Determinacy in Stochastic Games with Unbounded Payoff Functions*

Tomáš Brázdil**, Antonín Kučera**, and Petr Novotný**

Faculty of Informatics, Masaryk University
{xbrazdil,kucera,xnovot18}@fi.muni.cz

Abstract. We consider infinite-state turn-based stochastic games of two players, $\Box$ and $\Diamond$, who aim at maximizing and minimizing the expected total reward accumulated along a run, respectively. Since the total accumulated reward is unbounded, the determinacy of such games cannot be deduced directly from Martin's determinacy result for Blackwell games. Nevertheless, we show that these games *are* determined both for unrestricted (i.e., history-dependent and randomized) strategies and deterministic strategies, and the equilibrium value is the same. Further, we show that these games are generally *not* determined for memoryless strategies. Then, we consider a subclass of $\Diamond$-*finitely-branching* games and show that they are determined for all of the considered strategy types, where the equilibrium value is always the same. We also examine the existence and type of (ε-)optimal strategies for both players.

1 Introduction

Turn-based stochastic games of two players are a standard model of discrete systems that exhibit both non-deterministic and randomized choice. One player (called $\Box$ or Max in this paper) corresponds to the controller who wishes to achieve/maximize some desirable property of the system, and the other player (called $\Diamond$ or Min) models the environment which aims at spoiling the property. Randomized choice is used to model events such as system failures, bit-flips, or coin-tossing in randomized algorithms.

Technically, a turn-based stochastic game (SG) is defined as a directed graph where every vertex is either stochastic or belongs to one of the two players. Further, there is a fixed probability distribution over the outgoing transitions of every stochastic vertex. A *play* of the game is initiated by putting a token on some vertex. Then, the token is moved from vertex to vertex by the players or randomly. A *strategy* specifies how a player should play. In general, a strategy may depend on the sequence of vertices visited so far (we say that the strategy is *history-dependent (H)*), and it may specify a probability distribution over the outgoing transitions of the currently visited vertex rather than a single outgoing transtion (we say that the strategy is *randomized (R)*). Strategies that do not depend on the history of a play are called *memoryless (M)*, and strategies that do not randomize (i.e., select a single outgoing transition) are called *determinisctic (D)*. Thus, we obtain the MD, MR, HD, and HR strategy classes, where HR are unrestricted strategies and MD are the most restricted memoryless deterministic strategies.

* The full version of this paper can be found at http://arxiv.org/abs/1208.1639

** The authors are supported by the Czech Science Foundation, grant No. P202/12/G061.

A. Kučera et al. (Eds.): MEMICS 2012, LNCS 7721, pp. 94–105, 2013.

A *game objective* is usually specified by a *payoff function* which assigns some real value to every run (infinite path) in the game graph. The aim of Player □ is to *maximize* the expected payoff, while Player ◇ aims at *minimizing* it. It has been shown in [22] that for *bounded* and *Borel* payoff functions, Martin's determinacy result for Blackwell games [23] implies that

$$\sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}[\mathit{Payoff}] \quad = \quad \inf_{\pi \in \mathrm{HR}_\Diamond} \sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi}[\mathit{Payoff}] \tag{1}$$

where $\mathrm{HR}_\Box$ and $\mathrm{HR}_\Diamond$ are the classes of HR strategies for Player □ and Player ◇, respectively. Hence, every vertex v has a *HR-value* $\mathrm{Val}_{\mathrm{HR}}(v)$ specified by (1). A HR strategy is *optimal* if it achieves the outcome $\mathrm{Val}_{\mathrm{HR}}(v)$ or better against every strategy of the other player. In general, optimal strategies are not guaranteed to exist, but (1) implies that both players have ε-*optimal* HR strategies for every $\varepsilon > 0$ (see Section 2 for precise definitions).

The determinacy results of [23,22] cannot be applied to *unbounded* payoff functions, i.e., these results do not imply that (1) holds if *Payoff* is unbounded, and they do not say anything about the existence of a value for restricted strategy classes such as MD or MR. In the context of performance analysis and controller synthesis, these questions rise naturally; in some cases, the players cannot randomize or remember the history of a play, and some of the studied payoff functions are not bounded. In this paper, we study these issues for the *total accumulated reward* payoff function and *infinite-state* games.

The total accumulated reward payoff function, denoted by *Acc*, is defined as follows. Assume that every vertex v is assigned a fixed non-negative reward $r(v)$. Then *Acc* assigns to every run the sum of rewards of all vertices visited along the run. Obviously, *Acc* is unbounded in general, and may even take the ∞ value. A special case of a total accumulated reward is the *termination time*, where all vertices are assigned reward 1, except for terminal vertices that are assigned reward 0 (we also assume that the only outgoing transition of every terminal vertex t is a self-loop on t). Then, $\mathbb{E}_v^{\sigma,\pi}[\mathit{Acc}]$ corresponds to the expected termination time under the strategies σ, π. Another special (and perhaps the simplest) case of a total accumulated reward is *reachability*, where the target vertices are assigned reward 1 and the other vertices have zero reward (here we assume that every target vertex has a single outgoing transition to a special state s with zero reward, where $s \to s$ is the only outgoing transition of s). Although the reachability payoff is bounded, some of our negative results about the total accumulated reward hold even for reachability (see below).

