Noname manuscript No.
(will be inserted by the editor)

Computing Uniformly Optimal Strategies in Two-Player Stochastic Games

Eilon Solan and Nicolas Vieille

Received: date / Accepted: date

Abstract We provide a computable algorithm to calculate uniform ε-optimal strategies in two-player zero-sum stochastic games. Our approach can be used to construct algorithms that calculate uniform ε-equilibria and uniform correlated ε-equilibria in various classes of multi-player non-zero-sum stochastic games. JEL codes: C63, C73.

Keywords optimal strategies · stochastic games · computation

1 Introduction

Stochastic games model dynamic interactions in which the environment changes in response to the behavior of the players. These games were introduced in Shapley [20] (1953), who proved the existence of the discounted value and of stationary discounted optimal strategies in two-player zero-sum games with finite state and action spaces. This existence result was later generalized to the existence of a stationary discounted equilibrium in multi-player games (Fink, 1964). Bewley and Kohlberg (1976) proved that the limit of the discounted values exists. Mertens and Neyman (1981) proved the existence of uniform ε-optimal strategies in two-player zero-sum games: for every $\varepsilon > 0$ each of the two players has a strategy that guarantees the discounted value, up to ε, for every discount factor sufficiently close to 0. The limit of the discounted values is termed *the uniform value*. Thus, the players can play well, and approximately obtain the uniform value, even if they do not know their discount factor, provided their discount factor is sufficiently low.

E. Solan
The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: eilons@post.tau.ac.il

N. Vieille
Département Finance et Economie and GREGHEC, HEC Paris, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. E-mail: vieille@hec.fr

Computing the discounted or the uniform value of a stochastic game is a difficult problem. Linear programming methods were used to calculate the discounted and uniform value of several classes of stochastic games, see Filar and Vrieze (1996). However, Parthasarathy and Raghavan (1981) provides a game in which all the data are rational (including the discount factor), but the discounted value is irrational. Therefore it is not clear whether linear programming methods, which successfully calculate the value of general two-player zero-sum matrix games, can be used to calculate the value of stochastic games. Other methods that were used to calculate the value or equilibria in discounted stochastic games include fictitious play (Vrieze and Tijs, 1982), value iteration and general methods to find the maximum of a function (Filar and Vrieze, 1996), and an homotopy method (Herings and Peeters, 2004).

Recently Chatterjee, Majumdar and Henzinger (2008) provided a finite algorithm for approximating the uniform value. This algorithm relies on the following insight. By Bewley and Kohlberg (1976), the function $\lambda \mapsto v^\lambda$ that assigns to each λ the value of the λ-discounted game, is a semi-algebraic function of λ. It therefore can be expressed as a Taylor series in fractional powers of λ, and is monotonic, in a neighborhood of $\lambda = 0$. As mentioned before, identifying the uniform value amounts to finding the *limit* v of this semi-algebraic function. Relying on Bewley and Kohlberg (1976), Chatterjee, Majumdar and Henzinger note that, for a given α, determining whether $v > \alpha$ is equivalent to finding the truth value of a sentence in the theory of real-closed fields. Tarski's quantifier elimination algorithm (or Basu, 1999, see also Basu, Pollack and Roy, 2003) can be used to compute this truth value. Since the uniform value v is bounded by the payoffs of the game, it is sufficient to repeat this algorithm for finitely many different values of α, to get an approximation of v.

In this note we show how this approach can be extended to calculate uniform ε-optimal strategies. We will rely on ε-optimal strategies devised in Mertens and Neyman (1981). These strategies use an *unbounded* memory. This is a necessary feature, since in general uniformly ε-optimal strategies *cannot* be implemented by finite automata. A famous example of this is the game called "The Big Match", see Blackwell and Ferguson (1968), and also Fortnow and Kimmel (1998).

We provide an algorithm that terminates in finite time, and can be executed before the game starts. The output of this algorithm can be used to compute, for any given stage n, the mixed action to be played at that stage. The operations performed at stage n include addition, and calculating the root of a real number (an accurate approximation of the square root is sufficient), where the real numbers increase to infinity, at a rate which is at most linear in n, as the game evolves. Hence the memory space required for the calculation, as well as the per-stage processing time, increase with n.

The main purpose of the preprocessing algorithm is the following. By Bewley and Kohlberg (1976), there is a semi-algebraic function $\lambda \mapsto x^\lambda$, that associates to any discount factor a stationary, λ-discounted optimal strategy (of player 1). Therefore, the probability $x^\lambda(s, a)$ assigned to action a in state s

can be expressed as a Taylor series in fractional powers of λ, in a neighborhood of zero. The preprocessing stage consists in finding the leading exponent and an approximation of the leading coefficient of this series, for all states s and actions a, which are then used to implement Mertens and Neyman's ε-optimal strategies. As in Chatterjee, Majumdar and Henzinger (2008) we rely on the theory of real-closed fields for this approximation. It should be noted that we know of no *a priori* estimate for the leading coefficients. This prevents us from getting a complexity bound on this algorithm, in contrast to Chatterjee, Majumdar and Henzinger (2008).

The paper is organized as follows. Section 2 contains the model of stochastic games. Section 3 provides a self-contained reminder of Mertens and Neyman (1981), together with related material. We also introduce all the sentences that are used for the algorithm. The algorithm itself is presented in section 4.

In section 5, we briefly argue that such constructions can be used in non-zero-sum stochastic games as well. The existence of a uniform ε-equilibrium in multi-player games, that is, a strategy profile that is ε-equilibrium for every discount factor sufficiently low, was proved for two-player games (Vieille, 2000a,b), and for some classes of games with more than two players (see, e.g., Filar and Vrieze (1996), Solan (1999), Solan and Vieille (2001), Flesch, Thuijsman and Vrieze (2007), Flesch, Schoenmakers and Vrieze (2008)). In most cases, the ε-equilibria share the feature that each player behaves in a "simple" manner, as long as the empirical play of other players meets some predefined constraints, and switches to an ε-optimal strategy in a related zero-sum game as soon as one of these constraints is not met.

2 The Model

A *two-player zero-sum stochastic game* is given by a 5-tuple (S, A, B, r, q), where S is a finite set of states, A and B are two finite sets of actions for the two players, $r : S \times A \times B \to \mathbf{R}$ is a payoff function, and $q : S \times A \times B \to \Delta(S)$ is a transition function.[1] We assume w.l.o.g. that payoffs are bounded by 1: $|r(s, a, b)| \leq 1$ for every $(s, a, b) \in S \times A \times B$. The mixed extensions of r and q are still denoted by r and q respectively. Thus, and given $x \in \Delta(A)$, $y \in \Delta(B)$, $r(s, x, y)$ is the expected payoff to player 1, when the state is s and players use the mixed actions x and y respectively.

