Reduction of Stochastic Parity to Stochastic Mean-payoff Games

Krishnendu Chatterjee
Thomas A. Henzinger

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-140
http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-140.html

November 6, 2006

Acknowledgement

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

Reduction of Stochastic Parity to Stochastic Mean-payoff Games

Krishnendu Chatterjee[1] and Thomas A. Henzinger[1,2]

[1] EECS, UC Berkeley, USA
[2] EPFL, Switzerland
{c_krish,tah}@eecs.berkeley.edu

Abstract. A stochastic graph game is played by two players on a game graph with probabilistic transitions. We consider stochastic graph games with ω-regular winning conditions specified as parity objectives, and mean-payoff (or long-run average) objectives. These games lie in NP $\cap$ coNP. We present a polynomial time Turing reduction of stochastic parity games to stochastic mean-payoff games.

1 Introduction

Graph games. A stochastic graph game [Con92] is played on a directed graph with three kinds of states: player-1, player-2, and probabilistic states. At player-1 states, player 1 chooses a successor state; at player-2 states, player 2 chooses a successor state; at probabilistic states, a successor state is chosen according to a given probability distribution. The outcome of playing the game forever is an infinite path through the graph. If there are no probabilistic states, we refer to the game as a *2-player graph game*; otherwise, as a *2 1/2-player graph game*.

Parity objectives. The theory of graph games with ω-regular winning conditions is the foundation for modeling and synthesizing reactive processes with fairness constraints. In the case of 2 1/2-player graph games, the two players represent a reactive system and its environment, and the probabilistic states represent uncertainty. The *parity* objectives provide an adequate model, as the fairness constraints of reactive processes are ω-regular, and every ω-regular winning condition can be specified as a parity objective [Tho97]. The solution problem for a 2 1/2-player game with parity objective Φ asks for each state s, for the maximal probability with which player 1 can ensure the satisfaction of Φ if the game is started from s (this probability is called the *value* of the game at s). An *optimal strategy* for player 1 is a strategy that enables player 1 to win with that maximal probability. The existence of *pure memoryless* optimal strategies for 2 1/2-player games with parity objectives was established in [CJH04] (a pure memoryless strategy chooses for each player-1 state a unique successor state). The existence of pure memoryless optimal strategies implies that the solution problem for 2 1/2-player games with parity objectives lies in NP $\cap$ coNP.

Mean-payoff objectives. An important class of quantitative objectives is the class of mean-payoff (or long-run average) objectives. In case of mean-payoff objectives there is a real-valued reward at each state and the payoff of player 1 for a play is the long-run average of the rewards appearing in the play. The objective of player 1 is to maximize the long-run average, and values are defined in a similar way as for parity objectives. In $2^1\!/_2$-player games with mean-payoff objectives pure memoryless optimal strategies exist [LL69]. Again, the existence of pure memoryless optimal strategies implies that the solution problem for $2^1\!/_2$-player games with mean-payoff objectives lies in NP $\cap$ coNP.

Our result. We present a polynomial time Turing reduction of $2^1\!/_2$-player parity games to $2^1\!/_2$-player mean-payoff games for computation of values. Similar reduction was known for the special case of 2-player games [Jur98]. As a consequence of our reduction all algorithms for $2^1\!/_2$-player mean-payoff games [FV97,Put94] can now be used for $2^1\!/_2$-player parity games.

2 Definitions

We consider turn-based probabilistic games and some of its subclasses.