The reason for considering infinite-state games is that many recent works study various algorithmic problems for games over classical automata-theoretic models, such as pushdown automata [15,16,17,14,9,8], lossy channel systems [3,2], one-counter automata [7,5,6], or multicounter automata [18,11,10,21,13,4], which are finitely representable, but for which the underlying game graph is infinite and sometimes even infinitely-branching (see, e.g., [11,10,21]). Since the properties of finite-state games do *not* carry over to infinite-state games in general (see, e.g., [20]), the above issues need to be revisited and clarified explicitly, which is the main goal of this paper.

Our Contribution: We consider general infinite-state games, which may contain vertices with infinitely many outgoing transitions, and ◇-finitely-branching games, where

every vertex controlled by player $\diamond$ has finitely many outgoing transitions, with the total accumulated reward objective. For *general* games, we show the following:

- Every vertex has both a HR and a HD-value, and these values are equal[1].
- There is a vertex v of a game G with a reachability objective such that v has neither MD nor MR-value. Further, the game G has only one vertex (belonging to Player $\diamond$) with infinitely many outgoing transitions.

It follows from previous works (see, e.g., [8,20]) that optimal strategies in general games may not exist, and even if they do exist, they may require infinite memory. Interestingly, we observe that an optimal strategy for Player $\Box$ (if it exists) may also require randomization in some cases.

For $\diamond$-*finitely-branching* games, we prove the following results:

- Every vertex has a HR, HD, MR, and MD-value, and all of these values are equal.
- Player $\diamond$ has an optimal MD strategy in every vertex.

It follows from the previous works that Player $\Box$ may not have an optimal strategy and even if he has one, it may require infinite memory. Let us note that in finite-state games, both players have optimal MD strategies (see, e.g., [19]).

Our results are obtained by generalizing the arguments for reachability objectives presented in [8], but there are also some new observations based on original ideas and new counterexamples. In particular, this applies to the existence of a HD-value and the non-existence of MD and MR-values in general games.

Due to the space constraints, most proofs are omitted. They can be found in the full version of this paper [12].

2 Preliminaries

In this paper, the sets of all positive integers, non-negative integers, rational numbers, real numbers, and non-negative real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{R}^{\geq 0}$, respectively. We also use $\mathbb{R}^{\geq 0}_{\infty}$ to denote the set $\mathbb{R}^{\geq 0} \cup \{\infty\}$, where ∞ is treated according to the standard conventions. For all $c \in \mathbb{R}^{\geq 0}_{\infty}$ and $\varepsilon \in [0, \infty)$, we define the *lower* and *upper* ε-approximation of c, denoted by $c \ominus \varepsilon$ and $c \oplus \varepsilon$, respectively, as follows:

$$\begin{aligned} c \oplus \varepsilon &= c + \varepsilon \quad \text{for all } c \in \mathbb{R}^{\geq 0}_{\infty} \text{ and } \varepsilon \in [0, \infty), \\ c \ominus \varepsilon &= c - \varepsilon \quad \text{for all } c \in \mathbb{R}^{\geq 0} \text{ and } \varepsilon \in [0, \infty), \\ \infty \ominus \varepsilon &= 1/\varepsilon \quad \text{for all } \varepsilon \in (0, \infty), \\ \infty \ominus 0 &= \infty\,. \end{aligned}$$

Given a set V, the elements of $(\mathbb{R}^{\geq 0}_{\infty})^V$ are written as vectors $\boldsymbol{x}, \boldsymbol{y}, \ldots$, where $\boldsymbol{x}_v$ denotes the v-component of $\boldsymbol{x}$ for every $v \in V$. The standard component-wise ordering on $(\mathbb{R}^{\geq 0}_{\infty})^V$ is denoted by $\sqsubseteq$.

[1] For a given strategy type T (such as MD or MR), we say that a vertex v has a T *value* if $\sup_{\sigma \in T_{\Box}} \inf_{\pi \in T_{\diamond}} \mathbb{E}_v^{\sigma,\pi}[\mathit{Payoff}] = \inf_{\pi \in T_{\diamond}} \sup_{\sigma \in T_{\Box}} \mathbb{E}_v^{\sigma,\pi}[\mathit{Payoff}]$, where $T_{\Box}$ and $T_{\diamond}$ are the classes of all T strategies for Player $\Box$ and Player $\diamond$, respectively.

For every finite or countably infinite set M, a binary relation $\to \subseteq M \times M$ is *total* if for every $m \in M$ there is some $n \in M$ such that $m \to n$. A *finite path* in $\mathcal{M} = (M, \to)$ is a finite sequence $w = m_0, \ldots, m_k$ such that $m_i \to m_{i+1}$ for every i, where $0 \leq i < k$. The *length* of w, i.e., the number of transitions performed along w, is denoted by $|w|$. A *run* in $\mathcal{M}$ is an infinite sequence $\omega = m_0, m_1, \ldots$ every finite prefix of which is a path. We also use $\omega(i)$ to denote the element m_i of ω. Given $m, n \in M$, we say that n is *reachable* from m, written $m \to^* n$, if there is a finite path from m to n. The sets of all finite paths and all runs in $\mathcal{M}$ are denoted by $Fpath(\mathcal{M})$ and $Run(\mathcal{M})$, respectively. For every finite path w, we use $Run(\mathcal{M}, w)$ and $Fpath(\mathcal{M}, w)$ to denote the set of all runs and finite paths, respectively, prefixed by w. If $\mathcal{M}$ is clear from the context, we write just Run, $Run(w)$, $Fpath$ and $Fpath(w)$ instead of $Run(\mathcal{M})$, $Run(\mathcal{M}, w)$, $Fpath(\mathcal{M})$ and $Fpath(\mathcal{M}, w)$, respectively.

