The Complexity of Solving Stochastic Games on Graphs*

Daniel Andersson and Peter Bro Miltersen

Department of Computer Science, Aarhus University, Denmark

Abstract. We consider some well-known families of two-player zero-sum perfect-information stochastic games played on finite directed graphs. Generalizing and unifying results of Liggett and Lippman, Zwick and Paterson, and Chatterjee and Henzinger, we show that the following tasks are polynomial-time (Turing) equivalent.

- Solving *stochastic parity games*,
- Solving *simple stochastic games*,
- Solving *stochastic terminal-payoff games* with payoffs and probabilities given in unary,
- Solving stochastic terminal-payoff games with payoffs and probabilities given in binary,
- Solving *stochastic mean-payoff games* with rewards and probabilities given in unary,
- Solving stochastic mean-payoff games with rewards and probabilities given in binary,
- Solving *stochastic discounted-payoff games* with discount factor, rewards and probabilities given in binary.

It is unknown whether these tasks can be performed in polynomial time. In the above list, "solving" may mean either *quantitatively* solving a game (computing the values of its positions) or *strategically* solving a game (computing an optimal strategy for each player). In particular, these two tasks are polynomial-time equivalent for all the games listed above. We also consider a more refined notion of equivalence between quantitatively and strategically solving a game. We exhibit a linear time algorithm that given a simple stochastic game or a terminal-payoff game *and* the values of all positions of that game, computes a pair of optimal strategies. Consequently, for any restriction one may put on the simple stochastic game model, quantitatively solving is polynomial-time equivalent to strategically solving the resulting class of games.

1 Introduction

We consider some well-known families of two-player zero-sum perfect-information stochastic games played on finite directed graphs.

* Work supported by *Center for Algorithmic Game Theory*, funded by the Carlsberg Foundation. A large fraction of the results of this paper appeared in a preprint [12], co-authored by Vladimir Gurvich and the second author of this paper. Vladimir Gurvich's contributions to that preprint will appear elsewhere.

- *Simple stochastic games* were introduced to the algorithms and complexity community by Condon [6], who was motivated by the study of randomized Turing machine models. A simple stochastic game is given by a finite directed graph $G = (V, E)$, with the set of vertices V also called *positions* and the set of arcs E also called *actions*. There is a partition of the positions into V_1 (positions belonging to Player 1), V_2 (positions belonging to Player 2), V_R (coin-toss positions), and a special terminal position $\mathbf{1}$. Positions of V_R have exactly two outgoing arcs, the terminal position $\mathbf{1}$ has none, while all positions in V_1, V_2 have at least one outgoing arc. Between moves, a pebble is resting at some vertex u. If u belongs to a player, this player should strategically pick an outgoing arc from u and move the pebble along this edge to another vertex. If u is a vertex in V_R, nature picks an outgoing arc from u uniformly at random and moves the pebble along this arc. The objective of the game for Player 1 is to reach $\mathbf{1}$ and should play so as to maximize his probability of doing so. The objective for Player 2 is to prevent Player 1 from reaching $\mathbf{1}$.
- *Stochastic terminal-payoff games* is the natural generalization of Condon's simple stochastic games where we allow
  1. each vertex in V_R to have more than two outgoing arcs and an arbitrary rational valued probability distribution on these,
  2. several terminals with different payoffs, positive or negative.

  In a stochastic terminal-payoff game, the *outcome* of the game is the payoff of the terminal reached, with the outcome being 0 if play is infinite. Generalizing the objectives of a simple stochastic game in the natural way, the objective for Player 1 is to maximize the expected outcome while the objective for Player 2 is to minimize it. Such generalized simple stochastic games have sometimes also been called simple stochastic games in the literature (e.g., [10]), but we shall refer to them as stochastic terminal-payoff games.
- *Stochastic parity games* were introduced by Chatterjee, Jurdziński, and Henzinger at SODA'04 [4] and further studied in [3,2]. They are a natural generalization of the non-stochastic parity games of McNaughton [17], the latter having central importance in the computer-aided verification community, as solving them is equivalent to model checking the μ-calculus [7]. As for the case of simple stochastic games, a stochastic parity game is given by a directed graph $G = (V, E)$ with a partition of the vertices into V_1 (vertices belonging to Player 1), V_2 (vertices belonging to Player 2), and V_R (random vertices). Vertices of V_R have exactly two outgoing arcs, while all vertices in V_1, V_2 have at least one outgoing arc. Also, each vertex is assigned an integral *priority*. Between moves, a pebble is resting at some vertex u. If u belongs to a player, this player should strategically pick an outgoing arc from u and move the pebble along this edge to another vertex. If u is a vertex in V_R, nature picks an outgoing arc from u uniformly at random and moves the pebble along this arc. The play continues forever. If the highest priority that appears infinitely often during play is odd, Player 1 wins the game; if it is even, Player 2 wins the game.

