Stochastic Limit-Average Games are in EXPTIME

Krishnendu Chatterjee
Rupak Majumdar
Thomas A. Henzinger

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-143

http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-143.html

November 8, 2006

Acknowledgement

This research was supported in part by the AFOSR MURI grant F49620-00-1-0327 and the NSF ITR grant CCR-0225610.

Stochastic Limit-Average Games are in EXPTIME

Krishnendu Chatterjee[†] Rupak Majumdar[§] Thomas A. Henzinger[†,‡]

[†] EECS, University of California, Berkeley, USA
[§] CS, University of California, Los Angeles, USA
[‡] EPFL, Switzerland
{c_krish,tah}@eecs.berkeley.edu, rupak@cs.ucla.edu

Abstract

The value of a finite-state two-player zero-sum stochastic game with limit-average payoff can be approximated to within ε in time exponential in polynomial in the size of the game times polynomial in logarithmic in $\frac{1}{\varepsilon}$, for all $\varepsilon > 0$.

Keywords. *Stochastic games, Limit-average payoff.*

1 Introduction

A zero-sum stochastic game [14] is a repeated game over a finite state space, played by two-players. Each player has a non-empty set of actions available at every state, and at each round each player chooses an action from the set of available actions at the current state simultaneously with and independent from the other player. The transition function is *probabilistic*, and the next state is given by a probability distribution depending on the current state and the actions chosen by the players. At each round, player 1 gets (and player 2 loses) a reward depending on the current state and the actions chosen by the players, and the players are informed of the history of the play consisting of the sequence of states visited and the actions of the players played so far in the play. A strategy for a player is a recipe to extend the play: given a finite sequence of states and pairs of actions representing the history of the play, a strategy specifies a probability distribution over the set of available actions at the last state of the history. The limiting average reward of a pair of strategies σ and π and a starting state s is defined as

$$v_1(s,\sigma,\pi) = \mathrm{E}_s^{\sigma,\pi} \liminf_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i,\Theta_{i,1},\Theta_{i,2})\right];$$

where X_i is the random variable for the state reached at round i of the game and $\Theta_{i,j}$ is the random variable for the action played by player j at round i of the game, under strategies σ and π, and $r(s,a,b)$ gives the reward at state s for actions a and b. The form of the objective explains the term *limit average*. First, the average is taken with respect to the expected rewards in the first n rounds of the game. Then the objective is defined as the liminf of these averages. A stochastic game with a limit-average objective is called a limit-average game. The fundamental question in stochastic games is the existence of a *value*, that is, whether

$$\sup_\sigma \inf_\pi v_1(s,\sigma,\pi) = \inf_\pi \sup_\sigma v_1(s,\sigma,\pi)$$

Stochastic games were introduced by Shapley [14], where he showed the existence of value in *discounted* games, where the game stops at each round with probability β for some $0 < \beta < 1$. Limit-average games were introduced by Gillette [7], who studied the special cases of perfect information (at each round, at most one player has a choice of moves) and irreducible stochastic games. Existence of value for the perfect information case was proved in [9]. Gillette's paper also introduced a limit-average game called the Big Match, which was solved in [4]. Bewley and Kohlberg [3] then showed how Pusieux series expansions could be used for asymptotic analysis of discounted games. This, and the winning strategy in the Big Match, was used by Mertens and Neyman's result [10] to show the existence of value in limit-average games.

While the *existence* of a value in general limit-average stochastic games has been extensively studied, the *computation* of values has received less attention. In general, it may happen that a game with rational rewards and rational transition probabilities still has an irrational value. Hence, we can only hope to get approximation algorithms that compute the value of a game up to a given approximation ε, for $\varepsilon > 0$. Even the approximation of values is not simple, because in general the games only admit ε-optimal strategies, and strategies may require infinite memory. This precludes, for example, common techniques that enumerate over all (finite) strategies and (having fixed a strategy) solve the resulting Markov decision process using linear programming techniques. Most research has therefore characterized particular subclasses of games for which memoryless optimal strategies exist (a memoryless strategy is independent of the history of the play and depends only on the current state) [12, 8] (see [6] for a survey), and the main algorithmic tool has been value or policy iteration, which can be shown to terminate in exponential number of steps (but much better in practice) for many of these particular classes.

Our main technique is the characterization of values as semi-algebraic quantities [3, 10]. We show that the value of stochastic limit-average games can be expressed as a sentence in the theory of real-closed fields that is polynomial in the size of the game and has a constant number of quantifier alternations. The theory of real-closed fields is decidable in time exponential in the size of the formula and doubly exponential in the quantifier alternation depth [1]; this, together with binary search on the range of values gives an algorithm exponential in polynomial in the size of the game graph to approximate the value to any given $\varepsilon > 0$. Our techniques are simple and combine known results to provide the first complexity bound on the general problem of approximating the value of stochastic games with limit-average objectives. Further, the complexity of this algorithm lie in the same complexity class (EXPTIME) as the best known deterministic algorithm for the special case of perfect information games.