Now we recall basic notions of probability theory. Let A be a finite or countably infinite set. A *probability distribution* on A is a function $f : A \to \mathbb{R}^{\geq 0}$ such that $\sum_{a \in A} f(a) = 1$. A distribution f is *positive* if $f(a) > 0$ for every $a \in A$, *Dirac* if $f(a) = 1$ for some $a \in A$, and *uniform* if A is finite and $f(a) = \frac{1}{|A|}$ for every $a \in A$. A *σ-field* over a set X is a set $\mathcal{F} \subseteq 2^X$ that includes X and is closed under complement and countable union. A *measurable space* is a pair $(X, \mathcal{F})$ where X is a set called *sample space* and $\mathcal{F}$ is a σ-field over X. A *probability measure* over a measurable space $(X, \mathcal{F})$ is a function $\mathcal{P} : \mathcal{F} \to \mathbb{R}^{\geq 0}$ such that, for each countable collection $\{X_i\}_{i \in I}$ of pairwise disjoint elements of $\mathcal{F}$, $\mathcal{P}(\bigcup_{i \in I} X_i) = \sum_{i \in I} \mathcal{P}(X_i)$, and moreover $\mathcal{P}(X) = 1$. A *probability space* is a triple $(X, \mathcal{F}, \mathcal{P})$ where $(X, \mathcal{F})$ is a measurable space and $\mathcal{P}$ is a probability measure over $(X, \mathcal{F})$.

Definition 1. *A* stochastic game *is a tuple $G = (V, \to, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ where V is a finite or countably infinite set of* vertices*, $\to \subseteq V \times V$ is a total* transition relation*, $(V_\Box, V_\Diamond, V_\bigcirc)$ is a partition of V, and $Prob$ is a* probability assignment *which to each $v \in V_\bigcirc$ assigns a positive probability distribution on the set of its outgoing transitions. We say that G is* $\Diamond$-finitely-branching *if for each $v \in V_\Diamond$ there are only finitely many $u \in V$ such that $v \to u$.*

Strategies. A stochastic game G is played by two players, $\Box$ and $\Diamond$, who select the moves in the vertices of $V_\Box$ and $V_\Diamond$, respectively. Let $\odot \in \{\Box, \Diamond\}$. A *strategy* for Player $\odot$ in G is a function which to each finite path in G ending in a vertex $v \in V_\odot$ assigns a probability distribution on the set of outgoing transitions of v. We say that a strategy τ is *memoryless (M)* if $\tau(w)$ depends just on the last vertex of w, and *deterministic (D)* if it returns a Dirac distribution for every argument. Strategies that are not necessarily memoryless are called *history-dependent (H)*, and strategies that are not necessarily deterministic are called *randomized (R)*. Thus, we obtain the MD, MR, HD, and HR *strategy types*. The set of all strategies for Player $\odot$ of type T in a game G is denoted by $T^G_\odot$, or just by $T_\odot$ if G is understood (for example, $\mathrm{MR}_\Box$ denotes the set of all MR strategies for Player $\Box$).

Every pair of strategies $(\sigma, \pi) \in \mathrm{HR}_\Box \times \mathrm{HR}_\Diamond$ and an initial vertex v determine a unique probability space $(Run(v), \mathcal{F}, \mathcal{P}_v^{\sigma,\pi})$, where $\mathcal{F}$ is the smallest σ-field over $Run(v)$ containing all the sets $Run(w)$ such that w starts with v, and $\mathcal{P}_v^{\sigma,\pi}$ is the unique probability measure such that for every finite path $w = v_0, \ldots, v_k$ initiated in v we have

that $\mathcal{P}_v^{\sigma,\pi}(Run(w)) = \Pi_{i=0}^{k-1} x_i$, where x_i is the probability of $v_i \to v_{i+1}$ assigned either by $\sigma(v_0, \ldots, v_i)$, $\pi(v_0, \ldots, v_i)$, or $Prob(v_i)$, depending on whether v_i belongs to $V_\Box$, $V_\Diamond$, or $V_\bigcirc$, respectively (in the case when $k = 0$, i.e., $w = v$, we put $\mathcal{P}_v^{\sigma,\pi}(Run(w)) = 1$).

Determinacy, Optimal Strategies. In this paper, we consider games with the *total accumulated reward* objective and *reachability* objective, where the latter is understood as a restricted form of the former (see below).

Let $r : V \to \mathbb{R}^{\geq 0}$ be a *reward function*, and $Acc : Run \to \mathbb{R}_\infty^{\geq 0}$ a function which to every run ω assigns the *total accumulated reward* $Acc(\omega) = \sum_{i=0}^{\infty} r(\omega(i))$. Let T be a strategy type. We say that a vertex $v \in V$ *has a* T*-value* in G if

$$\sup_{\sigma \in T_\Box} \inf_{\pi \in T_\Diamond} \mathbb{E}_v^{\sigma,\pi}[Acc] \quad = \quad \inf_{\pi \in T_\Diamond} \sup_{\sigma \in T_\Box} \mathbb{E}_v^{\sigma,\pi}[Acc]\,,$$

where $\mathbb{E}_v^{\sigma,\pi}[Acc]$ denotes the expected value of Acc in $(Run(v), \mathcal{F}, \mathcal{P}_v^{\sigma,\pi})$. If v has a T-value, then $\mathrm{Val}_T(v, r, G)$ (or just $\mathrm{Val}_T(v)$ if G and r are clear from the context) denotes the T*-value of* v defined by this equality.

