau(N))$ such that $\mathbf{f}^\tau(s)$ is optimal for I in the matrix game $A^s(\mathbf{v}^\tau)$ for each s.

Step 3: Calculate $\mathbf{v}^{\tau+1} = \min_g \phi_\beta(f^\tau, g)$, the value vector of the MDP obtained by fixing player I's strategy at f^τ.

Step 4: If $\mathbf{v}^{\tau+1} \neq \mathbf{v}^\tau$ set $\tau := \tau + 1$ and return to step 2; otherwise stop.

Remark 3.4: It is shown in [43] that the above algorithm is finite, and that on termination $\mathbf{v}^{\tau}$ is the value vector of the stochastic game. Note that step 2 can be easily implemented because the AR-AT property ensures that $A^{s}(\mathbf{v}^{\tau})$ always decomposes into the sum of two matrices, one of which has identical rows while the other one has identical columns.

Remark 3.5: In a recent doctoral thesis Sinha [51] studied the ordered field property in "mixtures" of the above five structured classes. For instance, a mixture of the AR-AT and the switching-controller classes is a stochastic game whose data satisfy the AR-AT conditions in some states, and the switching-controller conditions in all of the remaining states.

4 Undiscounted Zerosum Games; The Issues

The undiscounted, or limiting average payoff stochastic games were introduced in 1957 by Gillette [24] who studied two special classes: games with perfect information that were already mentioned in Section 3, and *irreducible games* characterized by the property that $Q(f,g)$ is an irreducible matrix for every pair of stationary strategies. Gillette's proofs that the above classes possess stationary optimal strategies were later completed by Ligget and Lippmann [31]. One of the most significant contributions of Gillette's paper was the demonstration that in undiscounted games the players need not possess optimal stationary strategies, implying thereby that these games were inherently more complex than discounted stochastic games. It was Example 1 that exhibited the above complexity. This example under the name "Big Match" was later elegantly solved by Blackwell and Ferguson [5]; a solution that turned out to have a significant impact on the development of the underlying theory of undiscounted games. We shall now briefly return to this example.

Example 5: The Big Match Revisited (see Example 1). Assume the undiscounted payoff criterion. It is clear that this game is interesting only in the first state since trivially $v(2) = 0$ and $v(3) = 1$. Also, it seems natural that $v(1)$ "ought to" equal $\frac{1}{2}$. However, a moment's reflection shows that

$$\min_{g} \phi(f,g)(1) = 0 \tag{1}$$

for every stationary f, where $\phi(f,g)(s)$ denotes the undiscounted payoff of the game starting in state s, under the strategy pair (f,g). Validity of (1) is clear once we note that if $\mathbf{f}(1)=(\alpha,1-\alpha)$ with $\alpha\in(0,1)$, then $\phi(f,\hat{g})(1)=0$ if $\hat{\mathbf{g}}(1)=(1,0)$, and that for $\alpha=1$, $\phi(f,\tilde{g})(1)=0$ if $\tilde{\mathbf{g}}(1)=(0,1)$. It is even easier to observe that if g^0 is a stationary strategy such that $\mathbf{g}^0(1)=(\frac{1}{2},\frac{1}{2})$, then

$$\sup_f \phi(f,g^0)(1)\leq\tfrac{1}{2}\ . \tag{2}$$

Of course, (2) implies that $\inf_g \sup_f \phi(f,g)\leq\frac{1}{2}$. What is not at all obvious, however, is the fact that by using behavior strategies player I can attain payoffs from state 1 that approach $\frac{1}{2}$ arbitrarily closely. One sequence of strategies that demonstrates the latter was ingeniously constructed by Blackwell and Ferguson as follows: given an n-stage history h_n of the game, player I computes a "confidence coefficient" (provided the game is still in state 1); $k_n=(\#$ of times II selected first column of A^1 so far $-$ ($\#$ of times II selected column of A^1 so far). It is tempting to conjecture that, perhaps, negative values of k_n should encourage player I to risk choosing the second row of A^1. It is shown in [5] that if M is a fixed positive integer, and if a behavior strategy f_M for player I based on h_n selects the second row of A^1 with probability $\dfrac{1}{(k_n+M+1)^2}$, then

$$\inf_g \phi(f_M,g)(1)\geq\frac{M}{2(M+1)}\ . \tag{3}$$

Since the right side of (3) tends to $\frac{1}{2}$ as $M\to\infty$, it follows that

$$\sup_f \inf_g \phi(f,g)(1)\geq\tfrac{1}{2}\ , \tag{4}$$

which now implies that, as anticipated, the value from state 1 exists and $v(1)=\frac{1}{2}$.

The papers of Gillette [24] and Blackwell and Ferguson [5] served to clearly identify two important research questions namely:

(A) Do undiscounted stochastic games possess the value vector?
and
(B) What characterizes the class of undiscounted stochastic games solvable in optimal stationary strategies?

The research question (A) remained open for over twenty years (if we date it back to Gillette), and was eventually answered in the affirmitive in Mertens and

Neyman [32][7]. These results were based on an ingenious analysis by Bewley and Kohlberg [3] of the Shapley equation (see (6), Section 2) viewed as an equation defined over the field of Puiseux series. In the context of algorithms we are concerned with the status of the research question (B), and it is that issue which will be discussed next.

The trend set by Gillette in 1957 was followed by a number of authors who supplied sufficient conditions for the existence of stationary strategies; for instance, see Hoffmann and Karp [26], Stern [55], Parthasarathy and Raghavan [39], Bewley and Kohlberg [4], Filar [14], Parthasarathy et al. [41] and Raghavan et al. [43]. Among these, the conditions of Bewley and Kohlberg [4] should be particularly noted as they followed from a novel algebraic analysis of the "limit discount equation" for the discounted games. The latter is a version of the Shapley equation (6) of Section 2. In addition to the rather general sufficient conditions, Bewley and Kohlberg [4] supplied a separate set of necessary conditions, but stopped short of giving a single set of necessary and sufficient conditions for the existence of optimal stationary strategies. Indeed, these authors stated this to be one of the main open problems in the theory of undiscounted stochastic games. Two sets of necessary and sufficient conditions were obtained by Vrieze [59], and subsequently by Filar and Schulz [19]. These conditions were recently superseded by a nonlinear programming characterization (see Filar et al. [21]) that will be discussed in more detail in Section 7.