2 Definitions

Probability distributions. For a finite set A, a *probability distribution* on A is a function $\delta : A \to [0,1]$ such that $\sum_{a\in A} \delta(a) = 1$. We denote the set of probability distributions on A by $\mathcal{D}(A)$. Given a distribution $\delta \in \mathcal{D}(A)$, we denote by $\mathrm{Supp}(\delta) = \{x \in A \mid \delta(x) > 0\}$ the *support* of δ.

Definition 1 (Stochastic games) *A (two-player zero-sum)* stochastic game $G = \langle S, \mathsf{A}, \Gamma_1, \Gamma_2, \delta, r\rangle$ *consists of:*

- *A finite state space S.*

- *A finite set* A *of moves or actions.*

- *Two move assignments* $\Gamma_1, \Gamma_2 \colon S \to 2^{\mathsf{A}} \setminus \emptyset$. *For* $i \in \{1, 2\}$, *assignment* Γ_i *associates with each state* $s \in S$ *the non-empty set* $\Gamma_i(s) \subseteq \mathsf{A}$ *of moves available to player* i *at state* s.

- *A probabilistic transition function* $\delta : S \times \mathsf{A} \times \mathsf{A} \to \mathcal{D}(S)$, *that gives the probability* $\delta(s, a_1, a_2)(t)$ *of a transition from* s *to* t *when player 1 plays move* a_1 *and player 2 plays move* a_2, *for all* $s, t \in S$ *and* $a_1 \in \Gamma_1(s)$, $a_2 \in \Gamma_2(s)$.

- *A reward function* $r : S \times \mathsf{A} \times \mathsf{A} \to \mathbb{R}$ *that maps every state and pair of moves to a real valued reward.* ∎

Size of a stochastic game. Given a stochastic game G we use the following notations:

1. $n = |S|$ is the number of states;

2. $|\delta| = \sum_{s \in S} |\Gamma_1(s)| \cdot |\Gamma_2(s)|$ is the number of entries of the transition function;

3. $\mathsf{size}(\delta) = \sum_{t \in S} \sum_{a \in \Gamma_1(s)} \sum_{b \in \Gamma_2(s)} |\delta(s, a, b)(t)|$, where $|\delta(s, a, b)(t)|$ denotes the space to express $\delta(s, a, b)(t)$ in binary bits;

4. $\mathsf{size}(G) = |G| = n + \mathsf{size}(\delta) + \sum_{s \in S} \sum_{a \in \Gamma_1(s)} \sum_{b \in \Gamma_2(s)} |r(s, a, b)|$, where $|r(s, a, b)|$ denotes the space to express $r(s, a, b)$ in binary bits. The specification of a game structure G requires at least $O(\mathsf{size}(G))$-bits.

At every state $s \in S$, player 1 chooses a move $a_1 \in \Gamma_1(s)$, and simultaneously and independently player 2 chooses a move $a_2 \in \Gamma_2(s)$. The game then proceeds to the successor state t with probability $\delta(s, a_1, a_2)(t)$, for all $t \in S$. At the state t, for moves a for player 1 and b for player 2, player 1 wins and player 2 loses a reward of value $r(t, a, b)$. Each player wishes to maximize her own reward. A state s is called an *absorbing state* if for all $a_1 \in \Gamma_1(s)$ and $a_2 \in \Gamma_2(s)$ we have $\delta(s, a_1, a_2)(s) = 1$. In other words, at s for all choice of moves of the players the next state is always s. For all states $s \in S$ and moves $a_1 \in \Gamma_1(s)$ and $a_2 \in \Gamma_2(s)$, we indicate by $\mathrm{Dest}(s, a_1, a_2) = \mathrm{Supp}(\delta(s, a_1, a_2))$ the set of possible successors of s when moves a_1, a_2 are selected.

A *path* or a *play* ω of G is an infinite sequence $\omega = \langle s_0, (a_0, b_0), s_1, (a_1, b_1), s_2, (a_2, b_2), \ldots \rangle$ of states and pairs of moves such that $(a_i, b_i) \in \Gamma_1(s_i) \times \Gamma_2(s_i)$ and $s_{i+1} \in \mathrm{Dest}(s_i, a_i, b_i)$, for all $i \geq 0$. We denote by Ω the set of all paths and by Ω_s the set of all paths starting from state s.

Randomized strategies. A *strategy* for player 1 is a function $\sigma : (S \times \mathsf{A} \times \mathsf{A})^* \cdot S \to \mathcal{D}(A)$ that associates with every prefix of a play, representing the history of the play so far, and the current state a probability distribution from $\mathcal{D}(A)$ such that for all $w \in (S \times \mathsf{A} \times \mathsf{A})^*$ and all $s \in S$ we have $\mathrm{Supp}(\sigma(w \cdot s)) \subseteq \Gamma_1(s)$. Similarly we define strategies π for player 2. We denote by Σ and Π the set of all strategies for player 1 and player 2, respectively.