Let $\mathcal{G}$ be a class of games. If every vertex of every $G \in \mathcal{G}$ has a T-value for every reward function, we say that $\mathcal{G}$ is T*-determined*. Note that Acc is generally not bounded, and therefore we cannot directly apply the results of [23,22] to conclude that the class of all games is HR-determined. Further, these results do not say anything about the determinacy for the other strategy types even for bounded objective functions.

If a given vertex v has a T-value, we can define the notion of ε-optimal T strategy for both players.

Definition 2. *Let v be a vertex which has a T-value, and let $\varepsilon \geq 0$. We say that*

- $\sigma \in T_\Box$ *is* ε-T-optimal *in v if* $\mathbb{E}_v^{\sigma,\pi}[Acc] \geq \mathrm{Val}_T(v) \ominus \varepsilon$ *for all* $\pi \in T_\Diamond$*;*
- $\pi \in T_\Diamond$ *is* ε-T-optimal *in v if* $\mathbb{E}_v^{\sigma,\pi}[Acc] \leq \mathrm{Val}_T(v) \oplus \varepsilon$ *for all* $\sigma \in T_\Box$.

A 0-T*-optimal strategy is called* T-optimal.

In this paper we also consider *reachability* objectives, which can be seen as a restricted form of the total accumulated reward objectives introduced above. A "standard" definition of the reachability payoff function looks as follows: We fix a set $R \subseteq V$ of *target* vertices, and define a function $Reach : Run \to \{0, 1\}$ which to every run assigns either 1 or 0 depending on whether or not the run visits a target vertex. Note that $\mathbb{E}_v^{\sigma,\pi}[Reach]$ is the *probability* of visiting a target vertex in the corresponding play of G. Obviously, if we assign reward 1 to the target vertices and 0 to the others, and replace all outgoing transitions of target vertices with a single transition leading to a fresh stochastic vertex u with reward 0 and only one transition $u \to u$, then $\mathbb{E}_v^{\sigma,\pi}[Reach]$ in the original game is equal to $\mathbb{E}_v^{\sigma,\pi}[Acc]$ in the modified game. Further, if the original game was $\Diamond$-finitely-branching or finite, then so is the modified game. Therefore, all "positive" results about the total accumulated reward objective (e.g., determinacy, existence of T-optimal strategies, etc.) achieved in this paper carry over to the reachability objective, and all "negative" results about reachability carry over to the total accumulated reward.

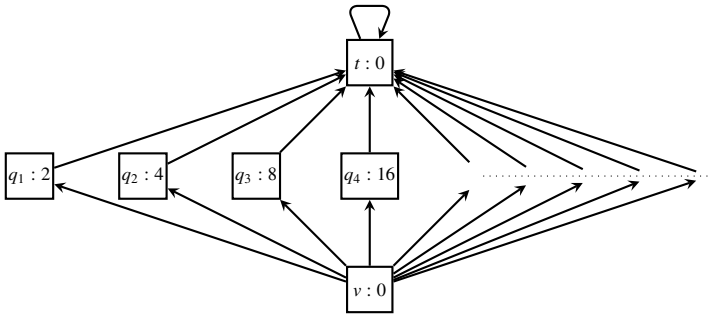

Fig. 1. Player □ has an MR-optimal strategy in v, but no HD-optimal strategy in v. All vertices are labelled by pairs of the form *vertex name*:*reward*.

3 Results

Our main results about the determinacy of general stochastic games with the total accumulated reward payoff function are summarized in the following theorem:

Theorem 3. *Let $\mathcal{G}$ be the class of all games. Then*

a) $\mathcal{G}$ is both HR-determined and HD-determined. Further, for every vertex v of every $G \in \mathcal{G}$ and every reward function r we have that $\mathrm{Val}_{\mathrm{HR}}(v) = \mathrm{Val}_{\mathrm{HD}}(v)$.
b) $\mathcal{G}$ is neither MD-determined nor MR-determined, and these results hold even for reachability objectives.

An optimal strategy for Player □ does not necessarily exist, even if G is a game with a reachability payoff function such that $V_{\diamond} = \emptyset$ and every vertex of $V_{\square}$ has at most two outgoing transitions (see, e.g., [8,20]). In fact, it suffices to consider the vertex v of Fig. 2 where the depicted game is modified by replacing the vertex u with a stochastic vertex u', where $u' \to u'$ is the only outgoing transition of u', and u' is the only target vertex (note that all vertices in the first two rows become unreachable and can be safely deleted). Clearly, $\mathrm{Val}_{\mathrm{HR}}(v) = 1$, but Player □ has no optimal strategy.

Similarly, an optimal strategy for Player ◇ may not exist even if $V_{\square} = \emptyset$ [8,20]. To see this, consider the vertex u of Fig. 2, where t is the only target vertex and the depicted game is modified by redirecting the only outgoing transition of p back to u (this makes all vertices in the last two rows unreachable). We have that $\mathrm{Val}_{\mathrm{HR}}(u) = 0$, but Player ◇ has no optimal strategy.

