ORIGINAL PAPER

Stochastic limit-average games are in EXPTIME

Krishnendu Chatterjee · Rupak Majumdar ·
Thomas A. Henzinger

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 a polynomial in the size of the game times polynomial in logarithmic in $1/\varepsilon$, for all $\varepsilon > 0$.

Keywords Stochastic games · Limit-average payoff · Computational complexity

1 Introduction

A zero-sum stochastic game 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 in 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 determined by a probability distribution depending on the current state and the actions chosen by the players. In each round, player 1 gets (and player 2 loses) a reward depending on the current state and the actions chosen by the players. 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

K. Chatterjee (✉) · T. A. Henzinger
EECS, University of California, Berkeley, CA, USA
e-mail: c_krish@eecs.berkeley.edu

R. Majumdar
CS, University of California, Los Angeles, CA, USA
e-mail: rupak@cs.ucla.edu

T. A. Henzinger
EPFL, Lausanne, Switzerland
e-mail: tah@eecs.berkeley.edu

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 *limit-average* for player 1 reward of a pair of strategies σ and π for player 1 and player 2, respectively, and a starting state s is defined as

$$v_1(s, \sigma, \pi) = \mathrm{E}_s^{\sigma, \pi}\left[\lim \inf_{n \to \infty} \frac{1}{n} \cdot \sum_{i=1}^{n} r\left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2}\right)\right],$$

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 starting state s, 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 reward is called a limit-average game. The fundamental question in stochastic games is the existence of a *value*, that is, whether

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

where Σ and Π denote the sets of strategies for player 1 and player 2, respectively.

Stochastic games were introduced by Shapley (1953), who showed the existence of a value in *discounted* games, where the game stops at each round with probability β, for some $0 < \beta < 1$, and the goal of a player is to maximize the expectation of the total sum of the rewards. Limit-average games were introduced by Gillete (1957), who studied the special cases of perfect information (in each round, at most one player has a choice of moves) and irreducible stochastic games. The existence of a value for the perfect information case was proved in Liggett and Lippman (1969). Gillette's paper also introduced a limit-average game called the Big Match, which was solved in Blackwell and Ferguson (1968). Bewley and Kohlberg (1976) then showed how Pusieux series expansions can be used for the asymptotic analysis of discounted games. This, and the winning strategy in the Big Match, was used by Mertens and Neyman (1981) to show the existence of a 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.[1] In general, it may happen that a game with rational rewards and rational transition probabilities still has an irrational value (Raghavan and Filar 1991). Hence, we can only hope to have approximation algorithms that compute the value of a game up to a given approximation ε, for a real $\varepsilon > 0$. Even the approximation of values is not simple, because in general limit-average games only admit η-optimal strategies, for all reals $\eta > 0$, rather than optimal strategies (Blackwell and Ferguson 1968), and the η-optimal strategies of Mertens and Neyman (1981) require infinite memory. This precludes, for

---

[1] In this paper we take the classical view of computation, where an algorithm either answers "Yes" or "No", or outputs a set of rational numbers.

example, common algorithmic techniques that enumerate over certain finite sets of strategies and, having fixed a strategy, solve the resulting Markov decision process using linear programming techniques (Filar and Vrieze 1997). Most research has therefore characterized particular subclasses of games for which *stationary* optimal strategies exist (a stationary strategy is independent of the history of a play and depends only on the current state; Parthasarathy and Raghavan 1981; Hoffman and Karp 1966; see Filar and Vrieze 1997 for a survey), and the main algorithmic tool has been value or policy iteration, which can be shown to terminate in an exponential number of steps (but often behaves better in practice) for many of these particular classes.

In this paper, we characterize the computational complexity of approximating the value of a limit-average game. We show that for any given real $\varepsilon > 0$, the value of a game G at a state can be computed to within ε-precision in time bounded by an exponential in a polynomial in the size of the game G times a polynomial function of $\log 1/\varepsilon$. This shows that approximating the value of limit-average games lies in the computational complexity class EXPTIME (Papadimitriou 1994). Our main technique is the characterization of values as semi-algebraic quantities (Bewley and Kohlberg 1976; Mertens and Neyman 1981). We show that for a real number α, whether the value of a stochastic limit-average game at a state s is strictly greater than α can be expressed as a sentence in the theory of real-closed fields. Moreover, this sentence 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 a formula and doubly exponential in the quantifier alternation depth (Basu 1999). This, together with binary search over the range of values, gives an algorithm exponential in polynomial in the size of the game graph times polynomial in logarithmic in $1/\varepsilon$ to approximate the value, for $\varepsilon > 0$. Our techniques combine several known results to provide the first complexity bound on the general problem of approximating the value of stochastic games with limit-average objectives. It may be noted that the best known deterministic algorithm for the special case of perfect information limit-average games also requires exponential time.

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 $\mathcal{D}(A)$. For 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* $G = \langle S, A, \Gamma_1, \Gamma_2, \delta, r \rangle$ consists of the following components.