Remark 4.1: It must be mentioned that the undiscounted payoff criterion (3) of Section 1 is only one of a number of related "long-run average" criteria which may be used when discounting is not appropriate. However, Bewley and Kohlberg [4] considered as many as six of such alternative criteria, and demonstrated that all six are equivalent in games that possess optimal stationary strategies. Since the latter class is the largest that this survey is concerned with, we shall not address this issue again.

5 Algorithms for Zerosum Undiscounted Games

In view of the fact that in general undiscounted games need not possess optimal stationary strategies, the algorithmic development for computing such strategies centered around identifying "natural" classes that possess optimal stationary strategies and on supplying algorithms for their computation. These classes of games can be roughly divided into two groups:

[7] Monash [36] communicated a similar result slightly earlier.

(C) Those that make assumptions on the structure of the game data (i.e., transitions and/or rewards),
and
(D) Those that make assumptions on the ergodic properties of the game.

Seemingly without an exception, the previously studied games that fall into the category (C) also belong to one of the five structured classes (i) – (v) discussed in Section 3, and this line of research can be regarded as descending from the paper of Parthasarathy and Raghavan [39] since for all of these classes the following analogue of Theorem 3.1 has been established (see [39], [14], [41] and [43]).

Theorem 5.1: Consider an undiscounted stochastic game with data that lie in an ordered Archimedean field, and belonging to any one of the five classes (i) – (v) of Section 3. Such a game possesses the ordered field property. Namely, there exists a pair of stationary optimal strategies that lies in the same field.

The above theorem has fulfilled part of its promise, in the sense that there are now finite algorithms for the computation of optimal stationary strategies for each one of the classes (i) – (v). However, in the case of the switching-controller and the AR-AT classes (iii) and (v) these algorithms could turn out to be computationally prohibitive. Consequently in the discussion below we shall emphasize the rather elegant solution algorithms for the classes (ii) and (iv).

(i) *Perfect Information Undiscounted Games.*

As mentioned in Section 3 despite its obvious appeal, this class has not been studied from an algorithmic point of view (of course, these games can be regarded as a special case of classes (iii) and (v) below).

(ii) *The Single-Controller Undiscounted Games.*

The first finite algorithm for this class of games was given in Filar [13], and the theory underlying this method can be found in Filar and Raghavan [18]. A significant portion of this analyses follows from the result given below.

Theorem 5.2: Let (f_σ, g_γ) denote a pair of pure stationary strategies in an undiscounted single-controller game. Let $H(s)$ be a matrix game whose (σ, γ)-th entry is $\phi(f_\sigma, g_\gamma)(s)$. Then, for every state s, $\operatorname{val}[H(s)] = v(s)$, the value of the stochastic game. Moreover, there exists a common optimal strategy for player I of the matrix games $H(s)$; $s = 1, 2, \ldots, N$. This common optimal strategy can be transformed to an optimal stationary strategy for player I in the stochastic game.

While the above theorem demonstrates that a solution can be obtained from a finite sequence of rather complex linear programs, it was Vrieze [58] and independently Hordijk and Kallenberg [27] who discovered the elegant "one-step" linear programming formulation presented below.

Algorithm 9: [For the Player II-Controlled Undiscounted Game].

Step 1: Let $(\mathbf{v}^0, \mathbf{u}^0, f^0)$ be any basic optimal solution to the linear program

$$\max \sum_s v(s) \ .$$

Subject to

$$\text{(a)}\ v(s) \geq \sum_t q(t|s,j) v(t) \qquad \text{for all} \quad s,j$$

$$\text{(b)}\ v(s) + u(s) \geq \sum_i r(s,i,j) f_i(s) + \sum_t q(t)(s,j) u(t) \qquad \text{for all} \quad s,j$$

$$\text{(c)}\ \sum_i f_i(s) = 1 \qquad \text{for all} \quad s$$

$$\text{(d)}\ f_i(s) \geq 0 \qquad \text{for all} \quad s,i \ .$$

Step 2: From the optimal basis obtained in Step 1, or otherwise, compute an optimal dual solution $(\mathbf{y}^0, \mathbf{x}^0, \mathbf{w}^0)$, where $\mathbf{y}^0, \mathbf{x}^0$, and $\mathbf{w}^0$ are the dual variable vectors corresponding to the constraints (a), (b) and (c) respectively. Let $x_s^0 := \sum_j x_j^0 x_j^0(s)$ and $d_s := \sum_j (x_j(s) + y_j(s))$ for every s, and set $S_0 := \{s | x_s^0 = 0\}$. Now we define a stationary strategy g^0 for player II according to

$$g_j^0(s) = \begin{cases} x_j^0(s)/x_s^0 & \text{for all} \quad j \ , \quad \text{if} \quad s \notin S_0 \\ y_j^0(s) + x_j^0(s)/d_s^0 & \text{for all} \quad j \ , \quad \text{if} \quad s \in S_0 \end{cases} ,$$

and stop; (f^0, g^0) is an optimal pair of stationary strategies and $\mathbf{v}^0$ is the value vector.

We shall now illustrate the preceding method with the following simple example.

Example 6: The data of this player II-controlled undiscounted game are the same as in Example 3 of Section 3. The linear program of Step 1 above is

$$\text{maximize } (v(1) + v(2)) \ .$$

Subject to

(a) $v(1)-v(1)\leq 0$

$v(1)-v(2)\leq 0$

$v(2)-v(1)\leq 0$

$v(2)-v(2)\leq 0$

(b) $v(1)+u(1)-1f_1(1)-5f_2(1)-0f_3(1)-u(1)\leq 0$

$v(1)+u(1)-2f_1(1)-0f_2(1)-4f_3(1)-u(2)\leq 0$

$v(2)+u(2)-0f_1(2)-6f_2(2)-u(1)\leq 0$

$v(2)+u(2)-3f_1(2)-2f_2(2)-u(2)\leq 0$

$v(2)+u(2)-6f_1(2)-0f_2(2)-u(1)\leq 0$

(c) $f_1(1)+f_2(1)+f_3(1)=1$

$f_1(2)+f_2(2)=1$

(d) $f_1(1), f_2(1), f_3(1), f_1(2), f_2(2)\geq 0$.