– *Stochastic mean-payoff games* and *stochastic discounted-payoff games* were first studied in the game theory community by Gillette [9] as the perfect information special case of the stochastic games of Shapley [18]. A stochastic mean-payoff or discounted-payoff game G is given by a finite set of positions V, partitioned into V_1 (positions belonging to Player 1) and V_2 (positions belonging to Player 2). To each position u is associated a finite set of possible actions. To each such action is associated a real-valued *reward* and a probability distribution on positions. At any point in time of play, the game is in a particular position i. The player to move chooses an action strategically and the corresponding reward is paid by Player 2 to Player 1. Then, nature chooses the next position at random according to the probability distribution associated with the action. The play continues forever and the sum of rewards may therefore be unbounded. Nevertheless, one can associate a finite payoff to the players in spite of this, in more ways than one (so G is not just one game, but really a family of games): For a stochastic discounted-payoff game, we fix a *discount factor* $\beta \in (0, 1)$ and define the outcome of the play (the payoff to Player 1) to be

$$\sum_{i=0}^{\infty} \beta^i r_i$$

where r_i is the reward incurred at stage i of the game. We shall denote the resulting game G_β. For a stochastic mean-payoff game we define the outcome of the play (the payoff to Player 1) to be the *limiting average* payoff

$$\liminf_{n\to\infty} (\sum_{i=0}^{n} r_i)/(n + 1).$$

We shall denote the resulting game G_1. A natural restriction of stochastic mean-payoff games is to *deterministic* transitions (i.e., all probability distributions put all probability mass on one position). This class of games has been studied in the computer science literature under the names of cyclic games [11] and mean-payoff games [19]. We shall refer to them as deterministic mean-payoff games in this paper.

A *strategy* for a game is a (possibly randomized) procedure for selecting which arc or action to take, given the history of the play so far. A *pure positional strategy* is the very special case of this where the choice is deterministic and only depends on the current position, i.e., a pure positional strategy is simply a map from positions to actions. If Player 1 plays using strategy x and Player 2 plays using strategy y, and the play starts in position i, a random play $P(x, y, i)$ of the game is induced. We let $u^i(x, y)$ denote the expected outcome of this play (for stochastic terminal-payoff, discounted-payoff, and mean-payoff games) or the winning probability of Player 1 (for simple stochastic games and stochastic parity games). A strategy x^* for Player 1 is called *optimal* if for any position i:

$$\inf_{y\in S_2} u^i(x^*, y) \geq \sup_{x\in S_1} \inf_{y\in S_2} u^i(x, y) \tag{1}$$

where S_1 (S_2) is the set of strategies for Player 1 (Player 2). Similarly, a strategy y^* for Player 2 is said to be optimal if

$$\sup_{x \in S_1} u^i(x, y^*) \leq \inf_{y \in S_2} \sup_{x \in S_1} u^i(x, y). \tag{2}$$

For all games described here, the references above (a proof of Liggett and Lippman [15] fixes a bug of a proof of Gillette [9] for the mean-payoff case) show:

- Both players have pure positional optimal strategies x^*, y^*.
- For such optimal x^*, y^* and for all positions i,

$$\inf_{y \in S_2} u^i(x^*, y) = \sup_{x \in S_1} u^i(x, y^*).$$

  This number is called the *value* of position i. We shall denote it $\mathrm{val}(i)$.

These facts imply that when testing whether conditions (1) and (2) hold, it is enough to take the infima and suprema over the finite set of pure positional strategies of the players.

In this paper, we consider *solving* games. By solving a game G, we may refer to two distinct tasks.

- *Quantitatively solving* G is the task of computing the values of all positions of the game, given an explicit representation of G.
- *Strategically solving* G is the task of computing a pair of optimal positional strategies for the game, given an explicit representation of G.