Game graphs. A *turn-based probabilistic game graph* ($2^1\!/_2$*-player game graph*) $G = ((S,E),(S_1,S_2,S_{\bigcirc}),\delta)$ consists of a directed graph (S,E), a partition $(S_1$, S_2, $S_{\bigcirc})$ of the finite set S of states, and a probabilistic transition function δ: $S_{\bigcirc} \to \mathcal{D}(S)$, where $\mathcal{D}(S)$ denotes the set of probability distributions over the state space S. The states in S_1 are the *player*-1 states, where player 1 decides the successor state; the states in S_2 are the *player*-2 states, where player 2 decides the successor state; and the states in $S_{\bigcirc}$ are the *probabilistic* states, where the successor state is chosen according to the probabilistic transition function δ. We assume that for $s \in S_{\bigcirc}$ and $t \in S$, we have $(s,t) \in E$ iff $\delta(s)(t) > 0$, and we often write $\delta(s,t)$ for $\delta(s)(t)$. For technical convenience we assume that every state in the graph (S,E) has at least one outgoing edge. For a state $s \in S$, we write $E(s)$ to denote the set $\{t \in S \mid (s,t) \in E\}$ of possible successors. The *turn-based deterministic game graphs* (*2-player game graphs*) are the special case of the $2^1\!/_2$-player game graphs with $S_{\bigcirc} = \emptyset$. The *Markov decision processes* ($1^1\!/_2$*-player game graphs*) are the special case of the $2^1\!/_2$-player game graphs with $S_1 = \emptyset$ or $S_2 = \emptyset$. We refer to the MDPs with $S_2 = \emptyset$ as *player*-1 MDPs, and to the MDPs with $S_1 = \emptyset$ as *player*-2 MDPs.

Plays and strategies. An infinite path, or a *play*, of the game graph G is an infinite sequence $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ of states such that $(s_k, s_{k+1}) \in E$ for all $k \in \mathbb{N}$. We write Ω for the set of all plays, and for a state $s \in S$, we write $\Omega_s \subseteq \Omega$ for the set of plays that start from the state s. A *strategy* for player 1 is a function σ: $S^* \cdot S_1 \to \mathcal{D}(S)$ that assigns a probability distribution to all finite sequences $\boldsymbol{w} \in S^* \cdot S_1$ of states ending in a player-1 state (the sequence represents a prefix of a play). Player 1 follows the strategy σ if in each player-1 move, given that the current history of the game is $\boldsymbol{w} \in S^* \cdot S_1$, she chooses the next state according to the probability distribution $\sigma(\boldsymbol{w})$. A strategy must prescribe only available moves, i.e., for all $\boldsymbol{w} \in S^*$, $s \in S_1$, and $t \in S$, if $\sigma(\boldsymbol{w} \cdot s)(t) > 0$, then

$(s,t) \in E$. The strategies for player 2 are defined analogously. We denote by Σ and Π the set of all strategies for player 1 and player 2, respectively.

Once a starting state $s \in S$ and strategies $\sigma \in \Sigma$ and $\pi \in \Pi$ for the two players are fixed, the outcome of the game is a random walk $\omega_s^{\sigma,\pi}$ for which the probabilities of events are uniquely defined, where an *event* $\mathcal{A} \subseteq \Omega$ is a measurable set of paths. For a state $s \in S$ and an event $\mathcal{A} \subseteq \Omega$, we write $\Pr_s^{\sigma,\pi}(\mathcal{A})$ for the probability that a path belongs to $\mathcal{A}$ if the game starts from the state s and the players follow the strategies σ and π, respectively. For a measurable function $f : \Omega \to \mathbb{R}$ we denote by $\mathbb{E}_s^{\sigma,\pi}[f]$ the *expectation* of the function f under the probability measure $\Pr_s^{\sigma,\pi}(\cdot)$.

Strategies that do not use randomization are called pure. A player-1 strategy σ is *pure* if for all $\boldsymbol{w} \in S^*$ and $s \in S_1$, there is a state $t \in S$ such that $\sigma(\boldsymbol{w} \cdot s)(t) = 1$. A *memoryless* player-1 strategy does not depend on the history of the play but only on the current state; it can be represented as a function σ: $S_1 \to \mathcal{D}(S)$. A *pure memoryless strategy* is a strategy that is both pure and memoryless. A pure memoryless strategy for player 1 can be represented as a function σ: $S_1 \to S$. We denote by Σ^{PM} the set of pure memoryless strategies for player 1. The pure memoryless player-2 strategies Π^{PM} are defined analogously.