It can now be easily checked that $\mathbf{v}^0=(\frac{5}{2},\frac{5}{2})$ is the value vector, and the strategies f^0 and g^0 with $\mathbf{f}^0(1)=(0,\frac{1}{2},\frac{1}{2})$, $\mathbf{f}^0(2)=(\frac{1}{2},\frac{1}{2})$ and $\mathbf{g}^0(1)=(\frac{4}{9},\frac{5}{9})$, $\mathbf{g}^0(2)=(\frac{1}{2},0,\frac{1}{2})$ form a pair of stationary optimal strategies for players I and II.

(iii) *The Switching-Controller Undiscounted Games.*

The first finite algorithm for this class of games was outlined in Filar and Raghavan [17]. This algorithm was subsequently improved and streamlined in Vrieze et al. [62]. The proof of convergence of the latter version also contained a proof of the existence of stationary optimal strategies, and of the ordered field property. However, given the considerable complexity of this algorithm, and of the related notation we shall not present it here. Instead, we refer the reader to [62], and mention only that this method is conceptually related to Algorithm 6 for the switching-controller discounted games.

(iv) *The SER-SIT Undiscounted Games.*

Just as in the discounted SER-SIT games it turns out that a closed-form "one-step" finite algorithm is possible for the undiscounted games with this structure as demonstrated in Parthasarathy et al. [41]. Below we shall use the same notation as in Section 3.

Algorithm 10: [for the SER-SIT Undiscounted Games].

Step 1: Set up, and solve the single matrix game $U(\mathbf{c}) = [p(i,j) + \sum_t g(t|i,j)c(t)]$ of the same dimension as A^s. Let $u = \text{val}\,[U(\mathbf{c})]$, and $\mathbf{x}^0$, $\mathbf{y}^0$ denote a pair of optimal strategies for players I and II.

Step 2: Set $\mathbf{v}^0 := u\mathbf{1}$, where $\mathbf{1} = (1,\ldots,1) \in \mathbb{R}^N$, $\mathbf{f}^0(s) := \mathbf{x}^0$ and $\mathbf{g}^0(s) := \mathbf{y}^0$ for all s. Now, $\mathbf{v}^0$ is the value vector of the undiscounted game, and $f^0 = (\mathbf{f}^0(1),\ldots,\mathbf{f}^0(N))$ and $g^0 = (\mathbf{g}^0(1),\ldots,\mathbf{g}^0(N))$ are a pair of optimal stationary strategies for players I and II.

(v) *The AR-AT Undiscounted Games.*

As in case (iii) it is possible to develop a finite algorithm for this case that is conceptually related to the method of Vrieze et al. [62] for the undiscounted switching-controller games. Indeed, [43] contains a prescription for adapting the method in [62] to the AR-AT case.

Open Problem: Find an efficient finite step algorithm that computes pure stationary optimal strategies in AR-AT and perfect information games. In particular, can this be achieved by solving a single linear program?

We shall now mention some algorithms that fall into the category (D) above; namely those that converge under appropriate assumptions on the ergodic properties of the game. For every state s, let $A^s(\mathbf{u},\alpha) := [r(s,i,j) + \left(\frac{1}{1+\alpha}\right) \times \sum_{t=1}^{N} q(t|s,i,j)u(t)]$, where $\mathbf{u} = (u(1), u(2),\ldots,u(N))$. Here $\left(\frac{1}{1+\alpha}\right)$ can be thought of as a discount factor whenever $\alpha \in (0,\infty)$. Similarly, $A^s(\mathbf{u},0)$ will denote the game whose entries are as above, but with $\left(\frac{1}{1+\alpha}\right)$ replaced by 1.

Perhaps, the earliest algorithm in the category (D) is due to Hoffman and Karp [26]. It converges for the class of *irreducible* games, which contains only those games for which $Q(f,g)$ is an irreducible Markov Chain for every pair of stationary strategies. The algorithm is analogous to Algorithm 2 of Section 2, except that the matrix games $A^s(\mathrm{v}^\tau,0)$ are solved in Step 2, and the average reward MDP is solved in Step 3.

It should be noted that the approach of Hoffman and Karp [26] is to implicitly treat the undiscounted game as the limiting case of discounted games, as the discount factor tends to 1. This approach is taken much further by Federgruen [12], [11] who selects an appropriate sequence of discount factors β_τ approaching 1 as follows: Let $\beta_\tau = \frac{1}{1+\alpha_\tau}$ for each $\tau = 1,2,\ldots$, where the sequence $\{\alpha_\tau\}_{\tau=1}^{\infty}$ satisfies

$$\lim_{\tau\to\infty} [(1-\alpha_\tau)(1-\alpha_{\tau-1})\dots(1-\alpha_1)] = 0 \tag{1}$$

and

$$\lim_{\tau\to\infty} \left\{ \sum_{k=2}^{\tau} [(1-\alpha_\tau)\dots(1-\alpha_k)] \left| \alpha_k^{M-1} - \alpha_{k-1}^{M-1} \right| \right\} = 0 \tag{2}$$

with M being an appropriate positive integer. Note that for $\gamma \in (0,1]$ the sequence $\alpha_\tau = \tau^{-\gamma}$ for $\tau = 1, 2, \dots$ satisfies (1) and (2). In [12] the following algorithm is presented

Algorithm 11:

Step 1: Take any $\mathbf{u}^1$ in $\mathbb{R}^N$ such that $u^1(1) = 0$. Set $\tau = 1$.

Step 2: At iteration τ, $\mathbf{u}^\tau$ is known. Find $v^{\tau+1} := \text{val}\,[A^1(\mathbf{u}^\tau, \alpha_\tau)]$, and $u^{\tau+1}(s)$ $:= \text{val}\,[A^s(\mathbf{u}^\tau, \alpha_\tau)] - v^{\tau+1}$, for every s.

Step 3: Set $\tau := \tau + 1$ and return to Step 2.

Federgruen [12] showed that the above algorithm converges for a class of games possessing the following two properties: (a) Both players possess optimal stationary strategies, and (b) the value of the game is independent of the initial state. In particular, $v^{\tau+1}$ converges to the above (scalar) values as τ tends to infinity, and for τ sufficiently large, ε-optimal stationary strategies can be made up from the optimal strategies of the matrix games $A^s(\mathbf{u}^\tau, \alpha_\tau)$. It should be mentioned that the above algorithm can also be viewed as an extension of the modified value-iteration method of Hordijk and Tijms [28] to stochastic games.