The game starts at the initial[2] state $s_1 \in S$. At every stage n, the game is in some state $s_n \in S$, the players choose independently and simultaneously actions $a_n \in A$ and $b_n \in B$, and a new state s_{n+1} is chosen according to the probability distribution $q(\cdot \mid s_n, a_n, b_n)$. Throughout, we let a two-player, zero-sum stochastic game Γ be given.

The information that is available to each player at stage n is the sequence of states visited so far, and the sequence of past actions that were chosen by both players. Therefore, the set of *finite histories* is $\bigcup_{n=1}^{\infty} S \times (S \times A \times$

[1] Given any finite set Ω, $\Delta(\Omega)$ denotes the set of probability distributions over Ω.

[2] We find it convenient to let the initial state be a parameter, and not a data of the game.

$B)^{n-1}$. A *behavior strategy* for player 1 is a function $\sigma : \bigcup_{n=1}^{\infty} S \times (S \times A \times B)^{n-1} \to \Delta(A)$ that assigns, to every finite history h, a mixed action to play if h occurs. Behavior strategies τ for players 2 are defined analogously. A *stationary strategy* for player 1 (resp. player 2) is a function $x : S \to \Delta(A)$ (resp. $y : S \to \Delta(B)$). We denote by $x_s(a)$ the probability to play the action $a \in A$ in state $s \in S$, when the stationary strategy x is used. The quantity $y_s(b)$ is defined analogously.

For every discount factor $\lambda \in (0,1]$ and every pair of strategies (σ, τ) for the two players, define the λ-discounted payoff by:

$$\gamma^{\lambda}_{s_1}(\sigma, \tau) = \mathbf{E}_{s_1,\sigma,\tau}\left[\lambda \sum_{n=1}^{\infty}(1-\lambda)^{n-1} r(s_n, a_n, b_n)\right],$$

where the expectation is computed under the probability distribution $\mathbf{P}_{s_1,\sigma,\tau}$ over plays induced by the initial state s_1 and the strategy pair (σ, τ). The quantity $v^{\lambda}_{s_1}$ is the λ-*discounted value* at the initial state s_1 if

$$v^{\lambda}_{s_1} = \sup_{\sigma} \inf_{\tau} \gamma^{\lambda}_{s_1}(\sigma, \tau) = \inf_{\tau} \sup_{\sigma} \gamma^{\lambda}_{s_1}(\sigma, \tau). \tag{1}$$

A strategy σ (resp. τ) that attain the supremum in the middle term of (1) (resp. the infimum in the right term of (1)) is called λ-*discounted optimal.* Shapley (1953) proved the existence of the discounted value, and that both players have stationary optimal strategies. Bewley and Kohlberg (1976) proved that $v^0_{s_1} := \lim_{\lambda \to 0} v^{\lambda}_{s_1}$ exists for every $s_1 \in S$. Mertens and Neyman (1981) proved that $v^0_{s_1}$ is the *uniform value* in the following sense. For every $\varepsilon > 0$ player 1 has a strategy σ_ε that *uniformly guarantees* $v^0 - 2\varepsilon$: there is $\lambda_0 > 0$ such that

$$\gamma^{\lambda}_{s_1}(\sigma_\varepsilon, \tau) \geq v^{\lambda}_{s_1} - \varepsilon \geq v^0_{s_1} - 2\varepsilon,$$

for every strategy τ of player 2, for every discount factor $\lambda \in (0, \lambda_0)$, and for every initial state $s_1 \in S$. Similarly, for every $\varepsilon > 0$ player 2 has a strategy τ_ε that *uniformly guarantees* $v^0 + 2\varepsilon$: there is $\lambda_0 > 0$ such that

$$\gamma^{\lambda}_{s_1}(\sigma, \tau_\varepsilon) \leq v^{\lambda}_{s_1} + \varepsilon \geq v^0_{s_1} - 2\varepsilon,$$

for every strategy σ of player 1, for every discount factor $\lambda \in (0, \lambda_0)$, and for every initial state $s_1 \in S$.

Unlike for discounted games, in which stationary optimal strategies exist, in general there are no stationary uniform ε-optimal strategies, and the ε-optimal strategies designed by Mertens and Neyman (1981) are history dependent. In their construction, at every stage n the player calculate a fictitious discount factor λ_n as a function of past play, and plays according to any optimal strategy in the λ_n-discounted game.

3 Preparations

We here collect diverse material. We first recall the main features of the construction of Mertens and Neyman (1981). We next state a result on the continuity of discounted payoffs with respect to strategies, and recall a few definitions related to Puiseux functions. Next, we introduce and discuss the sentences in the theory of real-closed fields that our algorithm will use.

3.1 Uniform ε-optimal strategies – Background

For completeness, we here describe the uniform ε-optimal strategies constructed in Mertens and Neyman (1981, see also Mertens, Sorin and Zamir (1994, Ch. VII), or Neyman (2003)). Our algorithm uses this construction.

For every discount factor λ, let z^λ be a stationary strategy for player 1, and let $w^\lambda : S \to \mathbf{R}$ be a function, such that the following inequality holds, for every state $s \in S$, and every action $b \in B$:

$$\lambda r(s, z^\lambda_s, b) + (1-\lambda) \sum_{s' \in S} q(s' \mid s, z^\lambda_s, b) w^\lambda_{s'} \geq w^\lambda_s. \tag{2}$$

That is, if the continuation payoff is given by w^λ, the strategy z^λ guarantees that the discounted payoff for player 1 when the state is s is at least $w^\lambda(s)$. Suppose that the limit $w^0_s := \lim_{\lambda \to 0} w^\lambda_s$ exists. Let $\delta < 1$, let $\lambda : (0, \infty) \to (0,1)$ be a strictly decreasing function, and let $u_* \geq 1/\delta$ such that the following holds: for every $\theta \in [-3, 3]$, every $u \geq u_*$ and every $s \in S$,

$$|\lambda(u+\theta) - \lambda(u)| \leq \delta\lambda(u), \tag{3}$$

$$\left| w_s^{\lambda(u+\theta)} - w_s^{\lambda(u)} \right| < 4\lambda(u), \tag{4}$$

$$\int_{u_*}^{\infty} \lambda(u) \leq \delta. \tag{5}$$

Mertens and Neyman (1981) proved the following.

Theorem 1 *Denote $\varepsilon = 12\delta$. In the notations given above, set*

$$u_0 = u_*, \quad u_{n+1} = \max\{u_*, r(s_n, a_n, b_n) - w_{s_n}^{\lambda(u_n)} + 2\varepsilon\}. \tag{6}$$

Let σ be the strategy that plays at each stage n the mixed action $z_{s_n}^{\lambda(u_n)}$. Then σ uniformly guarantees $w^0_{s_0} - 5\varepsilon$.

