Bidding Games on Markov Decision Processes*

Guy Avni[1], Thomas A. Henzinger[1], Rasmus Ibsen-Jensen[2], and Petr Novotný[3]

[1] IST Austria
[2] University of Liverpool
[3] Masaryk University

Abstract. In two-player games on graphs, the players move a token through a graph to produce an infinite path, which determines the qualitative winner or quantitative payoff of the game. In *bidding games*, in each turn, we hold an auction between the two players to determine which player moves the token. Bidding games have largely been studied with concrete bidding mechanisms that are variants of a first-price auction: in each turn both players simultaneously submit bids, the higher bidder moves the token, and pays his bid to the lower bidder in *Richman bidding*, to the bank in *poorman bidding*, and in *taxman bidding*, the bid is split between the other player and the bank according to a predefined constant factor. Bidding games are deterministic games. They have an intriguing connection with a fragment of stochastic games called *random-turn* games. We study, for the first time, a combination of bidding games with probabilistic behavior; namely, we study bidding games that are played on Markov decision processes, where the players bid for the right to choose the next action, which determines the probability distribution according to which the next vertex is chosen. We study parity and mean-payoff bidding games on MDPs and extend results from the deterministic bidding setting to the probabilistic one.

1 Introduction

Two-player infinite-duration games on graphs are a central class of games in formal verification [2], where they are used, for example, to solve the problem of reactive synthesis [12], and they have deep connections to foundations of logic [14]. A graph game proceeds by placing a token on a vertex in the graph, which the players move throughout the graph to produce an infinite path ("play") π. The game is zero-sum and π determines the winner or payoff.

A graph game is equipped with a set of rules, which we call the "mode of moving", that determine how the token is moved in each turn. The simplest mode of moving is *turn based* in which the vertices are partitioned between the two players, and when the token is placed on a vertex v, the player who owns v decides to which neighbor of v it proceeds to. Turn-based games are used to

* This research was supported in part by the Austrian Science Fund (FWF) under grants S11402-N23 (RiSE/SHiNE), Z211-N23 (Wittgenstein Award), and M 2369-N33 (Meitner fellowship), and the Czech Science Foundation grant no. GJ19-15134Y.

model antagonistic behavior and are appropriate in worst-case analysis. On the other hand, probabilistic transitions conveniently model lack of information and are appropriate for average-case analysis. In *Markov chains*, the token proceeds from each vertex according to a probability distribution on neighboring vertices. A *Markov decision process* (MDP, for short) is associated with a set of actions Γ, and each vertex v is associated with a probability distribution $\delta(v, \gamma)$ on neighboring vertices, for each action $\gamma \in \Gamma$. Thus, an MDP can be thought of as a 1.5-player game in which, assuming the token is placed on a vertex v, the single player chooses an action γ, and *Nature* chooses the vertex to move the token to according to the distribution $\delta(v, \gamma)$. *Stochastic games*, a.k.a. 2.5-player games, combine turn-based games and probabilistic transitions [7]. The vertices in a stochastic game are partitioned between two players and a Nature player. Whenever the token is placed on a vertex that is controlled by a player, we proceed as in turn-based games, and whenever it is placed on a vertex that is controlled by Nature, we proceed randomly as in Markov chains.

Bidding is another mode of moving. In bidding games, both players have budgets and an auction is held in each turn to determine which player moves the token. Bidding games where introduced in [9,10], where several concrete bidding rules were defined. In *Richman* bidding (named after David Richman), each player has a budget, and before each turn, the players submit bids simultaneously, where a bid is legal if it does not exceed the available budget. The player who bids higher wins the bidding, pays the bid to the other player, and moves the token. A second bidding rule called *poorman* bidding in [9], is similar except that the winner of the bidding pays the "bank" rather than the other player. Thus, the bid is deducted from his budget and the money is lost. A third bidding rule called *taxman* in [9], spans the spectrum between poorman and Richman bidding. Taxman bidding is parameterized by a constant $\tau \in [0, 1]$: the winner of a bidding pays portion τ of his bid to the other player and portion $1 - \tau$ to the bank. Taxman bidding with $\tau = 1$ coincides with Richman bidding and taxman bidding with $\tau = 0$ coincides with poorman bidding.

We study for the first time, a combination of the bidding and probabilistic modes of moving by studying bidding games that are played on MDPs; namely, the bidding game is played on an MDP, and in each turn we hold a bidding to determine which player chooses an action. One motivation for the study of bidding games on MDPs is practical; the extension expands the modelling capabilities of bidding games. A second motivation is theoretical and aims at a better understanding of a curious connection between bidding games and stochastic games, which we describe below.