Once the starting state s and the strategies σ and π for the two players have been chosen, the game is reduced to an ordinary stochastic process. Hence, the probabilities of events are uniquely defined, where an *event* $\mathcal{A} \subseteq \Omega_s$ is a measurable set of paths. For an event $\mathcal{A} \subseteq \Omega_s$, we denote by $\mathrm{Pr}_s^{\sigma,\pi}(\mathcal{A})$ the probability that a path belongs to $\mathcal{A}$ when the game starts from s and the players follows the strategies σ and π. For $i \geq 0$, we also denote by $X_i : \Omega \to S$ the random variable denoting the i-th state along a path, and for $j \in \{ 1, 2 \}$ we denote by $\Theta_{i,j} : \Omega_s \to \mathsf{A}$ the random variable denoting the move of player j in the i-th round of a play. A *valuation* is a mapping $v : S \to \mathbb{R}$, associating a real number $v(s)$ with each state s.

Limit-average payoff. Let σ and π be strategies of player 1 and player 2 respectively. The *limit-average* payoff $v_1(s, \sigma, \pi)$ for player 1 at a state s, for the strategies σ and π is defined as

$$v_1(s, \sigma, \pi) = \mathrm{E}_s^{\sigma,\pi} \liminf_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right];$$

Similarly, for player 2, the payoff $v_2(s, \sigma, \pi)$ is defined as

$$v_2(s, \sigma, \pi) = \mathrm{E}_s^{\sigma,\pi} \limsup_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} -r(X_i, \Theta_{i,1}, \Theta_{i,2})\right].$$

In other words, player 1 wins and player 2 looses the “long-run” average of the rewards of the play. A stochastic game G with limit-average payoff is called a limit-average game.

Given a state $s \in S$ and we are interested in finding the maximal payoff that player 1 can ensure against all strategy for player 2, and the maximal payoff that player 2 can ensure against all strategies for player 1. We call such payoff the *value* of the game G at s for player $i \in \{1, 2\}$. The value for player 1 and player 2 are given by the function $v_1 : S \to \mathbb{R}$ and $v_2 : S \to \mathbb{R}$, defined for all $s \in S$ by

$$v_1(s) = \sup\nolimits_{\sigma\in\Sigma} \inf\nolimits_{\pi\in\Pi} v_1(s, \sigma, \pi) \quad \text{and} \quad v_2(s) = \sup\nolimits_{\pi\in\Pi} \inf\nolimits_{\sigma\in\Sigma} v_2(s, \sigma, \pi).$$

Mertens and Neyman [10] establish the determinacy of stochastic limit-average games.

Theorem 1 ([10]) *For all stochastic limit-average games, for all state s, we have $v_1(s)+v_2(s) = 0$.*

Stronger notion of existence of values [10]. The value for stochastic games exists in a strong sense [10]: $\forall \varepsilon > 0$, $\exists \sigma^* \in \Sigma, \exists \pi^* \in \Pi$ such that $\forall \sigma \in \Sigma$ and $\forall \pi \in \Pi$ the following conditions hold:

1.

$$-\varepsilon + \mathrm{E}_s^{\sigma,\pi^*} \limsup_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right] \leq \mathrm{E}_s^{\sigma^*,\pi} \liminf_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right] + \varepsilon; \quad (1)$$

2. for all $\varepsilon_1 > 0$, there exists $n_0 = n(\varepsilon_1)$ such that for all σ and π, for all $n \geq n_0$ we have

$$-\varepsilon_1 + \mathrm{E}_s^{\sigma,\pi^*} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right] \leq \mathrm{E}_s^{\sigma^*,\pi} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right] + \varepsilon_1. \quad (2)$$

Let $\overline{v_1}(s, \sigma, \pi) = \mathrm{E}_s^{\sigma,\pi} \limsup_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} r(X_i, \Theta_{i,1}, \Theta_{i,2})\right]$, then (1) is equivalent to the following equality

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} v_1(s, \sigma, \pi) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \overline{v_1}(s, \sigma, \pi).$$

3 Theory of Real-closed Fields and Quantifier Elimination

Our main technique is to represent the value of a game as a formula in the theory of real-closed fields. An ordered field H is real-closed if no proper algebraic extension of H is ordered. We denote by $\mathbf{R}$ the real-closed field $(\mathbb{R}, +, \cdot, 0, 1, \leq)$ of the reals with addition and multiplication. An *atomic formula* is an expression of the form $p > 0$ or $p = 0$ where p is a (possibly) multi-variate polynomial with integer coefficients. An *elementary formula* is constructed from atomic formulas by the grammar

$$\varphi ::= a \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \exists x.\varphi \mid \forall x.\varphi,$$

where a is an atomic formula, $\wedge$ denotes conjunction, $\vee$ denotes disjunction, $\neg$ denotes complementation, and $\exists$ and $\forall$ denote existential and universal quantification respectively. From this basic syntax, we derive additional defined expressions $p \geq 0$ (for $p > 0 \vee p = 0$), $p < 0$ (for $\neg(p > 0) \vee \neg(p = 0)$), $p \leq 0$ (for $\neg(p > 0)$), and $p \sim q$ (for $p - q \sim 0$) for polynomials p and q, and $\sim \in \{ =, > \}$ in the usual way. The semantics of elementary formulas are given in a standard way [5]. A variable x is *free* in the formula φ if it is not in the scope of a quantifier $\exists x$ or $\forall x$. An *elementary sentence* is a formula with no free variables. A famous theorem of Tarski states that the theory of real-closed fields is decidable.