If z^λ is a λ-discounted optimal stationary strategy of player 1, and if w^λ is the λ-discounted value, then Eq. (2) holds, and, by finding a function $u \mapsto \lambda(u)$ and a constant u_* such that the requirements (3)-(5) are satisfied, Mertens and Neyman (1981) proved that v^0_s is the uniform value at the initial state s.

Our algorithm proceeds by computing functions w^λ and z^λ that approximate the value and optimal strategies of the λ-discounted game, and such that (2) is satisfied. Finally, we will find a function $\lambda(u)$ and a natural number u_* such that the requirements (3)-(5) are satisfied as well.

3.2 Perturbations

The following result follows from Solan (2003) or Solan and Vieille (2003). It says that small perturbations in the strategies of player 1 do not affect the discounted payoff by much. Observe that the bound that the theorem provides is independent of λ.

Theorem 2 *Let (S, A, B, r, q) be a stochastic game, let $\lambda \in (0, 1]$ be a discount factor, and let $\varepsilon > 0$. Let x, z be two stationary strategies of player 1 that satisfy*

$$\left|\frac{x_s(a)}{z_s(a)} - 1\right| \leq \varepsilon, \quad \forall s \in S, a \in A \tag{7}$$

(with the convention $\frac{0}{0} = 1$). Then for every stationary strategy y of player 2,

$$\left|\frac{\gamma_s^\lambda(x, y)}{\gamma_s^\lambda(z, y)} - 1\right| \leq |S|\varepsilon.$$

As a corollary we deduce that every perturbation of a λ-discounted optimal strategy is a λ-discounted $|S|\varepsilon$-optimal strategy, for every discount factor.

Corollary 1 *Let (S, A, B, r, q) be a stochastic game, let $\lambda \in (0, 1)$ be a discount factor, and let $\varepsilon > 0$. Let x be a λ-discounted optimal stationary strategy of player 1, and let z be a strategy that satisfies (7). Then z is a λ-discounted $|S|\varepsilon$-optimal strategy of player 1.*

3.3 Puiseux Functions

A key result that we use in our construction is that the function that assigns to every discount factor the discounted value at a given state has a representation as a Taylor series in a fractional power of λ in a neighborhood of 0 (Bewley and Kohlberg, 1976). Such a function is called a *Puiseux function.* In the sequel we will not use any property of Puiseux functions, except that such a representation exists.

Formally, a function $\lambda \mapsto f(\lambda)$ is a *Puiseux function* if there is $\lambda_0 \in (0, 1)$, a natural number $M \in \mathbf{N}$, an integer $K \in \mathbf{Z}$, and real numbers $(a_k)_{k=K}^{\infty}$ such that

$$f(\lambda) = \sum_{k=K}^{\infty} a_k \lambda^{k/M}, \quad \forall \lambda \in (0, \lambda_0].$$

If f is not identically zero, we can assume w.l.o.g. that $a_K \neq 0$. In this case we call $\frac{K}{M}$ the *leading exponent* of f, and a_K the *leading coefficient.* If f is a Puiseux function, then the limit $\lim_{\lambda \to 0} f(\lambda)$ exists. Note that $\lim_{\lambda \to 0} f(\lambda) = 0$ if and only if $K > 0$, and $\lim_{\lambda \to 0} f(\lambda) \in \mathbf{R}$ if and only if $K \geq 0$.

By Bewley and Kohlberg (1976) the function $\lambda \mapsto v_s^\lambda$ is a Puiseux function, for each initial state $s \in S$. Moreover, there is a function $\lambda \mapsto x^\lambda$ that assigns a λ-discounted optimal strategy x^λ to each discount factor λ, such that for every $s \in S$ and every $a \in A$, the function $\lambda \mapsto x_s^\lambda(a)$ is a Puiseux function.

By Corollary 1, for our purposes it will be sufficient to consider *only* the leading term of the Puiseux expansion of the functions $\lambda \mapsto x_s^\lambda(a)$. This makes the computation of an optimal strategy feasible.

3.4 Quantifier Elimination

The field of the real numbers $\mathbf{R}$, supplemented with the usual order $>$, is a real closed field. An *atomic formula* over $\mathbf{R}$ is an expression of the form $(p > 0)$ or $(p = 0)$, where p is a multi-variate polynomial over $\mathbf{R}$. The *set of all formulae* is the smallest set that contains all the atomic formulae, and is closed under conjunction, disjunction and negation (if ϕ_1 and ϕ_2 are formulae, so are $\phi_1 \wedge \phi_2$, $\phi_1 \vee \phi_2$ and $\neg\phi_1$), and under the existential and universal quantifiers (if ϕ is a formula, and x is a variable, then $\exists x(\phi)$ and $\forall x(\phi)$ are formulae; in this case we say that ϕ is the *scope* of x). A variable x is *free* if it is not in the scope of a quantifier $\exists x(\phi)$ or $\forall x(\phi)$. A *sentence* is a formula without free variables. With every sentence we attach its *truth value* over the field $\mathbf{R}$.

Tarski (1951) provided an algorithm that solves every formula in finite time; that is, the input of the algorithm is a sentence, and the output is “True” if the sentence is true, and “False” otherwise. Basu (1999, see also Basu, Pollack and Roy, 2003) provided a more efficient algorithm to determine the truth value of a sentence.

We are now going to present several formulae and sentences that will be used in our algorithm.

3.4.1 Stationary strategies

The formula $\phi_1(x)$ below expresses the property that x is a stationary strategy of player 1. That is, $\phi_1(x)$ is true if and only if x is a stationary strategy for player 1.

$$\phi_1(x) = (\wedge_{s,a} x_s(a) \geq 0) \wedge \left(\wedge_s \sum_{a \in A} x_s(a) = 1 \right).$$

The formula $\phi_2(y)$ below expresses the property that y is a stationary strategy of player 2.

$$\phi_2(y) = (\wedge_{s,b} y_s(b) \geq 0) \wedge \left(\wedge_s \sum_{b \in B} y_s(b) = 1 \right).$$

3.4.2 Optimal discounted stationary strategies

The formula $\phi_3(x, y, v, \lambda)$ below expresses the property that v is the λ-discounted value and x, y are λ-discounted optimal strategies (see Shapley, 1953).

$$\begin{aligned}
&\phi_3(x, y, v, \lambda) = \phi_1(x) \wedge \phi_2(y) \wedge (0 < \lambda < 1) \wedge \\
&\left(\wedge_{s,a} \left(\sum_b y_s(b) \left(\lambda r(s, a, b) + (1 - \lambda) \sum_{t \in S} q(t \mid s, a, b) v(t) \right) \leq v(s) \right) \right) \wedge \\
&\left(\wedge_{s,b} \left(\sum_a x_s(a) \left(\lambda r(s, a, b) + (1 - \lambda) \sum_{t \in S} q(t \mid s, a, b) v(t) \right) \geq v(s) \right) \right).
\end{aligned}$$

The formulas ϕ_1, ϕ_2 and ϕ_3 are the only formulas used in Chatterjee, Majumdar and Henzinger (2008). The computation of ε-optimal strategies require more complex formulas, that we now introduce.