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

- A probabilistic transition function $\delta : S \times A \times A \to \mathcal{D}(S)$ that gives the probability $\delta(s, a, b)(t)$ of a transition from s to t when player 1 plays move a and player 2 plays move b, for all $s, t \in S$ and $a \in \Gamma_1(s)$, $b \in \Gamma_2(s)$.
- A reward function $r : S \times A \times A \to \mathbb{R}$ that maps every state and pair of moves to a real-valued reward. $\qquad\square$

The special class of *perfect-information* games can be obtained from stochastic games with the restriction that for all $s \in S$ either $|\Gamma_1(s)| = 1$ or $|\Gamma_2(s)| = 1$, i.e., at every state at most one player can influence the transition. If the transition function δ is deterministic rather than probabilistic then we call the game a *deterministic* game. The class of *rational* stochastic games are the special class of stochastic games such that all rewards and transition probabilities are rational.

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.

Given a rational stochastic game we use the following notations:

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

The specification of a game G requires $O(|G|)$ bits. Given a stochastic game with n states, 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.

At every state $s \in S$, player 1 chooses a move $a \in \Gamma_1(s)$, and simultaneously and independently player 2 chooses a move $b \in \Gamma_2(s)$. The game then proceeds to the successor state t with probability $\delta(s, a, b)(t)$, for all $t \in S$. At the state s, for moves a for player 1 and b for player 2, player 1 wins and player 2 loses a reward of value $r(s, a, b)$.

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 $s_{i+1} \in \mathsf{Supp}(\delta(s_i, a_i, b_i))$, 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 $\sigma : (S \times A \times A)^* \cdot S \to \mathcal{D}(A)$ 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 $w \in (S \times A \times A)^*$ and all $s \in S$, we have $\mathsf{Supp}(\sigma(w \cdot s)) \subseteq \Gamma_1(s)$. Observe that the strategies can be randomized (i.e., not necessarily deterministic) and history-dependent (i.e., not necessarily stationary). Similarly we define strategies π for player 2. We denote by Σ and Π the sets of 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 a stochastic process. Hence, the probabilities of events are uniquely defined, where an *event* $\mathcal{E} \subseteq \Omega_s$ is a measurable set of paths. For an

event $\mathcal{E} \subseteq \Omega_s$, we denote by $\Pr_s^{\sigma,\pi}(\mathcal{E})$ the probability that a path belongs to $\mathcal{E}$ when the game starts from s and the players follow the strategies σ and π. We denote by $\mathrm{E}_s^{\sigma,\pi}[\cdot]$ the associated expectation operator with the probability measure $\Pr_s^{\sigma,\pi}(\cdot)$. For $i \geq 0$, we denote by $\mathsf{X}_i : \Omega \to S$ the random variable denoting the ith state along a path, and for $j \in \{1, 2\}$, we denote by $\Theta_{i,j} : \Omega_s \to A$ the random variable denoting the move of player j in the ith round of a play.

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

$$v_1(s, \sigma, \pi) = \mathrm{E}_s^{\sigma,\pi}\left[\liminf_{N \to \infty} \frac{1}{N} \cdot \sum_{i=1}^{N} r\left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2}\right)\right].$$

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

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

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 stochastic 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 strategies 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 values for player 1 and player 2 are defined for all $s \in S$ by

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

Mertens and Neyman (1981) established the determinacy of stochastic limit-average games.

Theorem 1 Mertens and Neyman (1981) *For all stochastic limit-average games G and for all states s of G, we have $v_1(s) + v_2(s) = 0$.*

Stronger notion of existence of values Mertens and Neyman (1981). The values for stochastic limit-average games exist in a strong sense: for all reals $\varepsilon > 0$, there exist strategies $\sigma^* \in \Sigma, \pi^* \in \Pi$ such that the following conditions hold:

1. for all $\sigma \in \Sigma$ and $\pi \in \Pi$, we have

$$-\varepsilon + \mathrm{E}_s^{\sigma,\pi^*}\left[\limsup_{N \to \infty} \frac{1}{N} \cdot \sum_{i=1}^{N} r\left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2}\right)\right]$$

$$\leq \mathrm{E}_s^{\sigma^*,\pi}\left[\liminf_{N \to \infty} \frac{1}{N} \cdot \sum_{i=1}^{N} r\left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2}\right)\right] + \varepsilon; \tag{1}$$

2. there exists an integer N_0 such that for all $\sigma \in \Sigma$ and $\pi \in \Pi$, for all integers $N \geq N_0$, we have

$$- \varepsilon + \mathrm{E}_s^{\sigma,\pi^*} \left[\frac{1}{N} \cdot \sum_{i=1}^{N} r \left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2} \right) \right] \leq \mathrm{E}_s^{\sigma^*,\pi} \left[\frac{1}{N} \cdot \sum_{i=1}^{N} r \left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2} \right) \right] + \varepsilon; \tag{2}$$

3. there exists $0 < \beta_0 < 1$ such that for all $\sigma \in \Sigma$ and $\pi \in \Pi$, for all $0 < \beta \leq \beta_0$, we have

$$-\varepsilon + \mathrm{E}_s^{\sigma,\pi^*} \left[\beta \cdot \sum_{i=1}^{\infty} (1 - \beta)^{i-1} r \left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2} \right) \right]$$

$$\leq \mathrm{E}_s^{\sigma^*,\pi} \left[\beta \cdot \sum_{i=1}^{\infty} (1 - \beta)^{i-1} r \left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2} \right) \right] + \varepsilon. \tag{3}$$