Theorem 2 ([15]) *The theory of real-closed fields in the language of ordered fields is decidable.*

Complexity of quantifier elimination. For a formula F we denote by $\mathsf{len}(F)$ the length of F. We also denote by size of F, denoted as $\mathsf{size}(F)$, the length of F plus the space required to specify the coefficients of the formula in binary. We state a result of Basu [1] (Theorem 1 of [1]; also see Theorem 14.16 of [2]) on the complexity of quantifier elimination over the real-closed fields.

Theorem 3 ([1]) *Let $\mathcal{P} = \{ P_1, P_2, \dots, P_s \}$ be a set of s polynomials each of degree at most d and in $k + \ell$ variables with coefficients in real-closed fields. Let*

$$\Phi(Y) = Q_r X^{[r]} Q_{r-1} X^{[r-1]} \dots Q_1 X^{[1]}.F(P_1, P_2, \dots, P_s)$$

be a first-order formula with r alternating quantifiers $Q_i \in \{ \exists, \forall \}$ (i.e., $Q_{i+1} \neq Q_i$), $Y = (Y_1, Y_2, \dots, Y_\ell)$ is a block of ℓ free variables, $X^{[i]}$ is a block of k_i variables with $\sum_{i=1}^{r} k_i = k$, and $F(P_1, P_2, \dots, P_s)$ is a quantifier free boolean formula with atomic predicates of the form $P_i(Y, X^{[r]}, X^{[r-1]}, \dots X^{[1]}) \bowtie 0$, where $\bowtie \in \{ <, >, = \}$. Let D denote the ring generated by the coefficients of the polynomials in $\mathcal{P}$. If every polynomial depends on at most τ variables of Y_j's, then the following assertions hold.

1. *There exists an equivalent quantifer free formula $\Psi(Y)$ of $\Phi(Y)$ with*

$$\mathsf{len}(\Psi(Y)) = s^{\prod_i (k_i+1)} \cdot d^{\ell' \prod_i O(k_i)} \cdot \mathsf{len}(F),$$

where $\ell' = \min\{ \ell, \tau \cdot \prod_i (k_i + 1) \}$.

2. *The degree of the polynomials in $\Psi(Y)$ are bounded by $d^{\prod_i O(k_i)}$ and there is an algorithm to compute $\Psi(Y)$ using*

$$s^{\prod_i (k_i+1)} \cdot d^{\ell' \prod_i O(k_i)} \cdot \mathsf{len}(F)$$

arithmetic operations (multiplication, addition and sign determinations) in D.

3. *If $D = \mathbb{Z}$ (the set of integers) and the bitsizes of the coefficients of the polynomials is bounded by γ, then the bitsize of integers appearing in the intermediate computations and the output is bounded by*
$$\gamma \cdot d^{O(\ell) \prod_i O(k_i)}.$$

Remark 1 *The result of part 3 of Theorem 3 follows from the results of [1] though not explicitly stated as a theorem; for an explicit statement as a theorem see Theorem 14.16 of [2]. Given two integers a and b, let $|a|$ and $|b|$ denote the space to express a and b in binary, respectively. Then the following assertions hold:*

1. *given signs of a and b the sign determination of $a + b$ can be done in $O(|a| + |b|)$ time, i.e., in linear time, and sign determination of $a \cdot b$ can be done in constant time;*

2. *addition of a and b can be done in $O(|a| + |b|)$ time, i.e., in linear time; and*

3. *multiplication of a and b can be done in $O(|a| \cdot |b|)$ time, i.e., in quadratic time.*

It follows from the above observations, along with part 2 and part 3 of Theorem 3 that if $D = \mathbb{Z}$ and $\ell = 0$, then the truth of $\Phi(Y)$ can be determined in time

$$s^{\prod_i O(k_i+1)} \cdot d^{\prod_i O(k_i)} \cdot \mathsf{len}(F) \cdot \gamma^2, \tag{3}$$

i.e., there is an algorithm to determine the truth of $\Phi(Y)$ in time $s^{\prod_i O(k_i+1)} \cdot d^{\prod_i O(k_i)} \cdot \mathsf{len}(F) \cdot \gamma^2$; (also see remark 14.17 of [2]).

4 Computation of Values in Stochastic Games

The values in stochastic limit-average games can be irrational even if the rewards and the transition probability function only take rational values [13]. Hence, we can algorithmically only approximate the value to within an ε. To approximate the values of stochastic games with limit-average objectives we restrict our attention to stochastic *positive* limit-average games. Since there is a simple reduction from all stochastic limit-average games to stochastic positive limit-average games, this is sufficient.