Up to now, we have only discussed modes of moving the token. A second classification for graph games is according to the players' objectives. The simplest objective is *reachability*, where Player 1 wins iff an infinite play visits a designated target vertex. Bidding reachability games were studied in [9,10], and these are the only objectives studied there. A central quantity in bidding games is the *initial ratio* of the players' budgets. The central question that was studied in [9] regards the existence of a necessary and sufficient initial ratio to guarantee winning the game. Formally, assuming that, for $i \in \{1, 2\}$, Player i's initial budget is B_i, we say that Player 1's initial ratio is $B_1/(B_1 + B_2)$. The *threshold ratio* in a vertex v, denoted $\texttt{Thresh}(v)$, is such that if Player 1's initial ratio exceeds $\texttt{Thresh}(v)$, he can guarantee winning the game, and if his initial ratio is less than $\texttt{Thresh}(v)$, Player 2 can guarantee winning the game[4]. Existence of threshold ratios in reachability games for all three bidding mechanisms was shown in [9].

Moreover, the following probabilistic connection was shown for reachability games with Richman-bidding and only for this bidding rule. *Random-turn* games are a fragment of stochastic games. A random-turn game is parameterized by $p \in [0, 1]$. In each turn, the player who moves is determined according to a (possibly) biased coin toss: with probability p, Player 1 chooses how to move the token, and Player 2 chooses with probability $1 - p$. Consider a reachability Richman-bidding game $\mathcal{G}$. We construct a "uniform" random-turn game on top of $\mathcal{G}$, denoted $\texttt{RT}^{0.5}(\mathcal{G})$, in which we toss an unbiased coin in each turn. The objective of Player 1 remains reaching his target vertex. It is well known that each vertex in $\texttt{RT}^{0.5}(\mathcal{G})$ has a *value*, which is, informally, the probability of reaching the target when both players play optimally, and which we denote by $val(\texttt{RT}^{0.5}(\mathcal{G}), v)$. The probabilistic connection that is observed in [10] is the following: For every vertex v in the reachability Richman-bidding game $\mathcal{G}$, the threshold ratio in v equals $1 - val(\texttt{RT}(\mathcal{G}), v)$. We note that such a connection is not known and is unlikely to exist in reachability games with neither poorman nor taxman bidding. Indeed, very simple poorman games have irrational threshold ratios [4]. Random-turn games have been extensively studied in their own right, mostly with unbiased coin tosses, since the seminal paper [11].

Infinite-duration bidding games were studied with Richman- [3], poorman- [4], and taxman-bidding [5]. The most interesting results in these papers regards an extended probabilistic connection for *mean-payoff* bidding games. Mean-payoff games are quantitative games; an infinite play is associated with a payoff that is Player 1's reward and Player 2's cost. Accordingly, we refer to the players in a mean-payoff game as Max and Min, respectively. Consider a strongly-connected mean-payoff taxman-bidding game $\mathcal{G}$ with taxman parameter $\tau \in [0, 1]$ and initial ratio $r \in (0, 1)$. The probabilistic connection is the following: the value of $\mathcal{G}$ w.r.t. τ and r, namely the optimal payoff Max can guarantee assuming his budget exceeds r, equals the *value* of the mean-payoff random-turn game $\texttt{RT}^{F(\tau,r)}(\mathcal{G})$ for $F(\tau, r) = \frac{r+\tau(1-r)}{1+\tau}$, where the value of $\texttt{RT}^{F(\tau,r)}(\mathcal{G})$ is the expected payoff when both players play optimally. Specifically, for Richman-bidding, the value does not depend on the initial ratio and equals the value of $\texttt{RT}^{0.5}(\mathcal{G})$. For poorman bidding, the value of $\mathcal{G}$ equals the value of $\texttt{RT}^{r}(\mathcal{G})$. We highlight the point that bidding games are deterministic. One way to understand the probabilistic connection is as a "derandomization"; namely, Max has a deterministic

[4] When the initial ratio is exactly $\texttt{Thresh}(v)$, the winner depends on the mechanism with which ties are broken. Our results do not depend on a specific tie-breaking mechanism.Tie-breaking mechanisms are particularly important in discrete-bidding games [1].

bidding strategy in $\mathcal{G}$ that ensures a behavior that mimics the probabilistic behavior of $\mathtt{RT}^{F(\tau,r)}(\mathcal{G})$.

For qualitative objectives, we show existence of surely-winning threshold ratios in Richman-bidding reachability games. We then focus on strongly-connected games and show that in a strongly-connected parity taxman-bidding game, one of the players wins almost-surely with any positive initial budget. For mean-payoff objectives, we extend the probabilistic connection for strongly-connected mean-payoff taxman-bidding games from the deterministic setting to the probabilistic one. Namely, we show that the optimal expected payoff in a taxman-bidding game $\mathcal{G}$ w.r.t. τ and r equals the value of $\mathtt{RT}^{F(\tau,r)}(\mathcal{G})$. The proof is constructive and we show an optimal bidding strategy for the two players.

2 Preliminaries

A *Markov decision process* (MDP, for short) is $\mathcal{M} = \langle V, \Gamma, \delta \rangle$, where V is a set of vertices, Γ is a set of actions, and $\delta : V \times \Gamma \to [0,1]^V$ is a probabilistic transition function, where for every $v \in V$ and $\gamma \in \Gamma$, we have $\sum_{u \in V} \delta(v,\gamma)(u) = 1$. We say that an MDP $\mathcal{M}$ is *strongly-connected* if from every two vertices v and u, both players have a strategy that forces the game from v to u with probability 1. We focus on strongly-connected MDPs, where the initial position of the token is not crucial and we sometimes omit it.