One may be also tempted to think that if Player □ (or Player ◇) has *some* optimal strategy, then he also has an optimal MD strategy. However, optimal strategies generally require *infinite memory* even for reachability objectives (this holds for both players). Since the corresponding counterexamples are not completely trivial, we refer to [20] for details. Interestingly, an optimal strategy for Player □ may also require *randomization*. Consider the vertex v of Fig. 1. Let $\sigma^* \in \mathrm{MR}_{\square}$ be a strategy selecting $v \to q_n$ with probability $1/2^n$. Since $V_{\diamond} = \emptyset$, we have that $\inf_{\pi \in \mathrm{HR}_{\diamond}} \mathbb{E}_v^{\sigma^*,\pi}[Acc] = \infty = \mathrm{Val}_{\mathrm{HR}}(v)$. However, for every $\sigma \in \mathrm{HD}_{\square}$ we have that $\inf_{\pi \in \mathrm{HR}_{\diamond}} \mathbb{E}_v^{\sigma,\pi}[Acc] < \infty$.

For $\diamond$-finitely-branching games, the situation is somewhat different, as our second main theorem reveals.

Theorem 4. *Let $\mathcal{G}$ be the class of all $\diamond$-finitely-branching games. Then $\mathcal{G}$ is HR-determined, HD-determined, MR-determined, and MD-determined, and for every vertex v of every $G \in \mathcal{G}$ and every reward function r we have that*

$$\mathrm{Val}_{\mathrm{HR}}(v) = \mathrm{Val}_{\mathrm{HD}}(v) = \mathrm{Val}_{\mathrm{MR}}(v) = \mathrm{Val}_{\mathrm{MD}}(v)\,.$$

Further, for every $G \in \mathcal{G}$ there exists an MD strategy for Player $\diamond$ which is optimal in every vertex of G.

An optimal strategy for Player $\Box$ may not exist in $\diamond$-finitely-branching games, and even if it does exist, it may require infinite memory [20].

Theorems 3 and 4 are proven by a sequence of lemmas presented below. For the rest of this section, we fix a stochastic game $G = (V, \to, (V_\Box, V_\diamond, V_\bigcirc), \mathit{Prob})$ and a reward function $r\colon V \to \mathbb{R}^{\geq 0}$. We start with the first part of Theorem 3 (a), i.e., we show that every vertex has a HR-value. This is achieved by defining a suitable Bellman operator L and proving that the least fixed-point of L is the tuple of all HR-values. More precisely, let $L\colon (\mathbb{R}^{\geq 0}_\infty)^V \to (\mathbb{R}^{\geq 0}_\infty)^V$, where $\boldsymbol{y} = L(\boldsymbol{x})$ is defined as follows:

$$\boldsymbol{y}_v = \begin{cases} r(v) + \sup_{v \to v'} \boldsymbol{x}_{v'} & \text{if } v \in V_\Box \\ r(v) + \inf_{v \to v'} \boldsymbol{x}_{v'} & \text{if } v \in V_\diamond \\ r(v) + \sum_{v \to v'} \boldsymbol{x}_{v'} \cdot \mathit{Prob}(v)(v, v') & \text{if } v \in V_\bigcirc. \end{cases}$$

A proof of the following lemma can be found in the full version of this paper. Some parts of this proof are subtle, and we also need to make several observations that are useful for proving the other results.

Lemma 5. *The operator L has the least fixed point $\boldsymbol{K}$ (w.r.t. $\sqsubseteq$) and for every $v \in V$ we have that*

$$\boldsymbol{K}_v \;=\; \sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma,\pi}[\mathit{Acc}] \;=\; \inf_{\pi \in \mathrm{HR}_\diamond} \sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi}[\mathit{Acc}] \;=\; \mathrm{Val}_{\mathrm{HR}}(v)\,.$$

Moreover, for every $\varepsilon > 0$ there is $\pi_\varepsilon \in \mathrm{HD}_\diamond$ such that for every $v \in V$ we have that $\sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi_\varepsilon} \leq \mathrm{Val}_{\mathrm{HR}}(v) \oplus \varepsilon$.

To complete our proof of Theorem 3 (a), we need to show the existence of a HD-value in every vertex, and demonstrate that HR and HD values are equal. Due to Lemma 5, for every $\varepsilon > 0$ there is $\pi_\varepsilon \in \mathrm{HD}_\diamond$ such that π_ε is ε-HR-optimal in every vertex. Hence, it suffices to show the same for Player $\Box$. The following lemma is proved in the full version.

Lemma 6. *For every $\varepsilon > 0$, there is $\sigma_\varepsilon \in \mathrm{HD}_\Box$ such that σ_ε is ε-HR-optimal in every vertex.*

The next lemma proves Item (b) of Theorem 3.

Lemma 7. *Consider the vertex v of the game shown in Fig. 2, where t is the only target vertex and all probability distributions assigned to stochastic states are uniform. Then*

(a) $\sup_{\sigma\in \mathrm{MD}_\Box} \inf_{\pi\in \mathrm{MD}_\Diamond} \mathbb{E}_v^{\sigma,\pi}[Reach] = \sup_{\sigma\in \mathrm{MR}_\Box} \inf_{\pi\in \mathrm{MR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}[Reach] = 0$;
(b) $\inf_{\pi\in \mathrm{MD}_\Diamond} \sup_{\sigma\in \mathrm{MD}_\Box} \mathbb{E}_v^{\sigma,\pi}[Reach] = \inf_{\pi\in \mathrm{MR}_\Diamond} \sup_{\sigma\in \mathrm{MR}_\Box} \mathbb{E}_v^{\sigma,\pi}[Reach] = 1$.

