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

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 $\Box$ is to *maximize* the expected payoff, while Player $\Diamond$ 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 $\Box$ and Player $\Diamond$, 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 all vertices visited along the run. Obviously, *Acc* is unbounded in general, and may even take the ∞ value. A special case of total accumulated reward is *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 simplest) case of 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 [18,11,10,21,12,4], which are finitely representable but 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 $\Diamond$-finitely-branching games,

where every vertex of $V_\diamond$ 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 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 $\square$ (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 $\square$ 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.

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:

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

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 $\sqsubseteq$.

For every finite or countably infinite set M, a binary relation $\rightarrow \subseteq M \times M$ is *total* if for every $m \in M$ there is some $n \in M$ such that $m \rightarrow n$. A *finite path* in $\mathcal{M} = (M, \rightarrow)$

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

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, and ω_i to denote the run $m_i, m_{i+1}, \ldots$ 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 *rational* if $f(a) \in \mathbb{Q}$ for every $a \in A$, *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 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 $(Run(v), \mathcal{F}, \mathcal{P}^{\sigma,\pi}_v)$, where $\mathcal{F}$ is the σ-field over $Run(v)$ generated by all $Run(w)$ such that w starts with v, and $\mathcal{P}^{\sigma,\pi}_v$ is the unique probability measure such that for every finite path $w = v_0, \ldots, v_k$ initiated in v we have that $\mathcal{P}^{\sigma,\pi}_v(Run(w)) = \Pi_{i=0}^{k-1} x_i$, 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 $Acc : Run \to \mathbb{R}^{\geq 0}_{\infty}$ 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 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.

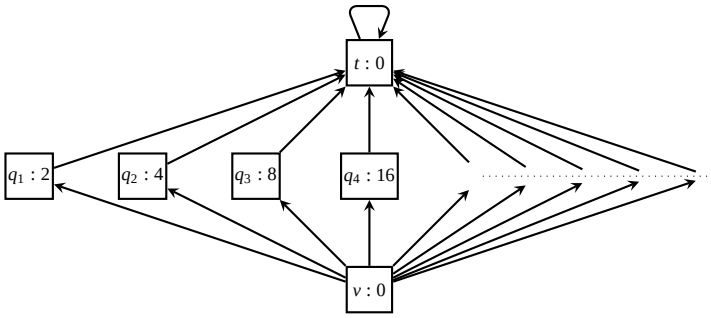

Fig. 1. Player □ has an MR-optimal strategy in v, but no HD-optimal strategy in v. All vertices are labelled by pairs of the form *vertex name*:*reward*.

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 a 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 Appendix A. 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 Appendix B.

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

Proof. We start by proving item (a) for MD strategies. Let $\sigma^* \in \mathrm{MD}_\Box$. We show that $\inf_{\pi \in \mathrm{MD}_\Diamond} \mathbb{E}_v^{\sigma^*,\pi}[\mathit{Reach}] = 0$. Let us fix an arbitrarily small $\varepsilon > 0$. We show that there is a suitable $\pi^* \in \mathrm{MD}_\Diamond$ such that $\mathbb{E}_v^{\sigma^*,\pi^*}[\mathit{Reach}] \leq \varepsilon$. 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

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

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 $\mathbb{E}_v^{\sigma^*,\pi^*}[\mathit{Reach}] \geq 1 - \varepsilon$. 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 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 $\mathbb{E}_v^{\sigma^*,\pi^*}[\mathit{Reach}] \geq 1 - \varepsilon$. 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 $\Box$. 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. □

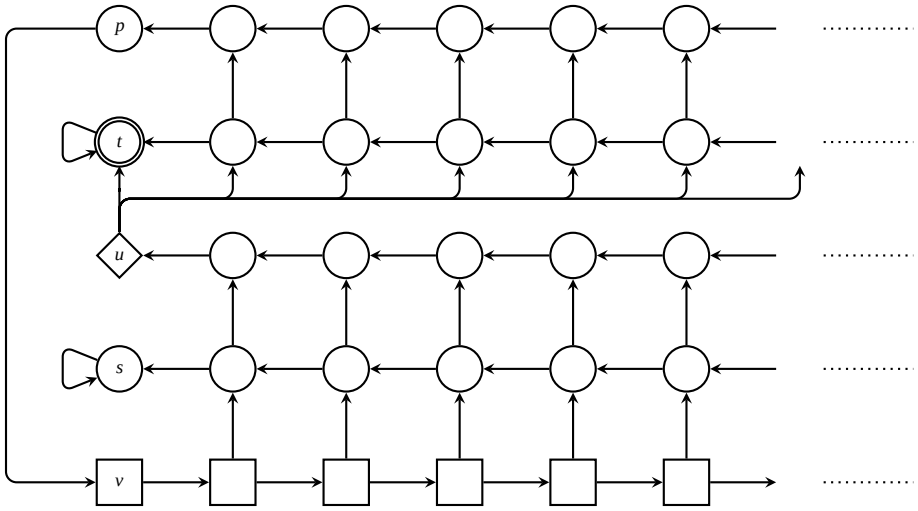

Fig. 2. A game whose vertex v has neither MD-value nor MR-value.