To be able to explicitly represent the games, we assume that the discount factor, rewards and probabilities are rational numbers and given as fractions in binary or unary. The notion of "quantitatively solving" is standard terminology in the literature (e.g., [4]), while the notion of "strategically solving" is not standard. We believe the distinction is natural, in particular since recent work of Hansen *et al.* [13] shows that for some classes of games more general than the ones considered in this paper, the two notions are not equivalent. Still, it is relatively easy to see that for all the games considered here, once the game has been strategically solved, we may easily solve it quantitatively.

With so many distinct computational problems under consideration, we shall for this paper introduce some convenient notation for them. We will use superscripts q/s to distinguish quantitative/strategic solutions and subscripts b/u to distinguish binary/unary input encoding (when applicable). For instance, $\textsc{Mean}^{\mathrm{s}}_{\mathrm{b}}$ is the problem of solving a stochastic mean-payoff game strategically with probabilities and rewards given in binary, and $\textsc{Simple}^{\mathrm{q}}$ is the problem of solving a simple stochastic game quantitatively. We use "$\preceq$" to express polynomial-time (Turing) reducibility.

Solving the games above in polynomial time are all celebrated open problems (e.g., [6,4]). Also, some polynomial time reductions between these challenging tasks were known. Some classes of games are obviously special cases of others (e.g., simple stochastic games and stochastic terminal-payoff games), leading to trivial reductions, but more intricate reductions were also known. In particular, a recent paper by Chatterjee and Henzinger [2] shows that solving stochastic parity games reduces to solving stochastic mean-payoff games. Earlier, Zwick and Paterson [19] showed that solving *deterministic* mean-payoff games reduces to solving simple stochastic games. However, in spite of these reductions and the fact that similar kinds of (non-polynomial time) algorithms, such as value iteration and strategy iteration are used to actually solve all of these games in practice, it does not seem that it was believed or even suggested that the tasks of solving these different classes of games were all polynomial-time equivalent. Our first main result is that they are, unifying and generalizing the previous reductions (and also using them in a crucial way):

Theorem 1. *The following tasks are polynomial-time (Turing) equivalent.*

- *Solving stochastic parity games,*
- *Solving simple stochastic games,*
- *Solving stochastic terminal-payoff games with unary payoffs and probabilities,*
- *Solving stochastic terminal-payoff games with binary payoffs and probabilities,*
- *Solving stochastic mean-payoff games with unary rewards and probabilities,*
- *Solving stochastic mean-payoff games with binary rewards and probabilities,*
- *Solving stochastic discounted-payoff games with binary discount factor, rewards and probabilities.*

Solving here may mean either "quantitatively" or "strategically". In particular, the two tasks are polynomial-time equivalent for all these classes of games.

Note that the equivalence between solving games with unary input encoding and solving games with binary input encoding means that there are pseudopolynomial-time algorithms for solving these games if and only there are polynomial-time algorithms. Note also that a "missing bullet" in the theorem is solving stochastic discounted-payoff games given in unary representation. It is in fact known that this can be done in polynomial time (even if only the discount factor is given in unary), see Littman [16, Theorem 3.4].

One obvious interpretation of our result is that Condon "saw right" when she singled out simple stochastic games as a representative model. In fact, our result implies that if simple stochastic games can be solved in polynomial time, then the same is true for essentially all important classes of two-player zero-sum perfect-information stochastic games considered in the literature.

Our second main result takes a closer look at the equivalence between quantitatively solving the games we consider and strategically solving them. Even though Theorem 1 shows these two tasks to be polynomial-time equivalent, the reductions from strategically solving games to quantitatively solving games in general changes the game under consideration, i.e., strategically solving one game reduces to quantitatively solving another. In other words, Theorem 1 does *not* mean *a priori* that once a game has been quantitatively solved, we can easily solve it strategically. However, for the case of discounted-payoff games, we trivially can: An optimal action can be determined "locally" by comparing sums of local rewards and values. Our second result shows that we (less trivially) can do such "strategy recovery" also for the case of terminal-payoff games (and therefore also for simple stochastic games).

