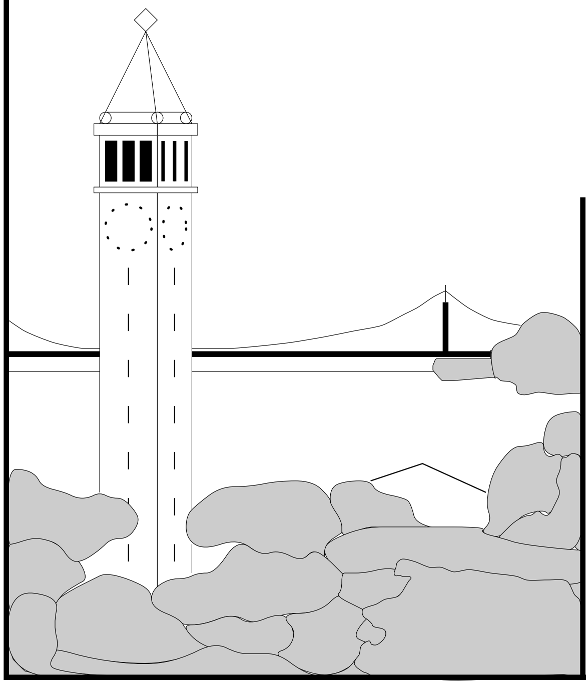
Stochastic Limit-Average Games are in EXPTIME

Krishnendu Chatterjee, Rupak Majumdar and Thomas A. Henzinger



Report No. UCB/CSD-5-1405

August 2005

Computer Science Division (EECS)
University of California
Berkeley, California 94720

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

August 2005

Abstract

The value of a finite-state stochastic game with limit-average objectives can be approximated to within ϵ in time exponential in the size of the game and logarithmic in $\frac{1}{\epsilon}$.

1 Introduction

A stochastic game [12] 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 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 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

*This research was supported in part by the ONR grant N00014-02-1-0671, the AFOSR MURI grant F49620-00-1-0327, and the NSF ITR grant CCR-0225610.

reward of a pair of strategies σ and π and a starting state s is defined as

$$v_1(s, \sigma, \pi) = \lim_{n\to\infty}\inf \mathrm{E}_s^{\sigma,\pi}[\frac{1}{n}\sum_{i=0}^{n} r(\Theta_i)];$$

where Θ_i is the random state reached at round i of the game under strategies σ and π, and $r(s)$ gives the reward at state s. 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 [12], 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 [6], 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 [8]. Gillette's paper also introduced a limit-average game called the Big Match, which was solved in [3]. Bewley and Kohlberg [2] 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 [9] 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. 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) [10, 7] (see [5] 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 [2, 9]. 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 linear 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 exponential algorithm to approximate the value to any given ε. 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 matches the complexity of the best known deterministic algorithm for the special case of perfect information games.

2 Definitions

Stochastic games. For a finite set A, a *probability distribution* on A is a function $\delta : A \mapsto [0,1]$ 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.

A (two-player) *stochastic game* $G = \langle S, \mathit{Moves}, \Gamma_1, \Gamma_2, \delta, r\rangle$ consists of:

- A finite state space S.
- A finite set *Moves* of moves.
- Two move assignments $\Gamma_1, \Gamma_2 : S \mapsto 2^{\mathit{Moves}} \setminus \emptyset$. For $i \in \{1,2\}$, assignment Γ_i associates with each state $s \in S$ the non-empty set $\Gamma_i(s) \subseteq \mathit{Moves}$ of moves available to player i at state s.
- A probabilistic transition function $\delta : S \times \mathit{Moves} \times \mathit{Moves} \to \mathcal{D}(S)$, that gives the probability $\delta(s, a_1, a_2)(t)$ of a transition from s to t when player 1 plays 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 $r : S \to \mathbb{R}$ that maps every state to a real valued reward.

The *size* of a stochastic game G is equal to the sum of the size of the transition function δ and reward function r, that is,

$$|G| = \sum_{s \in S} \big((\sum_{t \in S} \sum_{a \in \Gamma_1(s)} \sum_{b \in \Gamma_2(s)} |\delta(s,a,b)(t)|) + |r(s)| \big),$$

where $|\delta(s,a,b)(t)|$ and $|r(s)|$ denotes the space to specify the probability distribution and the reward, respectively.

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, player 1 wins and player 2 loses a reward of value $r(t)$. 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. We assume that the players act *non-cooperatively*, i.e., each player chooses her strategy independently and secretly from the other player, and is only interested in maximizing her own reward. 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 $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ of states and such that there exists $(a_i, b_i) \in \Gamma_1(s_i) \times \Gamma_2(s_i)$ 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 $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ such that $s_0 = s$, i.e., the set of plays starting from state s.