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 Appendix C.

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 a 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 makes the same choices as 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 Appendix D 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 $\square$ 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., Henda, N., de Alfaro, L., Mayr, R., Sandberg, S.: Stochastic games with lossy channels. In: Proceedings of FoSSaCS 2008. Lecture Notes in Computer Science, vol. 4962, pp. 35–49. Springer (2008)
3. Baier, C., Bertrand, N., Schnoebelen, P.: On computing fixpoints in well-structured regular model checking, with applications to lossy channel systems. In: Proceedings of LPAR 2006. Lecture Notes in Computer Science, vol. 4246, pp. 347–361. Springer (2006)
4. Bouyer, P., Fahrenberg, U., Larsen, K., Markey, N., Srba, J.: Infinite runs in weighted timed automata with energy constraints. In: Proceedings of FORMATS 2008. Lecture Notes in Computer Science, vol. 5215, pp. 33–47. Springer (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: Proceedings of ICALP 2011, Part II. Lecture Notes in Computer Science, vol. 6756, pp. 332–343. Springer (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: Proceedings of CAV 2012. Lecture Notes in Computer Science, vol. 7358, pp. 23–38. Springer (2012)
11. Brázdil, T., Jančar, P., Kučera, A.: Reachability games on extended vector addition systems with states. In: Proceedings of ICALP 2010, Part II. Lecture Notes in Computer Science, vol. 6199, pp. 478–489. Springer (2010)

12. 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
13. Cousot, P., Cousot, R.: Constructive versions of Tarski's fixed point theorems. Pacific Journal of Mathematics 82, 43–57 (1979)
14. Etessami, K., Wojtczak, D., Yannakakis, M.: Recursive stochastic games with positive rewards. In: Proceedings of ICALP 2008, Part I. Lecture Notes in Computer Science, vol. 5125, pp. 711–723. Springer (2008)
15. Etessami, K., Yannakakis, M.: Recursive Markov decision processes and recursive stochastic games. In: Proceedings of ICALP 2005. Lecture Notes in Computer Science, vol. 3580, pp. 891–903. Springer (2005)
16. Etessami, K., Yannakakis, M.: Efficient qualitative analysis of classes of recursive Markov decision processes and simple stochastic games. In: Proceedings of STACS 2006. Lecture Notes in Computer Science, vol. 3884, pp. 634–645. Springer (2006)
17. Etessami, K., Yannakakis, M.: Recursive concurrent stochastic games. In: Proceedings of ICALP 2006. Lecture Notes in Computer Science, vol. 4052, pp. 324–335. Springer (2006)
18. Fahrenberg, U., Juhl, L., Larsen, K., Srba, J.: Energy games in multiweighted automata. In: Proceedings of the 8th International Colloquium on Theoretical Aspects of Computing (ICTAC'11). Lecture Notes in Computer Science, vol. 6916, pp. 95–115. Springer (2011)
19. Filar, J., Vrieze, K.: Competitive Markov Decision Processes. Springer (1996)
20. Kučera, A.: Turn-based stochastic games. In: K.R. Apt, E. Grädel (Eds.): Lectures in Game Theory for Computer Scientists. pp. 146–184. Cambridge University Press (2011)
21. Kučera, A.: Playing games with counter automata. In: Reachability Problems. Lecture Notes in Computer Science, Springer (2012), To Appear.
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)

Technical Appendix

A Proof of Lemma 5

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*

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

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

The partially ordered set $((\mathbb{R}^{\geq 0}_\infty)^V, \sqsubseteq)$, where $\sqsubseteq$ is a standard componentwise ordering, is a complete lattice. Moreover, from the definition of L we can easily see that L is monotonic, i.e. $L(\boldsymbol{x}) \sqsubseteq L(\boldsymbol{x}')$ whenever $\boldsymbol{x} \sqsubseteq \boldsymbol{x}'$. Thus, by the Knaster-Tarski theorem the operator L has the least fixed point, which we denote by $\boldsymbol{K}$.

In order to prove that $\boldsymbol{K}_v = \mathrm{Val}_{\mathrm{HR}}(v)$ for every $v \in V$, it suffices to prove the following:

$$\forall v \in V: \quad \boldsymbol{K}_v \leq \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc) \leq \inf_{\pi\in \mathrm{HR}_\Diamond}\sup_{\sigma\in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi}(Acc) \leq \boldsymbol{K}_v. \tag{3}$$

The second inequality holds trivially, so it suffices to prove the remaining ones.