3.4.3 Finding leading exponents

We now provide few sentences that, when true, imply that there are optimal discounted stationary strategies with certain features. The sentences ϕ_4 and ϕ_5 are not used in the sequel, and are only given for expositional purposes.

Given $s \in S$ and $a \in A$, the formula $\phi_4(s, a)$ expresses the property that for λ small enough, there is an optimal stationary strategy in the λ-discounted game that assigns probability 0 to action a at state s.

$$\begin{aligned}
\phi_4(s, a) = \; &\exists \lambda_0 \in (0, 1). \forall \lambda \in (0, \lambda_0). \exists x \in \mathbf{R}^{S \times A}. \exists y \in \mathbf{R}^{S \times B}. \exists v \in \mathbf{R}^S \\
&(\phi_3(x, y, v, \lambda) \wedge (x_s(a) = 0)).
\end{aligned}$$

By the theory of real algebraic sets (see, e.g., Bochnak, Coste and Roy, 1998, or Mertens, Sorin and Zamir, 2004, ch. VII), if the sentence $\phi_4(s, a)$ is true, then there is a Puiseux function $\lambda \mapsto x^\lambda$ such that (i) x^λ is a λ-discounted optimal stationary strategy, for every discount factor sufficiently small, and (ii) $x_s^\lambda(a) = 0$ for all λ in a neighborhood of zero.

Given $s \in S, a \in A$, a nonnegative integer K and a natural number M, the sentence $\phi_5(s, a, K, M)$ expresses the property that there is a Puiseux function $\lambda \mapsto x^\lambda$ such that (i) x^λ is a λ-discounted optimal stationary strategy, for every discount factor sufficiently small, and (ii) the leading exponent of this function is $\frac{K}{M}$.

$$\begin{aligned}
\phi_5(s, a, K, M) = \; &\forall \varepsilon > 0. \exists c \in (0, +\infty). \exists \lambda_0 \in (0, 1). \forall \lambda \in (0, \lambda_0). \exists \mu \in (0, 1] \\
&\exists x \in \mathbf{R}^{S \times A}. \exists y \in \mathbf{R}^{S \times B}. \exists v \in \mathbf{R}^S \\
&\left(\phi_3(x, y, v, \lambda) \wedge (\mu^M = \lambda^K) \wedge \left(-\varepsilon < \frac{x_s(a)}{\mu} - c < \varepsilon \right) \right).
\end{aligned}$$

We now extend ϕ_4 and ϕ_5 to collections of pairs (state,action). Let D, E be two disjoint subsets of $S \times A$, and for every $(s', a') \in D$ let $K_{s',a'} \in \mathbf{N} \cup \{0\}$ be a nonnegative integer and $M_{s',a'} \in \mathbf{N}$ be a natural number. The formula $\phi_6(D, E, (K_{s',a'}, M_{s',a'})_{(s',a') \in D})$ below expresses the property that there is a Puiseux function $\lambda \mapsto x^\lambda$ such that

(i) x^λ is a λ-discounted optimal stationary strategy, for every discount factor sufficiently small,
(ii) the leading exponent of $\lambda \mapsto x^\lambda_{s'}(a')$ is $\frac{K_{s',a'}}{M_{s',a'}}$ for every $(s', a') \in D$, and
(iii) $x^\lambda_{s'}(a') = 0$ for all λ in a neighborhood of zero and every $(s', a') \in E$.

$$\begin{aligned}
&\phi_6(D, E, (K_{s',a'}, M_{s',a'})_{(s',a') \in D}) = \forall \varepsilon > 0 . \exists (c_{s',a'}) \in (0, +\infty)^D \\
&\quad \exists \lambda_0 \in (0,1) . \forall \lambda \in (0, \lambda_0) . \exists x \in \mathbf{R}^{S \times A} . \exists y \in \mathbf{R}^{S \times B} . \exists v \in \mathbf{R}^S \\
&\quad \big(\phi_3(x, y, v, \lambda) \wedge \big(\wedge_{(s',a') \in E} x_{s'}(a') = 0\big) \wedge \\
&\quad \left(\wedge_{(s',a') \in D} \left(\exists \mu \in (0,1] \left((\mu^{M_{s',a'}} = \lambda^{K_{s',a'}}) \wedge \left(-\varepsilon < \frac{x_{s'}(a')}{\mu} - c_{s',a'} < \varepsilon \right)\right)\right)\right) .
\end{aligned}$$

3.4.4 Approximating leading coefficients

In the algorithm we will allocate recursively every pair $(s, a) \in S \times A$ to either D or E. Whenever (s, a) is allocated to D, we will also compute the leading exponent $\frac{K_{s,a}}{M_{s,a}}$ of $x^\lambda_s(a)$. This will be achieved through a repeated computation of the truth value of ϕ_6. We will then find recursively the leading coefficients of the functions $(x^\lambda_s(a))_{(s,a) \in D}$. We now introduce the formula ϕ_7 that is used for that purpose.

Let a positive number ε, let a partition (D, E) of $S \times A$, and let a subset $D' \subseteq D$ be given. In addition, for each $(s', a') \in D$, let a nonnegative integer $K_{s',a'}$ and a natural number $M_{s',a'}$ be given, and for each $(s', a') \in D'$, let a positive number $c_{s',a'}$ be given.