Theorem 2. *Given a stochastic terminal-payoff game and the values of all its vertices, optimal pure positional strategies can be computed in linear time.*

Theorem 2 means that algorithms that efficiently and quantitatively solve simple stochastic games satisfying certain conditions (such as the algorithm of Gimbert and Horn [10], which is efficient when the number of coin-toss vertices is small) can also be used to strategically solve those games, in essentially the same time bound. We leave as an open problem whether a similar algorithm (even a polynomial-time one) can be obtained for stochastic mean-payoff games.

2 Proof of Theorem 1

Figure 1 shows a minimal set of reductions needed to establish all equivalences. We first enumerate a number of trivial and known reductions and afterwards fill in the remaining "gaps".

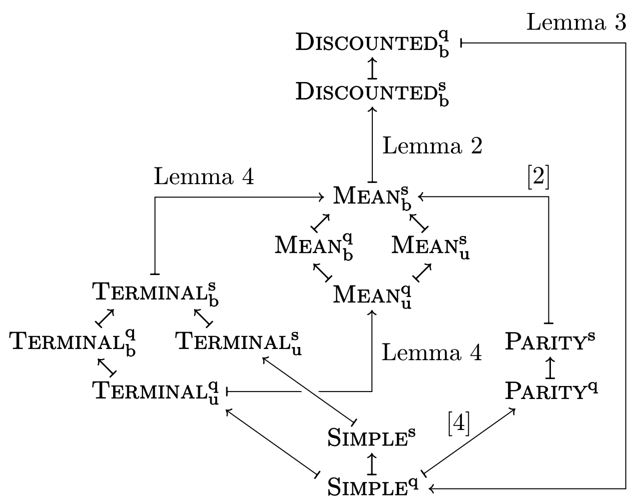

Fig. 1. Reductions used in the proof of Theorem 1.

For all games considered here, it is well-known that quantitatively solving them reduces to strategically solving them: Once the strategies have been fixed, a complete analysis of the resulting finite-state random process can be obtained using linear algebra and the theory of Markov chains [14]. Also, for the case of stochastic discounted-payoff games, the converse reduction is also easy: A strategy is optimal if and only it in each position chooses an action that maximizes the sum of the reward obtained by this action and the discounted expected value of the next position. Thus, $\text{Discounted}^{\text{s}}_{\text{b}} \preceq \text{Discounted}^{\text{q}}_{\text{b}}$.

Of course, each problem with unary input encoding trivially reduces to the corresponding binary version. Also, it is obvious that stochastic terminal-payoff games generalize simple stochastic games, and the only numbers that appear are 0, 1 and $\frac{1}{2}$, so $\text{Simple}^{\text{q}} \preceq \text{Terminal}^{\text{q}}_{\text{u}}$ and $\text{Simple}^{\text{s}} \preceq \text{Terminal}^{\text{s}}_{\text{u}}$.

Quantitatively solving simple stochastic games easily reduces to quantitatively solving stochastic parity games, as was noted in the original work on stochastic parity games [4]. Also, Chatterjee and Henzinger [2] show that strategically solving stochastic parity games reduces to strategically solving stochastic mean-payoff games.

Thus, to "complete the picture" and establish all equivalences, we only have to show: $\text{Mean}^{\text{s}}_{\text{b}} \preceq \text{Discounted}^{\text{s}}_{\text{b}}$,$\text{Discounted}^{\text{q}}_{\text{b}} \preceq \text{Simple}^{\text{q}}$,$\text{Terminal}^{\text{s}}_{\text{b}} \preceq \text{Mean}^{\text{s}}_{\text{b}}$ and $\text{Terminal}^{\text{q}}_{\text{u}} \preceq \text{Mean}^{\text{q}}_{\text{u}}$. These reductions are provided by the following lemmas.

Lemma 1. *Let G be a stochastic discounted-payoff/mean-payoff game with n positions and all transition probabilities and rewards being fractions with integral numerators and denominators, all of absolute value at most M. Let $\beta^* = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$ and let $\beta \in [\beta^*, 1)$. Then, any optimal pure positional strategy (for either player) in the discounted-payoff game G_β is also an optimal strategy in the mean-payoff game G_1.*