Normalized positive limit-average games. A stochastic limit-average game G is a normalized positive limit-average game if the reward function r maps every state to a non-negative reward between 0 and 1, i.e., $r : S \to [0, 1]$. Given a stochastic limit-average game G, let $c_{\min} = \min_{s \in S} r(s)$ and $c_{\max} = \max_{s \in S} |r(s)|$. Consider the reward function r^+ such that $r^+(s) = \frac{r(s)+|c_{\min}|+\eta}{c_{\max}+|c_{\min}|+\eta}$, with $\eta > 0$. Consider the normalized positive limit-average game G^+ derived from G where the reward function r is replaced by r^+. Let v_1 and v_1^+ be the value functions in the game G and G^+, respectively. It follows easily that $v_1^+(s) = \frac{v_1(s)+|c_{\min}|+\eta}{c_{\max}+|c_{\min\,}|+\eta}$, for all state s. Hence without loss of generality we consider only normalized positive limit-average games to compute the values. Observe that the value function v_1^+ only takes values in the interval $[0, 1]$ for normalized positive limit-average games.

Discounted version of a game. Let G be a normalized positive limit-average game with reward function r. Let $0 < \beta < 1$. A β-discounted version of the game G, denoted G_β, is a game that halts with probability β at each round, and proceeds as game G with probability $1 - \beta$. The process of halting can be interpreted as going to an absorbing state halt, such that $r(\mathsf{halt}, a, b) = 0$, for all $a \in \Gamma_1(\mathsf{halt})$ and for all $b \in \Gamma_2(\mathsf{halt})$. We denote by $v_1^\beta(\cdot)$ the value function of a β-discounted game. It may be noted that for normalized positive limit-average games G, the value function of the corresponding β-discounted game v_1^β is monotonic with respect to β in a neighborhood of 0, i.e., there exists $\beta > 0$ such that for all $\beta_1, \beta_2 \in (0, \beta)$ if $\beta_1 \leq \beta_2$, then $v_1^{\beta_1} \geq v_1^{\beta_2}$.

We assume without loss of generality that the state space of the stochastic game structure is enumerated as natural numbers, $S = \{1, 2, \ldots, n\}$, i.e., the states are numbered from 1 to n.

4.1 Quantifier free sentence for value of stochastic games

We first show how the values of β-discounted stochastic games can be expressed as a quantifier free formula over the theory of real-closed fields. We then extend the result to all stochastic games.

Formula for value of β-discounted games. We first present a formula in the theory of real-closed fields to characterize the values of a β-discounted stochastic game, with $0 < \beta < 1$. Given a valuation $v \in \mathbb{R}^n$, for every pure strategy of player 2, for every state in S, we write a polynomial for player 1 expressing the β-discounted value as a function of a randomized strategy x for player 1 and the value subtracted from the valuation. For a state $i \in S$, $b \in \Gamma_2(i)$, $x \in \mathcal{D}(\Gamma_1(i))$, $v \in \mathbb{R}^n$ and $0 < \beta < 1$, we have

$$u_{(i,b,1)}(x, v, \beta) = \beta \sum_{a \in \Gamma_1(i)} x(a) r(i, a, b) + (1 - \beta) \sum_{a \in \Gamma_1(i)} x(a) \sum_{i' \in S} \delta(i, a, b)(i') v(i') - v(i),$$

is a polynomial indexed by $(i, b, 1)$ (a state $i \in S$, and move $b \in \Gamma_2(i)$ and by player 1). The polynomial $u_{(i,b,1)}$ consists of variables β, $x(a)$ for $a \in \Gamma_1(i)$, $v(i)$ for $i \in S$. Observe that given the β-discounted stochastic game, $r(i, a, b)$ for $a \in \Gamma_1(i)$ and $\delta(i, a, b)(i')$ for $i' \in S$ and $a \in \Gamma_1(i)$ are specified constants and not variables. The coefficients of the polynomials are $r(i, a, b)$ for $a \in \Gamma_1(i)$ and $\delta(i, a, b)$ for $a \in \Gamma_1(i)$. Hence the polynomial has degree 3 and has $1 + |\Gamma_1(i)| + n$ variables. Similarly, for $i \in S$, $a \in \Gamma_1(i)$, $y \in \mathcal{D}(\Gamma_2(i))$, $v \in \mathbb{R}^n$ and $\beta < 1$, we have polynomials for player 2 as

$$u_{(i,a,2)}(y, v, \beta) = \beta \sum_{b \in \Gamma_2(i)} y(b) r(i, a, b) + (1 - \beta) \sum_{b \in \Gamma_2(i)} y(b) \sum_{i' \in S} \delta(i, a, b)(i') v(i') - v(i).$$