We study bidding games that are played on MDPs. The game proceeds as follows. Initially, a token is placed on some vertex and the players start with budgets, which are real numbers. Suppose the token is placed on $v \in V$ in the beginning of a turn. We hold a bidding in which both players simultaneously submit bids, where a bid is legal if it does not exceed the available budget. The player who bids higher wins the bidding and chooses an action $\gamma \in \Gamma$, and the next position of the token is chosen at random according to the distribution $\delta(v, \gamma)$. The bidding rules that we consider differ in the update to the players' budget, and specifically, in how the winning bid is distributed.

Definition 1. *Suppose the players budgets are B_1 and B_2 and Player 1 wins the bidding with a bid of b. The budgets in the next turn are obtained as follows.*

- **Richman bidding**: *Player 1 pays Player 2, thus $B_1' = B_1 - b$ and $B_2' = B_2 + b$.*
- **Poorman bidding**: *Player 1 pays the bank, thus $B_1' = B_1 - b$ and $B_2' = B_2$.*
- **Taxman bidding** *with parameter $\tau \in [0,1]$: Player 1 pays portion τ to Player 2 and portion $(1-\tau)$ to the bank, thus $B_1' = B_1 - b$ and $B_2' = B_2 + b \cdot \tau$.*

Note that fixing the taxman parameter to $\tau = 1$ gives Richman bidding and fixing $\tau = 0$ gives poorman bidding.

A finite *play* of a bidding game is in $(V \times \Gamma \times \mathbb{R} \times \{1,2\})^* \cdot V$. A strategy is a function that takes a finite player and prescribes a bid as well as an action to perform upon winning the bid. Two strategies f_1 and f_2 for the two players and an initial vertex v_0 give rise to a distribution over plays of length $n \in \mathbb{N}$, which we denote by $Dist_n(v_0, f_1, f_2)$ and define inductively. For $n = 0$, the probability of the play v_0 is 1. Consider a finite play π that visits $n-1$ vertices. For $i \in \{1, 2\}$, let $\langle b_i, \gamma_i \rangle = f_i(\pi)$. If $b_i > b_{3-i}$, then Player i wins the bidding, and the next action to be played is γ_i. For $u \in V$, the probability of the n-lengthed play $\pi \cdot \langle v, \gamma_i, b_i, i \rangle \cdot u$ is $\Pr[\pi] \cdot \delta(v, \gamma_i)(u)$. The issue of draws, i.e., the case in which $b_i = b_{3-i}$, needs to be handled with a tie-breaking mechanism, and our results are not affected by which mechanism is used. The extension of the distribution $D_n(v_0, f_1, f_2)$ to infinite paths is standard.

Random-turn games *Stochastic games* generalize MDPs; while an MDP can be thought of as a player playing against Nature, in a stochastic game, a player is playing against a second adversarial player as well as against Nature. We consider a fragment of stochastic games called *random-turn* games, which are similar to bidding games except that, in each turn, rather than bidding, the player who chooses an action is selected according to some fixed probability. Formally, let $\mathcal{G} = \langle V, \Gamma, \delta, w \rangle$ be a mean-payoff bidding game and $p \in (0, 1)$, then the random-turn game that is associated with $\mathcal{G}$ and p is $\mathtt{RT}^p(\mathcal{G}) = \langle V \cup (V \times \{1, 2\}), \Gamma, \delta', w \rangle$, where vertices in V are controlled by Nature and model coin tosses and a vertex $\langle v, i \rangle$, for $i \in \{1, 2\}$, models the case that Player i is chosen to play. Thus, for every $v \in V$ and $\gamma \in \Gamma$, we have $\delta'(v, \gamma)(\langle v, 1 \rangle) = p$ and $\delta'(v, \gamma)(\langle v, 2 \rangle) = 1 - p$. Also, Player i controls every vertex in $V \times \{i\}$, and we have $\delta'(\langle v, i \rangle, \gamma) = \delta(v, \gamma)$. Finally, it is technically convenient to assume that vertices in $V \times \{1, 2\}$ do not contribute to the energy of a play.

3 Qualitative Bidding Games on MDPs

In this section we study infinite-duration games with qualitative objectives. We adapt the concept of *surely winning* to bidding games played on MDPs.

Definition 2. *Let $\mathcal{G}$ be a game that is played on an MDP $\langle V, \Gamma, \delta \rangle$, let $O \subseteq V^\omega$ be an objective for Player 1, and let $v \in V$. The* surely-winning threshold ratio *in v, denoted $\mathtt{Thresh}(v)$, is such that*

- *If Player 1's initial ratio exceeds $\mathtt{Thresh}(v)$, then Player 1 has a strategy such that no matter how Player 2 plays, the resulting play is in O.*
- *If Player 2's initial ratio exceeds $1 - \mathtt{Thresh}(v)$, then he has a strategy such that no matter how Player 1 plays, the resulting play is not in O.*

In reachability games, Player 1 has a target vertex and an infinite play is winning for him iff it visits the target. We show existence of surely-winning threshold ratios in reachability Richman-bidding games.