Given a pure memoryless strategy $\sigma \in \Sigma^{PM}$, let G_σ be the game graph obtained from G under the constraint that player 1 follows the strategy σ. The corresponding definition G_π for a player-2 strategy $\pi \in \Pi^{PM}$ is analogous, and we write $G_{\sigma,\pi}$ for the game graph obtained from G if both players follow the pure memoryless strategies σ and π, respectively. Observe that given a $2^1\!/\!_2$-player game graph G and a pure memoryless player-1 strategy σ, the result G_σ is a player-2 MDP. Similarly, for a player-1 MDP G and a pure memoryless player-1 strategy σ, the result G_σ is a Markov chain. Hence, if G is a $2^1\!/\!_2$-player game graph and the two players follow pure memoryless strategies σ and π, the result $G_{\sigma,\pi}$ is a Markov chain.

Objectives. We specify objectives for the players by providing a set of *winning* plays $\Phi \subseteq \Omega$ for each player, or a measurable function $f : \Omega \to \mathbb{R}$ for each player. We say that a play ω *satisfies* the objective Φ if $\omega \in \Phi$. We study only zero-sum games, where the objectives of the two players are complementary; i.e., if player 1 has the objective Φ, then player 2 has the objective $\Omega \setminus \Phi$; or if the objective for player 1 is f, then the objective for player 2 is $-f$. We consider ω-*regular objectives* [Tho97], specified as parity conditions, and mean-payoff (or long -run average) objective. We also define the special case of reachability objectives.

- *Reachability objectives.* Given a set $T \subseteq S$ of "target" states, the reachability objective requires that some state of T be visited. The set of winning plays is $\text{Reach}(T) = \{ \omega = \langle s_0, s_1, s_2, \ldots \rangle \in \Omega \mid s_k \in T \text{ for some } k \geq 0 \}$.
- *Parity objectives.* For $c, d \in \mathbb{N}$, we write $[c..d] = \{ c, c+1, \ldots, d \}$. Let p: $S \to [0..d]$ be a function that assigns a *priority* $p(s)$ to every state $s \in S$, where $d \in \mathbb{N}$. For a play $\omega = \langle s_0, s_1, \ldots \rangle \in \Omega$, we define $\text{Inf}(\omega) = \{ s \in S \mid s_k = s \text{ for infinitely many } k \}$ to be the set of states that occur infinitely often in ω. The *even-parity objective* is defined as $\text{Parity}(p) = \{ \omega \in \Omega \mid$

$\max\big(p(\mathrm{Inf}(\omega))\big)$ is even $\}$, and the *odd-parity objective* as $\mathrm{coParity}(p) = \{\, \omega \in \Omega \mid \max\big(p(\mathrm{Inf}(\omega))\big)$ is odd $\}$.

- *Mean-payoff objectives.* Let $r : S \to \mathbb{R}$ be a real-valued reward function that assigns to every state s the reward $r(s)$ assigned to s. The mean-payoff objective MP assigns to every play the long-run average of the rewards appearing in the play. Formally, for a play $\omega = \langle s_1, s_2, s_3, \ldots \rangle$ we have

$$\mathsf{MP}(r)(\omega) = \lim\inf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r(s_i).$$

The complementary objective $-\mathsf{MP}$ is defined as follows

$$-\mathsf{MP}(r)(\omega) = \lim\sup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} -(r(s_i)).$$