The total number of polynomials is $s_1 = \sum_{i \in S}(|\Gamma_1(i) + |\Gamma_2(i)|) = O(|\delta|)$. The formula stating that the value at state 1 is at least α is as follows:

$$\begin{array}{rcll}
\Phi_\beta(\alpha) & = & \exists x_1, \ldots, x_n.\ \exists y_1, \ldots, y_n.\ \exists v(1), \ldots, v(n). & \\
& & \left(\bigwedge_{i \in S} \Big(\sum_{a \in \Gamma_1(i)} x_i(a) - 1 = 0 \Big) \right) & \wedge\ \left(\bigwedge_{i \in S} \Big(\sum_{b \in \Gamma_2(i)} y_i(b) - 1 = 0 \Big) \right) \\
& \wedge & \left(\bigwedge_{i \in S, a \in \Gamma_1(i)} x_i(a) \geq 0 \right) & \wedge\ \left(\bigwedge_{i \in S, b \in \Gamma_2(i)} y_i(b) \geq 0 \right) \\
& \wedge & \left(\bigwedge_{i \in S, b \in \Gamma_2(i)} u_{(i,b,1)}(x, v, \beta) \geq 0 \right) & \wedge\ \left(\bigwedge_{i \in S, a \in \Gamma_1(i)} u_{(i,a,2)}(y, v, \beta) \leq 0 \right) \\
& \wedge & \Big(v(1) - \alpha > 0 \Big). &
\end{array}$$

The correctness of the above formula to specify $v_1^\beta(1) > \alpha$ can be proved from the results of [14].

Value of a game as limit of discounted games. The result of Mertens-Neyman [10] establishes the equivalence of the value of a stochastic limit-average game as the limit of the β-discounted games as β goes to 0. Formally, we have

$$v_1(s) = \lim_{\beta \to 0, 0 < \beta < 1} v_1^{\beta}(s).$$

Sentence for value of stochastic games. From the characterization of the value of a stochastic limit-average game as the limit of the β-discounted games and the monotonicity property of normalized positive limit-average games in a neighborhood of 0 we obtain the following sentence stating that the value at state 1 is at least α:

$$\begin{array}{rcll}
\Phi(\alpha) & = & \multicolumn{2}{l}{\exists \beta_1 .\ \forall \beta .\ \exists x_1, \ldots, x_n .\ \exists y_1, \ldots, y_n .\ \exists v(1), \ldots, v(n).} \\
& & \Big(\bigwedge\limits_{i \in S} \big(\sum\limits_{a \in \Gamma_1(i)} x_i(a) - 1 = 0 \big) \Big) & \wedge\ \Big(\bigwedge\limits_{i \in S} \big(\sum\limits_{b \in \Gamma_2(i)} y_i(b) - 1 = 0 \big) \Big) \\
& \wedge & \Big(\bigwedge\limits_{i \in S, a \in \Gamma_1(i)} x_i(a) \geq 0 \Big) & \wedge\ \Big(\bigwedge\limits_{i \in S, b \in \Gamma_2(i)} y_i(b) \geq 0 \Big) \\
& \wedge & \Big(\beta_1 > 0 \Big) \bigwedge \Big[(\beta \leq 0) \bigvee \quad \Big((\beta_1 - \beta > 0) & \wedge\ \Big(\bigwedge\limits_{i \in S, b \in \Gamma_2(i)} u_{(i,b,1)}(x, v, \beta) \geq 0 \Big) \\
& & & \wedge\ \Big(\bigwedge\limits_{i \in S, a \in \Gamma_1(i)} u_{(i,a,2)}(y, v, \beta) \leq 0 \Big) \Big) \Big] \\
& \wedge & \Big(v(1) - \alpha > 0 \Big). &
\end{array}$$

The total number of polynomials in the above formula in addition to the polynomials s_1 is $s_2 = \sum_{i \in S} (2 + |\Gamma_1(i)| + |\Gamma_2(i)|) + 4$. Hence the total number of polynomials is $O(n + |\delta|) = O(|\delta|)$. The formula $\Phi(\alpha)$ contains no free variables (i.e., each varaible x, y, v and β's are quantified). In the setting of Theorem 3 we obtain the following bounds for $\Phi(\alpha)$:

$$s = O(|\delta|); \qquad k = O(n); \qquad \prod_i (k_i + 1) = O(n); \qquad \ell = 0; \tag{4}$$

$$\tau = 0; \qquad r = O(1); \qquad d = 3; \qquad \ell' = 0; \tag{5}$$

and hence we have

$$s^{\prod_i (k_i+1)} \cdot d^{\ell' \prod_i O(k_i)} = O(|\delta|)^{O(n)} = 2^{O\left(n \cdot \log(|\delta|)\right)}.$$

Also observe that for a stochastic game G, the sum of the length of the polynomials appearing in the formula is $O(|\delta|)$ and the size of the polynomials appearing in the formula is $O(\mathsf{size}(G)) + |\alpha| = O(|G|) + O(|\alpha|)$, where $|\alpha|$ is the space required to express α in binary. The present analysis along with Theorem 3 yields the following theorem.