Let $\overline{v}_1(s, \sigma, \pi) = \mathrm{E}_s^{\sigma,\pi} \left[\limsup_{N \to \infty} 1/N \cdot \sum_{i=1}^{N} r \left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2} \right) \right]$, 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. We denote by **R** the real-closed field $(\mathbb{R}, +, \cdot, 0, 1, \leq)$ of the reals with addition and multiplication. In the sequel we write "real-closed field" to denote the real-closed field **R**. An *atomic formula* is an expression of the form $p < 0$ or $p = 0$, where p is a (possibly) multi-variate polynomial with coefficients in the real-closed field. Coefficients are rationals or symbolic constants (e.g., the symbolic constant e stands for $2.71828\ldots$). We will consider the special case when only rational coefficients of the form q_1/q_2, where q_1, q_2 are integers, are allowed. A *formula* is constructed from atomic formulas by the grammar

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

where a is an atomic formula, $\neg a$ denotes complement of a, $\varphi_1 \wedge \varphi_2$ denotes conjunction of φ_1 and φ_2, $\varphi_1 \vee \varphi_2$ denotes disjunction of φ_1 and φ_2, and $\exists$ and $\forall$ denote existential and universal quantification, respectively. We use the standard abbreviations such as $p \leq 0$, $p \geq 0$ and $p > 0$ that are derived as follows:

$$p \leq 0 \ (\text{for } p < 0 \vee p = 0), \quad p \geq 0 \ (\text{for } \neg(p < 0)), \quad \text{and} \ p > 0 \ (\text{for } \neg(p \leq 0)).$$

The semantics of formulas are given in a standard way. A variable x is *free* in the formula φ if it is not in the scope of a quantifier $\exists x$ or $\forall x$. A *sentence* is a formula with

no free variables. A formula is *quantifier-free* if it does not contain any existential or universal quantifier. Two formulas φ_1 and φ_2 are *equivalent* if the set of free variables of φ_1 and φ_2 are the same, and for every assignment to the free variables the formula φ_1 is true if and only if the formula φ_2 is true. A formula φ admits *quantifier elimination* if there is an algorithm to convert it to an equivalent quantifier-free formula. A quantifier elimination algorithm takes as input a formula φ and returns an equivalent quantifier-free formula, if one exists.

Tarski proved that every formula in the theory of real-closed fields admits quantifier elimination, and (by way of quantifier elimination) that there is an algorithm to decide the truth of a sentence φ in the theory of real-closed fields (see Tarski 1951 for algorithms that decide the truth of a sentence φ in the theory of real-closed fields). The complexity of the algorithm of Tarski has subsequently improved, and we now present a result of Basu (1999) on the complexity of quantifier elimination for formulas in the theory of the real-closed field.

Complexity of quantifier elimination. We first define the length of a formula φ, and then define the size of a formula with rational coefficients. We denote the length and size of φ as $\mathsf{len}(\varphi)$ and $\mathsf{size}(\varphi)$, respectively. The length of a polynomial p is defined as the sum of the length of its constituent monomials plus the number of monomials in the polynomial. The length of a monomial is defined as its degree plus the number of variables plus 1 (for the coefficient). For example, for the monomial $1/4 \cdot x^3 \cdot y^2 \cdot z$, its length is $6 + 3 + 1 = 10$. Given a polynomial p, the length of both $p < 0$ and $p = 0$ is $\mathsf{len}(p) + 2$. This defines the length of an atomic formula a. The length of a formula φ is inductively defined as follows:

$$\mathsf{len}(\neg a) = \mathsf{len}(a) + 1;$$
$$\mathsf{len}(\varphi_1 \wedge \varphi_2) = \mathsf{len}(\varphi_1) + \mathsf{len}(\varphi_2) + 1;$$
$$\mathsf{len}(\varphi_1 \vee \varphi_2) = \mathsf{len}(\varphi_1) + \mathsf{len}(\varphi_2) + 1;$$
$$\mathsf{len}(\exists x.\varphi) = \mathsf{len}(\varphi) + 2;$$
$$\mathsf{len}(\forall x.\varphi) = \mathsf{len}(\varphi) + 2.$$

Observe that the length of a formula is defined for formulas that may contain symbolic constants as coefficients. For formulas with rational coefficients we define its size as follows: the size of φ, i.e., $\mathsf{size}(\varphi)$, is defined as the sum of $\mathsf{len}(\varphi)$ and the space required to specify the rational coefficients of the polynomials appearing in φ in binary. We state a result of Basu (1999) on the complexity of quantifier elimination for the real-closed field. The following theorem is a specialization of Theorem 1 of Basu (1999); also see Theorems 14.14 and 14.16 of Basu et al. (2003).

Theorem 2 Basu (1999) *Let d, k, m be non-negative integers, $X = \{X_1, X_2, \ldots, X_k\}$ be a set of k variables, and $\mathcal{P} = \{p_1, p_2, \ldots, p_m\}$ be a set of m polynomials over the set X of variables, each of degree at most d and with coefficients in the real-closed field. Let $X_{[r]}, X_{[r-1]}, \ldots, X_{[1]}$ denote a partition of the set X of variables into r subsets such that the set $X_{[i]}$ of variables has size k_i, i.e., $k_i = |X_{[i]}|$ and $\sum_{i=1}^{r} k_i = k$. Let*

$$\Phi = (Q_r X_{[r]}).\, (Q_{r-1} X_{[r-1]}).\, \ldots .(Q_2 X_{[2]}).\, (Q_1 X_{[1]}).\, \varphi(p_1, p_2, \ldots, p_m)$$

be a sentence with r alternating quantifiers $Q_i \in \{\exists, \forall\}$ (i.e., $Q_{i+1} \neq Q_i$), and $\varphi(p_1, p_2, \ldots, p_m)$ is a quantifier-free formula with atomic formulas of the form $p_i \bowtie 0$, where $\bowtie \in \{<, >, =\}$. Let D denote the ring generated by the coefficients of the polynomials in $\mathcal{P}$. Then the following assertions hold.