The formula $\phi_7(\varepsilon, D, E, D', (K_{s',a'}, M_{s',a'})_{(s',a') \in D}, (c_{s',a'})_{(s',a') \in D'})$ below expresses the property that there is a Puiseux function $\lambda \mapsto x^\lambda$ such that

(i) x^λ is a λ-discounted optimal stationary strategy, for every discount factor sufficiently small,
(ii) for every $(s', a') \in D$, the leading exponent of $\lambda \mapsto x^\lambda_{s'}(a')$ is $\frac{K_{s',a'}}{M_{s',a'}}$,
(iii) for every $(s', a') \in E$ and λ close to zero, $x^\lambda_{s'}(a') = 0$, and
(iv) for every $(s', a') \in D'$, the leading coefficient of $\lambda \mapsto x^\lambda_{s'}(a')$ is within ε of $c_{s',a'}$.

$$\begin{aligned}&\phi_7(\varepsilon, D, E, D', (c_{s',a'})_{(s',a')\in D'}, (K_{s',a'}, M_{s',a'})_{(s',a')\in D}) =\\ &\exists \lambda_0 \in (0,1).\forall \lambda \in (0,\lambda_0).\exists x \in \mathbf{R}^{S\times A}.\exists y \in \mathbf{R}^{S\times B}.\exists v \in \mathbf{R}^S\\ &\big(\phi_1(x) \wedge \phi_2(y) \wedge \phi_3(x,y,v,\lambda) \wedge \phi_6(D,E,(K_{s',a'}, M_{s',a'})_{(s',a')\in D})\\ &\wedge \left(\wedge_{(s',a')\in D'} \left(\exists \mu \in (0,1) \left(\mu^{K_{s',a'}} = \lambda^{M_{s',a'}}\right) \wedge \left(-\varepsilon < \frac{x_{s'}(a')}{\mu} - c_{s',a'} < \varepsilon\right)\right)\right)\right).\end{aligned}$$

The algorithm will use one additional sentence ϕ_8, whose definition is postponed to the next section.

4 The Algorithm

4.1 An overview

We first describe the main ideas of the algorithm. As mentioned before, there is a Puiseux function $\lambda \mapsto x^\lambda$ that assigns an optimal λ-discounted stationary strategy for every discount factor λ. In particular, one has the representation

$$x^\lambda_s(a) = \sum_{k=K}^{\infty} a_k \lambda^{k/M}. \tag{8}$$

Using sentences over a real closed field we will find the leading exponent $\frac{K_{s,a}}{M_{s,a}}$ and the leading coefficient $c_{s,a}$ in the representation (8), for every $(s,a) \in S \times A$. We then define

$$z^\lambda_s(a) := \frac{c_{s,a}\lambda^{K_{s,a}/M_{s,a}}}{\sum_{s',a'} \lambda^{K_{s,a}/M_{s,a}}}.$$

By Corollary 1 it follows that z^λ is an ε-optimal discounted stationary strategy, provided λ is sufficiently small. When player 1 follows z^λ, the optimization problem faced by player 2 reduces to a Markov decision problem, and so the highest payoff that z^λ guarantees, $(w^\lambda_s)_{s\in S}$, can be found using a linear program with coefficients that are rational functions of λ. Thus, Eq. (2) is satisfied w.r.t. z^λ and w^λ that were just defined, and to complete the construction it is left to find u_* that satisfies Eq. (3)-(5).

The algorithm is divided into three phases as follows. All computations below are done once, at the beginning of the game. At each stage, a small number of additional arithmetic computations are needed to determine the mixed move used in that stage.

Phase 1: For every $(s,a) \in S \times A$, we find the leading exponent of the Puiseux expansion of some discounted optimal strategy, by enumerating over the possible values of the leading exponent.

Phase 2: We approximate the leading coefficients of the Puiseux expansion of some discounted optimal strategy x^λ. These leading exponents and coefficients are used to define an auxiliary strategy z^λ.

Phase 3: By solving a linear program we determine the highest quantity w^λ that is guaranteed by z^λ in the λ-discounted game. Next, we determine the function $\lambda(u)$ and the constant u_* that are used in Mertens and Neyman's (1981) construction, and conclude that the resulting strategy is uniform $C\varepsilon$-optimal, for $C = |S| + 5$.

A uniform $C\varepsilon$-optimal strategy for player 2 can be calculated analogously.

4.2 Detailed presentation

We now describe in detail the three phases of the algorithm.

Phase 1: Finding the leading exponents of a Puiseux expansion of discounted optimal strategies.

To find the leading exponent, the algorithm divides the set of pairs (state, action) into two disjoint sets D and E. Intuitively, the set E will contain all pairs (s, a) such that the action a is *not* played in the state s by an optimal strategy, while the set D will contain the remaining pairs (s, a).

The algorithm does this as follows. Suppose that we have two disjoint subsets of pairs D and E such that, for every discount factor sufficiently small, there exists an optimal strategy that does not play any action pair in E. By taking a pair (s, a) that is not in $D \cup E$, and by using the formula ϕ_6, we can determine whether for every discount factor sufficiently small there exists an optimal strategy that does not play any action pair in $E \cup \{(s, a)\}$. If this is true, we add (s, a) to E. Otherwise, we add it to D, and using the formula ϕ_6 we find the leading coefficient of the Puiseux expansion of $\lambda \mapsto x_s^\lambda(a)$.

Step 1 : Set $D = E = \emptyset$.

Step 2 : Choose a pair $(s, a) \notin D \cup E$. If there is none, continue to **Phase 2**. Otherwise, continue to Step 3.

Step 3 : Determine the truth value of the sentence

$$\phi_6(D, E \cup \{(s, a)\}, (K_{s',a'}, M_{s',a'})_{(s',a') \in D}).$$

If it is true, add (s, a) to E, and go to Step 2. If it is false, continue to Step 4.

Step 4 : Set $M = 1$.

Step 5 : For every $K = 0, 1, 2 \ldots, M$, set $K_{s,a} = K$ and $M_{s,a} = M$, and determine the truth value of the sentence

$$\phi_6(D \cup (s, a), E, (K_{s',a'}, M_{s',a'})_{(s',a') \in D \cup \{(s,a)\}}).$$

Step 6 : If there is $K_0 \in \{0, 1, 2, \ldots, M\}$ for which the sentence is true, set $K_{s,a} = K_0$, $M_{s,a} = M$, add (s, a) to D, and go to Step 2.

Step 7 : If there is no such K_0, increase M by 1, and go to Step 5.

Since there is a function $\lambda \mapsto x^\lambda$ that assigns to each discount factor a stationary λ-discounted optimal strategy, this stage is bound to stop in finite time. An upper bound on $M_{s,a}$ can be calculated by the results of Benedetti and Risler (1990).

Phase 2: Approximating the leading coefficients of a Puiseux expansion of discounted optimal strategies.

In the second phase we approximate up to ε, for every $(s,a) \in D$, the leading coefficient of the Puiseux function $\lambda \mapsto x_s^\lambda(a)$. This is done by enumerating on values in the set $\{l\varepsilon, l \in \mathbf{Z}\}$, starting at $l = 1$. For each $(s,a) \in S \times A$ and every l, one should calculate the truth value of the sentence ϕ_7. Denote by $(c_{s,a})_{(s,a)\in D}$ the output of this phase.