Theorem 1. *Let $\mathcal{G}$ be a reachability Richman-bidding game. Surely-winning threshold ratios exist in $\mathcal{G}$ and can be found using a linear reduction to a stochastic reachability game.*

Proof. Recall that the random-turn game $\mathtt{RT}^{0.5}(\mathcal{G})$ is a stochastic game that models the following process: in each turn, we toss a fair coin, and if it turns "heads" Player 1 determines the next action and otherwise Player 2 determines the next action. The action gives rise to a probability distribution with which the following vertex is chosen. We construct $\mathcal{G}'$ similarly, only that we replace the last probabilistic choice with a deterministic choice of Player 2. Formally, the vertices of $\mathcal{G}'$ are $V \cup (V \times \{1, 2\}) \cup (V \times \Gamma)$. The transition function δ' restricted to V is the same as in $\mathtt{RT}^{0.5}(\mathcal{G})$, namely, for every action, we proceed from $v \in V$ to $\langle v, i\rangle$, for $i \in \{1, 2\}$, with probability 0.5. The vertex $\langle v, i\rangle$ is controlled by Player i. A vertex $u \in V$ is a neighbor of $\langle v, 2\rangle$ iff there exists $\gamma \in \Gamma$ with $\delta(v, \gamma)(u) > 0$. The neighbors of $\langle v, 1\rangle$ are $\{v\} \times \Gamma$, where moving to $\langle v, \gamma\rangle$ models Player 1 choosing the action γ at v. Each vertex $\langle v, \gamma\rangle$ is controlled by Player 2 and a vertex $u \in V$ is a neighbor of $\langle v, \gamma\rangle$ iff $\delta(v, \gamma)(u) > 0$.

Let $v \in V$. The *value* of v in $\mathcal{G}'$, denoted $val(\mathcal{G}', v)$ is the probability of reaching the target when both players play optimally. We claim that the surely-winning threshold ratio in v equals $1 - val(\mathcal{G}', v)$. Note that when $val(\mathcal{G}', v) = 0$, no matter how Player 1 plays, there is no path from v to t, thus Player 1 cannot win and we have $\mathtt{Thresh}(v) = 1$. Suppose $val(\mathcal{G}', v) = 1$ and we claim that $\mathtt{Thresh}(v) = 0$. We follow the construction in the deterministic setting [3,9]. Let $n = |V|$. It is not hard to show that if Player 1 wins n biddings in a row, he wins the game. Suppose Player 1's initial ratio is $\epsilon > 0$. He follows a strategy that guarantees that he either wins n biddings in a row or, if he loses, his budget increases by a constant that depends on ϵ and n. Thus, by repeatedly playing according to this strategy, he either wins the game or increases his budget arbitrarily close to 1, where he can force n bidding wins. The proof for vertices with $val(\mathcal{G}', v) \in (0, 1)$ is similar only that Player 1's strategy maintains the invariant that his budget exceeds $1 - val(\mathcal{G}', v)$ and his *surplus*, namely the difference between his budget and $1 - val(\mathcal{G}', v)$, increases every time he loses a bidding. The proof for Player 2 is dual. $\square$

Theorem 1 shows a reduction from the problem of finding threshold ratios to the problem of solving a stochastic reachability game. The complexity of the later is known to be in NP and coNP [7], thus we obtain the following corollary.

Corollary 1. *The problem of deciding, given a reachability Richman-bidding game on an MDP $\mathcal{G}$ and a vertex v in $\mathcal{G}$, whether the surely-winning threshold ratio is at least* 0.5*, is in NP and coNP.*

The solution to strongly-connected games is the key in the deterministic setting. We show that Player 1 almost-surely wins reachability games that are played on strongly-connected MDPs.

Proposition 1. *Let $\mathcal{G} = \langle V, \Gamma, \delta, w\rangle$ be a strongly-connected taxman-bidding game with taxman parameter τ. For every positive initial budget, initial vertex $v \in V$, and target vertex $u \in V$, Player i has a strategy that guarantees that u is reached from v with probability* 1.

Proof. Let f_i be a strategy for Player i in the MDP $\langle V, \Gamma, \delta \rangle$ that guarantees that u is reached from v with probability 1. Let $\epsilon > 0$ be an initial budget or Player i in the bidding game $\mathcal{G}$. It is shown in [5] that, for every $n \in \mathbb{N}$, there is a bidding strategy that guarantees that Player i eventually wins n biddings in a row. Intuitively, Player i splits his budget into n exponentially increasing parts $\epsilon_1, \ldots, \epsilon_n$ such that if Player i loses the j-th bidding, for $1 \leq j \leq n$, his budget increases by a constant factor. By repeatedly following such a strategy, Player i's ratio approaches 1, which guarantees n consecutive wins. Player i splits his budget into infinitely many parts $\epsilon_1, \epsilon_2, \ldots$, and, for $n \geq 1$, he plays as if his budget is ϵ_n until he wins n consecutive biddings. Upon winning a bidding, he chooses actions according to f_i. Thus, Player i essentially follows f_i for growing sequences thereby ensuring visiting u with a probability that approaches 1. □