To prove the first inequality, it suffices to show that the vector $\boldsymbol{S} \in (\mathbb{R}^{\geq 0}_\infty)^V$ defined by $\boldsymbol{S}_v = \sup_{\sigma\in \mathrm{HR}_\Box} \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc)$ is a fixed point of L. Since $\boldsymbol{K}$ is the least fixed point of L, the inequality then follows. So let $v \in V$ be arbitrary. We will show that $L(\boldsymbol{S})_v = \boldsymbol{S}_v$.

If $v \in V_\Box$, then we have to show that

$$L(\boldsymbol{S})_v = r(u) + \sup_{v\to v'}\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_{v'}^{\sigma,\pi}(Acc) = \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc) = \boldsymbol{S}_v.$$

Assume, for the sake of contradiction, that the equality does not hold, i.e. that either $L(\boldsymbol{S})_v < \boldsymbol{S}_v$ or $L(\boldsymbol{S})_v > \boldsymbol{S}_v$. If $L(\boldsymbol{S})_v > \boldsymbol{S}_v$, then there is a transition $v \to v'$ and a strategy $\sigma' \in \mathrm{HR}_\Box$ such that $r(u) + \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_{v'}^{\sigma',\pi}(Acc) > \sup_{\sigma\in \mathrm{HR}_\Box} \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc)$. If we denote by σ'' the strategy that moves from the initial vertex v to v' with probability 1 and then starts to behave exactly like the strategy σ', then we obtain

$$\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma'',\pi}(Acc) = r(u) + \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_{v'}^{\sigma',\pi}(Acc) > \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc) \geq \inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma'',\pi}(Acc),$$

a contradiction. So assume that $L(\boldsymbol{S})_v < \boldsymbol{S}_v$. Then there is some $\delta > 0$ and some function $f : \mathrm{HR}_\Box \times V \to \mathrm{HR}_\Diamond$ such that for every transition $v \to v'$ and every $\sigma \in \mathrm{HR}_\Box$ we have $r(u) + \mathbb{E}_{v'}^{\sigma, f(\sigma,v')} < \boldsymbol{S}_v \ominus \delta$. For any strategy σ we denote by $p_\sigma^{v'}$ the probability the strategy σ assigns to transition $v \to v'$ in a game starting in v. Then we can write

$$\begin{aligned}
\sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc) &= r(u) + \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \sum_{v\to v'} p_\sigma^{v'} \cdot \mathbb{E}_{v'}^{\sigma,\pi}(Acc) \\
&\leq r(u) + \sup_{\sigma\in \mathrm{HR}_\Box} \sum_{v\to v'} p_\sigma^{v'} \cdot \mathbb{E}_{v'}^{\sigma, f(\sigma,v')}(Acc) < \boldsymbol{S}_v \ominus \delta \\
&\leq \boldsymbol{S}_v = \sup_{\sigma\in \mathrm{HR}_\Box}\inf_{\pi\in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc),
\end{aligned}$$

again a contradiction.

For $v \in V_\Diamond$ the proof is dual to the proof for $v \in V_\Box$, so we omit it. Finally, for $v \in V_\bigcirc$ we have

$$L(\boldsymbol{S}_v) = r(u) + \sum_{v \to v'} Prob(v)(v,v') \cdot \left(\sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_{v'}^{\sigma,\pi}(Acc) \right)$$
$$= \sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \left(r(u) + \sum_{v \to v'} Prob(v)(v,v') \cdot \mathbb{E}_{v'}^{\sigma,\pi}(Acc) \right) = \sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc) = \boldsymbol{S}_v.$$

This concludes the proof that $\boldsymbol{S}$ is a fixed point of L and thus also the proof of the first inequality in (3).

It remains to prove the third inequality in (3). To this end we prove that for every $\varepsilon > 0$ there is a strategy $\pi_\varepsilon \in \mathrm{HD}_\Diamond$ such that for every $v \in V$ we have $\sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi_\varepsilon} \leq \boldsymbol{K}_v + \varepsilon$. Note that this will also prove the second part of the lemma.

If $\boldsymbol{K}_v = \infty$, then the desired inequality holds trivially for any strategy of player $\Diamond$ (and particularly for every $\pi \in \mathrm{HD}_\Diamond$). So assume that $\boldsymbol{K}_v$ is finite and fix arbitrary $\varepsilon > 0$. We define the strategy π_ε as follows: let wu be any finite path with $u \in V_\Diamond$. Since $\boldsymbol{K}$ is a fixed point of L, there must be a successor u' of u such that $r(u) + \boldsymbol{K}_{u'} \leq \boldsymbol{K}_u + \varepsilon/2^{|wu|+1}$. We set $\pi_\varepsilon(w)$ to be a Dirac distribution that selects the transition $u \to u'$ with probability 1.