Proof. The fact that *some* β^* with the desired property exists is explicit in the proof of Theorem 1 of Liggett and Lippman [15]. We assume familiarity with that proof in the following. Here, we derive a concrete value for β^*. From the proof of Liggett and Lippman, we have that for x^* to be an optimal pure positional strategy (for Player 1) in G_1, it is sufficient to be an optimal pure positional strategy in G_β for all values of β sufficiently close to 1, i.e., to satisfy the inequalities

$$\min_{y \in S'_2} u^i_\beta(x^*, y) \geq \max_{x \in S'_1} \min_{y \in S'_2} u^i_\beta(x, y)$$

for all positions i and for all values of β sufficiently close to 1, where S'_1 (S'_2) is the set of pure positional, strategies for Player 1 (2) and u^i_β is the expected payoff when game starts in position i and the discount factor is β. Similarly, for y^* to be an optimal pure positional strategy (for Player 2) in G_1, it is sufficient to be an optimal pure positional strategy in G_β for all values of β sufficiently close to 1, i.e., to satisfy the inequalities

$$\max_{x \in S'_1} u^i_\beta(x, y^*) \leq \min_{y \in S'_2} \max_{x \in S'_1} u^i_\beta(x, y).$$

So, we can prove the lemma by showing that for all positions i and *all* pure positional strategies x, y, x', y', the sign of $u^i_\beta(x, y) - u^i_\beta(x', y')$ is the same for all $\beta \geq \beta^*$. For fixed strategies x, y we have that $v_i = u^i_\beta(x, y)$ is the expected total reward in a *discounted Markov process* and is therefore given by the formula (see [14])

$$v = (I - \beta Q)^{-1} r, \tag{3}$$

where v is the vector of $u_\beta(x, y)$ values, one for each position, Q is the matrix of transition probabilities and r is the vector of rewards (note that for *fixed* positional strategies x, y, rewards can be assigned to positions in the natural way). Let $\gamma = 1-\beta$. Then, (3) is a system of linear equations in the unknowns v, where each coefficient is of the form $a_{ij}\gamma + b_{ij}$ where a_{ij}, b_{ij} are rational numbers with numerators with absolute value bounded by $2M$ and with denominators with absolute value bounded by M. By multiplying the equations with all denominators, we can in fact assume that a_{ij}, b_{ij} are integers of absolute value less than $2M^n$. Solving the equations using Cramer's rule, we may write an entry of v as a quotient between determinants of $n \times n$ matrices containing terms of the form $a_{ij}\gamma + b_{ij}$. The determinant of such a matrix is a polynomial in γ of degree n with the coefficient of each term being of absolute value at most $n!(2M^n)^n = n!2^n M^{n^2}$. We denote these two polynomials p_1, p_2. Arguing similarly about $u_\beta(x', y')$ and deriving corresponding polynomials p_3, p_4, we have that $u^i_\beta(x, y) - u^i_\beta(x', y') \geq 0$ is equivalent to $p_1(\gamma)/p_2(\gamma) - p_3(\gamma)/p_4(\gamma) \geq 0$, i.e., $p_1(\gamma)p_4(\gamma) - p_3(\gamma)p_2(\gamma) \geq 0$. Letting $q(\gamma) = p_1(\gamma)p_4(\gamma) - p_3(\gamma)p_2(\gamma)$, we have that q is a polynomial in γ, with integer coefficients, all of absolute value at most $R = 2(n!)^2 2^{2n} M^{2n^2}$. Since $1 - \beta^* < 1/(2R)$, the sign of $q(\gamma)$ is the same for all $\gamma \leq 1 - \beta^*$, i.e., for all $\beta \geq \beta^*$. This completes the proof.

Lemma 2. $\textsc{Mean}^{\mathrm{s}}_{\mathrm{b}} \preceq \textsc{Discounted}^{\mathrm{s}}_{\mathrm{b}}$.

Proof. This follows immediately from Lemma 1 by observing that the binary representation of the number $\beta^* = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$ has length polynomial in the size of the representation of the given game.

Lemma 3. $\textsc{Discounted}^{\mathrm{q}}_{\mathrm{b}} \preceq \textsc{Simple}^{\mathrm{q}}$.

Proof. Zwick and Paterson [19] considered solving *deterministic* discounted-payoff games, i.e., games where the action taken deterministically determines the transition taken and reduced these to solving simple stochastic games. It is natural to try to generalize their reduction so that it also works for stochastic discounted-payoff games. We find that such reduction indeed works, even though the correctness proof of Zwick and Paterson has to be modified slightly compared to their proof. The details follow.

The redution proceeds in two steps: First we reduce to stochastic terminal-payoff games with 0/1 payoffs, and then to simple stochastic games.