A special, and a simpler case of Algorithm 11 arises when the sequence $\{\alpha_\tau\}_{\tau=1}^{\infty}$ is set to be identically zero, since the algorithm now resembles the standard successive approximations scheme. Both Federgruen [12] and Van der Wal [63] show that this simplified scheme can be successfully applied to the *unichain stochastic game.* The latter is a game in which the transition matrix $Q(f, g)$ contains only a single ergodic class (plus, perhaps, some transient states), for every pair of stationary strategies f and g for players I and II. An important technical device in the analysis of this simplified algorithm is the data transformation of Schweitzer [48] which ensures the "strong aperiodicity" property, that is, $q(t|s,i,j) > 0$ for all t, s, i and j.

6 Nonzerosum Stochastic Games

Nonzerosum stochastic games can be defined in a natural manner by considering a vector of immediate payoffs, one for each player, based on the current state and current actions of the players. The transitions among states remain the same as in zerosum games. For the sake of clarity we shall consider only two person games. Let $\phi_{\beta}^{k}(f,g)$ be the expected β-discounted payoff for player k, $k=1,2$ when stationary strategy pair (f,g) is used. Independently, Fink [23], Takahashi [56], Rogers [46] and Sobel [52] proved that nonzerosum discounted stochastic games have stationary equilibria. One can define also nonzerosum undiscounted games in the obvious manner. Although the problem of existence of ε-Nash equilibria is still open, recently Vrieze and Thuijsman [60] (also see [57]), proved that nonzerosum repeated games with absorbing states have ε-equilibria in behavior strategies. From an algorithmic point of view it would be desirable to characterize games possessing stationary equilibria.

Theorem 6.1: The following classes of nonzerosum undiscounted stochastic games have equilibria in stationary strategies.

(i) The transition matrix is irreducible for any set of stationary strategy choices by the players ([46]).
(ii) There is a special state s^* which is visited infinitely often with probability 1 ([55]).
(iii) There exists a non-empty set of states A and a positive integer N such that from any s the game moves to some state in A with an expected waiting time of at most N ([12]).
(iv) Single-controller stochastic games ([39]).
(v) SER-SIT games ([53], [41]).

Among discounted games the single-controller, SER-SIT, and AR-AT games are known to possess the ordered field property. Among the undiscounted nonzerosum games the single-controller and SER-SIT games possess the ordered field property.

We shall now outline a finite step algorithm to compute a stationary equilibrium point for discounted single-controller, and also for undiscounted irreducible single-controller games. The algorithm is similar in spirit to the method proposed by Filar and Raghavan [18] for solving the single-controller zerosum games.

Theorem 6.2: [Nowak and Rahavan [38]]. In a player II-controlled game let $f_1,f_2,\ldots,f_m$ $(g_1,g_2,\ldots,g_n)$ be an enumeration of all pure stationary strategies for player I (player II). Let (A, B) be an $m\times n$ bimatrix game defined by A: =

$[\sum_s r^1(s,f_i(s),g_j(s))]$, B: = $[(\sum_s \phi_\beta^2(f_i,g_j)(s)]$, and (ξ^*,η^*) be a Nash equilibrium point for the game (A, B). Then

(a) the stationary strategies $f^* = \sum_i \xi_i^* f_i$ and $g^* = \sum_j \eta_j^* g_j$ constitute a Nash equilibrium pair to the discounted game,

(b) in the case of the undiscounted irreducible single-controller games if we replace the above matrix B by the matrix $C = [\sum_s \phi^2(f_i,g_i)]$, then any equilibrium point (ξ^*,η^*) of the bimatrix game (A, C) induces analogously a stationary equilibrium point (f^*,g^*) as in the discounted case. Further, for the irreducible case Nash equilibrium payoffs are independent of the starting state.

Example 7: We illustrate the above algorithm with an example taken from [38] with 3 states and two actions at each state for both players, and $\beta = 0.8$. In this example the entries of the payoff matrices are treated as *costs* rather than as rewards.

s = 1

(6.3) 1	(0,8) 2
(0,5) 1	(7,1) 2

s = 2

(0,10) 2	(9,2) 3
(7,5) 2	(0,8) 3

s = 3

(3,0) 3	(0,5) 1
(0,4) 3	(4,0) 1

There will now be eight pure stationary strategies for player I lexicographically enumerated as (111), (112), . . . , (222) with the understanding that (ijk) corresponds to choosing the i^{th} row in state 1, the j^{th} row in state 2, and the k^{th} row in state 3. Similarly one can define pure stationary strategies for player II. The bimatrix (A,B) is given by

$$\begin{bmatrix}
(9,65) & (6,82) & (18,17) & (15,47.6) & (3,98) & (0,141.1) & (12,11.6) & (9,75) \\
(6,85) & (10,77) & (15,53) & (19,38.6) & (0,118) & (4,136.4) & (9,60.4) & (13,50) \\
(16,40) & (13,57) & (9,23) & (6,53.6) & (10,53) & (7,80.4) & (3,22.4) & (0,105) \\
(13,60) & (17,52) & (6,59) & (10,44.6) & (7,73) & (11,75.4) & (0,71.2) & (4,80) \\
(3,75) & (7,100) & (12,27) & (9,72) & (10,91) & (7,128.8) & (19,4.6) & (14,67.8) \\
(0,95) & (4,95) & (9,63) & (13,63) & (7,111) & (11,123.8) & (16,53.4) & (20,15) \\
(10,50) & (7,75) & (3,33) & (0,78) & (17,46) & (14,67.8) & (10,15.4) & (7,70) \\
(7,70) & (11,70) & (0,69) & (4,69) & (14,66) & (18,62.8) & (7,64.2) & (11,45)
\end{bmatrix}$$

Using the Lemke-Howson algorithm [30] we computed the Nash equilibrium

$$(\xi^*,\eta^*)=\left[\left(0,0,\frac{192}{1613},\frac{408}{1613},0,0,0,\frac{1013}{1613}\right)\,,\;\left(0,0,\frac{10}{91},\frac{39}{91},0,0,\frac{42}{91},0\right)\right]\,.$$