Step 8 : Set $D' = \emptyset$.
Step 9 : Choose a pair $(s,a) \in D \cup D'$. If there is none, continue to **Phase 3**. Otherwise, continue to Step 10.
Step 10 : Set $l = 1$.
Step 11 : Set $c_{s,a} = l\varepsilon$.
Step 12 : Determine the truth value of the sentence $\phi_7(\varepsilon, D, E, D', (K_{s',a'}, M_{s',a'})_{(s',a')\in D}, (c_{s',a'})_{(s',a')\in D'})$.
Step 13 : If the truth value is true, add (s,a) to D' and go to Step 9. If it is false and $l > 0$, multiply l be -1. If it is false and $l < 0$, multiply l be -1, increase the resulting l by 1, and go to Step 11.

Since for some $l \in \mathbf{Z}$ the sentence ϕ_7 will hold, this phase is finite. However, we do not know of an upper bound to the coefficients $(c_{s,a})_{(s,a)\in D}$, as a function of the date of the game.[3] Consequently, we have no complexity bound for this phase.

Phase 3: The functions $(w_s^\lambda)_{s\in S}$, $\lambda(u)$, and u_* are determined in turn, in steps.

Define for every $\lambda > 0$ a stationary strategy z^λ as follows:

$$z_s^\lambda(a) = \frac{c_{s,a}\lambda^{K_{s,a}/M_{s,a}}}{\sum_{a':(s,a')\in D} c_{s,a'}\lambda^{K_{s,a'}/M_{s,a'}}} \text{ if } (s,a) \in D,$$

and $z_s^\lambda(a) = 0$ if $(s,a) \in E$. That is, the strategy z^λ is defined using only the (approximation of the) leading coefficients and the leading exponents of x^λ. In particular, $\lim_{\lambda\to 0} \frac{z_s^\lambda(a)}{x_s^\lambda(a)} = 1$ for every $(s,a) \in S \times A$.[4] By Corollary 1, the strategy z^λ is λ-discounted $|S|\varepsilon$-optimal for every λ sufficiently small. Denote

$$w_s^\lambda := \inf_y \gamma_s^\lambda(z^\lambda, y).$$

[3] This issue is further discussed in the next section.

[4] By convention, $\frac{0}{0} = 1$.

This is the amount that is guaranteed by the strategy z^λ in the λ-discounted game with initial state s. By Corollary 1, $w_s^\lambda \geq v_s^\lambda - |S|\varepsilon$ for λ sufficiently small. By taking the limit $\lambda \to 0$ we obtain $w_s^0 \geq v_s^0 - |S|\varepsilon$. Since w^λ is the minimal payoff that z^λ guarantees, it follows that (2) holds.

The quantity w_s^λ is the value of the Markov decision problem with initial state s, in which player 1 uses the stationary strategy z^λ and player 2 minimizes the λ-discounted payoff. Therefore the function $\lambda \mapsto w^\lambda$ is the solution of a linear program with coefficients that are polynomial functions of $(z_s^\lambda(a))_{s\in S, a\in S}$, see Eaves and Rothblum (1985) or Altman et al. (1999). Since $z_s^\lambda(a)$ is a rational function of λ, the coefficients of this linear program all lie in the ordered field of rational functions of λ. Therefore, the determination of the functions $\lambda \mapsto w_s^\lambda$, $s \in S$, can be done by solving a linear program over a real closed field.

Write the Puiseux expansion of w_s^λ as

$$w_s^\lambda = w_s^0 + w_s \lambda^{\widehat{K}_s/\widehat{M}_s} + \sum_{k=\widehat{K}_s+1}^{\infty} w_s^k \lambda^{k/\widehat{M}_s},$$

where $w_s \lambda^{\widehat{K}_s/\widehat{M}_s}$ is the first non-zero non-constant term, and the fraction $\widehat{K}_s/\widehat{M}_s$ is irreducible. Using the expression of w_s^λ as a rational function, the value of $\widehat{M}_s \in \mathbf{N}$ is readily obtained in finitely many arithmetic steps.

Mertens and Neyman (1981) show that, setting $\widehat{M}$ to be the g.c.d. of $(\widehat{M}_s)_s$, the function $\lambda(u) := \frac{1}{u^{\widehat{M}+1}}$ satisfies (4) for u sufficiently large, and it also satisfies (3). Finally we have to choose u_* to be sufficiently large. Since $\int_{u_*}^{\infty} u^{-\widehat{M}-1} du = \frac{u_*^{-\widehat{M}}}{\widehat{M}}$, requirement (5) is satisfied once $u_* \geq \frac{1}{\widehat{M}\delta}$. To ensure that u_* is sufficiently large so that (4) is satisfied for every $u \geq U_0$, we need the following sentence to be true:

$$\begin{aligned}\phi_8(s, u_*) = \forall u \geq u_*. \exists \mu > 0. \exists \nu > 0 \\ \Big((\mu^{1+\widehat{M}} = u) \wedge (\nu^{1+\widehat{M}} = u+3) \wedge (|w_s^\mu - w_s^\nu| < 4\mu) \Big).\end{aligned}$$

We find u_* that satisfies this condition for all s, by calculating the truth value of $\phi_8(s, u_*)$ successively for $u_* = l$, $l \geq \frac{1}{\widehat{M}\delta}$.

To summarize, **Phase 3** consists of the following three steps:

- Compute the rational function w^λ by solving a linear program with coefficients in an ordered field.
- Compute $\widehat{M}$.
- Find a suitable value for u_*.

The strategy of Mertens and Neyman (1981), relative to (z^λ, w^λ), guarantees $w_{s_1}^0 - 5\varepsilon \geq v_{s_1}^0 - (|S| + 5)\varepsilon$. Thus, at every stage of the game one has to compute u_n by (6), $\lambda(u_n)$ and $z_{s_n}^{\lambda_n}$.

4.3 Running Time and Space Requirements

To bound the running time of the algorithm, as well as the amount of space required by the computations, one needs to bound (1) the time and space required to determine the truth value of a sentence, (2) the denominator of the leading exponent of the Puiseux expansion of the value function, (3) the leading coefficient $(c_{s,a})_{(s,a)\in S\times A}$ of the Puiseux functions $\lambda \mapsto x_s^\lambda(a)$, for every $s, a \in S \times A$, and (4) the parameter u_*. A bound for (1) can be calculated using Basu (1999, Theorem 2; see also Chatterjee, Majumdar and Henzinger (2008, Theorem 2)). However, it depends on the bound for (2): e.g., to determine the truth value of the sentence $\phi_6(D, E, (K_{s,a}, M_{s',a'})_{(s',a')\in D})$ requires $|S|(|A| + |B|)m^{O(|S|^2|A|(|A|+|B|))}$ arithmetic operations, where $m = O(|S|(|A| + |B|)\max_{s,a} M_{s,a})$, and the determination of the truth value of the sentence ϕ_7 requires $|S|(|A|+|B|)m^{O(|S|(|A|+|B|))}$ arithmetic operations, where $m = O(|S|(|A| + |B|)\max_s \widehat{M}_s)$.