We will now prove the following lemma, that not only shows that the strategy π_ε has the desired property, but it will also be useful later.

Lemma 12. *Let $\varepsilon \geq 0$ be arbitrary and let π_ε be any deterministic strategy of player $\Diamond$ that has the following property: for every finite path wu starting in v and ending in $u \in V_\Diamond$, the transition $u \to u'$ selected by $\pi_\varepsilon(wu)$ satisfies $r(u) + \boldsymbol{K}_{u'} \leq \boldsymbol{K}_u + \varepsilon/2^{|wu|+1}$. Then $\sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi_\varepsilon}(Acc) \leq \boldsymbol{K}_v + \varepsilon$.*

Proof. We will prove that for every v, every $n \in \mathbb{N}_0$ and every strategy σ of player $\Box$ we have $\mathbb{E}_v^{\sigma,\pi_\varepsilon}(\sum_{i=0}^n \omega(i)) \leq \boldsymbol{K}_v + \varepsilon$. By the monotone convergence theorem this means that $\mathbb{E}_v^{\sigma,\pi_\varepsilon}(Acc) \leq \boldsymbol{K}_v + \varepsilon$ for every σ, and thus also $\sup_{\sigma \in \mathrm{HR}_\Box} \mathbb{E}_v^{\sigma,\pi_\varepsilon}(Acc) \leq \boldsymbol{K}_v + \varepsilon$.

So let us fix arbitrary v, n and σ. Recall that $\mathbb{E}_v^{\sigma,\pi}[X|Y]$ denotes the conditional expectation of random variable X given the event Y. We show that for every $0 \leq k \leq n$ and every finite path $w = v_0, \ldots, v_k$ we have

$$\mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=k}^{n} r(\omega(i)) \mid Run(w)] \leq \boldsymbol{K}_{v_k} + \sum_{i=k}^{n} \varepsilon/2^{k+1}.$$

In particular, this means that $\mathbb{E}_v^{\sigma,\pi_\varepsilon}(\sum_{i=0}^n \omega(i)) = \mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=0}^n r(\omega(i)) \mid Run(v)] \leq \boldsymbol{K}_v + \varepsilon$.

We proceed by downward induction on k. If $n = k$, then we trivially have

$$\mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=k}^{n} r(\omega(i)) \mid Run(w)] = r(v_k) \leq L(\boldsymbol{K})_{v_k} = \boldsymbol{K}_{v_k},$$

where the inequality follows from the definition of L.

Now suppose that $k < n$. We distinguish two cases. If $v_k \in V_\diamond$, denote by u the successor of v_k chosen by π_ε. Then we have

$$\begin{aligned}\mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=k}^{n} r(\omega(i)) \mid \mathit{Run}(w)] &= r(v_k) + \mathbb{E}^{\sigma,\pi_\varepsilon}[\sum_{i=k+1}^{n} r(\omega(i)) \mid \mathit{Run}(wu)] \\ &\leq r(v_k) + \boldsymbol{K}_u + \sum_{i=k+1}^{n} \varepsilon/2^{i+1} \\ &\leq \boldsymbol{K}_{v_k} + \sum_{i=k}^{n} \varepsilon/2^{i+1},\end{aligned}$$

where the inequality on the second line follows from induction hypothesis and the inequality on the third line follows from the definition of π_ε.

If $v_k \in V_\Box \cup V_\bigcirc$, then we can see that $\mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=k}^{n} r(\omega(i)) \mid \mathit{Run}(w)] = \sum_{v_k \to u} p_u \cdot \mathbb{E}^{\sigma,\pi_\varepsilon}[\sum_{i=k+1}^{n} r(\omega(i)) \mid \mathit{Run}(wu)]$ for some sequence of real numbers $(p_u)_{v_k \to u}$ s.t. $p_u \geq 0$ for every u and $\sum_{v_k \to u} p_u = 1$. By induction hypothesis we have $\mathbb{E}^{\sigma,\pi_\varepsilon}[\sum_{i=k+1}^{n} r(\omega(i)) \mid \mathit{Run}(wu)] \leq K_u + \sum_{i=k+1}^{n} \varepsilon/2^{i+1}$ for every $v_k \to u$. Finally, from the definition of L we obtain $\boldsymbol{K}_{v_k} = L(\boldsymbol{K})_{v_k} \geq \sum_{v_k \to u} p_u \cdot \boldsymbol{K}_u$ (the inequality can be strict only if $v \in V_\Box$). Together, we have

$$\mathbb{E}_v^{\sigma,\pi_\varepsilon}[\sum_{i=k}^{n} r(\omega(i)) \mid \mathit{Run}(w)] \leq \boldsymbol{K}_{v_k} + \sum_{i=k+1}^{n} \varepsilon/2^{i+1} < \boldsymbol{K}_{v_k} + \sum_{i=k}^{n} \varepsilon/2^{i+1}.$$

□

This finishes the proof of Lemma 5.