Hence an equilibrium point (f^*,g^*) for the stochastic game is given by

$$f^*=\begin{cases}\left(\dfrac{600}{1613},\dfrac{1013}{1613}\right) & s=1\\ (0,1) & s=2\\ \left(\dfrac{192}{1613},\dfrac{1421}{1613}\right) & s=3\end{cases}$$

and

$$g^*=\begin{cases}\left(\dfrac{7}{13},\dfrac{6}{13}\right) & s=1\\ (0,1) & s=2\\ \left(\dfrac{4}{7},\dfrac{3}{7}\right) & s=3\,.\end{cases}$$

Open Problem: Even though in the above the problem is reduced to solving for a Nash equilibrium point of a bimatrix game (for both the discounted and the irreducible undiscounted games), the full enumeration of the entire matrix can be prohibitive. This leads to the natural question: Is it possible to solve the discounted and the irreducible undiscounted single-controller games by solving only one linear complementarity problem?

Remark 6.1: In [16] Filar formulates an indefinite quadratic program for finding the equilibria of single-controller stochastic games. See [16] for an example with two states solved by a package for non-linear programming SOL/NPSOL [25]. It must be mentioned that Theorem 6.2 is not applicable to Filar's example with reducible transitions.

The next theorem gives a recipe for computing Nash equilibria for SER-SIT games by solving a single bimatrix game.

Theorem 6.3: [Sobel [53], and Parthasarathy et al. [41]].

Let Γ be a nonzerosum SER-SIT stochastic game with the immediate rewards $r^k(s,i,j) = c^k(s) + p^k(i,j)$ and the transitions $q(t|s,i,j) = q(t|i,j)$. Then any equilibrium (f^β, g^β) of the bimatrix game $[(p^k(i,j) + \beta \sum_t p(t|i,j)c^k(t), k = 1,2)]$ yields $(f^\beta, f^\beta, \ldots, f^\beta)$, $(g^\beta, g^\beta, \ldots, g^\beta)$ as a stationary equilibrium point. A similar solution for the undiscounted game can be constructed out of a Nash equilibrium point of the above bimatrix game with $\beta = 1$.

7 Mathematical Programming; A Unified Viewpoint

The linear programming (matrix game) formulations of the single-controller (SER-SIT) stochastic games that were presented in Sections 3 and 5 indicate that at least in some special cases the solutions of stochastic games can be derived from optimal solutions of suitably constructed mathematical programs. While the lack of the ordered field property eliminates the possibility that general stochastic games could be solved by linear programming, the question of how closely are these games related to problems of mathematical programming is an extremely important one from the algorithmic standpoint. This importance stems from the fact that there is now a vast pool of mathematical programming techniques that could be brought to bear on stochastic games once they can be interpreted as linear/nonlinear programs.

We shall demonstrate that Nash Equilibria in stationary strategies can always be characterized as global optima of certain explicit mathematical programs. Insofar as these formulations include the discounted, undiscounted, zerosum and nonzerosum K-person stochastic games, this approach indeed represents a unified viewpoint.

We shall require the following notation. Let $m_s(n_s)$ be the number of actions available to player I (II) in a state s, and set $m := \sum_s m_s$ $(n := \sum_s n_s)$. Given an arbitrary vector $\mathbf{v} \in \mathbb{R}^N$ and $m_s \times n_s$ matrices $A^s = [r(s,i,j)]$ and $Q^s(\mathbf{v}) := [\sum_t q(t|s,i,j)v(t)]$, define the corresponding $m \times n$ block diagonal matrices $A := \text{diag}\,[A^1, A^2, \ldots, A^N]$ and $Q(\mathbf{v}) := \text{diag}\,[Q^1(\mathbf{v}), Q^2(\mathbf{v}), \ldots, Q^N(\mathbf{v})]$. Now recall that a stationary strategy $f(g)$ for player I (II) can be regarded as a block m-vector (n-vector) whose s-th block is $\mathbf{f}(s)(\mathbf{g}(s))$. Let $F(G)$ denote the sets of all stationary strategies for player I (II) and note that these sets are characterized by sets of only linear constraints. Let $\mathbf{1}_k = (1,1,\ldots,1) \in \mathbb{R}^k$ and with each $\mathbf{v} \in \mathbb{R}^N$ associate a block m-vector (n-vector) $\mathbf{v}_{(m)}$ $(\mathbf{v}_{(n)})$ whose s-block is $v(\mathrm{s})\mathbf{1}_{m_s}(v(\mathrm{s})\mathbf{1}_{n_s})$.

A. Zerosum Stochastic Games and Mathematical Programming

Perhaps, the earliest nonlinear programming formulation of a stochastic game was due to Rothblum [47] who proposed to consider the following mathematical program.

Mathematical Program MP1. [The variables are v, and g]

$$\min (\mathbf{1}.\mathbf{v}) \ .$$

Subject to

(a) $\mathbf{v}_{(m)} \geq [A + \beta Q(\mathbf{v})] g$

(b) $g \in G$.

Presupposing Shapley's existence theorem [49], Rothblum showed that every global minimum of MP1 yields the value vector, and an optimal stationary strategy for player II in the above discounted stochastic game. The next theorem substantially extends Rothblum's results.

Theorem 7.1: [Vrieze [59]].

(i) Every local minimum of MP1 is also a global minimum.
(ii) For each minimum $(\mathbf{v}^0, g^0)$ of MP1, the $\mathbf{v}^0$-part is unique and equals the value vector of the discounted stochastic game, and g^0 is an optimal stationary strategy for player II.
(iii) Each minimum $(\mathbf{v}^0, g^0)$ of MP1 is a Kuhn-Tucker point. If the variables $\lambda_i(s)$ are the Lagrange multipliers corresponding to the constraints *(a)*, then the stationary strategy $f_i^0 = \lambda_i(s)/\sum_i \lambda_i(s)$ for each s and i, is well defined and is optimal for player I.

Remark 7.1: Note that the above theorem supplies a new proof of Shapley's existence theorem (see [59]), without using the contraction property.