Proof. We start by proving item (a) for MD strategies. Let $\sigma^* \in \mathrm{MD}_\Box$. We show that $\inf_{\pi\in \mathrm{MD}_\Diamond} \mathbb{E}_v^{\sigma^*,\pi}[Reach] = 0$. Let us fix an arbitrarily small $\varepsilon > 0$. We show that there is a suitable $\pi^* \in \mathrm{MD}_\Diamond$ such that $\mathbb{E}_v^{\sigma^*,\pi^*}[Reach] \leq \varepsilon$. If the probability of reaching the vertex u from v under the strategy σ^* is at most ε, we are done. Otherwise, let p_s be the probability of visiting the vertex s from v under the strategy σ *without* passing through the vertex u. Note that $p_s > 0$ and p_s does not depend on the strategy chosen by Player $\Diamond$. The strategy π^* selects a suitable successor of u such that the probability p_t of visiting the vertex t from u without passing through the vertex v satisfies $p_t/p_s < \varepsilon$ (note that p_t can be arbitrarily small but positive). Then

$$\mathbb{E}_v^{\sigma^*,\pi^*}[Reach] \quad \leq \quad \sum_{i=1}^{\infty}(1-p_s)^i p_t \quad = \quad \frac{(1-p_s)p_t}{p_s} \quad \leq \quad \varepsilon\,.$$

For MR strategies, the argument is the same.

Item (b) is proven similarly. We show that for all $\pi^* \in \mathrm{MD}_\Diamond$ and $0 < \varepsilon < 1$ there exists a suitable $\sigma^* \in \mathrm{MD}_\Box$ such that $\mathbb{E}_v^{\sigma^*,\pi^*}[Reach] \geq 1-\varepsilon$. Let p_t be the probability of visiting t from u without passing through the vertex v under the strategy π^*. We choose the strategy σ^* so that the probability p_s of visiting the vertex s from v without passing through the vertex u satisfies $p_s/p_t < \varepsilon$. Note that almost all runs initiated in v eventually visit either s or t under (σ^*,π^*). Since the probability of visiting s is bounded by ε (the computation is similar to the one of item (a)), we obtain $\mathbb{E}_v^{\sigma^*,\pi^*}[Reach] \geq 1-\varepsilon$. For MR strategies, the proof is almost the same. □

We continue by proving Theorem 4. This theorem follows immediately from Lemma 5 and the following proposition:

Proposition 8. *If G is $\Diamond$-finitely-branching, then*

1. *for all $v \in V$ and $\varepsilon > 0$, there is $\sigma_\varepsilon \in \mathrm{MD}_\Box$ such that σ_ε is ε-*HR*-optimal in v;*
2. *there is $\pi \in \mathrm{MD}_\Diamond$ such that π is* HR*-optimal in every vertex.*

As an immediate corollary to Proposition 8, we obtain the following result:

Corollary 9. *If G is $\Diamond$-finitely-branching, $V_\Box$ is finite, and every vertex of $V_\Box$ has finitely many successors, then there is $\sigma \in \mathrm{MD}_\Box$ such that σ is* HR*-optimal in every vertex.*

Proof. Due to Proposition 8, for every vertex v and every $\varepsilon > 0$, there is $\sigma_\varepsilon \in \mathrm{MD}_\Box$ such that σ_ε is ε-HR-optimal in v. Since $V_\Box$ is finite and every vertex of $V_\Box$ has only finitely many successors, there are only finitely many MD strategies for Player □. Hence, there is a MD strategy σ that is ε-HR-optimal in v for infinitely many ε from the set $\{1, 1/2, 1/4, \ldots\}$. Such a strategy is clearly HR-optimal in v. Note that σ is HR-optimal in every vertex which can be reached from v under σ and some strategy π for Player $\Diamond$. For the remaining vertices, we can repeat the argument, and thus eventually produce a MD strategy that is HR-optimal in every vertex. □

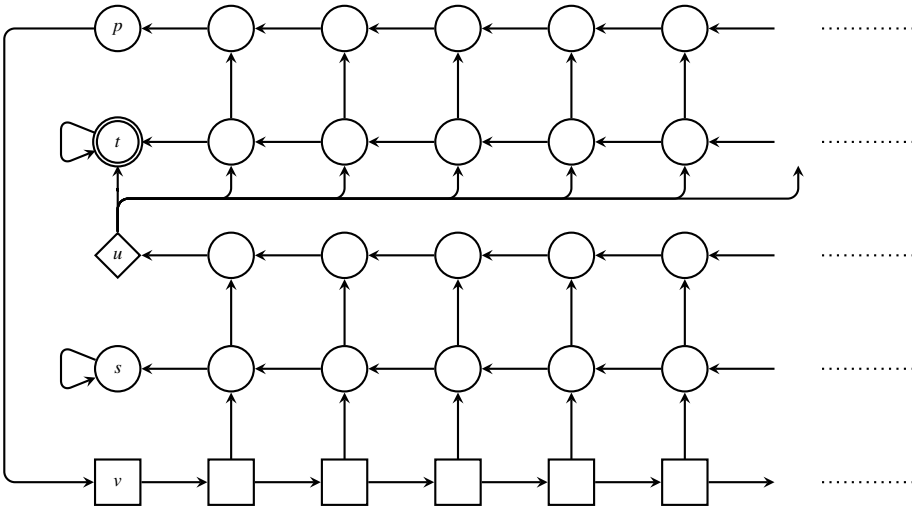

Fig. 2. A game whose vertex v has neither MD-value nor MR-value

Hence, if all non-stochastic vertices have finitely many successors and $V_\Box$ is finite, then both players have HR-optimal MD strategies. This can be seen as a (tight) generalization of the corresponding result for finite-state games [19].