Optimal strategies. Given objectives $\Phi \subseteq \Omega$ for player 1 and $\Omega \setminus \Phi$ for player 2, and measurable functions f and $-f$ for player 1 and player 2, respectively, we define the *value* functions $\langle\langle 1\rangle\rangle_{val}$ and $\langle\langle 2\rangle\rangle_{val}$ for the players 1 and 2, respectively, as the following functions from the state space S to the set $\mathbb{R}$ of reals: for all states $s \in S$, let

$$\langle\langle 1\rangle\rangle_{val}(\Phi)(s) = \sup_{\sigma\in\Sigma}\inf_{\pi\in\Pi} \mathrm{Pr}_s^{\sigma,\pi}(\Phi); \qquad \langle\langle 1\rangle\rangle_{val}(f)(s) = \sup_{\sigma\in\Sigma}\inf_{\pi\in\Pi} \mathbb{E}_s^{\sigma,\pi}[f];$$

$$\langle\langle 2\rangle\rangle_{val}(\Omega\setminus\Phi)(s) = \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathrm{Pr}_s^{\sigma,\pi}(\Omega\setminus\Phi); \qquad \langle\langle 2\rangle\rangle_{val}(-f)(s) = \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}_s^{\sigma,\pi}[-f].$$

In other words, the value $\langle\langle 1\rangle\rangle_{val}(\Phi)(s)$ and $\langle\langle 1\rangle\rangle_{val}(f)(s)$ gives the maximal probability and expectation with which player 1 can achieve her objective Φ and f from state s, and analogously for player 2. The strategies that achieve the value are called optimal: a strategy σ for player 1 is *optimal* from the state s for the objective Φ if $\langle\langle 1\rangle\rangle_{val}(\Phi)(s) = \inf_{\pi\in\Pi} \mathrm{Pr}_s^{\sigma,\pi}(\Phi)$; and σ is optimal from the state s for f if $\langle\langle 1\rangle\rangle_{val}(f)(s) = \inf_{\pi\in\Pi} \mathbb{E}_s^{\sigma,\pi}[f]$. The optimal strategies for player 2 are defined analogously. We now state the classical determinacy results for $2^1\!/\!_2$-player parity and mean-payoff games.

Theorem 1 (Quantitative determinacy). *For all $2^1\!/\!_2$-player game graphs $G = ((S,E),(S_1,S_2,S_\bigcirc),\delta)$ the following assertions hold.*

1. [LL69] *For all reward functions $r : S \to \mathbb{R}$, and all states s, we have $\langle\langle 1\rangle\rangle_{val}(\mathsf{MP}(r))(s) + \langle\langle 2\rangle\rangle_{val}(-\mathsf{MP}(r))(s) = 0$. Pure memoryless optimal strategies exist for both players from all states s.*
2. [CJH04,MM02,Zie04] *For all parity objectives Φ, and all states s, we have $\langle\langle 1\rangle\rangle_{val}(\Phi)(s) + \langle\langle 2\rangle\rangle_{val}(\Omega\setminus\Phi)(s) = 1$. Pure memoryless optimal strategies exist for both players from all states s.*

Since in $2^1\!/\!_2$-player games with parity and mean-payoff objectives pure memoryless strategies suffice for optimality, in the sequel we consider only pure memoryless strategies.

3 Reduction of $2^1/_2$ Player Parity to Mean-payoff Games

In this section we present a polynomial time Turing reduction of $2^1/_2$-player parity games to $2^1/_2$-player mean-payoff games. The reduction will be obtained in two stages. The first stage consists of computation of set of states with value 1 for a parity objective (or the set of *almost-sure* winning states). The second stage consists of the reduction after the computation of almost-sure winning states. We first define the set of almost-sure winning states for parity objectives.

Almost-sure winning states. Given a $2^1/_2$-player game graph G with a parity objective Φ for player 1 we denote by

$$W_1^G(\Phi) = \{s \in S \mid \langle\langle 1\rangle\rangle_{val}(\Phi)(s) = 1\}; \quad W_2^G(\Omega\setminus\Phi) = \{s \in S \mid \langle\langle 2\rangle\rangle_{val}(\Omega\setminus\Phi)(s) = 1\};$$

the set of states such that the values for player 1 and player 2 are 1, respectively. These sets of states are also referred as the almost-sure winning states for the players.