1. *There is an algorithm to decide the truth of Φ using*

$$m^{\prod_i (k_i+1)} \cdot d^{\prod_i O(k_i)} \cdot \mathsf{len}(\varphi)$$

   arithmetic operations (multiplication, addition, and sign determination) in D.
2. *If $D = \mathbb{Z}$ (the set of integers) and the bit sizes of the coefficients of the polynomials are bounded by γ, then the bit sizes of the integers appearing in the intermediate computations of the truth of Φ is bounded by*

$$\gamma \cdot d^{\prod_i O(k_i)}.$$

The result of part 1 of Theorem 2 holds for sentences with symbolic constants as coefficients. The result of part 2 of Theorem 2 is for the special case of sentences with only integer coefficients. Part 2 of Theorem 2 follows from the results of Basu (1999), but is not explicitly stated as a theorem there; for an explicit statement as a theorem, see Theorems 14.14 and 14.16 of Basu et al. (2003).

Remark 1 Given two integers a and b, let $|a|$ and $|b|$ denote the space to express a and b in binary, respectively. The following assertions hold: given integers a and b,

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 the sign determination of $a \cdot b$ can be done $O(1)$ time, i.e., 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 Theorem 2, that if $D = \mathbb{Z}$ and the bit sizes of the coefficients of the polynomials appearing in Φ are bounded by γ, then the truth of Φ can be determined in time

$$m^{\prod_i O(k_i+1)} \cdot d^{\prod_i O(k_i)} \cdot O(\mathsf{len}(\varphi) \cdot \gamma^2). \tag{4}$$

4 Computation of values in stochastic games

The values in stochastic limit-average games can be irrational even if all rewards and transition probability values are rational (Raghavan and Filar 1991). Hence, we can algorithmically only approximate the values to within a precision ε, for $\varepsilon > 0$.

Discounted value functions. Let G be a stochastic game with reward function r. For a real β, with $0 < \beta < 1$, the β-discounted value function v_1^β is defined as follows:

$$v_1^\beta(s) = \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \ \beta \cdot \mathrm{E}_s^{\sigma,\pi} \left[\sum_{i=1}^{\infty} (1-\beta)^i \cdot r\left(\mathsf{X}_i, \Theta_{i,1}, \Theta_{i,2}\right) \right].$$

For a stochastic game G, the β-discounted value function v_1^β is monotonic with respect to β in a neighborhood of 0 (Bewley and Kohlberg 1976).

4.1 Sentence for the value of a stochastic game

We now describe how we can obtain a sentence in the theory of the real-closed field that states that the value of a stochastic limit-average game at a given state is strictly greater than α, for a real α. The sentence applies to the case where the rewards and the transition probabilities are specified as symbolic or rational constants.

Formula for β-discounted value functions. Given a real α and a stochastic limit-average game G, we present a formula in the theory of the real-closed field to express that the β-discounted value $v_1^\beta(s)$ at a given state s is strictly greater than α, for $0 < \beta < 1$. A *valuation* $v \in \mathbb{R}^n$ is a vector of reals, and for $1 \le i \le n$, the ith component of v represents the value $v(i)$ for state i. For every state $s \in S$ and for every move $b \in \Gamma_2(s)$ we define a polynomial $u_{(s,b,1)}$ for player 1 as a function of $x \in \mathcal{D}(\Gamma_1(s))$, a valuation v and $0 < \beta < 1$ as follows:

$$u_{(s,b,1)}(x, v, \beta) = \beta \cdot \sum_{a \in \Gamma_1(s)} x(a) \cdot r(s, a, b) + (1 - \beta) \cdot \sum_{a \in \Gamma_1(s)} x(a)$$
$$\cdot \sum_{t \in S} \delta(s, a, b)(t) \cdot v(t) - v(s).$$

The polynomial $u_{(s,b,1)}$ consists of the variables β, and $x(a)$ for $a \in \Gamma_1(s)$, and $v(t)$ for $t \in S$. Observe that given a stochastic limit-average game, $r(s, a, b)$ for $a \in \Gamma_1(s)$, and $\delta(s, a, b)(t)$ for $t \in S$ and $a \in \Gamma_1(s)$ are rational or symbolic constants given by the game graph, not variables. The coefficients of the polynomial are $r(s, a, b)$ for $a \in \Gamma_1(s)$, and $\delta(s, a, b)(t)$ for $a \in \Gamma_1(s)$ and $t \in S$. Hence the polynomial has degree 3 and has $1 + |\Gamma_1(s)| + n$ variables. Similarly, for $s \in S$, $a \in \Gamma_1(s)$, $y \in \mathcal{D}(\Gamma_2(s))$, $v \in \mathbb{R}^n$, and $0 < \beta < 1$, we have polynomials $u_{(s,a,2)}$ defined by