B Proof of Lemma 6

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.*

Let $\varepsilon > 0$ be arbitrary. It suffices to fix an arbitrary initial vertex v, define choices of the strategy σ_ε only on the finite paths starting in v and verify, that the resulting strategy is ε-HR-optimal in v. By repeating this construction for every $v \in V$ we obtain a strategy that is ε-HR-optimal in every vertex.

For the sake of better readability, we first present the detailed construction of the deterministic ε-HR-optimal strategy σ_ε for games in which the HR-value is finite in every vertex. Almost identical construction can be used for games with arbitrary HR-values; there are some subtle technical differences that will be presented in the second part of the proof.

We already know that the least fixed point $\mathbf{K}$ of the operator L is equal to the vector of HR-values. Moreover, from the standard results of the fixed-point theory (see, e.g., Theorem 5.1 in [13]) we know that $\mathbf{K} = L^\alpha(\mathbf{0})$ for some ordinal number α (where $\mathbf{0}$ is the vector of zeros and where the transfinite iteration of L is defined in a standard way, i.e. we put $L^\beta(\mathbf{0}) = \sup_{\gamma<\beta} L^\gamma(\mathbf{0})$ for every limit ordinal β). The following lemma is instrumental in the construction of σ_ε.

Lemma 13. *Let $\varepsilon > 0$ be arbitrary. Denote by α the ordinal number α such that $L^\alpha(\mathbf{0})_v = \mathrm{Val}_{\mathrm{HR}}(v)$ and denote by Ord_α the set of all ordinal numbers lesser than or equal to α. Then there is a labeling function $d\colon Fpath(v) \to Ord_\alpha$ satisfying the following conditions:*

(a) $d(v) = \alpha$.
(b) For every $wu \in Fpath(v)$ it holds either $d(w) = 0$ or $d(wu) < d(w)$.
(c) For every $wu \in Fpath(v)$, we have

$$L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|+1}} \le \begin{cases} r(u) + L^{d(wuu')}(\mathbf{0})_{u'}, \text{ for some } u \to u' & \text{if } u \in V_\Box \\ r(u) + \inf_{u\to u'} L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\Diamond \\ r(u) + \sum_{u\to u'} Prob(u)(u,u') \cdot L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\bigcirc. \end{cases}$$

Proof. We define the labeling d inductively, proceeding from the shorter paths to the longer ones. Obviously we set $d(v) = \alpha$. Now suppose that $d(wu)$ has already been defined. We will define $d(wuu')$ for all successors u' of u simultaneously. First let us assume that $d(wu)$ is a successor ordinal of the form $\beta + 1$. Then it suffices to put $d(wuu') = \beta$ for all successors u' of u. From the definition of L we can easily see that for every $\delta > 0$ it then holds

$$L^{\beta+1}(\mathbf{0})_u - \delta \le \begin{cases} r(u) + L^{\beta}(\mathbf{0})_{u'}, \text{ for some } u \to u' & \text{if } u \in V_\Box \\ r(u) + \inf_{u\to u'} L^{\beta}(\mathbf{0})_{u'} & \text{if } u \in V_\Diamond \\ r(u) + \sum_{u\to u'} Prob(u)(u,u') \cdot L^{\beta}(\mathbf{0})_{u'} & \text{if } u \in V_\bigcirc, \end{cases}$$

so in particular the inequality in (c) holds for wu.

Now let us assume that $d(wu)$ is a limit ordinal. Then $L^{d(wu)}(\mathbf{0})_u = \sup_{\gamma<d(wu)} L^\gamma(\mathbf{0})_u$. This means that there is $\gamma < d(wu)$ such that $L^{d(wu)}(\mathbf{0})_u - \varepsilon/2^{|wu|+2} \le L^\gamma(\mathbf{0})_u$. Clearly, we can assume that $\gamma = \beta + 1$ fore some ordinal β. Now we again set $d(wuu') = \beta$ for all successors u' of u. Using the argument from the previous paragraph with $\delta = \varepsilon/2^{|wu|+2}$ we obtain

$$L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|+1}} \le L^\gamma(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|+2}} \le \begin{cases} r(u) + L^{\beta}(\mathbf{0})_{u'}, \text{ for some } u \to u' & \text{if } u \in V_\Box \\ r(u) + \inf_{u\to u'} L^{\beta}(\mathbf{0})_{u'} & \text{if } u \in V_\Diamond \\ r(u) + \sum_{u\to u'} Prob(u)(u,u') \cdot L^{\beta}(\mathbf{0})_{u'} & \text{if } u \in V_\bigcirc, \end{cases}$$

so (c) again holds for wu.