In the case of undiscounted zerosum games, the fact that stationary strategies need not exist implies that a complete analogue of Theorem 7.1 is not possible. Nonetheless, it is possible to give a mathematical programming characterization of optimal stationary strategies whenever they exist. In addition, it will be seen that a mathematical programming approach can also characterize the "best" stationary strategies. Towards this aim we need to develop a measure of the "distance" from optimality for an arbitrary pair of stationary strategies. The following natural measure of such a distance is proposed.

Let $(\hat{f},\hat{g})$ be an arbitrary and fixed pair of stationary strategies, and define

$$\delta(\hat{f},\hat{g}) := \sum_s \, [\max_f \phi(f,\hat{g})(s) - \min_g \phi(\hat{f},g)(s)] \ . \tag{1}$$

Note that every term in (1) is nonnegative, and that $\delta(\hat{f},\hat{g})=0$ if and only if $\hat{f}$ and $\hat{g}$ are optimal for players I and II respectively. Furthermore, if *ε-optimality* of $\hat{f}$ and $\hat{g}$ is defined by the condition

$$\phi(f,\hat{g})(s) - \varepsilon \le \phi(\hat{f},\hat{g})(s) \le \phi(\hat{f},g)(s) + \varepsilon \ , \tag{2}$$

for all s,f and g, then it is evident that whenever $\delta(\hat{f},\hat{g})>0$ the strategies $\hat{f}$ and $\hat{g}$ are ε-optimal with ε no greater than $\delta(\hat{f},\hat{g})$.

Mathematical Program MP2. [The variables are $\mathbf{v}^1$, $\mathbf{v}^2$, $\mathbf{u}^1$, $\mathbf{u}^2$, f and g]

$$\inf\,[\mathbf{1}.(\mathbf{v}^1+\mathbf{v}^2)] \ .$$

Subject to

(a) $\mathbf{v}^1_{(m)} \ge Q(\mathbf{v}^1)g$

(b) $\mathbf{v}^1_{(m)} + \mathbf{u}^1_{(m)} \ge [A + Q(\mathbf{u}^1)]\,g$

(c) $\mathbf{v}^2_{(n)} \ge fQ(\mathbf{v}^2)$

(d) $\mathbf{v}^2_{(n)} + \mathbf{u}^2_{(n)} \ge f[-A + Q(\mathbf{u}^2)]$

(e) $(f,g) \in F \times G$.

Theorem 7.2: [Filar et al. [21]]. Consider a two person, undiscounted zerosum, stochastic game and the corresponding mathematical program MP2. The following hold for every $\varepsilon \ge 0$:

(i) If $(\hat{\mathbf{v}}^1, \hat{\mathbf{v}}^2, \hat{\mathbf{u}}^1, \hat{\mathbf{u}}^2, \hat{f}, \hat{g})$ is feasible for MP2 and has objective function value of ε or less, then $\hat{f}$ and $\hat{g}$ are ε-optimal. If $\hat{f}$ and $\hat{g}$ are ε-optimal, then there exist vectors $\hat{\mathrm{v}}^1, \hat{\mathrm{v}}^2, \hat{\mathrm{u}}^1, \hat{\mathrm{u}}^2$ which together with $\hat{f}$ and $\hat{g}$ form a feasible point of *MP2* with the objective value of $2N\varepsilon$ or less.

(ii) If the point $(\hat{\mathrm{v}}^1, \hat{\mathrm{v}}^2, \hat{\mathrm{u}}^1, \hat{\mathrm{u}}^2, \hat{f}, \hat{g})$ is a global minimum of MP2 with the objective function value of $\hat{w}$, then $\hat{w} = \delta(\hat{f}, \hat{g}) \leq \delta(f, g)$, for all stationary strategy pairs (f, g). That is, $(\hat{f}, \hat{g})$ is the best stationary strategy pair with respect to the measure of the distance from optimality (1).

B. General-Sum Stochastic Games and Mathematical Programming

We shall now extend the mathematical programming approach to nonzerosum stochastic games. For the sake of clarity we assume there are only two players.

We shall require the following additional notation: the superscript 1 (respectively 2) will in general refer to quantities connected with player I (respectively player II). For instance, $r^2(s, \hat{i}, j)$ and $\phi^2_\beta(f, g)(s)$ denote player II's immediate and overall payoffs at state s resulting from the use of actions i and j (respectively strategies f and g), and A^2 is the associated $m \times n$ block-diagonal matrix. Similarly for player I. For the nonzerosum, two-person, discounted stochastic game we now construct the following nonlinear program.

Mathematical Program MP3. [The variables are $\mathrm{v}^1, \mathrm{v}^2, f$, and g]

$$\min \sum_{k=1}^{2} [\mathbf{1}.\mathrm{v}^k - f[A^k + \beta Q(\mathrm{v}^k)]g]$$

Subject to:

(a) $\mathrm{v}^1_{(m)} \geq [A^1 + \beta Q(\mathrm{v}^1)]g$

(b) $\mathrm{v}^2_{(n)} \geq f[A^2 + \beta Q(\mathrm{v}^2)]$

(c) $(f, g) \in F \times G$.

Theorem 7.3: [[21]]. Consider a two-person, general-sum, discounted stochastic game, and the corresponding mathematical program MP3. The following hold:

(i) The objective function of MP3 possesses a global minimum of zero. The stationary strategy pair $(\hat{f}, \hat{g})$ forms a Nash Equilibrium with equilibrium payoff vectors $\hat{\mathrm{v}}^1$ and $\hat{\mathrm{v}}^2$ if and only if the point $(\hat{\mathrm{v}}^1, \mathrm{v}^2, \hat{f}, \hat{g})$ is a global minimum of MP3.

(ii) If $(\hat{v}^1, \hat{v}^2, \hat{f}, \hat{g})$ is an arbitrary feasible point of *MP3* with the objective function value of $\gamma(>0)$, then $(\hat{f}, \hat{g})$ in an ε-equilibrium with some $\varepsilon \leq \frac{\lambda}{1-\beta}$.

Remark 7.2: Note that the importance of part (ii) of the above theorem stems from the fact that any heuristic that will produce a feasible point of MP3 with a low objective value will yield a "near-equilibrium" for the stochastic game. In Breton [6] and Breton et al. [7] a number of such near-equilibria are computed in small-sized problems by solving a related, but more complex, nonlinear program within a prespecified precision bound.