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

The sentence stating that $v_1^\beta(s)$ is strictly greater than α is as follows. We have variables $x_s(a)$ for $s \in S$ and $a \in \Gamma_1(s)$, $y_s(b)$ for $s \in S$ and $b \in \Gamma_2(s)$, and variables $v(1), v(2), \ldots, v(n)$. For simplicity we write x_s for the vector of variables $x_s(a_1), x_s(a_2), \ldots, x_s(a_j)$, where $\Gamma_1(s) = \{ a_1, a_2, \ldots, a_j \}$, y_s for the vector of variables $y_s(b_1), y_s(b_2), \ldots, y_s(b_l)$, where $\Gamma_2(s) = \{ b_1, b_2, \ldots, b_l \}$, and v for the vector of variables $v(1), v(2), \ldots, v(n)$. The sentence is as follows:

$$\Phi_\beta(s,\alpha) = \exists x_1, \ldots, x_n. \, \exists y_1, \ldots, y_n. \, \exists v. \qquad \Psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$$
$$\wedge \bigwedge_{s \in S, b \in \Gamma_2(s)} \left(u_{(s,b,1)}(x_s, v, \beta) \geq 0 \right) \wedge \bigwedge_{s \in S, a \in \Gamma_1(s)} \left(u_{(s,a,2)}(y_s, v, \beta) \leq 0 \right)$$
$$\wedge \, (v(s) - \alpha > 0),$$

where $\Psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$ specify the constraints that $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are valid randomized strategies and is defined as follows:

$$\Psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n) = \bigwedge_{s \in S} \left(\left(\sum_{a \in \Gamma_1(s)} x_s(a) \right) - 1 = 0 \right) \wedge$$
$$\bigwedge_{s \in S, a \in \Gamma_1(s)} (x_s(a) \geq 0)$$
$$\wedge \bigwedge_{s \in S} \left(\left(\sum_{b \in \Gamma_2(s)} y_s(b) \right) - 1 = 0 \right) \wedge$$
$$\bigwedge_{s \in S, b \in \Gamma_2(s)} (y_s(b) \geq 0).$$

The total number of polynomials in $\Phi_\beta(s,\alpha)$ is $1 + \sum_{s \in S}(3 \cdot |\Gamma_1(s)| + 3 \cdot |\Gamma_2(s)| + 2) = O(|\delta|)$. In the above formula we treat β as a variable; it is a free variable in $\Phi_\beta(s,\alpha)$. Given a stochastic limit-average game G, for all $0 < \beta < 1$, the correctness of $\Phi_\beta(s,\alpha)$ to specify that $v_1^\beta(s) > \alpha$ can be proved from the results of Shapley (1953).

Value of a game as limit of discounted games. The result of Mertens and Neyman (1981) established that the value of a stochastic limit-average game is the limit of the β-discounted values, as β goes to 0. Formally, we have

$$v_1(s) = \lim_{\beta \to 0^+} v_1^\beta(s).$$

Sentence for the value of a stochastic game. From the characterization of the value of a stochastic limit-average game as the limit of the β-discounted values and the monotonicity property of the β-discounted values in a neighborhood of 0, we obtain the following sentence $\Phi(s,\alpha)$ stating that the value at state s is strictly greater than α. In addition to variables for $\Phi_\beta(s,\alpha)$, we have the variables β and β_1. The sentence $\Phi(s,\alpha)$ specifies the expression

$$\exists \beta_1 > 0. \, \forall \beta \in (0, \beta_1). \, \Phi_\beta(s,\alpha),$$

and is defined as follows:

$$\Phi(s,\alpha) = \exists \beta_1. \, \forall \beta. \, \exists x_1, \ldots, x_n. \, \exists y_1, \ldots, y_n. \, \exists v. \quad \Psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$$
$$\wedge (\beta_1 > 0) \wedge \Big[(\beta_1 - \beta \leq 0) \vee (\beta \leq 0) \vee \Big((\beta_1 - \beta > 0)$$
$$\wedge \bigwedge_{s \in S, b \in \Gamma_2(s)} \left(u_{(s,b,1)}(x_s, v, \beta) \geq 0 \right)$$
$$\wedge \bigwedge_{s \in S, a \in \Gamma_1(s)} \left(u_{(s,a,2)}(y_s, v, \beta) \leq 0 \right) \Big) \Big]$$
$$\wedge (v(s) - \alpha > 0),$$

where $\Psi(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$ specify the constraints that $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are valid randomized strategies (the same formula used for $\Phi_\beta(s, \alpha)$).[2] Observe that $\Phi(s, \alpha)$ contains no free variable (i.e., the variables x_s, y_s, v, β_1, and β are quantified). A similar sentence was used in Bewley and Kohlberg (1976) for values of discounted games. The total number of polynomials in $\Phi(s, \alpha)$ is $O(|\delta|)$; in addition to the $O(|\delta|)$ polynomials of $\Phi_\beta(s, \alpha)$ there are four more polynomials in $\Phi(s, \alpha)$. In the setting of Theorem 2 we obtain the following bounds for $\Phi(s, \alpha)$:

$$m = O(|\delta|); \quad k = O(|\delta|); \quad \prod_i (k_i + 1) = O(|\delta|); \quad r = O(1); \quad d = 3; \quad (5)$$

and hence we have

$$m^{\prod_i (k_i+1)} \cdot d^{\prod_i O(k_i)} = O(|\delta|)^{O(|\delta|)} = 2^{O(|\delta| \cdot \log(|\delta|))}.$$

Also observe that for a stochastic game G, the sum of the lengths of the polynomials appearing in the sentence is $O(|\delta|)$. The present analysis along with Theorem 2 yields Theorem 3. The result of Theorem 3 holds for stochastic limit-average games where the transition probabilities and rewards are specified as symbolic constants.