Finally, if $d(wu) = 0$, then we set $d(wuu') = 0$ for all successors u' of u. In this way, we eventually define $d(w)$ for every finite path starting in v. It is obvious that d satisfies (a)–(c). □

We use the labeling d provided by the previous lemma to define the ε-HR-optimal HD strategy σ_ε of player $\Box$. For a given finite path wu the strategy σ_ε selects a transition $u \to u'$ such that $L^{d(wu)}(\mathbf{0})_u - \varepsilon/2^{|wu|+1} \le r(u) + L^{d(wuu')}(\mathbf{0})_{u'}$. Such a transition always exists due to the previous lemma. We now prove that the strategy σ_ε is ε-HD-optimal in v. We will actually prove a more general statement, that we will reuse later.

Lemma 14. *For every run ω denote by $\tau(\omega)$ the least k such that $d(\omega(0),\dots,\omega(k)) = 0$ and denote by S^τ_k the random variable defined by $S^\tau_k(\omega) = \sum_{i=k}^{\tau(\omega)} r(\omega(i))$. Then the following holds for every $wu \in Fpath(v)$:*

$$\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wu|} \mid Run(wu)] \geq L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|}}. \tag{4}$$

In particular, we have

$$\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}(Acc) \geq \inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_0 \mid Run(v)] \geq L^\alpha(\mathbf{0})_v - \varepsilon = \mathrm{Val}_{\mathrm{HR}}(v) - \varepsilon.$$

Proof. We proceed by transfinite induction on $d(wu)$. If $d(wu) = 0$, then the inequality (4) clearly holds. Now suppose that $d(wu) > 0$ and that the inequality (4) holds for every $\beta < d(wu)$. We distinguish three cases depending on the type of u.

(1.) $u \in V_\Box$. Denote by u' the successor of u selected by $\sigma_\varepsilon(wu)$. Then we have

$$\begin{aligned}\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wu|} \mid Run(wu)] &= r(u) + \inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wuu'|} \mid Run(wuu')] \\ &\geq r(u) + L^{d(wuu')}(\mathbf{0})_{u'} - \frac{\varepsilon}{2^{|wu|+1}} \\ &\geq L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|}},\end{aligned}$$

where the second line follows from the induction hypothesis and from the fact that $d(wuu') < d(wu)$, and the third line follows from the definition of σ_ε.

(2.) $u \in V_\diamond$. Then we have

$$\begin{aligned}\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wu|} \mid Run(wu)] &= r(u) + \inf_{u\to u'} \inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wuu'|} \mid Run(wuu')] \\ &\geq r(u) + \inf_{u\to u'} L^{d(wuu')}(\mathbf{0})_{u'} - \frac{\varepsilon}{2^{|wu|+1}} \\ &\geq L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|}},\end{aligned}$$

where the first line is easy, the second line again follows from the induction hypothesis and the third line follows from Lemma 13.

(3.) $u \in V_\bigcirc$. We denote by $u \xrightarrow{x} u'$ the fact that $Prob(u)(u,u') = x$. We have

$$\begin{aligned}\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wu|} \mid Run(wu)] &= r(u) + \sum_{u\xrightarrow{x}u'} x \cdot \left(\inf_{\pi\in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^\tau_{|wuu'|} \mid Run(wuu')]\right) \\ &\geq r(u) + \left(\sum_{u\xrightarrow{x}u'} x \cdot L^{d(wuu')}(\mathbf{0})_{u'}\right) - \frac{\varepsilon}{2^{|wu|+1}} \\ &\geq L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|}},\end{aligned}$$

where again the second and the third line follows from induction hypothesis and Lemma 13, respectively.

□

It remains to show how to handle the case when there are vertices with infinite HR-values. The idea is the same, but the proof is more technical. We need to slightly generalize the previous two lemmas. The following lemma generalizes Lemma 13. We denote by last(w) the last vertex on a nonempty path w.

Lemma 15. *Under the assumptions of Lemma 13 there exists a labeling function $d\colon Fpath(v) \to Ord_\alpha$ satisfying the following conditions:*