Randomized strategies. A *selector* ξ for player $i \in \{1, 2\}$ is a function $\xi : S \mapsto \mathcal{D}(\mathit{Moves})$ such that for all $s \in S$ and $a \in \mathit{Moves}$, if $\xi(s)(a) > 0$ then $a \in \Gamma_i(s)$. We denote by Λ_i the set of all selectors for player $i \in \{1, 2\}$. A *strategy* for player 1 is a function $\sigma : S^+ \to \Lambda_1$ associates a selector with every finite non-empty sequence of states representing the history of the play so far. 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 $\Pr_s^{\sigma,\pi}(\mathcal{A})$ 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 $\Theta_i : \Omega_s \to S$ the random variable denoting the i-th state along a path.

A *valuation* is a mapping $v : S \to \mathbb{R}$, associating a real number $v(s)$ with each state s. Given a valuation v and two selectors $\xi_1 \in \Lambda_1$ and $\xi_2 \in \Lambda_2$, we define the expectation $Pre^{\xi_1,\xi_2}(v) : S \to \mathbb{R}$ by $Pre^{\xi_1,\xi_2}(v)(s) = \sum_{a,b \in Moves} \sum_{t \in S} v(t)\, \delta(s,a,b)(t)\, \xi_1(a)\, \xi_2(b)$: intuitively, $Pre^{\xi_1,\xi_2}(v)(s)$ is the expected value of v when player 1 chooses a move according to ξ_1 and player 2 chooses a move according to ξ_2. We denote by $Pre_1(v) : S \to \mathbb{R}$ the maximal valuation player 1 can achieve, defined by $Pre_1(v)(s) = \sup_{\xi_1 \in \Lambda_1} \inf_{\xi_2 \in \Lambda_2} Pre^{\xi_1,\xi_2}(v)(s)$ for all $s \in S$; and we write $Pre_1(v) = \sup_{\xi_1 \in \Lambda_1} \inf_{\xi_2 \in \Lambda_2} Pre^{\xi_1,\xi_2}(v)$, where inf and sup are interpreted pointwise. We define $Pre_2 : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ symmetrically.

Limit-average objective. 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

$$v_1(s, \sigma, \pi) = \lim\inf_{n \to \infty} \mathrm{E}_s^{\sigma,\pi}[\frac{1}{n} \sum_{i=0}^{n} r(\Theta_i)];$$

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

$$v_2(s, \sigma, \pi) = \lim\sup_{n \to \infty} \mathrm{E}_s^{\sigma,\pi}[\frac{1}{n} \sum_{i=0}^{n} -r(\Theta_i)].$$

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* G at s for player $i \in \{ 1, 2 \}$. The value for player 1 and player 2 are given by the function $v_1 : S \mapsto [0, 1]$ and $v_2 : S \mapsto [0, 1]$, 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 [9] establish the determinacy of stochastic limit-average games.

Theorem 1 ([9]) *For all stochastic limit-average games, for all state s, we have $v_1(s) + v_2(s) = 0$.*

3 Theory of Real-closed Fields

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 [4]. 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 ([14]) *The theory of real-closed fields in the language of ordered fields is decidable.*

The operator Pre_1 can be interpreted as an operator to obtain the optimal (or minmax) value of a matrix game. We start with the following classical observation [15] that this minmax value can be written as an elementary formula in the theory of ordered fields.

Lemma 1 *For a valuation $v : S \to \mathbb{R}$, $Pre_1(v)$ can be written as an existential elementary formula in the language of real-closed fields.*

4 Complexity of Approximating the Value

The values in stochastic limit-average games can be irrational even if the reward at every state and the transition probability function only take rational values [11]. 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 $r(\text{halt}) = 0$. 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, i.e., if $\beta_1 \le \beta_2$, then $v_1^{\beta_1} \ge v_1^{\beta_2}$.

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. We write x, y, z to denote variables and $\vec{x}, \vec{y}, \vec{z}$ to denote vector of variables of length n. We denote by $\vec{x}(i)$ the i-th component of the vector $\vec{x}$.

For $0 < \beta < 1$, the value function of a β-discounted game G_β can be characterized as the least fixed-point of the following *optimality* equation [12]:

$$v_1^\beta(i) = [\beta \cdot r(i) + (1-\beta) Pre_1(\vec{x})(i)], \tag{1}$$

where $\vec{x}$ is a vector of variables, with variable $\vec{x}(i)$ for state i. The least fixed-point can be interpreted as follows:

$$\begin{array}{rcll} \vec{x}^0(j) & = & 0, & \forall j. 1 \le j \le n; \\ \vec{x}^{i+1}(j) & = & \beta \cdot r(j) + (1-\beta) Pre_1(\vec{x})(j), & \forall j. 1 \le j \le n; \\ v_1^\beta(j) & = & \lim_{i \to \infty} \vec{x}^i(j), & \forall j. 1 \le j \le n. \end{array}$$

The existence of the limit follows from monotonicity of the operator Pre_1.