Reduction for almost-sure winning states. The computation of almost-sure winning states in $2^1/_2$-player games with parity objectives by computation of values in mean-payoff games can be achieved as follows. The results of [CJH03] shows that the computation of almost-sure winning states in a $2^1/_2$-player game graph $G = ((S, E), (S_1, S_2, S_\bigcirc), \delta)$ with a parity objective with d priorities can be achieved by a reduction to a 2-player game graph with $|S| \cdot d$ states, and a parity objective with $d+1$ parities. The result of [Jur98] establishes a polynomial time reduction of 2-player games with parity objectives to 2-player games with mean-payoff objectives. The above two reduction ensures that the computation of almost-sure winning states in $2^1/_2$-player games with parity objectives can be reduced to the computation of 2-player games with mean-payoff objectives.

Reduction for value computation. We now present a reduction of $2^1/_2$-player parity games to $2^1/_2$-player mean-payoff games for value computation. Note that the computation of almost-sure winning states can be achieved by solving 2-player (and hence $2^1/_2$-player) mean-payoff games. Theorem 2 presents the reduction for value computation. We first present a lemma that will be used in the proof of Theorem 2.

Lemma 1. *Let C be a closed connected recurrent set of states in a Markov chain with minimum non-zero transition probability as $\delta_{\min} > 0$. For $s, s_0 \in C$, let*

$$\mathsf{freq}(s, s_0) = \lim\inf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \Pr_{s_0}(X_t = s),$$

where X_t is a random variable denoting the t-th state of a path, denote the "long-run" frequency of state s with starting state s_0. Then for all $s, s_0 \in C$ we have

$$\mathsf{freq}(s, s_0) \geq \frac{1}{n} \cdot \left(\delta_{\min}\right)^n,$$

where $n = |C|$.

Proof. For a state $t \in C$, let $\mathsf{In}(t) = \{ s \in C \mid \delta(s)(t) > 0 \}$ be the set of states with incoming edges to t. We start with two simple facts.

– Fact 1. For a state $t \in C$, for all $s_0 \in C$ we have

$$\mathsf{freq}(t, s_0) \geq \mathsf{freq}(s, s_0) \cdot \delta(s)(t) \geq \mathsf{freq}(s, s_0) \cdot \delta_{\min}; \qquad \text{for } s \in \mathsf{In}(t).$$

– Fact 2. We have $\sum_{t \in C} \mathsf{freq}(t, s_0) = 1$.

The first fact relates the "long-run" frequency of a state to the "long-run" frequency of the predecessors, and since C is a closed connected recurrent set of states, the sum of the "long-run" frequencies of states in C is 1. We prove the desired result by an argument by contradiction. Assume towards contradiction that there exist $t, s_0 \in C$ with $\mathsf{freq}(t, s_0) < \dfrac{1}{n} \cdot \big(\delta_{\min}\big)^n$. It follows from fact 1, that for all states $s \in \mathsf{In}(t)$ we have $\mathsf{freq}(s, s_0) < \dfrac{1}{n} \cdot \big(\delta_{\min}\big)^{n-1}$. Again for a state $s \in \mathsf{In}(t)$, for all $s' \in \mathsf{In}(s)$ we have $\mathsf{freq}(s', s_0) < \dfrac{1}{n} \cdot \big(\delta_{\min}\big)^{n-2}$, and so on. Since $|C| = n$, it follows that for all states $s \in C$ we have $\mathsf{freq}(s, s_0) < \frac{1}{n}$. Again as $|C| = n$, this contradicts fact 2 that $\sum_{s \in C} \mathsf{freq}(s, s_0) = 1$. Hence the desired result follows. ∎