(a) $d(v) = \alpha$.
(b) For every $wu \in Fpath(v)$ it holds either $d(w) = 0$ or $d(wu) < d(w)$.
(c) For every $wu \in Fpath(v)$, such that $L^{d(wu)}(\mathbf{0})_u < \infty$, we have

$$L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|+1}} \leq \begin{cases} r(u) + L^{d(wuu')}(\mathbf{0})_{u'}, \text{ for some } u \to u' & \text{if } u \in V_\Box \\ r(u) + \inf_{u\to u'} L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\Diamond \\ r(u) + \sum_{u\to u'} Prob(u)(u,u') \cdot L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\bigcirc, \end{cases}$$

and for every $wu \in Fpath(v)$, such that $L^{d(wu)}(\mathbf{0})_u = \infty$, we have

$$\frac{1}{\varepsilon} + \varepsilon \cdot (|wu| + 1) + F(w) \leq \begin{cases} r(u) + L^{d(wuu')}(\mathbf{0})_{u'}, \text{ for some } u \to u' & \text{if } u \in V_\Box \\ r(u) + \inf_{u\to u'} L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\Diamond \\ r(u) + \sum_{u\to u'} Prob(u)(u,u') \cdot L^{d(wuu')}(\mathbf{0})_{u'} & \text{if } u \in V_\bigcirc, \end{cases}$$

$$\text{where } F(w) = \begin{cases} L^{d(w)}(\mathbf{0})_{\mathrm{last(w)}} & \text{if } w \text{ is nonempty and } L^{d(w)}(\mathbf{0})_{\mathrm{last(w)}} < \infty \\ 0 & \text{otherwise}. \end{cases}$$

Proof. We again define the function d inductively, starting by putting $d(v) = \alpha$. Now let wu be an arbitrary finite path such that $L^{d(wu)}(\mathbf{0})_u = \infty$. If $d(wu) = \beta + 1$ for some ordinal β, then we can put $d(wuu') = \beta$ for all successors u' of u. From the definition of L it then easily follows that the inequality in (c) holds for wu. (For example, if $u \in V_\Box$, then we have $\infty = r(u) + \sup_{u\to u'} L^\beta(\mathbf{0})_{u'}$ and there is surely $u \to u'$ s.t. $r(u) + L^\beta(\mathbf{0})_{u'} \geq 1/\varepsilon + \varepsilon \cdot (|wu| + 1) + F(w)$. It is of course possible that $L^\beta(\mathbf{0})_{u'} = \infty$.)

If $d(wu)$ is an limit ordinal, then there is a successor ordinal $\beta + 1 < d(wu)$ s.t. $L^{\beta+1}(\mathbf{0})_u \geq 2/\varepsilon + \varepsilon \cdot (|wu| + 1) + F(w)$. We set $d(wuu') = \beta$ for all successors u' of u. If $L^{\beta+1}(\mathbf{0})_u = \infty$, then from the previous paragraph we get that (c) holds for wu. If $L^{\beta+1}(\mathbf{0})_u < \infty$, then the same argument as in the proof of Lemma 13 shows, that for every $\delta > 0$ the right-hand side of the inequality in (c) is δ-close to $L^{\beta+1}(\mathbf{0})_0$. If we set $\delta = 1/\varepsilon$, we get that (c) holds for wu.

For wu with $L^{d(wu)}(\mathbf{0})_u < \infty$ we can use the same construction as in the Lemma 13. □

For every wu let us set

$$A^{wu}_\varepsilon = \begin{cases} L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|+1}} & \text{if } L^{d(wu)}(\mathbf{0})_u < \infty \\ \frac{1}{\varepsilon} + \varepsilon \cdot (|wu| + 1) + F(w) & \text{otherwise,} \end{cases}$$

and

$$B^{wu}_\varepsilon = \begin{cases} L^{d(wu)}(\mathbf{0})_u - \frac{\varepsilon}{2^{|wu|}} & \text{if } L^{d(wu)}(\mathbf{0})_u < \infty \\ \frac{1}{\varepsilon} + \varepsilon \cdot |wu| + F(w) & \text{otherwise.} \end{cases}$$

Note that $A^{wu}_{\varepsilon} - \delta \geq B^{wu}_{\varepsilon}$ for every $0 \leq \delta \leq \varepsilon/2^{|wu|+1}$. We now define the ε-HR-optimal deterministic strategy σ_ε as follows: for a given $wu \in Fpath(v)$, the $\sigma(wu)$ selects a transition $u \to u'$ such that $A^{wu}_{\varepsilon} \leq r(u) + L^{d(wuu')}(\mathbf{0})_{u'}$. It remains to prove that σ_ε is ε-HR-optimal in v. We generalize Lemma 14 as follows:

Lemma 16. *The following holds for every $wu \in Fpath(v)$:*

$$\inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wu|} \mid Run(wu)] \geq B^{wu}_{\varepsilon}. \tag{5}$$

Proof. The proof again proceeds by transfinite induction on $d(wu)$. The base case is the same as in Lemma 14, because if $d(wu) = 0$, then $B^{wu}_{\varepsilon} = -\frac{\varepsilon}{2^{|wu|+1}}$. So assume that $d(wu) > 0$ and that (5) hols for all $\alpha < d(wu)$. If $L^{d(wu)}(\mathbf{0})_u < \infty$, then we can basically proceed in exactly the same way as in the Lemma 14. The only difference here is the case when $u \in V_\diamond$, $L^{d(wu)}(\mathbf{0})_u < \infty$ and $L^{d(wuu')}(\mathbf{0})_{u'} = \infty$ for some $u \to u'$. But in this case we have $\mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wuu'|} \mid Run(wuu')] \geq B^{wuu'}_{\varepsilon} > 1/\varepsilon + F(wu) = 1/\varepsilon + L^{d(wu)}(\mathbf{0})_u \geq 1/\varepsilon + \inf_{u\to u'} L^{d(wuu')}(\mathbf{0})_{u'}$, so the computation in part (2.) of the proof of Lemma 14 is still valid.