For the undiscounted, general-sum stochastic games a result analogous to part (i) of Theorem 7.3 has been established in [21]. The absence of an analog of part (ii) of the above theorem is significant since it points to an inherent difficulty that is illustrated by an example given in [21].

Open Problem: What natural conditions on the data of the game would ensure that points with small values of the objective function of an appropriate mathematical program, yield near-equilibria of the undiscounted stochastic game?

C. Miscellaneous Results for Special Classes

We conclude this survey with a brief mention of a number of results in the general spirit of mathematical programming which, however, apply only to special classes of stochastic games. In particular, it should be mentioned that for the two-person nonzerosum, single-controller stochastic games a single stationary equilibrium can be seen as a solution of a quadratic program that is a generalization of algorithms 9 (or of 5 in the discounted case), as demonstrated in [16]. In the zerosum single-controller, undiscounted games Baykal-Gursoy and Ross [2] have decomposed the state space into a number of "strongly communicating classes", and one transient class in order to obtain a disaggregated linear programming type algorithm for computing a pair of stationary optimal strategies and the value vector.

For nonzerosum, switching-controller, or even AR-AT undiscounted games the existence of a stationary equilibrium point is still unknown. An interesting result for the discounted AR-AT games (see [43]) says that there always exists one stationary equilibrium in which each player has to use at most two pure actions with positive probability, in every state. Further, in Parthasarathy et al. [41] it is shown that in the SER-SIT case one stationary equilibrium can always be obtained by solving a set of N bimatrix games. In Filar and Shultz [20], it is shown that in the zerosum switching-controller, or AR-AT games (discounted or undiscounted), stationary optimal strategies from the global minima of suitably constructed bilinear objectives over polyhedral feasible regions. While these can be reformulated as linear complementarity problems, it is not known whether they can be solved by Lemke's algorithm (see [29]), or by some natural modification of the latter. In addition, Mizuno [33], [34] studied a version of the Lagrangian problem associated with stationary Nash equilibria of stochastic games, and derived a set of necessary conditions that these equilibria must satisfy.

References

[1] Aumann RJ (1964) Mixed and Behaviour Strategies in Infinite Extensive Games. In: Dresher M, Shapley LS (eds) Advances in Game Theory. Princeton University Press. Annals of Mathematics Studies 52

[2] Baykal-Gursoy M, Ross KW (1989) A Sample Path Approach to Stochastic Games. Techn Rep, Rutgers University

[3] Bewley T, Kohlberg E (1976) The asymptotic theory of stochastic games. Math Oper Res 1:197-208

[4] Bewley T, Kohlberg E (1978) On Stochastic Games with Stationary Optimal Strategies. Math Oper Res 3:104-125

[5] Blackwell D, Ferguson T (1968) The Big Match. Ann Math Statistics 39:159-163

[6] Breton M (1987) Equilibre pour des Jeux Sequentiel. PhD Thesis, University of Montreal

[7] Breton M, Filar JA, Haurie A, Shultz TA (1985) On the Computation of Equilibria in Discounted Stochastic Games. In: Basar T (ed) Dynamic Games and Applications in Economics. Springer, Lecture Notes in Economics and Mathematical Systems 265

[8] Brown GW (1951) Iterative Solutions of Games by Fictitious Play. In: Koopmans TC (ed) Activity Analysis of Production and Allocation. Wiley

[9] Charnes A, Schroeder R (1967) On Some Tactical Antisubmarine Games. Naval Res Log Qtly 14:291-311

[10] Derman C (1970) Finite State Markovian Decision Processes. Academic Press, New York

[11] Federgruen A (1980) Successive Approximation Methods in Undiscounted Stochastic Games. Oper Res 28:794-810

[12] Federgruen A (1983) Markovian Control Problems. Mathematical Centre Tracts 97, Amsterdam

[13] Filar JA (1980) Algorithms for Solving Some Undiscounted Stochastic Games. PhD Thesis, University of Illinois at Chicago, Chicago, Illinois

[14] Filar JA Ordered Field Property for Stochastic Games When the Player Who Controls Transitions Changes from State to State. J Optim Theory Appl 34:503-513

[15] Filar JA (1985) Player Aggregation in the Traveling Inspector Model. IEEE Trans on Aut Control AC-30:723-729

[16] Filar JA (1986) Quadratic Programming and the Single-Controller Stochastic Game. J Math Anal Appl 113:136-147

[17] Filar JA, Raghavan TES (1980) Two Remarks Concerning Two Undiscounted Stochastic Games. Tech Rep 392. John Hopkins University, Department of Mathematical Sciences

[18] Filar JA, Raghavan TES (1984) A Matrix Game Solution of the Single-Controller Stochastic Game. Math Oper Res 9:356-362

[19] Filar JA, Shultz TA (1986) Nonlinear Programming and Stationary Strategies in Stochastic Games. Math Progr 35:243 – 247

[20] Filar JA, Shultz TA (1987) Bilinear Programming and Structured Stochastic Games. J Optim Theory Appl 53:85 – 104

[21] Filar JA, Shultz TA, Thuijsman F, Vrieze OJ (1987) Nonlinear Programming and Stationary Equilibria in Stochastic Games. Math Program (to appear)

[22] Filar JA, Tolwinski B (1988) On the Algorithm of Pollatschek and Avi-Itzhak. In: Ferguson T et al. (eds) Stochastic Games and Related Topics, Shapley Honor volume. Kluwer, Dordrecht, The Netherlands (to appear)

[23] Fink AM (1964) Equilibrium in a Stochastic N-Person Game. J Sci in Hiroshima Univ, Series A-I. 28:89 – 93

[24] Gilette D (1957) Stochastic Games with Zero Stop Probabilities. In: Dresher AWTM, Wolfe P (eds) Contributions to the Theory of Games. Princeton University Press, Annals of Mathematics Studies 39

[25] Gill PE, Murray W, Saunders MA, Wright MH (1983) User's Guide for SOL/NPSOL: A Fortran Package for Nonlinear Programming. Tech Rep SOL 83 – 12. Stanford University, Stanford, California

[26] Hoffman AJ, Karp RM (1966) On Non-terminating Stochastic Games. Management Sci 12:359 – 370