The rest of this section is devoted to a proof of Proposition 8. We start with Item 1. The strategy σ_ε is constructed by employing discounting. Assume, w.l.o.g., that rewards are bounded by 1 (if they are not, we may split every state v with a reward $r(v)$ into a sequence of $\lceil r(v) \rceil$ states, each with the reward $r(v)/\lceil r(v) \rceil$). Given $\lambda \in (0,1)$, define $Acc^\lambda : Run \to \mathbb{R}^{\geq 0}$ to be a function which to every run ω assigns $Acc^\lambda(\omega) = \sum_{i=0}^{\infty} \lambda^i \cdot r(\omega(i))$.

Lemma 10. *For λ sufficiently close to one we have that*

$$\sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) \quad \geq \quad \mathrm{Val}_{\mathrm{HR}}(v) \ominus \frac{\varepsilon}{2}\,.$$

Proof. We show that for every $\varepsilon > 0$ there is $n \geq 0$ such that the expected reward that Player $\Box$ may accumulate up to n steps is ε-close to $\mathrm{Val}_{\mathrm{HR}}(v)$ no matter what Player $\Diamond$ is doing. Formally, define $Acc_k : Run \to \mathbb{R}^{\geq 0}$ to be a function which to every run ω assigns $Acc_k(\omega) = \sum_{i=0}^{k} r(\omega(i))$. The following lemma is proved in the full version of this paper.

Lemma 11. *If G is $\Diamond$-finitely-branching, then for every $v \in V$ there is $n \in \mathbb{N}$ such that*

$$\sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc_n) \quad > \quad \mathrm{Val}_{\mathrm{HR}}(v) \ominus \frac{\varepsilon}{4}\,.$$

Clearly, if λ is close to one, then for every run ω we have that

$$Acc^\lambda(\omega) \quad \geq \quad Acc_n(\omega) - \frac{\varepsilon}{4}\,.$$

Thus,

$$\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc_n) - \frac{\varepsilon}{4} \quad \geq \quad \mathrm{Val}_{\mathrm{HR}}(v) \ominus \frac{\varepsilon}{2}\,.$$

This proves Lemma 10. □

So, it suffices to find an MD strategy σ_ε satisfying

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}(Acc^\lambda) \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{2}\,.$$

We define such a strategy as follows. Let us fix some $\ell \in \mathbb{N}$ satisfying

$$\frac{\lambda^\ell}{1-\lambda}\cdot \max_{v\in V} r(v) \quad < \quad \frac{\varepsilon}{8}\,.$$

Intuitively, the discounted reward accumulated after ℓ steps can be at most $\frac{\varepsilon}{8}$. In a given vertex $v \in V_\Box$, the strategy σ_ε chooses a fixed successor vertex u satisfying

$$\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_u^{\sigma,\pi}(Acc^\lambda) \quad \geq \quad \sup_{v\to u'}\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_{u'}^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{\ell\cdot 4}$$

Now we show that

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}(Acc^\lambda) \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{2}\,.$$

which finishes the proof of Item 1 of Proposition 8.

For every $k \in \mathbb{N}$ we denote by σ_k a strategy for Player □ defined as follows: For the first k steps the strategy behaves similarly to σ_ε, i.e., chooses, in each state $v \in V_\Box$, a next state u satisfying

$$\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_u^{\sigma,\pi}(Acc^\lambda) \quad \geq \quad \sup_{v\to u'}\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_{u'}^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{k\cdot 4}\,.$$

From k+1-st step on, say in a state u, the strategy follows some strategy ζ satisfying

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_u^{\zeta,\pi}(Acc^\lambda) \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_u^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{8}\,.$$

A simple induction reveals that σ_k satisfies

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma_k,\pi}(Acc^\lambda) \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) - \frac{3\varepsilon}{8}\,. \tag{2}$$

(Intuitively, the error of each of the first k steps is at most $\frac{\varepsilon}{k\cdot 4}$ and thus the total error of the first k steps is at most $k\cdot\frac{\varepsilon}{k\cdot 4} = \frac{\varepsilon}{4}$. The rest has the error at most $\frac{\varepsilon}{8}$ and thus the total error is at most $\frac{3\varepsilon}{8}$.)

We consider $k = \ell$ (recall that $\frac{\lambda^\ell}{1-\lambda}\cdot\max_{v\in V} r(v) < \frac{\varepsilon}{8}$). Then

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}(Acc^\lambda) \quad \geq \quad \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma_k,\pi}(Acc^\lambda) - \frac{\varepsilon}{8} \quad \geq \quad \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc^\lambda) - \frac{\varepsilon}{2}\,.$$

Here the first equality follows from the fact that σ_k behaves similarly to σ_ε on the first $k = \ell$ steps and the discounted reward accumulated after k steps is at most $\frac{\varepsilon}{8}$. The second inequality follows from Equation (2).

It remains to prove Item 2 of Proposition 8. The MD strategy π can be easily constructed as follows: In every state $v \in V_\diamond$, the strategy π chooses a successor u minimizing $\mathrm{Val}_{\mathrm{HR}}(u)$ among all successors of v. We show in the full version that this is indeed an optimal strategy.