Consider a strongly-connected parity taxman-bidding game $\mathcal{G}$ in which the highest parity index is odd. A corollary of the above proposition is that Player 1 almost-surely wins in $\mathcal{G}$ with any positive initial budget. Indeed, in $\texttt{RT}^p(\mathcal{G})$, by repeatedly playing according to a strategy f_i that forces a visit to the vertex v with the highest parity index, Player 1 forces infinitely many visits to v with probability 1. A bidding strategy proceeds as in the proof of the proposition above and forces increasingly longer sequences of bidding winnings, which in turn implies following f_i for increasingly longer sequences.

Theorem 2. *Let $\mathcal{G}$ be a strongly-connected parity game. If the maximal parity index in $\mathcal{G}$ is odd, then Player* 1 *almost-surely wins in $\mathcal{G}$ with any positive initial budget, and if the maximal parity index in $\mathcal{G}$ is even, Player* 2 *almost-surely wins in $\mathcal{G}$ with any positive initial budget.*

4 Mean-Payoff Bidding Games on Strongly-Connected MDPs

Mean-payoff bidding games are played on a *weighted* MDP $\langle V, \Gamma, \delta, w \rangle$, where $\langle V, \Gamma, \delta \rangle$ is an MDP and $w : V \to \mathbb{Q}$ is a weight function. The *energy* of a finite play π, denoted $E(\pi)$, refers to the accumulated weights, thus $E(\pi) = \sum_{1 \leq i \leq n} w(v_i)$. Consider two strategies f_1 and f_2, and an initial vertex v_0. The *payoff* w.r.t f_1, f_2, and v_0, is $\texttt{MP}(v_0, f_1, f_2) = \liminf_{n \to \infty} \mathbb{E}_{\pi \sim Dist_n(v_0, f_1, f_2)}[E(\pi)/n]$. A mean-payoff game is a zero-sum game. The payoff is Player 1's reward and Player 2's cost. Accordingly, we refer to Player 1 as Max and Player 2 as Min.

We focus on strongly-connected mean-payoff games. Since the mean-payoff objective is prefix independent, Proposition 1 implies that the optimal payoff from each vertex in a strongly-connected game is the same.

Definition 3. *(***Mean-payoff values***) Consider a strongly-connected mean-payoff taxman-bidding game $\mathcal{G} = \langle V, \Gamma, \delta, w \rangle$, a ratio $r \in (0, 1)$, and a taxman parameter $\tau \in [0, 1]$. We say that $c \in \mathbb{R}$ is the* value *of $\mathcal{G}$ w.r.t. r and τ, denoted $\texttt{MP}^{\tau,r}(\mathcal{G})$, if for every $\epsilon > 0$,*

- *when Max's initial ratio is $r + \epsilon$, he can guarantee an expected payoff of at least c, and*
- *when Max's initial ratio is $r - \epsilon$, Min can guarantee an expected payoff of at most c.*

We describe an optimal bidding strategy for Max in $\mathcal{G}$ w.r.t. τ and r. The construction consists of two components. The first component assigns an "importance" to each vertex, which we call the *strength* of a vertex and denote by $\mathrm{St}^p(v)$, for every $v \in V$. Intuitively, if $\mathrm{St}^p(v) > \mathrm{St}^p(u)$, then it is more important to move in v than it is in u. The second ingredient is a "normalization scheme" for the strengths, which consists of a sequence $(r_x)_{x\geq 1}$ and associating normalization factors $(\beta_x)_{x\geq 1}$, where $\beta_x, r_x \in [0,1]$. Max keeps track of a position on the sequence, where he maintains the invariant that when the position is x, his ratio exceeds r_x. One property of the sequence is that the invariant implies that position $x = 1$ is never reached. Assuming the token is placed on $v \in V$ and the position on the sequence is x, Max's bid is roughly $\beta_x \cdot \mathrm{St}^p(v)$. The outcome of the bidding determines the next position on the sequence, where winning means that we proceed up on the sequence and losing means that we proceed down on the sequence. A normalization scheme for Richman bidding was devised in [3], for poorman bidding in [4], and we use a unified normalization scheme that was devised in [5] for taxman bidding.

We start with assigning importance to vertices. Our definition relies on a solution to random-turn games.

Definition 4. *(***Values***) For a strongly-connected mean-payoff bidding game $\mathcal{G}$ and $p \in (0,1)$, the* mean-payoff value *of $\mathtt{RT}^p(\mathcal{G})$, denoted $\mathtt{MP}(\mathtt{RT}^p(\mathcal{G}))$, is the maximal expected payoff that Max guarantee from every vertex.*

A *positional strategy* is a strategy that always chooses the same action in a vertex. It is well known that there exist optimal positional strategies for both players in stochastic mean-payoff games. For some $p \in (0,1)$, consider two optimal positional strategies f and g in $\mathtt{RT}^p(\mathcal{G})$, for Min and Max, respectively. For a vertex $v \in V$, let $\gamma^+(v), \gamma^-(v) \in \Gamma$ denote the actions that f and g prescribe, thus $\gamma^+(v) = f(\langle v, 1\rangle)$ and $\gamma^-(v) = g(\langle v, 2\rangle)$.