A bound for (2) can be derived by the results of Benedetti and Risler (1990).

Unfortunately, we do not know how to bound the quantities $(c_{s,a})_{(s,a)\in S\times A}$. Whereas for every state $s \in S$, the leading coefficient of the value function $\lambda \mapsto v_s^\lambda$ is bounded by the maximal payoff, as the following example shows this is not the case for the quantities $(c_{s,a})_{(s,a)\in S\times A}$.

Example 1

Consider the stochastic game described in Figure 1. Player 1 is the row player, and player 2 is the column player. There are three states, s_*, s_0 and s_1. States s_0 and s_1 are absorbing, with absorbing payoff 0 and 1 respectively; the asterisks that appear in states s_0 and s_1 in Figure 1 indicate that these states are absorbing.[5] State s_* is not absorbing; when at that state, the payoff is 0 whatever the players play, and the transition is given in Figure 1.

	L	R
T	s_*	s_1
B	$[p(s_1), (1-p)(s_*)$	s_0

state s_*

1 *

state s_1

0 *

state s_0

Figure 1

Plainly $v_{s_0}^\lambda = 0$ and $v_{s_1}^\lambda = 1$. In states s_0 and s_1 player 1 has a unique action, and therefore the optimal strategy in these states are trivial. We concentrate on the case that the initial state is s_*.

The value $v_{s_*}^\lambda$ at s_* satisfies the following recursive equation (see Shapley, 1953, and the formula ϕ_3):

$$v^\lambda = \mathrm{val}\left(\begin{array}{c|c} (1-\lambda)v_{s_*}^\lambda & 1 \\ \hline p + (1-p)(1-\lambda)v_{s_*}^\lambda & 0 \end{array}\right). \tag{9}$$

[5] Formally, the action sets of the players are the same in all states. Here we assume that in states s_0 and s_1 all action pairs yield the same payoff and the same transition.

Denote by $x^\lambda_{s_*}$ the probability by which player 1 chooses T at s_* under the stationary λ-discounted optimal strategy. Then

$$v^\lambda_{s_*} = x^\lambda_{s_*} = x^\lambda_{s_*}(1-\lambda)v^\lambda_{s_*} + (1-x^\lambda_{s_*}) + (1-x^\lambda_{s_*})(1-\lambda)(1-p)v^\lambda_{s_*}.$$

The solution of this system of equations is:

$$x^\lambda_{s_*} = \frac{1+\lambda+p-\lambda p \pm \sqrt{(1+\lambda+p-\lambda p)^2 - 4p}}{2p}.$$

Using the Taylor expansion $\sqrt{a+x} = \sqrt{a} + \frac{x}{2\sqrt{a}} + o(x)$, we obtain that

$$x^\lambda_{s_*} = 1 - \frac{1-p}{2p}\lambda + o(\lambda).$$

In other words, the leading coefficient of $x^\lambda_{s_*}$ is $\frac{1-p}{2p}$, which goes to infinity as p goes to 0.

We do not know either how to bound the parameter u_*, as a function of ε, and of the data of the game. We end this discussion by commenting that the proof of Mertens and Neyman (1981) implies that it is enough to approximate the quantity $\lambda(u_n) = \frac{1}{u_n^{\widehat{M}+1}}$ to within $\lambda(u_n)\varepsilon$; in that case Eq. (2) holds for the approximating λ, provided one adds the term $-3\lambda\varepsilon$. This change will imply that the resulting strategy is uniform C+3-optimal (see Neyman, 2003, Lemma 1).

4.4 Example

We illustrate the algorithm using the example provided in the previous section, with $p = 1$. The solution of Eq. (9) is $v^\lambda = \dfrac{1-\sqrt{\lambda}}{1-\lambda}$. As calculated in the previous section, the λ-discounted optimal strategy of player 1 at s_* is given by $\left[\frac{1}{1+\sqrt{\lambda}}(T), \frac{\sqrt{\lambda}}{1+\sqrt{\lambda}}(B)\right]$. Therefore, the optimal strategy of player 1 assigns probabilities $1-\sqrt{\lambda}+o(\sqrt{\lambda})$ to T and $1-v^\lambda = \sqrt{\lambda}+o(\sqrt{\lambda})$ to B.

When applied to player 1, the algorithm will yield the following results.

Phase 1 will consider each of the two actions T and B in turn. Since both actions are played with positive probability by the unique optimal strategy, the algorithm assigns both of them to the set D. In addition, $K_T = 0$ and $M_T = 1$, while $K_B = 1$ and $M_B = 2$, since the leading exponents of the optimal strategy are 0 for T and 1/2 for B.

Phase 2 executes repeatedly an algorithm that computes the truth value of ϕ_7 for various values of c_B and c_T, using the values of K_T, M_T, K_B, M_B obtained in Phase 1. It yields values c_B and c_T such that $|c_B - 1| < \varepsilon$ and $|c_T - 1| < \varepsilon$. For knife-edge values of ε, the outcome of Phase 2 may depend on the order of enumeration of the possible values.

Phase 3: The stationary strategy z^λ assigns probabilities

$$z^\lambda(T) = \frac{c_T}{c_T + c_B\sqrt{\lambda}},$$

and

$$z^\lambda(B) = \frac{c_B\sqrt{\lambda}}{c_T + c_B\sqrt{\lambda}}$$

to the actions T and B. When facing z^λ, player 2 gets an expected payoff of $z^\lambda(T)$ if he plays R in the first round of the game, and an expected payoff which does not exceed $z^\lambda(B) + (1-\lambda)z^\lambda(T)w^\lambda$ if he plays L in the first round. Thus, w^λ is the highest number which satisfies both inequalities

$$\begin{aligned} w^\lambda &\leq z^\lambda(T) \\ w^\lambda &\leq z^\lambda(B) + (1-\lambda)z^\lambda(T)w^\lambda \end{aligned}$$

This yields

$$w^\lambda = \min\left\{ z^\lambda(T), \frac{z^\lambda(B)}{1-(1-\lambda)z^\lambda(T)} \right\} = \frac{\min\{c_T, c_B\}}{c_T + c_B\sqrt{\lambda}}.$$

Thus, $w^\lambda = \min\left\{1, \frac{c_B}{c_T}\right\}\left(1 - \frac{c_B}{c_T}\sqrt{\lambda}\right) + o(\sqrt{\lambda})$, so that $\hat{K} = 1$, $\hat{M} = 2$. Consequently, $\lambda(u) = \frac{1}{u^3}$. Phase 3 then executes repeatedly an algorithm that finds the truth value of ϕ_8, for increasing values of u_*. This concludes the preprocessing stage of the algorithm.