We are given as input a stochastic discounted-payoff game G with discount factor β and must first produce a stochastic terminal-payoff game G' whose values can be used to construct the values for the stochastic discounted-payoff game G_β. First, we affinely scale and translate all rewards of G so that they are in the interval $[0, 1]$. This does not influence the optimal strategies, and all values are transformed accordingly. Vertices of G' include all positions of G (belonging to the same player in G' as in G), and, in addition, a random vertex $w_{u,A}$ for each possible action A of each position u of G. We also add terminals $\mathbf{0}$ and $\mathbf{1}$. We construct the arcs of G' by adding, for each (position,action) pair (u, A) the "gadget" indicated in Figure 2. To be precise, if the action has reward r

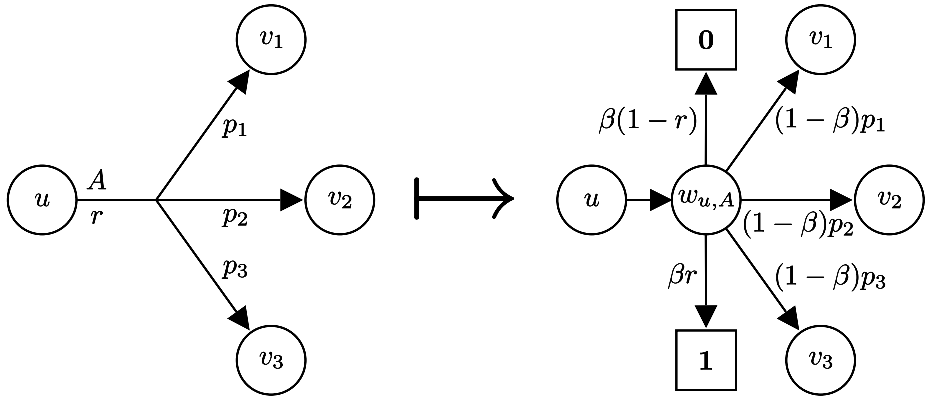

Fig. 2. Reducing discounted-payoff games to terminal-payoff games

and leads to positions $v_1, v_2, \ldots, v_k$ with probability weights $p_1, p_2, \ldots, p_k$, we include in G' an arc from u to $w_{u,A}$, arcs from $w_{u,A}$ to $v_1, \ldots, v_k$ with probability weights $(1-\beta)p_1, \ldots, (1-\beta)p_k$, an arc from $w_{u,A}$ to $\mathbf{0}$ with probability weight $\beta(1-r)$ and finally an arc from $w_{u,A}$ to the terminal $\mathbf{1}$ with probability weight βr.

There is clearly a one-to-one correspondence between strategies in G and in G'. To see the correspondance between values, fix a strategy profile and consider any play. By construction, if the expected reward of the play in G is h, the probability that the play in G' reaches $\mathbf{1}$ is exactly βh.

The second step is identical to that of Zwick and Paterson [19]. They describe how arbitrary rational probability distributions can be implemented using a polynomial number of coin-toss vertices. Thus, we can transform G' into an equivalent simple stochastic game.

Lemma 4. $\textsc{Terminal}^{\mathrm{s}}_{\mathrm{b}} \preceq \textsc{Mean}^{\mathrm{s}}_{\mathrm{b}}$ *and* $\textsc{Terminal}^{\mathrm{q}}_{\mathrm{u}} \preceq \textsc{Mean}^{\mathrm{q}}_{\mathrm{u}}$.

Proof. We are given a simple stochastic game G and must construct a stochastic mean-payoff game G'. Positions of G' will coincide with vertices of G, with the positions of G' including the terminals. Positions u belonging to a player in G belongs to the same player in G'. For each outgoing arc of u, we add an action in G' with reward 0, and with a deterministic transition to the endpoint of the arc of G. Random vertices of G can be assigned to either player in G', but he will only be given a single "dummy choice": If the random vertex has arcs to v_1 and v_2, we add a single action in G' with reward 0 and transitions into v_1, v_2, both with probability weight $1/2$. Each terminal can be assigned to either player in G', but again he will be given only a dummy choice: We add a single action with reward equal to the payoff of the terminal and with a transition back into the same terminal with probability weight 1.