The *potential* of v, denoted $\mathrm{Po}^p(v)$, is a known concept in probabilistic models and was originally used in the context of the strategy iteration algorithm for MDPs [8]. We use the potential to define the *strength* of v, denoted $\mathrm{St}^p(v)$, which intuitively measures how much the expected potentials of the neighbors of v differ. The potential and strengths of v are functions that satisfy the following:

$$\mathrm{Po}^p(v) = p \cdot \sum_{u \in V} \delta(v, \gamma^+(v))(u) \cdot \mathrm{Po}^p(u) + (1-p) \cdot \sum_{u \in V} \delta(v, \gamma^-(v))(u) \cdot \mathrm{Po}^p(u) - \mathtt{MP}(\mathtt{RT}^p(\mathcal{G})) \text{ and}$$

$$\mathrm{St}^p(v) = p(1-p)\big(\sum_{u \in V} \delta(v, \gamma^+(v))(u) \cdot \mathrm{Po}^p(u) - \sum_{u \in V} \delta(v, \gamma^-(v))(u) \cdot \mathrm{Po}^p(u)\big)$$

The existence of the potential and thus the strength is known to be guaranteed [13].

Consider a finite path $\eta = \langle v_1, \gamma_1 \rangle, \ldots, \langle v_{n-1}, \gamma_{n-1} \rangle, v_n$. Consider a partition of $\{1, \ldots, n-1\}$ to $W(\eta) \cup L(\eta)$ such that $i \in W(\eta)$ iff $\gamma_i = \gamma^+(v_i)$. Intuitively, we think of η as a play and the indices in $W(\eta)$ are the ones that Max wins whereas the ones in $L(\eta)$ represent the ones in which he loses. The probability of η is $\prod_{1 \leq i < n} \delta(v_i, \gamma_i(v_i))(v_{i+1})$. The energy of η, denoted $E(\eta)$, is $\sum_{1 \leq i < n} w(v_i)$. We define a random variable Ψ_n over paths of length n. Let η be such a path that ends in a vertex v, then

$$\Psi^p_n(\eta) = \mathrm{Po}^p(v) + E(\eta) - \sum_{i \in W(\eta)} \mathrm{St}^p(v_i)/p + \sum_{i \in L(\eta)} \mathrm{St}^p(v_i)/(1-p) - (n-1) \cdot \mathtt{MP}(\mathtt{RT}^p(\mathcal{G})).$$

Lemma 1. *For every game $\mathcal{G}$, $p \in [0,1]$, and $n \in I\!N$, we have $\mathbb{E}[\Psi^p_n - \Psi^p_{n+1}] \geq 0$. Thus, $\mathbb{E}[\Psi_n] \geq \mathbb{E}[\Psi_1] \geq \min_v Po^p(v)$.*

Proof. Let $\eta = \langle v_1, \gamma_1 \rangle, \ldots, \langle v_{n-1}, \gamma_{n-1} \rangle, v_n$ and $\gamma \in \Gamma$. We show that $\mathbb{E}[\Psi_n(\eta) - \Psi_{n+1}(\eta')] \leq 0$, where η' is obtained from η by extending it with a last vertex that is chosen according to the distribution $\delta(v_n, \gamma)$. We prove for the case of $\gamma = \gamma^+(v_n)$. Since Max wins the last bidding, we have $W(\eta) = W(\eta') \cup \{n\}$ and $I(\eta) = I(\eta')$. In addition, we have $E(\eta) + w(v_n) = E(\eta')$. Thus,

$$\mathbb{E}[\Psi_n(\eta) - \Psi_{n+1}(\eta')] =$$

$$= \mathrm{Po}^p(v_n) - \Big(\sum_{u \in V} \mathrm{Po}^p(u) \cdot \delta(v_n, \gamma^+(v_n))(u) + w(v_n) - \mathrm{St}^p(v_n)/p - \mathtt{MP}(\mathtt{RT}^p(\mathcal{G})) \Big) =$$

$$= \mathrm{Po}^p(v_n) - \big((1-p) \sum_{u \in V} \mathrm{Po}^p(u) \cdot \delta(v_n, \gamma^-(v_n))(u) +$$

$$+ p \sum_{u \in V} \mathrm{Po}^p(u) \cdot \delta(v_n, \gamma^+(v_n))(u) + w(v_n) - \mathtt{MP}(\mathtt{RT}^p(\mathcal{G})) \big) =$$

$$= \mathrm{Po}^p(v_n) - \mathrm{Po}^p(v_n) = 0$$

The proof for the case that $\gamma \neq \gamma^+(v_n)$ is similar. Since we define $\gamma^-(v)$ to be the action that minimizes $\min_a \sum_{u \in V} \delta(v_n, a)(u) \cdot \mathrm{Po}^p(u)$, we get $\mathbb{E}[\Psi_n(\eta) - \Psi_{n+1}(\eta')] \geq 0$. □

We continue to describe the properties of a normalization scheme as well as show its existence.