At each stage of the game, the algorithm first updates u_n according to (6). As long as no absorbing state has been reached the stage payoff is 0, hence

$$u_{n+1} = \max\left\{ u_*, u_n - \frac{\min\{c_T, c_B\}}{c_T + c_B\sqrt{\lambda(u_n)}} + 2\varepsilon \right\}.$$

The algorithm finally computes $\lambda(u_{n+1)}) = 1/u_{n+1}^3$, and the probabilities $z^{\lambda(u_{n+1})}(T)$ and $z^{\lambda(u_{n+1})}(B)$ assigned to the two actions.

5 Extensions

Several existence results of uniform equilibria in multi-player stochastic games use the vanishing discount approach: one considers a sequence of stationary discounted equilibria as the discount factor goes to 0, and, using the sequence, calculates a uniform equilibrium. This method was used, among others, for two-player non-zero-sum absorbing games (Vrieze and Thuijsman, 1989), two-player non-zero-sum games (Vieille, 2000a,b), three-player absorbing games (Solan, 1999), normal-form correlated equilibrium in multi-player absorbing games (Solan and Vohra, 2001), and extensive-form correlated equilibrium in multi-player stochastic games (Solan and Vieille, 2002). Our approach can be used in these cases to calculate a uniform ε-equilibrium/correlated ε-equilibrium. We omit the details, as the presentation of such algorithms requires intimate knowledge of the various constructions.

Acknowledgements This research was supported by a grant from the Ministry of Science and Technology, Israel, and the Ministry of Research, France. Nicolas Vieille also gratefully acknowledges the financial support of the Fondation HEC. We thank Bernhard von Stengel and three anonymous referees for their comments on an earlier version of the paper.

References

1. Altman E., Avrachenkiv K.E. and Filar J.A. (1999) Asymptotic Linear Programming and Policy Improvement for Singularly Perturbed Markov Decision Processes. *Math. Meth. Oper. Res.*, **49**, 97-109.
2. Basu S. (1999) New Results on Quantifier Elimination over Real-Closed Fields and Applications to Constraint Databases. *Journal ACM*, **46**, 537-555.
3. Basu S., Pollack R. and Roy M.F. (2003) Algorithms in Real Algebraic Geometry. Springer.
4. Benedetti R. and Risler J.J. (1990) Real Algebraic and Semi-Algebraic Sets. Hermann.
5. Bewley T. and Kohlberg E. (1976) The Asymptotic Theory of Stochastic Games, *Math. Oper. Res.*, **1**, 197-208
6. Blackwell D. and Ferguson T.S. (1968) The Big Match, *The Annals of Math. Stat.*, **39**, 159-163.
7. Bochnak J., Coste M. and Roy M.-F.(1998) Real Algebraic Geometry. Springer.
8. Chatterjee K., Majumdar R. and Henzinger T.A. (2008) Stochastic Limit-Average Games are in EXPTIME. *Int. J. Game Theory*, **37**, 219-234.
9. Eaves B.C. and Rothblum U.G. (1989) A Theory on Extending Algorithms for Parametric Problems. *Mathematics of Operations Research*, **14**, 502-533.
10. Filar J.A. and Vrieze K. (1996) Competitive Markov Decision Processes. Springer.
11. Fink A.M. (1964) Equilibrium in a Stochastic n-Person Game, *J. Sci. Hiroshima Univ.*, **28**, 89-93
12. Flesch J., Thuijsman F. and Vrieze K. (2007) Stochastic Games with Additive Transitions. *European J. Oper. Res.*, forthcoming.
13. Flesch J., Schoenmakers G. and Vrieze K. (2008) Stochastic Games on a Product State Space. *Math. Oper. Research*, **33**, 403-420.
14. Fortnow L. and Kimmel P. (1998) Beating a Finite Automaton in the Big Match. In Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge, pages 225-234. Morgan Kaufmann, San Francisco.
15. Herings J.J.P and Peeters R.J.A.P (2004) Stationary Equilibria in Stochastic Games: Structure, Selection, and Computation. *Journal of Economic Theory*, **118**, 32-60.
16. Mertens J.F. and Neyman A. (1981) Stochastic Games. *Int. J. Game Th.*, **10**, 53-66.
17. Mertens J.F., Sorin S. and Zamir S. (1994) Repeated Games, CORE Discussion Papers 9420-9422
18. Neyman A. (2003) Stochastic Games: Existence of the Minmax. In Stochastic Games and Applications, Neyman A. and Sorin S. (eds.), NATO Science Series. Kluwer Academic Publishers.
19. Parthasarathy T. and Raghavan T.E.S. (1981) An Orderfield Property for Stochastic Games when one Player Controls Transition Probabilities. *J. Optim. Theory Appl.*, **33**, 375-392.
20. Shapley L.S. (1953) Stochastic Games. *Proc. Nat. Acad. Sci. U.S.A.*, **39**, 1095-1100.
21. Solan E. (1999) Three-Person Absorbing Games, *Math. Oper. Res.*, **24**, 669-698
22. Solan E. (2003) Continuity of the Value of Competitive Markov Decision Processes. *J. Theoret. Probab.* **16**, 831–845.
23. Solan E. and Vieille N. (2001), Quitting Games, *Math. Oper. Res.*, **26**, 265-285

24. Solan E. and Vieille N. (2002), Correlated Equilibrium in Stochastic Games, *Games Econ. Behavior*, **38**, 362-399
25. Solan E. and Vieille N. (2003) Perturbed Markov Chains. *J. Applied Prob.*, **40**, 107–122.
26. Solan E. and Vohra R. (2002), Correlated Equilibrium Payoffs and Public Signalling in Absorbing Games, *Int. J. Game Th.*, **31**, 91-122
27. Tarski A. (1951) A Decision Method for Elementary Algebra and Geometry. University of California Press.
28. Vieille N. (2000a) Equilibrium in 2-Person Stochastic Games I: A Reduction, *Israel J. Math.*, **119**, 55-91
29. Vieille N. (2000b) Equilibrium in 2-Person Stochastic Games II: The Case of Recursive Games, *Israel J. Math.*, **119**, 93-126
30. Vrieze O.J. and Thuijsman F. (1989) On Equilibria in Repeated Games With Absorbing States, *Int. J. Game Th.*, **18**, 293-310
31. Vrieze O.J. and Tijs S.H. (1982) Fictitious Play Applied to Sequences of Games and Discounted Stochastic Games. *Int. J. Game Th.*, **12**, 71-85.