Theorem 3 *Given a stochastic limit-average game G with reward function r, a state s of G, and a real α, there is an algorithm to decide whether $v_1(s) > \alpha$ using $2^{O(|\delta| \cdot \log(|\delta|))} \cdot O(|\delta|)$ arithmetic operations (addition, multiplication, and sign determination) in the ring generated by the set*

$$\{\, r(s, a, b) \mid s \in S, a \in \Gamma_1(s), b \in \Gamma_2(s) \,\} \cup \{\, \delta(s, a, b)(t) \mid s, t \in S, a \in \Gamma_1(s),$$
$$b \in \Gamma_2(s) \,\} \cup \{\, \alpha \,\}.$$

4.2 Algorithmic analysis

For algorithmic analysis we consider rational stochastic games, i.e., stochastic games such that $r(s, a, b)$ and $\delta(s, a, b)(t)$ are rational for all states $s, t \in S$, and moves $a \in \Gamma_1(s)$ and $b \in \Gamma_2(s)$. In the sequel we will only consider rational stochastic games. Given the sentence $\Phi(s, \alpha)$ to specify that $v_1(s) > \alpha$, we first reduce it to an equivalent sentence $\widehat{\Phi}(s, \alpha)$ as follows.

- For every rational coefficient $\ell = \frac{q_1}{q_2}$, where $q_1, q_2 \in \mathbb{Z}$, appearing in $\Phi(s, \alpha)$ we apply the following procedure:
  1. introduce a new variable z_ℓ;
  2. replace ℓ by z_ℓ in $\Phi(s, \alpha)$;
  3. add a polynomial $q_2 \cdot z_\ell - q_1 = 0$ as a conjunct to the quantifier-free body of the formula; and

---

[2] Our detailed formulas $\Phi_\beta(s, \alpha)$ and $\Phi(s, \alpha)$ can be shortened, however, the present formulas make it easier to understand the bound on parameters required for complexity bounds.

4. existentially quantify z_ℓ 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 $\Phi(s, \alpha)$ by 1. Also observe that the coefficients in $\widehat{\Phi}(s, \alpha)$ are integers, and hence the ring $\widehat{D}$ generated by the coefficients in $\widehat{\Phi}(s, \alpha)$ is $\mathbb{Z}$. Similar to the bounds obtained in (5), in the setting of Theorem 2 we obtain the following bounds for $\widehat{\Phi}(s, \alpha)$:

$$\widehat{m} = O(|\delta|); \quad \widehat{k} = O(|\delta|); \quad \prod_i (\widehat{k}_i + 1) = O(|\delta|); \quad \widehat{r} = O(1); \quad \widehat{d} = 4;$$

and hence

$$\widehat{m}^{\prod_i O(\widehat{k}_i+1)} \cdot \widehat{d}^{\prod_i O(\widehat{k}_i)} = O(|\delta|)^{O(|\delta|)} = 2^{O(|\delta| \cdot \log(|\delta|))}.$$

Also observe that the length of the sentence $\widehat{\Phi}(s, \alpha)$ can be bounded by $O(|\delta|)$, and the sum of the bit sizes of the coefficients in $\widehat{\Phi}(s, \alpha)$ can be bounded by $O(|G| + |\alpha|)$, where $|\alpha|$ is the space required to express α in binary. This along with (4) of Remark 1 yields the following result.

Theorem 4 *Given a rational stochastic limit-average game G, a state s of G, and a rational α, there is an algorithm that decides whether $v_1(s) > \alpha$ in time*

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

4.3 Approximating the value of a stochastic game

We now present an algorithm that approximates the value $v_1(s)$ within a tolerance of $\varepsilon > 0$. The algorithm (Algorithm 1) is obtained by a binary search technique along with the result of Theorem 4. Algorithm 1 works for the special case of *normalized* rational stochastic games. We first define normalized rational stochastic games and then present a reduction of rational stochastic games to normalized rational stochastic games.

Normalized rational stochastic games. A rational stochastic game is *normalized* if the reward function satisfies the following two conditions: (1) min$\{ r(s, a, b) \mid s \in S, a \in \Gamma_1(s), b \in \Gamma_2(s) \} \geq 0$; and (2) max$\{ r(s, a, b) \mid s \in S, a \in \Gamma_1(s), b \in \Gamma_2(s) \} \leq 1$.