Lemma 2. *[5] Let $S \subseteq \mathbb{Q}_{\geq 0}$, a ratio $r \in (0,1)$, and a taxman parameter $\tau \in [0,1]$. For every $K > \frac{\tau r^2 + r(1-r)}{\tau(1-r)^2 + r(1-r)}$ there exist sequences $(r_x)_{x \geq 1}$ and $(\beta_x)_{x \geq 1}$ with the following properties.*

1. *For each position $x \in \mathbb{R}_{\geq 1}$ and $s \in S$, we have $\beta_x \cdot s \cdot r \cdot (r-1) < r_x$.*
2. *For every $s \in S \backslash \{0\}$ and $1 \leq x < 1 + rs$, we have $\beta_x \cdot s \cdot r \cdot (r-1) > 1 - r_x$.*
3. *The ratios tend to r from above, thus for every $x \in \mathbb{R}_{\geq 1}$, we have $r_x \geq r$, and $\lim_{x \to \infty} r_x = r$.*

4. We have

$$\frac{r_x - \beta_x \cdot s \cdot r \cdot (r-1)}{1-(1-\tau)\cdot \beta_x \cdot s \cdot r \cdot (r-1)} \geq r_{x+(1-r)\cdot K \cdot s} \text{ and}$$

$$\frac{r_x + \tau \cdot \beta_x \cdot s \cdot r \cdot (r-1)}{1-(1-\tau)\cdot \beta_x \cdot s \cdot r \cdot (r-1)} \geq r_{x-s\cdot r}$$

We combine the two ingredients to obtain the following.

Theorem 3. *Let $\mathcal{G}$ be a strongly-connected mean-payoff taxman-bidding game, $r \in (0,1)$ an initial ratio, and $\tau \in [0,1]$ a taxman constant. Then, the mean-payoff value of $\mathcal{G}$ w.r.t. r and τ equals the value of the random-turn game $\mathtt{RT}^{F(r,\tau)}(\mathcal{G})$ in which Max is chosen to move with probability $F(\tau,r)$ and Min with probability $1-F(\tau,r)$, where $F(\tau,r)=\frac{r+\tau(1-r)}{1+\tau}$.*

Proof. Since the definition of payoff favors Min, it suffices to show an optimal strategy for Max. Let $\mathcal{G}$ such that $\mathtt{RT}^{F(\tau,r)}(\mathcal{G})=0$. For $\epsilon>0$, we describe a strategy for Max that guarantees a payoff that is greater than $-\epsilon$, assuming his initial ratio is strictly greater than r. Following [5], we consider a slight change of parameters; we choose $K > \frac{\tau r^2 + r(1-r)}{\tau(1-r)^2+r(1-r)}$, and define $\nu = r$, $\mu = K\cdot(1-r)$, and $p=\nu/(\nu+\mu)$, where we choose K such that $\mathtt{MP}(\mathtt{RT}^p(\mathcal{G})) > -\epsilon$, where this is possible due to the continuity of the mean-payoff value due to changes in the probabilities in the game structure [6,15]. We find potentials and strengths w.r.t. p and find a sequence $(r_x)_{x\geq 1}$ as in Lemma 2, where we set $S=\{\mathrm{St}^p(v): v\in V\}$.

Max maintains a position on the sequence. Recall that Max's ratio strictly exceeds r and that Point 3 implies that the sequence tends from above to r, thus Max can choose an initial position x_0 such that his initial ratio exceeds r_{x_0}. Whenever the token reaches a vertex v and the position on the sequence is x, Max bids $\mathrm{St}^p(v)\cdot r(1-r)\beta_x$, and chooses the action $\gamma^+(v)$ upon winning. If Max wins the bidding, the next position on the sequence is $x+\mu\mathrm{St}^p(v)$, and if he loses a bidding, the next position is $x-\nu\cdot\mathrm{St}^p(v)$. Note that Point 4 implies the invariant that whenever the position is x, Max's ratio exceeds r_x; indeed, the first part of the point takes care of winning a bidding, and the second part of losing a bidding. The invariant together with Point 1 implies that Max has sufficient funds for bidding. Suppose the current position is x following a play π, then $x = x_0 + \mu\sum_{i\in W(\pi)}\mathrm{St}^p(v) - \nu\sum_{i\in L(\pi)}\mathrm{St}^p(v)$. Point 2 implies that $x>1$; indeed, consider a position that is close to 1, i.e., a position such that if Min wins a bidding, the next position is $x\leq 1$, then Point 2 states that Max's bid is greater than Min's ratio, thus he necessarily wins the bidding and the next position is farther from 1. Rearranging, dividing by $\mu\cdot\nu$, and multiplying by (-1), we obtain $\sum_{i\in L(\pi)}\mathrm{St}^p(v)/\mu - \sum_{i\in W(\pi)}\mathrm{St}^p(v)/\nu = (x_0-x)/(\mu\cdot\nu) < (x_0-1)/(\mu\cdot\nu)$, where recall that x_0 is a constant.