Theorem 2. *Let $G = ((S, E), (S_1, S_2, S_{\bigcirc}), \delta)$ be a $2^1\!/_2$-player game graph. Let $p : S \to [0..d]$ be a priority function, and let $W_1 = W_1^G(\text{Parity}(p))$ and $W_2 = W_2^G(\text{coParity}(p))$ be the set of almost-sure winning states for the players. Let*

$$\delta_{\min} = \min\{ \delta(s)(t) \mid s \in S_{\bigcirc}, t \in S, \delta(s)(t) > 0 \} > 0.$$

Consider the reward function $r : S \to \mathbb{R}$ as follows:

$$r(s) = \begin{cases} 1 & s \in W_1; \\ -1 & s \in W_2; \\ (-1)^k \cdot (2 \cdot n)^k \cdot \left(\dfrac{1}{\delta_{\min}}\right)^{n \cdot k} & p(s) = k, s \in S \setminus (W_1 \cup W_2); \end{cases}$$

where $n = |S|$. Then for all $s \in S \setminus (W_1 \cup W_2)$ we have

$$\langle\!\langle 1 \rangle\!\rangle_{val}(\text{Parity}(p))(s) = \frac{1}{2} \cdot \Big(\langle\!\langle 1 \rangle\!\rangle_{val}(\mathsf{MP}(r))(s) + 1 \Big).$$

Proof. We prove the following two inequalities.

1. We first prove that for all $s \in S \setminus (W_1 \cup W_2)$ we have

$$\langle\!\langle 1 \rangle\!\rangle_{val}(\text{Parity}(p))(s) \leq \frac{1}{2} \cdot \Big(\langle\!\langle 1 \rangle\!\rangle_{val}(\mathsf{MP}(r))(s) + 1 \Big).$$

Consider a pure memoryless optimal strategy σ for player 1 for the parity objective Parity(p). Fix the strategy in the mean-payoff game, and consider a pure memoryless counter-optimal strategy π for player 2 in the MDP G_σ (i.e., the strategy π is optimal in G_σ for the objective $-\mathsf{MP}(r)$). We first show that

$$\Pr_s^{\sigma,\pi}(\text{Reach}(W_2)) \leq \langle\langle 2\rangle\rangle_{val}(\text{coParity}(p))(s) = 1 - \langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s).$$

Otherwise, if $\Pr_s^{\sigma,\pi}(\text{Reach}(W_2)) > \langle\langle 2\rangle\rangle_{val}(\text{coParity}(p))(s)$, then player 2 plays π to reach W_2 and an almost-sure winning strategy for coParity(p) from W_2 to ensure that the probability to satisfy coParity(p) given σ is greater than $\langle\langle 2\rangle\rangle_{val}(\text{coParity}(p))(s)$; this contradicts that σ is optimal. Now consider the Markov chain $G_{\sigma,\pi}$. Let C be a closed connected recurrent set of states in $G_{\sigma,\pi}$. If $C \cap (S \setminus (W_1 \cup W_2)) \neq \emptyset$, then there is a state $s' \in C \cap (S \setminus (W_1 \cup W_2))$ with $\langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s') > 0$. Since σ is optimal for player 1 for Parity(p) and in $G_{\sigma,\pi}$ from s' the set C is visited infinitely often with probability 1, it follows that $\max(p(C))$ is even. Let $z \in C$ be a state with $p(z) = \max(p(C))$. Then since the minimum transition probability is $\delta_{\min}$ and $|C| \leq |S|$, it follows from Lemma 1 that the long-run frequency for state z is at least $\frac{1}{n} \cdot (\delta_{\min})^n$. The reward assignment ensures that the long-run average for the closed connected recurrent set C is at least 1. This is obtained as follows. If $p(z) = 0$, then for all states $s \in C$ we must have $p(s) = p(z) = 0$, and then long-run average for C is $(2 \cdot n)^0 \cdot \left(\frac{1}{\delta_{\min}}\right)^{n \cdot 0} = 1$. We consider the case with $p(z) \geq 2$ and then long-run average contribution by z is at least

$$\frac{1}{n} \cdot (\delta_{\min})^n \cdot (2 \cdot n)^{p(z)} \cdot \left(\frac{1}{\delta_{\min}}\right)^{n \cdot p(z)} = 2 \cdot \left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}}\right)^{n \cdot (p(z)-1)}\right);$$