Reduction. We now present a reduction of rational stochastic games to normalized rational stochastic games, such that by approximating the values of normalized rational stochastic games we can approximate the values of rational stochastic games. Given a reward function $r : S \times A \times A \to \mathbb{R}$, let

$$M = \max\{ \operatorname{abs}(r(s, a, b)) \mid s \in S, a \in \Gamma_1(s), b \in \Gamma_2(s) \},$$

where $\operatorname{abs}(r(s, a, b))$ denotes the absolute value of $r(s, a, b)$. Without loss of generality we assume $M > 0$. Otherwise, $r(s, a, b) = 0$ for all states $s \in S$, and moves

---

Algorithm 1 Approximating the value of a stochastic game

Input: a normalized rational stochastic limit-average game G,
           a state s of G, and a rational value $\varepsilon > 0$ specifying the desired tolerance.
Output: a rational interval $[l, u]$ such that $u - l \le 2\varepsilon$ and $v_1(s) \in [l, u]$.

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

$a \in \Gamma_1(s)$ and $b \in \Gamma_2(s)$, and hence $v_1(s) = 0$ for all states $s \in S$ (i.e., the value function can be trivially computed). Consider the reward function $r^+ : S \times A \times A \to [0, 1]$ defined as follows: for $s \in S$, $a \in \Gamma_1(s)$, and $b \in \Gamma_2(s)$, we have

$$r^+(s, a, b) = \frac{r(s, a, b) + M}{2M}.$$

The reward function r^+ is normalized and the following assertion hold. Let v_1 and v_1^+ denote the value functions for the reward functions r and r^+, respectively. Then for all states $s \in S$ we have

$$v_1^+(s) = \frac{v_1(s) + M}{2M}.$$

Hence it follows that for rationals α, l, and u, such that $l \le u$, we have

$$v_1(s) > \alpha \text{ iff } v_1^+(s) > \frac{\alpha + M}{2M}; \quad \text{and}$$
$$v_1^+(s) \in [l, u] \text{ iff } v_1(s) \in [M \cdot (2l - 1), M \cdot (2u - 1)].$$

Given a rational $\varepsilon > 0$, to obtain an interval $[l_1, u_1]$ such that $u_1 - l_1 \le \varepsilon$ and $v_1(s) \in [l_1, u_1]$, we first obtain an interval $[l, u]$ such that $u - l \le \varepsilon/2M$ and $v_1^+(s) \in [l, u]$. From the interval $[l, u]$ we obtain the interval $[l_1, u_1] = [M \cdot (2l - 1), M \cdot (2u - 1)]$ such that $v_1(s) \in [l_1, u_1]$ and $u_1 - l_1 = 2 \cdot M \cdot (u - l) \le \varepsilon$. Hence we present the algorithm to approximate the values for normalized rational stochastic games.

   Running time of Algorithm 1. In Algorithm 1 we denote by $\Phi(s, m)$ the sentence to specify that $v_1(s) > m$, and by Theorem 4 the truth of $\Phi(s, m)$ can be decided in time

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

for a stochastic game G, where $|m|$ is the number of bits required to specify m. In Algorithm 1, the variables l and u are initially set to 0 and 1, respectively. Since the

game is normalized, the initial values of l and u clearly provide lower and upper bounds on the value, and provide starting bounds for the binary search. In each iteration of the algorithm, in Steps 2.1.1 and 2.2.1, there is a division by 2. It follows that after i iterations l, u, and m can be expressed as $q/2^i$, where q is an integer and $q \leq 2^i$. Hence l, u, and m can always be expressed in

$$O\left(\log\left(\frac{1}{\varepsilon}\right)\right)$$

bits. The loop in Step 4 runs for $\lceil \log(1/\varepsilon) \rceil = O(\log(1/\varepsilon))$ iterations, and every iteration can be computed in time $2^{O(|\delta| \cdot \log(|\delta|))} \cdot O\left(|G|^2 + \log^2(1/\varepsilon)\right)$. This gives the following theorem.

Theorem 5 *Given a normalized rational stochastic limit-average game G, a state s of G, and a rational $\varepsilon > 0$, Algorithm 1 computes an interval $[l, u]$ such that $v_1(s) \in [l, u]$ and $u - l \leq 2\varepsilon$, in time*

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

The reduction from rational stochastic games to normalized stochastic games suggest that for a rational stochastic game G and a rational tolerance $\varepsilon > 0$, to obtain an interval of length at most ε that contains the value $v_1(s)$, it suffices to obtain an interval of length of at most $\varepsilon/2M$ that contains the value in the corresponding normalized game, where $M = \max\{ \mathsf{abs}(r(s, a, b)) \mid s \in S, a \in \Gamma_1(s), b \in \Gamma_2(s) \}$. Since M can be expressed in $|G|$ bits, it follows that the size of the normalized game is $O(|G|^2)$. Given a tolerance $\varepsilon > 0$ for the rational stochastic game, we need to consider the tolerance $\varepsilon/2 \cdot M$ for the normalized game. The above analysis along with Theorem 5 yields the following corollary (the corollary is obtained from Theorem 5 by substituting $|G|$ by $|G|^2$, and $\log(1/\varepsilon)$ by $|G| \cdot \log(1/\varepsilon)$).