Let $n\in\mathbb{N}$. We adapt the notation in Lemma 1 from paths to plays in the straightforward manner. The lemma implies that $\mathbb{E}[\Psi_n]\geq c$, for some constant $c\in\mathbb{Q}$. On the other hand, recall that, for a play π of length n that ends in a vertex v, we have

$$\Psi_n(\pi) = \text{Po}^p(v) + E(\pi) - \sum_{i \in W(\pi)} \text{St}^p(v)/\nu + \sum_{i \in L(\pi)} \text{St}^p(v)/\mu - (n-1)\texttt{MP}(\texttt{RT}^p(\mathcal{G})).$$

For every vertex v, we have $\text{Po}^p(v) \leq \max_u \text{Po}^p(u)$. Also, as in the above, we have $\mathbb{E}[\sum_{i \in W(\pi)} \text{St}^p(v)/\nu - \sum_{i \in L(\pi)} \text{St}^p(v)/\mu]$ is bounded from above by a constant. Combining, we have that $\mathbb{E}[E(\pi)] \geq c' + (n-1) \cdot \texttt{MP}(\texttt{RT}^p(\mathcal{G}))$. We divide both sides by n and tend it to infinity, thus the constant c' vanishes, and we get a payoff that exceeds $-\epsilon$, as required. □

Theorem 3 shows a reduction from the problem of finding the value of a mean-payoff taxman-bidding game on a strongly-connected MDP to the problem of solving a stochastic mean-payoff game. The complexity of the later is known to be in NP and coNP, thus we obtain the following corollary.

Corollary 2. *The problem of deciding, given a mean-payoff taxman-bidding game $\mathcal{G}$ that is played on a strongly-connected MDP, an initial ratio r, a taxman parameter τ, and a value $k \in \mathbb{Q}$, whether $\texttt{MP}^{\tau,r}(\mathcal{G}) \geq k$, is in NP and coNP.*

5 Discussion

We study qualitative and mean-payoff bidding games on MDPs. For qualitative objectives, we show existence of surely-winning threshold ratios in reachability bidding games, and we study almost-surely winning in strongly-connected parity bidding games. For mean-payoff objectives, we extend the probabilistic connection from the deterministic setting to the probabilistic one. A problem that we leave open is a quantitative solution to reachability bidding games that are played on MDPs; namely, given an MDP with a target vertex t, an initial vertex v, and a probability p, find a necessary and sufficient budget with which Player 1 can guarantee that t is reached from v with probability at least p. We expect that a solution to this problem will imply a solution to parity and mean-payoff bidding games on general graphs.

References

1. M. Aghajohari, G. Avni, and T. A. Henzinger. Determinacy in discrete-bidding infinite-duration games. In *Proc. 30th CONCUR*, 2019.
2. K.R. Apt and E. Grädel. *Lectures in Game Theory for Computer Scientists*. Cambridge University Press, 2011.
3. G. Avni, T. A. Henzinger, and V. Chonev. Infinite-duration bidding games. *J. ACM*, 66(4):31:1–31:29, 2019.
4. G. Avni, T. A. Henzinger, and R. Ibsen-Jensen. Infinite-duration poorman-bidding games. In *Proc. 14th WINE*, volume 11316 of *LNCS*, pages 21–36. Springer, 2018.
5. G. Avni, T. A. Henzinger, and Đ. Žikelić. Bidding mechanisms in graph games. In *In Proc. 44th MFCS*, 2019.

6. K. Chatterjee. Robustness of structurally equivalent concurrent parity games. In *Proc. 15th FoSSaCS*, pages 270–285, 2012.
7. A. Condon. The complexity of stochastic games. *Inf. Comput.*, 96(2):203–224, 1992.
8. A. R. Howard. *Dynamic Programming and Markov Processes.* MIT Press, 1960.
9. A. J. Lazarus, D. E. Loeb, J. G. Propp, W. R. Stromquist, and D. H. Ullman. Combinatorial games under auction play. *Games and Economic Behavior*, 27(2):229–264, 1999.
10. A. J. Lazarus, D. E. Loeb, J. G. Propp, and D. Ullman. Richman games. *Games of No Chance*, 29:439–449, 1996.
11. Y. Peres, O. Schramm, S. Sheffield, and D. B. Wilson. Tug-of-war and the infinity laplacian. *J. Amer. Math. Soc.*, 22:167–210, 2009.
12. A. Pnueli and R. Rosner. On the synthesis of a reactive module. In *Proc. 16th POPL*, pages 179–190, 1989.
13. M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons, Inc., New York, NY, USA, 2005.
14. M.O. Rabin. Decidability of second order theories and automata on infinite trees. *Transaction of the AMS*, 141:1–35, 1969.
15. E. Solan. Continuity of the value of competitive markov decision processes. *Journal of Theoretical Probability*, 16:831–845, 2003.