Theorem 4 *Given a stochastic limit-average game G and a real α, the following assertions hold.*

1. *There is a quantifier free sentence $\Psi(\alpha)$ that specifies $v_1(1) > \alpha$ with*

$$\begin{array}{rcl}
\mathsf{len}\big(\Psi(\alpha)\big) & = & 2^{O\left(n \cdot \log(|\delta|)\right)} \cdot O(|\delta|); \\
\mathsf{size}\big(\Psi(\alpha)\big) & = & 2^{O\left(n \cdot \log(|\delta|)\right)} \cdot \big(O(\mathsf{size}(G)) + O(|\alpha|)\big).
\end{array}$$

2. *There is an algorithm to determine the truth of $\Psi(\alpha)$ using $2^{O\left(n\cdot\log(|\delta|)\right)} \cdot O(|\delta|)$ arithmetic operations.*

4.2 Algorithmic analysis

For algorithmic analysis we consider stochastic games with rational inputs, i.e., stochastic games such that $r(i,a,b)$ and $\delta(i,a,b)$ are rational for all $i \in S$, $a \in \Gamma_1(i)$ and $b \in \Gamma_2(i)$. Given the formula $\Phi(\alpha)$ to specify $v_1(1) > \alpha$ we first reduce it to an equivalent formula $\widehat{\Phi}(\alpha)$ as follows:

- for every rational coefficient $e = \frac{p}{q}$, where $p, q \in \mathbb{N}$, appearing in $\Phi(\alpha)$ we apply the following procedure
  1. introduce a variable z_e,
  2. replace e by z_e in $\Phi(\alpha)$,
  3. add a polynomial $q \cdot z_e - p = 0$, and
  4. existentially quantify z_e in the block of existential quantifiers after quantifying β_1 and β.

Thus we add $O(|\delta|)$ variables and polynomials and increase the degree of the polynomials in $\Phi(\alpha)$ by 1. Also observe that the coefficients in $\widehat{\Phi}(\alpha)$ are integers and hence the ring $\widehat{D}$ generated by the coefficients in $\widehat{\Phi}(\alpha)$ is a subset of $\mathbb{Z}$. Similar to the bounds obtained in (4) and (5), in the setting of Theorem 3 we obtain the following bounds for $\widehat{\Phi}(\alpha)$:

$$\widehat{s} = O(|\delta|); \qquad \widehat{k} = O(n + |\delta|); \qquad \prod_i (\widehat{k}_i + 1) = O(n + |\delta|); \qquad \widehat{\ell} = 0;$$

$$\widehat{\tau} = 0; \qquad \widehat{r} = O(1); \qquad \widehat{d} = 4; \qquad \widehat{\ell'} = 0;$$

and hence we have

$$\widehat{s}^{\prod_i O(\widehat{k}_i+1)} \cdot \widehat{d}^{\prod_i O(\widehat{k}_i)} = O(|\delta|)^{O(n+|\delta|)} = 2^{O\left(|\delta|\cdot\log(n+|\delta|)\right)} = 2^{O\left(|\delta|\cdot\log(|\delta|)\right)}$$

Also observe that the length of the formula $\widehat{\Phi}(\alpha)$ can be bounded by $O(|\delta|)$ and the bitsizes of the coefficients in $\widehat{\Phi}(\alpha)$ can be bounded by $\mathsf{size}(G) + O(|\alpha|) = O(|G| + |\alpha|)$. This along with (3) of Remark 1 yield the following result.

Theorem 5 *Given a stochastic limit-average game G and a real α, there is an algorithm that computes whether $v_1(1) > \alpha$ in time*

$$2^{O\left(|\delta|\cdot\log(|\delta|)\right)} \cdot O(|\delta|) \cdot O(|G|^2 + |\alpha|^2).$$

4.3 Approximating the value

We now present an algorithm that approximates the value within a tolerance of $\varepsilon > 0$. The algorithm (Algorithm 1) is obtained by a binary search technique along with the result of Theorem 5.

Running time of Algorithm 1. In Algorithm 1 we denote by $\Phi(m)$ the formula to specify $v_1(1) > m$ and by Theorem 5 the formula $\Phi(m)$ can be computed in time

$$2^{O\left(|\delta|\cdot\log(|\delta|)\right)} \cdot O(|\delta|) \cdot O(|G|^2 + |m|^2),$$

Algorithm 1 Approximating the value

```
Input: Normalized positive limit-average game $G$,
    and a rational value $\varepsilon$ as tolerance.
Output: An interval $[l, u]$ such that $u - l \leq 2\varepsilon$ and $v_1(1) \in [l, u]$.

1. $l := 0, u := 1, m := 1/2$.
2. repeat for $\lceil \log\left(\frac{1}{\varepsilon}\right) \rceil$ steps
   2.1 if $(\Phi(m))$
      2.1.a $l := m, u := u, m := \frac{l+u}{2}$;
   2.2 else
      2.2.a $l := l, u := m, m := \frac{l+u}{2}$.
3. return $[l, u]$.
```

for the stochastic game G, where $|m|$ is the number of bits required to specify m. In Algorithm 1, the variables l, u, m are initially set to $0, 1$ and $1/2$, respectively and can be expressed in 1-bit. In each iteration of the algorithm, after the division by 2 in Steps 2.1.a and 2.2.a the variables l, u and m can be expressed with one more bit w.r.t. to the previous iteration. Hence l, u and m can always be expressed in $\log\left(\frac{1}{\varepsilon}\right)$ bits. The loop in Step 2 runs for $\log\left(\frac{1}{\varepsilon}\right)$-steps and every iteration can be computed in time $2^{O\left(|\delta|\cdot\log(|\delta|)\right)} \cdot O(|\delta|) \cdot O\left(|G|^2 + \log^2\left(\frac{1}{\varepsilon}\right)\right)$. This gives us the following theorem.