(this obtained by multiplying the long-run frequency of z along with its reward). Since $p(z)$ is the greatest priority appearing in C, the long-run average contribution of all the other states in C is at least

$$-\left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}}\right)^{n \cdot (p(z)-1)}\right),$$

(in the worst case all other states have priority $p(z) - 1$). Hence the long-run average in C is at least

$$\left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}}\right)^{n \cdot (p(z)-1)}\right);$$

the claim follows. A lower bound on the long-run average payoff for player 1 is obtained as follows: we consider the maximum probability of reaching W_2 and consider the closed connected recurrent states C that intersect with W_2 is contained in W_2 (and the long-run average is -1 in this case) and with the rest of the probability the long-run average is at least 1. Hence we have

$$\begin{aligned}\langle\langle 1\rangle\rangle_{val}(\mathsf{MP}(r))(s) &\geq (-1) \cdot \left(1 - \langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s)\right) + 1 \cdot \langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s) \\ &= 2 \cdot \langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s) - 1.\end{aligned}$$

2. We now prove that for all $s \in S \setminus (W_1 \cup W_2)$ we have

$$\langle\langle 1 \rangle\rangle_{val}(\text{Parity}(p))(s) \geq \frac{1}{2} \cdot \left(\langle\langle 1 \rangle\rangle_{val}(\mathsf{MP}(r))(s) + 1 \right).$$

Consider a pure memoryless optimal strategy π for player 2 for the objective coParity(p). Fix the strategy in the mean-payoff game, and consider a pure memoryless counter-optimal strategy σ for player 1 in the MDP G_π (i.e., the strategy σ is optimal in G_σ for the objective $\mathsf{MP}(r)$). We first show that

$$\Pr_s^{\sigma,\pi}(\text{Reach}(W_1)) \leq \langle\langle 1 \rangle\rangle_{val}(\text{Parity}(p))(s).$$

Otherwise, if $\Pr_s^{\sigma,\pi}(\text{Reach}(W_1)) > \langle\langle 1 \rangle\rangle_{val}(\text{Parity}(p))(s)$, then player 1 plays σ to reach W_1 and an almost-sure winning strategy for Parity(p) from W_1 to ensure that the probability to satisfy Parity(p) given π is greater than $\langle\langle 1 \rangle\rangle_{val}(\text{Parity}(p))(s)$; this contradicts that π is optimal. Now consider the Markov chain $G_{\sigma,\pi}$. Let C be a closed connected recurrent set of states in $G_{\sigma,\pi}$. If $C \cap (S \setminus (W_1 \cup W_2)) \neq \emptyset$, then there is a state $s' \in C \cap (S \setminus (W_1 \cup W_2))$ with $\langle\langle 2 \rangle\rangle_{val}(\text{coParity}(p))(s') > 0$. Since π is optimal for player 2 for coParity(p) and in $G_{\sigma,\pi}$ from s' the set C is visited infinitely often with probability 1, it follows that $\max(p(C))$ is odd. Let $z \in C$ be a state with $p(z) = \max(p(C))$. Then since the minimum transition probability is $\delta_{\min}$ and $|C| \leq |S|$, it follows from Lemma 1 that the long-run frequency for state z is at least $\frac{1}{n} \cdot (\delta_{\min})^n$. The reward assignment ensures that the long-run average for the closed connected recurrent set C is at most -1. This is obtained as follows: the long-run average contribution by z is at most

$$\frac{1}{n} \cdot (\delta_{\min})^n \cdot (-1) \cdot (2 \cdot n)^{p(z)} \cdot \left(\frac{1}{\delta_{\min}} \right)^{n \cdot p(z)} = (-2) \cdot \left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}} \right)^{n \cdot (p(z)-1)} \right);$$