4 Conclusions

We have considered infinite-state stochastic games with the total accumulated reward objective, and clarified the determinacy questions for the HR, HD, MR, and MD strategy types. Our results are almost complete. One natural question which remains open is whether Player $\Box$ needs memory to play ε-HR-optimally in general games (it follows from the previous works, e.g., [8,20], that ε-HR-optimal strategies for Player $\diamond$ require infinite memory in general).

References

1. Proceedings of FST&TCS 2010, Leibniz International Proceedings in Informatics, vol. 8. Schloss Dagstuhl–Leibniz-Zentrum für Informatik (2010)
2. Abdulla, P.A., Ben Henda, N., de Alfaro, L., Mayr, R., Sandberg, S.: Stochastic Games with Lossy Channels. In: Amadio, R.M. (ed.) FoSSaCS 2008. LNCS, vol. 4962, pp. 35–49. Springer, Heidelberg (2008)
3. Baier, C., Bertrand, N., Schnoebelen, P.: On Computing Fixpoints in Well-Structured Regular Model Checking, with Applications to Lossy Channel Systems. In: Hermann, M., Voronkov, A. (eds.) LPAR 2006. LNCS (LNAI), vol. 4246, pp. 347–361. Springer, Heidelberg (2006)
4. Bouyer, P., Fahrenberg, U., Larsen, K.G., Markey, N., Srba, J.: Infinite Runs in Weighted Timed Automata with Energy Constraints. In: Cassez, F., Jard, C. (eds.) FORMATS 2008. LNCS, vol. 5215, pp. 33–47. Springer, Heidelberg (2008)
5. Brázdil, T., Brožek, V., Etessami, K.: One-counter stochastic games. In: Proceedings of FST&TCS 2010 [1], pp. 108–119
6. Brázdil, T., Brožek, V., Etessami, K., Kučera, A.: Approximating the Termination Value of One-Counter MDPs and Stochastic Games. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011, Part II. LNCS, vol. 6756, pp. 332–343. Springer, Heidelberg (2011)
7. Brázdil, T., Brožek, V., Etessami, K., Kučera, A., Wojtczak, D.: One-counter Markov decision processes. In: Proceedings of SODA 2010, pp. 863–874. SIAM (2010)
8. Brázdil, T., Brožek, V., Forejt, V., Kučera, A.: Reachability in recursive Markov decision processes. Information and Computation 206(5), 520–537 (2008)
9. Brázdil, T., Brožek, V., Kučera, A., Obdržálek, J.: Qualitative reachability in stochastic BPA games. Information and Computation 208(7), 772–796 (2010)
10. Brázdil, T., Chatterjee, K., Kučera, A., Novotný, P.: Efficient Controller Synthesis for Consumption Games with Multiple Resource Types. In: Madhusudan, P., Seshia, S.A. (eds.) CAV 2012. LNCS, vol. 7358, pp. 23–38. Springer, Heidelberg (2012)
11. Brázdil, T., Jančar, P., Kučera, A.: Reachability Games on Extended Vector Addition Systems with States. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010, Part II. LNCS, vol. 6199, pp. 478–489. Springer, Heidelberg (2010)

12. Brázdil, T., Kučera, A., Novotný, P.: Determinacy in stochastic games with unbounded payoff functions. CoRR abs/1208.1639 (2012)
13. Chatterjee, K., Doyen, L., Henzinger, T., Raskin, J.F.: Generalized mean-payoff and energy games. In: Proceedings of FST&TCS 2010 [1], pp. 505–516
14. Etessami, K., Wojtczak, D., Yannakakis, M.: Recursive Stochastic Games with Positive Rewards. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) ICALP 2008, Part I. LNCS, vol. 5125, pp. 711–723. Springer, Heidelberg (2008)
15. Etessami, K., Yannakakis, M.: Recursive Markov Decision Processes and Recursive Stochastic Games. In: Caires, L., Italiano, G.F., Monteiro, L., Palamidessi, C., Yung, M. (eds.) ICALP 2005. LNCS, vol. 3580, pp. 891–903. Springer, Heidelberg (2005)
16. Etessami, K., Yannakakis, M.: Efficient Qualitative Analysis of Classes of Recursive Markov Decision Processes and Simple Stochastic Games. In: Durand, B., Thomas, W. (eds.) STACS 2006. LNCS, vol. 3884, pp. 634–645. Springer, Heidelberg (2006)
17. Etessami, K., Yannakakis, M.: Recursive Concurrent Stochastic Games. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4052, pp. 324–335. Springer, Heidelberg (2006)
18. Fahrenberg, U., Juhl, L., Larsen, K.G., Srba, J.: Energy Games in Multiweighted Automata. In: Cerone, A., Pihlajasaari, P. (eds.) ICTAC 2011. LNCS, vol. 6916, pp. 95–115. Springer, Heidelberg (2011)
19. Filar, J., Vrieze, K.: Competitive Markov Decision Processes. Springer (1996)
20. Kučera, A.: Turn-based stochastic games. In: Apt, K.R., Grädel, E. (eds.). Lectures in Game Theory for Computer Scientists, pp. 146–184. Cambridge University Press (2011)
21. Kučera, A.: Playing Games with Counter Automata. In: Finkel, A., Leroux, J., Potapov, I. (eds.) RP 2012. LNCS, vol. 7550, pp. 29–41. Springer, Heidelberg (2012)
22. Maitra, A., Sudderth, W.: Finitely additive stochastic games with Borel measurable payoffs. International Journal of Game Theory 27, 257–267 (1998)
23. Martin, D.: The determinacy of Blackwell games. Journal of Symbolic Logic 63(4), 1565–1581 (1998)