Theorem 6 *Given a normalized positive limit-average game G, the state 1 of G, and a rational $\varepsilon > 0$, Algorithm 1 computes an interval $[l, u]$ such that $v_1(1) \in [l, u]$ and $u - l \leq 2\varepsilon$, in time*

$$2^{O\left(|\delta|\cdot\log(|\delta|)\right)} \cdot O(|\delta|) \cdot O\left(|G|^2 \cdot \log\left(\frac{1}{\varepsilon}\right) + \log^3\left(\frac{1}{\varepsilon}\right)\right).$$

By our reduction to normalized positive limit-average games, this also gives an algorithm for general limit-average games.

Corollary 1 *The value of a stochastic limit-average game G at a state i can be approximated to within $\varepsilon > 0$ in time*

$$2^{O\left(|\delta|\cdot\log(|\delta|)\right)} \cdot O(|\delta|) \cdot O\left(|G|^2 \cdot \log\left(\frac{1}{\varepsilon}\right) + \log^3\left(\frac{1}{\varepsilon}\right)\right).$$

The complexity class EXPTIME. A problem is in the complexity class EXPTIME, if there is an algorithm $\mathcal{A}$ that solves the problem, and there is a polynomial $p(\cdot)$ such that for all inputs I of $|I|$-bits, the running time of the algorithm $\mathcal{A}$ on input I can be bounded by $2^{p(|I|)}$. In case of stochastic games the input is the size of the game G, i.e., $\mathsf{size}(G)$. Hence from Corollary 1 and Theorem 5 we obtain the following result.

Theorem 7 *Given a rational $\varepsilon > 0$ and a rational α the following assertions hold:*

1. *the values of stochastic limit-average games can be computed within ε-precision in EXPTIME;*

2. whether $v_1(i) > \alpha$ for a state i can be decided in EXPTIME.

Unfortunately, the only lower bound on the complexity is PTIME-hardness (polytime hardness). The hardness follows from a simple reduction from alternating reachability. Even for the simpler case of perfect information deterministic games no polynomial time algorithm is known [16], and the best known algorithm for perfect information games is exponential in the size of the game [9]. In case of perfect information stochastic games, deterministic and stationary optimal strategies exist [9]. Since the number of deterministic stationary strategies can be at most exponential in the size of the game, there is an exponential time algorithm to compute the values exactly (not approximation) (also see Chapter by Raghavan in [11]).

Acknowledgements

We are grateful to Prof. Abraham Neyman for his comments on an earlier draft of the paper. His comments helped in improving the presentation of the paper vastly and making the proofs more precise. This research was supported in part by the AFOSR MURI grant F49620-00-1-0327 and the NSF ITR grant CCR-0225610.

References

[1] S. Basu. New results on quantifier elimination over real closed fields and applications to constraint d atabases. *Journal of the ACM*, 46(4):537–555, 1999.

[2] S. Basu, R. Pollack, and M.-F.Roy. *Algorithms in Real Algebraic Geometry.* Springer-Verlag.

[3] T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1, 1976.

[4] D. Blackwell and T.S. Ferguson. The big match. *Annals of Mathematical Statistics*, 39:159–163, 1968.

[5] C.C. Chang and H.J. Keisler. *Model Theory.* North Holland, 3rd edition, 1990.

[6] J. Filar and K. Vrieze. *Competitive Markov Decision Processes.* Springer-Verlag, 1997.

[7] D. Gillete. Stochastic games with zero stop probabilitites. In *Contributions to the Theory of Games III*, pages 179–188. Princeton University Press, 1957.

[8] A.J. Hoffman and R.M. Karp. On nonterminating stochastic games. *Management Sciences*, 12(5):359–370, 1966.

[9] T. A. Liggett and S. A. Lippman. Stochastic games with perfect information and time average payoff. *Siam Review*, 11:604–607, 1969.

[10] J.F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–66, 1981.

[11] A. Neyman and S. Sorin. *Stochastic Games and Applications.* Kluwer Academic Publishers, 2003.

[12] T. Parthasarathy and T.E.S. Raghavan. An orderfield property for stochastic games when one player controls transition probabilities. *Journal of Optimization Theory and Applications*, 33:375–392, 1981.

[13] T.E.S. Raghavan and J.A. Filar. Algorithms for stochastic games — a survey. *ZOR — Methods and Models of Op. Res.*, 35:437–472, 1991.

[14] L.S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.

[15] A. Tarski. *A Decision Method for Elementary Algebra and Geometry.* University of California Press, Berkeley and Los Angeles, 1951.

[16] U. Zwick and M.S. Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158:343–359, 1996.