[27] Hordijk A, Kallenberg LGM (1981) Linear Programming and Markov Games I, II. In: Moeschlin O, Pallaschke D (eds) Game Theory and Mathematical Economics. North Holland

[28] Hordijk A, Tijms HC (1975) A Modified Form of the Iterative Method of Dynamic Programming. Annals of Stat 3:203 – 208

[29] Lemke CE (1965) Bimatrix Equilibrium Points and Mathematical Programming. Management Sci 12:413 – 423

[30] Lemke CE, Howson JT (1964) Equilibrium Points of Bimatrix Games. J Soc Indust Appl Math 12:413 – 423

[31] Liggett T, Lipman S (1969) Stochastic Games with Perfect Information and Time Average Payoff. SIAM Rev 11:604 – 607

[32] Mertens JF, Neyman A (1981) Stochastic Games. International J Game Theory 10:53 – 56

[33] Mizuno N (1986) A New Algorithm for Non-Zerosum Markov Games. Tech Rep, New York University, Graduate School of Business

[34] Mizuno N (1987) A Necessary Condition for the Existence of Average Reward Equilibrium Points for Finite N-Person Markov Games. Tech Rep, New York University, Graduate School of Business

[35] Mohan SR, Raghavan TES (1987) An Algorithm for Discounted Switching Control Games. OR Spectrum 9:41 – 45

[36] Monash CA (1979) Stochastic Games. The Minimax Theorem. PhD Thesis, Harvard University

[37] von Neumann J (1928) Zur Theorie der Gesellschaftsspiele. Math Annal 100:295 – 320

[38] Nowak A, Raghavan TES (1989) A Finite Step Algorithm via a Bimatrix Game to a Single-Controller Non-zerosum Stochastic Game. Tech Rep 89 – 1. The University of Illinois at Chicago

[39] Parthasarathy T, Raghavan TES (1981) An Orderfield Property for Stochastic Games when One Player Controls Transition Probabilities. J Optim Theory Appl 33:375 – 392

[40] Parthasarathy T, Stern M (1977) Markov Games: A Survey. In: Roxin PLE, Sternberg R (eds) Differential Games and Control Theory. Marcel Dekker

[41] Parthasarathy T, Tijs SH, Vrieze OJ (1984) Stochastic Games with State Independent Transitions and Separable Rewards. In: Hammer G, Pallaschke D (eds) Selected Topics in OR and Mathematical Economics. Springer, Lecture Notes Series 226

[42] Pollatschek M, Avi-Itzhak B (1969) Algorithms for Stochastic Games with Geometrical Interpretation. Management Sci 15:399 – 415

[43] Raghavan TES, Tijs SH, Vrieze OJ (1985) On Stochastic Games with Additive Reward and Transition Structure. J Optim Theory Appl 47:451 – 464

[44] Rao S, Chandrasekaran R, Nair K (1973) Algorithms for Discounted Stochastic Games. J Optim Theory Appl 11:627 – 637

[45] Robinson J (1950) An Iterative Model of Solving a Game. Ann Math 54:296 – 301

[46] Rogers PD (1969) Non-zerosum Stochastic Games. PhD Thesis, University of California at Berkeley, Berkeley, California

[47] Rothblum UG (1978) Solving Stopping Stochastic Games by Maximizing a Linear Function Subject to Quadratic Constraints. In: Moeschlin O, Pallaschke D (eds) Game Theory and Related Topics. North Holland

[48] Schweitzer PJ (1968) Perturbation theory and finite Markov chains. J Appl Prob 5:401 – 413

[49] Shapley LS (1953) Stochastic Games. Proc Nat Acad Sci USA 39:1095 – 1100

[50] Shapley LS (1964) Some Topics in Two Person Games. In: Dresher LSSM, Tucker AW (eds) Advances in Game Theory. Princeton University Press

[51] Sinha S (1989) A Contribution to the Theory of Stochastic Games. PhD Thesis, Indian Statistical Institute, New Delhi

[52] Sobel MJ (1971) Non-cooperative Stochastic Games. Ann Math Stat 42:1930 – 1935

[53] Sobel MJ (1981) Myopic Solutions of Markov Decision Processes and Stochastic Games. Operat Res 29:995 – 1009

[54] Sobel MJ (1981) Stochastic Fishery Games with Myotopic Equilibria. In: Mirman LJ, Spulber D (eds) The Economics of Renewable Resources. Elsevier-North-Holland

[55] Stern M (1975) On Stochastic Games with Limiting Average Payoff. PhD Thesis, University of Illinois at Chicago

[56] Takahashi M (1964) Equilibrium Points of Stochastic Non-cooperative n-Person Games. J Sci Hiroshima University, Series A-I, 28:95 – 99

[57] Thuijsman F (1989) Optimality and Equilibria in Stochastic Games. PhD Thesis, Rijksuniversiteit Limburg, Maastricht

[58] Vrieze OJ (1981) Linear Programming and Undiscounted Stochastic Games. OR Spectrum 3:29 – 35

[59] Vrieze OJ (1987) Stochastic Games with Finite State and Action Spaces. CWI Tracts 33, Amsterdam

[60] Vrieze OJ, Thuijsman F (1986) On Equilibria in Repeated Games with Absorbing States. Tech Rep 8535. Catholic University, Nijmegen, Department of Mathematics

[61] Vrieze OJ, Tijs SH (1980) Fictitious Play Applied to Sequence of Games and Discounted Stochastic Games. Intern J Game Theory 11:71 – 85

[62] Vrieze OJ, Tijs SH, Raghavan TES, Filar JA (1983) A Finite Algorithm for the Switching Controller Stochastic Game. OR Spectrum 5:15 – 24

[63] van der Wal J (1981) Stochastic Dynamic Programming. Math Center Tracts 139, Amsterdam

[64] van der Wal J (1977) Discounted Markov Games: Successive Approximation and Stopping Times. Intern J Game Theory 6:11 – 22

[65] Winston W (1978) A Stochastic Game Model of a Weapons Development Competition. SIAM J Control Optim 16:411 – 419

[66] Winston WL, Cabot AV (1984) A Stochastic Game Model of Football Play Selection. Tech Rep, Indiana University, Paper presented at the TIMS/ORSA joint National meeting in Dallas

Received March 1988
Revised version received August 1989