There is clearly a one-to-one correspondence between pure positional strategies in G and strategies in G'. To see the correspondence between values, fix a strategy profile for the two players and consider play starting from some vertex. By construction, if the probability of the play reaching some particular terminal in G is q, then the probability of the play reaching the corresponding self-loop in G' is also q. Thus, the expected reward is the same.

3 Proof of Theorem 2

In this section, we consider a given stochastic terminal-payoff game. Our goal is an algorithm for computing optimal strategies, given the values of all vertices. To simplify the presentation, we will focus on strategies for Player 1. Because of symmetry, there is no loss of generality. In this section, "strategy" means "pure positional strategy". An arc (u, v) is called *safe* if $\mathrm{val}(u) = \mathrm{val}(v)$. A *safe strategy* is a strategy that only uses safe arcs. A strategy x for Player 1 is called *stopping* if for any strategy y for Player 2 and any vertex v with positive value, there is a non-zero probability that the play $P(x, y, v)$ reaches a terminal.

It is not hard to see that any optimal strategy for Player 1 must be safe and stopping, so these two conditions are necessary for optimality. We will now show that they are also sufficient.

Lemma 5. *If a strategy is safe and stopping, then it is also optimal.*

Proof. Let x be any safe and stopping strategy for Player 1, let y be an arbitrary strategy for Player 2, and let v_0 be an arbitrary vertex. Consider the play $P(x, y, v_0)$. Denote by $q_i(v)$ the probability that after i steps, the play is at v. Since x is safe,

$$\forall i : \quad \mathrm{val}(v_0) \leq \sum_{v \in V} \mathrm{val}(v) q_i(v) \leq \sum_{v \in T \cup V^+} \mathrm{val}(v) q_i(v), \tag{4}$$

where T denotes the set of terminal vertices and V^+ denotes the set of non-terminal vertices with positive value. Since x is stopping,

$$\forall v \in V^+ : \quad \lim_{i \to \infty} q_i(v) = 0. \tag{5}$$

Finally, note that

$$\begin{aligned} u^{v_0}(x, y) &= \sum_{t \in T} \mathrm{val}(t) \lim_{i \to \infty} q_i(t) = \sum_{v \in T \cup V^+} \mathrm{val}(v) \lim_{i \to \infty} q_i(v) && \text{(by (5))} \\ &= \lim_{i \to \infty} \sum_{v \in T \cup V^+} \mathrm{val}(v) q_i(v) \geq \mathrm{val}(v_0). && \text{(by (4))} \end{aligned}$$

Therefore, x is optimal.

Using this characterization of optimality, strategy recovery can be reduced to strategically solving a simple stochastic game without random vertices.

Theorem 2. Given a stochastic terminal-payoff game and the values of all its vertices, optimal pure positional strategies can be computed in linear time.

Proof. Construct from the given game G a simple stochastic game G' as follows:

1. Merge all terminals into one.
2. Remove all outgoing non-safe arcs from the vertices of Player 1.
3. Transfer ownership of all random vertices to Player 1.

Compute an optimal strategy x' for Player 1 in G' using linear-time retrograde analysis [1]. Let x be the interpretation of x' as a strategy in G obtained by restricting x' to the vertices of Player 1 in G. By construction, from any starting vertex v, Player 1 can ensure reaching the terminal in G' if and only if x ensures a non-zero probability of reaching a terminal in G. Let v be any vertex with positive value in G. Player 1 has a safe strategy that ensures a non-zero probability of reaching a terminal from v, specifically, the optimal strategy. This implies that there is a corresponding strategy for Player 1 in G' that ensures reaching the terminal from v. It follows that x is stopping, and it is safe by construction. Therefore, by Lemma 5, it is optimal.

4 Conclusions and Open Problems

Informally, we have shown that solving simple stochastic games is "complete" for a wide natural range of two-player zero-sum perfect-information games on graphs. However, in the logic and verification literature, even more general classes of winning conditions on infinite plays in finite graphs have been considered. Since we now know that simple stochastic games are more expressive than previously suggested, it would be interesting to fully characterize the class of such games for which solving simple stochastic game is complete. It is also interesting to ask whether solving some classes of *imperfect information* games reduces to solving simple stochastic games. It seems natural to restrict attention to cases where it is known that optimal positional strategies exists. This precludes general stochastic games (but see [5]). An interesting class of games generalizing stochastic mean-payoff games was considered by Filar [8]. Filar's games allow simultaneous moves by the two players. However, for any position, the probability distribution on the next position can depend on the action of one player only. Filar shows that his games are guaranteed to have optimal positional strategies. The optimal strategies are not necessarily pure, but the probabilities they assign to actions are guaranteed to be rational numbers if rewards and probabilities are rational numbers. So, we ask: Is solving Filar's games polynomial time equivalent to solving simple stochastic games?