Lemma 2 ([12]) *The value of the β-discounted stochastic game G_β is obtained as the least fixed-point of the optimality equation 1.*

We describe a procedure to express the fixed-point equation for v_1^β as a formula over the theory of reals over addition and multiplication. For $\sim \in \{ \leq, = \}$ we write $\sim_{vec}$ to denote the corresponding operator for vector comparison, i.e., $\vec{x} \sim_{vec} \vec{y}$ if and only if $\vec{x}(i) \sim \vec{y}(i)$, $\forall\ i.\ 1 \leq i \leq n$. Let us denote by $\vec{c}$ the vector $\langle 1, 2, \ldots, n \rangle$. We describe a formula $f_\beta(\vec{x})$ for the equation $f_\beta(\vec{x})(j) = \mu\vec{x}.[\beta r(j) + (1-\beta)Pre_1(\vec{x})(j)]$ as follows

$$\begin{array}{rll} f_\beta(\vec{x}) = \forall \vec{y}. & (& \big((\beta \cdot r(\vec{c}) + (1-\beta)Pre_1(\vec{y})) =_{vec} \vec{y}\big) \Rightarrow (\vec{x} \leq_{vec} \vec{y}) \quad) \\ & \wedge & \big((\beta \cdot r(\vec{c}) + (1-\beta)Pre_1(\vec{x})) =_{vec} \vec{x}\big). \end{array}$$

Intuitively, the formula describes for all vectors $\vec{y}$ such that $\vec{y}$ is a fixed-point of the optimality equation, we must have $\vec{x} \leq_{vec} \vec{y}$, and $\vec{x}$ is also a fixed-point of the optimality equation. It follows that $\vec{x}$ is the least-fixed point of the optimality equation. Moreover, it follows from Lemma 1 that $f_\beta(\vec{x})$ can be expressed as a formula over H.

Value of a game as limit of discounted games. The result of Mertens-Neyman [9] 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

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

We describe the procedure to express the limit of a function as formula over real-closed fields. Let f_β be a function that can be expressed as a formula over real-closed fields. Then the expression $\vec{z} =_{vec} \lim_{\beta \to 0, 0 < \beta < 1} f_\beta(\vec{x})$ can be expressed as

$$\Psi(\vec{x}, \vec{z}) = \forall \varepsilon > 0.\ \exists \beta_1.\ \forall \beta_2.\ \big((0 < \beta_1 < 1) \wedge (0 < \beta_2 \leq \beta_1) \Rightarrow (f_{\beta_2}(\vec{x}) - \vec{z}) \leq_{vec} \vec{\varepsilon}\big),$$

where $\vec{\varepsilon}$ is the constant vector with value ε for all components. Given a stochastic normalized positive limit-average game, the formula $\Psi(\vec{x}, \vec{z})$ expresses the value function $v_1(\cdot)$ of the game as the limit of the value of the discounted game $v_1^\beta(\cdot)$ as $\beta \to 0$, and the vector $\vec{z}$ gives the values function. It follows from above the formula $\Psi(\vec{x}, \vec{z})$ can be expressed over H. Moreover, the number of quantifier alternations is constant in Ψ, and the length

Algorithm 1 Approximating the value

Input: Normalized positive limit-average game G,
and a rational value ε as tolerance, a state i in G.
Output: An interval $[l, u]$ such that $u - l \leq 2\varepsilon$ and $v_1(i) \in [l, u]$.
1. $l := 0, u := 1, m := 1/2$.
2. **repeat** for $\lceil \log(\frac{1}{\varepsilon}) \rceil$ steps
   2.1 **if** $(\exists \vec{z}, \vec{x}, y.\ \ (y = \Psi(\vec{x}, \vec{z})(i) \wedge y \geq m))$
   then $l := m, u := u, m := \frac{l+u}{2}$;
   2.2 **else** $l := l, u := m, m := \frac{l+u}{2}$.
3. **return** $[l, u]$.

of the formula is linear in the size of the game. An algorithm that approximates the value within a tolerance of ε is obtained by a binary search, see Algorithm 1.

The result of [1] shows that quantifier elimination in the theory of reals over addition and multiplication can be achieved in time exponential in the size of the formula and double exponential in the number of quantifier alternations. Since $\Psi(\vec{x}, \vec{z})$ has constant number of quantifier alternations, and is linear in the size of the game graph, we get the following bound on Algorithm 1.

Theorem 3 *Given a normalized positive limit-average game G, a state i of G, and a rational ε, Algorithm 1 computes an interval $[l, u]$ such that $v_1(i) \in [l, u]$ and $u - l \leq 2\varepsilon$, in time exponential in the size of the game and logarithmic in $\frac{1}{\varepsilon}$.*

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 ε in time exponential in the size of the game and logarithmic in $\frac{1}{\varepsilon}$.*

Unfortunately, the only lower bound on the complexity is P-hardness (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 [8]. Thus our complexity bound matches the best known result for the simpler case of perfect information games.

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

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

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

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

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

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

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

[9] J.F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–56, 1981.

[10] 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.

[11] 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.

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

[13] M. Sobel. Noncooperative stochastic games. *Ann. Math. Stat.*, 42:1930–1935, 1971.

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

[15] H. Weyl. Elementary proof of a minmax theorem due to von Neumann. In *Contributions to the Theory of Games, I*, volume 24 of *Annals of Mathematical Studies*, pages 19–25. Princeton University Press, 1950.

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