Corollary 1 *Given a rational stochastic limit-average game G, a state s of G, and a rational $\varepsilon > 0$, an interval $[l, u]$ such that $v_1(s) \in [l, u]$ and $u - l \leq 2\varepsilon$, can be computed in time*

$$2^{O(|\delta| \cdot \log(|\delta|))} \cdot O\left(|G|^5 \cdot \log\left(\frac{1}{\varepsilon}\right) + |G|^3 \cdot \log^3\left(\frac{1}{\varepsilon}\right)\right).$$

The complexity class EXPTIME. A problem is in the complexity class EXPTIME (Papadimitriou 1994) 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 $2^{O(p(|I|))}$. In case of rational stochastic limit-average games, the input is the size of the game G, i.e., the input requires $|G|$ bits. Hence from Theorem 4 and Corollary 1 we obtain the following result.

Theorem 6 *Given a rational stochastic limit-average game G, a state s of G, rational $\varepsilon > 0$, and rational α, the following assertions hold.*

1. (Decision problem) *Whether $v_1(s) > \alpha$ can be decided in EXPTIME.*
2. (Approximation problem) *An interval $[l, u]$ such that $u - l \leq 2\varepsilon$ and $v_1(s) \in [l, u]$ can be computed in EXPTIME.*

Approximate analysis of games with approximate description. Let $G = \langle S, A, \Gamma_1, \Gamma_2, \delta, r \rangle$ and $G' = \langle S, A, \Gamma_1, \Gamma_2, \delta', r' \rangle$ be two stochastic games such that

1. for all $s, t \in S$ and for all $a \in \Gamma_1(s)$ and $b \in \Gamma_2(s)$, we have

$$\delta(s, a, b) < (1 + \eta) \cdot \delta'(s, a, b)(t) \quad \text{and} \quad \delta'(s, a, b) < (1 + \eta) \cdot \delta(s, a, b)(t),$$

   for $\eta < \frac{1}{2|S|}$; and
2. for all $s \in S$ and for all $a \in \Gamma_1(s)$ and all $b \in \Gamma_2(s)$ we have

$$\mathsf{abs}(r(s, a, b) - r'(s, a, b)) \leq \gamma.$$

Let $\rho(G, G')$ be defined as the infimum over $\left(\frac{2\eta \cdot |S|}{(1 - 2\eta \cdot |S|)} \cdot ||r|| + \gamma \right)$, where η, γ ranges over all pairs that satisfy the above two inequalities. From the result of Solan (2003) it follows that the absolute difference in the values of a player at all states in G and G' is bounded by $\rho(G, G')$. Hence given a game G and an auxiliary game G' that approximates G within η, i.e., $\rho(G, G') \leq \eta$, we can approximate the values of the game G' for $\varepsilon > 0$, and obtain a $\eta + \varepsilon$ approximation of the values of the game G. This enables us to approximate the values of stochastic games described approximately.

Unfortunately, the only lower bound we know on the complexity of the decision problem is PTIME-hardness (polynomial-time hardness). The hardness follows from a reduction from alternating reachability (Beeri 1980; Immerman 1981). Even for the simpler case of perfect-information deterministic games, no polynomial time algorithm is known (Zwick and Paterson 1996), and the best known deterministic algorithm for perfect information games is exponential in the size of the game. In case of perfect-information stochastic games, deterministic and stationary optimal strategies exist (Liggett and Lippman 1969). 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 approximately) (also see the survey; Neyman and Sorin 2003). From the polynomial time algorithm to compute values in Markov decision processes (Filar and Vrieze 1997) and the existence of pure stationary optimal strategies in perfect-information games (Liggett and Lippman 1969), it follows that the decision problem for perfect-information games lie in NP ∩ coNP. Better complexity bounds than EXPTIME to solve the decision and the approximation problem for stochastic games is an interesting open problem; and even for deterministic games no better bound is known.

References

Basu S (1999) New results on quantifier elimination over real-closed fields and applications to constraint databases. J ACM 46:537–555

Basu S, Pollack R, Roy M-F (2003) Algorithms in real algebraic geometry. Springer, Berlin

Beeri C (1980) On the membership problem for functional and multivalued dependencies in relational databases. ACM Trans Database Syst 5:241–259

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

Blackwell D, Ferguson TS (1968) The big match. Ann Math Stat 39:159–163

Filar J, Vrieze K (1997) Competitive Markov decision processes. Springer, Berlin

Gillete D (1957) Stochastic games with zero stop probabilities. In: Contributions to the theory of games III. Princeton University Press, Princeton, pp 179–188

Hoffman AJ, Karp RM (1966) On nonterminating stochastic games. Manage Sci 12:359–370

Immerman N (1981) Number of quantifiers is better than number of tape cells. J Comput Syst Sci 22:384–406

Liggett TA, Lippman SA (1969) Stochastic games with perfect information and time average payoff. Siam Rev 11:604–607

Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10:53–66

Neyman A, Sorin S (2003) Stochastic games and applications. Kluwer, Dordrecht

Papadimitriou CH (1994) Computational complexity. Addison-Wesley, Reading

Parthasarathy T, Raghavan TES (1981) An orderfield property for stochastic games when one player controls transition probabilities. J Optim Theory Appl 33:375–392

Raghavan TES, Filar JA (1991) Algorithms for stochastic games—a survey. ZOR—Methods Models Oper Res 35:437–472

Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39:1095–1100

Solan E (2003) Continuity of the value of competitive Markov decision processes. J Theor Probab 16:831–845

Tarski A (1951) A decision method for elementary algebra and geometry. University of California Press, California

Zwick U, Paterson MS (1996) The complexity of mean payoff games on graphs. Theor Comput Sci 158:343–359