Turning to strategy recovery, we mention again that we do not know whether the task of computing optimal strategies once values are known can be done in polynomial time for stochastic mean-payoff games. Recent work by Vladimir Gurvich (personal communication) indicates that the problem of computing optimal strategies in stochastic mean-payoff games is challenging even in the *ergodic* case where all positions have the same value. One may even hypothesise that strategy recovery for stochastic mean-payoff games is hard, in the (optimal) sense that it is polynomial time equivalent to strategically solving stochastic mean-payoff games. Establishing this would be most interesting.

Acknowledgements. We would like to thank Vladmir Gurvich for fruitful collaboration and Florian Horn for helpful comments.

References

1. D. Andersson, K. A. Hansen, P. B. Miltersen, and T. B. Sørensen. Deterministic graphical games revisited. In *CiE '08: Proceedings of the 4th conference on Computability in Europe*, pages 1–10, Berlin, Heidelberg, 2008. Springer-Verlag.
2. K. Chatterjee and T. A. Henzinger. Reduction of stochastic parity to stochastic mean-payoff games. *Inf. Process. Lett.*, 106(1):1–7, 2008.
3. K. Chatterjee, M. Jurdziński, and T. Henzinger. Simple stochastic parity games. In *CSL: Computer Science Logic*, LNCS 2803, pages 100–113. Springer, 2003.
4. K. Chatterjee, M. Jurdziński, and T. A. Henzinger. Quantitative stochastic parity games. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 121–130, 2004.
5. K. Chatterjee, R. Majumdar, and T. A. Henzinger. Stochastic limit-average games are in EXPTIME. International Journal of Game Theory, to appear.
6. A. Condon. The complexity of stochastic games. *Information and Computation*, 96:203–224, 1992.
7. E. A. Emerson, C. S. Jutla, and A. P. Sistla. On model checking for the mu-calculus and its fragments. *Theoretical Computer Science*, 258(1-2):491–522, May 2001.
8. J. Filar. Ordered field property for stochastic games when the player who controls transitions changes from state to state. *Journal of Optimization Theory and Applications*, 34:503–513, 1981.
9. D. Gillette. Stochastic games with zero stop probabilities. In *Contributions to the Theory of Games III*, volume 39 of *Annals of Math. Studies*, pages 179–187. 1957.
10. H. Gimbert and F. Horn. Simple Stochastic Games with Few Random Vertices are Easy to Solve. In *Proceedings of the 11th International Conference on the Foundations of Software Science and Computational Structures, FoSSaCS'08*, volume 4962 of *Lecture Notes in Computer Science*, pages 5–19. Springer-Verlag, 2008.
11. V. Gurvich, A. Karzanov, and L. Khachiyan. Cyclic games and an algorithm to find minimax cycle means in directed graphs. *USSR Computational Mathematics and Mathematical Physics*, 28:85–91, 1988.
12. V. Gurvich and P. B. Miltersen. On the computational complexity of solving stochastic mean-payoff games. `http://www.citebase.org/abstract?id=oai:arXiv.org:0812.0486`, 2008.
13. K. A. Hansen, M. Koucky, and P. B. Miltersen. Winning concurrent reachability games requires doubly-exponential patience. LICS'09, to appear., 2009.
14. R. A. Howard. *Dynamic Programming and Markov Processes.* M.I.T. Press, 1960.
15. T. M. Liggett and S. A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607, 1969.
16. M. L. Littman. *Algorithms for sequential decision making.* PhD thesis, Brown University, Department of Computer Science, 1996.
17. R. McNaughton. Infinite games played on finite graphs. *An. Pure and Applied Logic*, 65:149–184, 1993.
18. L. Shapley. Stochastic games. *Proc. Nat. Acad. Science*, 39:1095–1100, 1953.
19. U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theor. Comput. Sci.*, 158(1-2):343–359, 1996.