(this obtained by multiplying the long-run frequency of z along with its reward). Since $p(z)$ is the greatest priority appearing in C, the long-run average contribution of all the other states in C is at most

$$\left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}} \right)^{n \cdot (p(z)-1)} \right).$$

(in the worst case all other states have priority $p(z)-1$). Hence the long-run average in C is at most

$$-\left((2 \cdot n)^{p(z)-1} \cdot \left(\frac{1}{\delta_{\min}} \right)^{n \cdot (p(z)-1)} \right);$$

the claim follows. An upper bound on the long-run average payoff for player 1 is obtained as follows: we consider the maximum probability of reaching W_1 and consider the closed connected recurrent states C that intersect with W_1

is contained in W_1 (and the long-run average is 1 in this case) and with the rest of the probability the long-run average is at most -1. Hence we have

$$\begin{aligned}\langle\langle 1\rangle\rangle_{val}(\mathsf{MP}(r))(s) &\leq 1\cdot\langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s) + (-1)\cdot\left(1-\langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s)\right)\\ &= 2\cdot\langle\langle 1\rangle\rangle_{val}(\text{Parity}(p))(s) - 1\end{aligned}$$

The desired result follows. ∎

Remark. In the proof of Theorem 2 we used existence of pure memoryless optimal strategies in $2^1\!/\!_2$-player games graphs with parity objectives and existence of pure memoryless optimal strategies in MDPs with mean-payoff objectives. The proof does not rely on existence of pure memoryless optimal strategies in $2^1\!/\!_2$-player game graphs with mean-payoff objectives.

Reduction to mean-payoff games. The reduction of $2^1\!/\!_2$-player games with parity objectives to $2^1\!/\!_2$-player games with mean-payoff objectives is achieved in Theorem 2. We argue that the reduction is polynomial. The size of a game graph $G = ((S,E),(S_1,S_2,S_\bigcirc),\delta)$ is

$$|G| = |S| + |E| + \sum_{t\in S}\sum_{s\in S_\bigcirc} |\delta(s)(t)|;$$

where $|\delta(s)(t)|$ denotes the space to represent the transition probability $\delta(s)(t)$ in binary. The reduction of Theorem 2 is polynomial, since the reward at every state can be expressed in $n\cdot d\cdot |G|\cdot \log(n)$ bits, and $d\leq n$. Hence Theorem 2 achieves a polynomial time Turing reduction of $2^1\!/\!_2$-player parity games to $2^1\!/\!_2$-player mean-payoff games.

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

References

[CJH03] K. Chatterjee, M. Jurdziński, and T.A. Henzinger. Simple stochastic parity games. In *CSL'03*, volume 2803 of *LNCS*, pages 100–113. Springer, 2003.

[CJH04] K. Chatterjee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *SODA'04*, pages 121–130. SIAM, 2004.

[Con92] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

[FV97] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1997.

[Jur98] M. Jurdzinski. Deciding the winner in parity games is in UP ∩ co-UP. *Information Processing Letters*, 68(3):119–124, 1998.

[LL69] T. A. Liggett and S. A. Lippman. Stochastic games with perfect information and time average payoff. *Siam Review*, 11:604–607, 1969.

[MM02] A. McIver and C. Morgan. Games, probability, and the quantitative μ-calculus $q\mu$. In *LPAR'02*, pages 292–310. LNCS 2514 Springer, 2002.

[Put94] M.L. Puterman. *Markov Decision Processes*. John Wiley and Sons, 1994.

[Tho97] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.

[Zie04] W. Zielonka. Perfect-information stochastic parity games. In *FoSSaCS'04*, pages 499–513. LNCS 2987, Springer, 2004.