If $L^{d(wu)}(\mathbf{0})_u = \infty$, then we consider the following cases:

(1.) $u \in V_\Box$. Denote by u' the successor of u selected by $\sigma_\varepsilon(wu)$. Then

$$\begin{aligned}\inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wu|} \mid Run(wu)] &= r(u) + \inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wuu'|} \mid Run(wuu')] \\ &\geq r(u) + B^{wuu'}_{\varepsilon},\end{aligned}$$

where the second line comes from the induction hypothesis. There are two possibilities. Either

$$B^{wuu'}_{\varepsilon} = 1/\varepsilon + \varepsilon \cdot |wu| + \varepsilon + F(w) > 1/\varepsilon + \varepsilon \cdot |wu| + F(w) = B^{wu}_{\varepsilon}, \tag{6}$$

or

$$r(u) + B^{wuu'}_{\varepsilon} = r(u) + L^{d(wuu')}(\mathbf{0})_{u'} - \frac{\varepsilon}{2^{|wu|+1}} \geq A^{wu}_{\varepsilon} - \frac{\varepsilon}{2^{|wu|+1}} \geq B^{wu}_{\varepsilon}, \tag{7}$$

where the second inequality follows from Lemma 15 and from the definition of σ_ε. In both cases the equation (5) holds.

(2.) $u \in V_\diamond$. Then we have

$$\begin{aligned}\inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wu|} \mid Run(wu)] &= r(u) + \inf_{u\to u'} \inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wuu'|} \mid Run(wuu')] \\ &\geq \inf_{u\to u'} \left(r(u) + B^{wuu'}_{\varepsilon}\right).\end{aligned}$$

Exactly the same computation as in the case (1.) reveals that (6) or (7) holds for all $u \to u'$, and thus for all these transitions we have $r(u) + B^{wuu'}_{\varepsilon} \geq B^{wu}_{\varepsilon}$. Thus, $\inf_{u\to u'} \left(r(u) + B^{wuu'}_{\varepsilon}\right) \geq B^{wu}_{\varepsilon}$ and (5) holds for wu.

(3.) $u \in V_\bigcirc$. Then again from the induction hypothesis it follows that

$$\begin{aligned}\inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wu|} \mid Run(wu)] &= r(u) + \sum_{\substack{x \\ u\to u'}} x \cdot \left(\inf_{\pi \in \mathrm{HR}_\diamond} \mathbb{E}_v^{\sigma_\varepsilon,\pi}[S^{\tau}_{|wuu'|} \mid Run(wuu')]\right) \\ &\geq \sum_{\substack{x \\ u\to u'}} x \cdot \left(r(u) + B^{wuu'}_{\varepsilon}\right) \geq B^{wu}_{\varepsilon},\end{aligned}$$

where the last inequality can be justified in exactly the same way as in the previous two cases.

□

C Proof of Lemma 11

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

$$\sup_{\sigma \in \mathrm{HR}_\Box} \inf_{\pi \in \mathrm{HR}_\Diamond} \mathbb{E}_v^{\sigma,\pi}(Acc_n) \quad > \quad \mathrm{Val}_{\mathrm{HR}}(v) \ominus \frac{\varepsilon}{4} \tag{8}$$

Let $v \in V$ be arbitrary. Without loss of generality, we can assume that $v \in V_\bigcirc$ and that v has only one outgoing transition. If this is not the case, we can simply add a new stochastic vertex v' with a zero reward and a single new transition $v \to v'$. It is clear, that if the statement of the lemma holds for v' in this new game, then it holds for v in the original game.

Observe that if every vertex of player $\Diamond$ has only finitely many successors, then the operator L is Scott-continuous.

Lemma 17. *Let $D \subseteq (\mathbb{R}_\infty^{\geq 0})^V$ be an arbitrary directed set (i.e. such a set that each pair of elements in D has an upper bound in D.) Then $L(\sup_{\boldsymbol{d} \in D} \boldsymbol{d}) = \sup_{\boldsymbol{d} \in D} L(\boldsymbol{d})$.*

Proof. The inequality $\geq$ follows immediately from the monotonicity of L. So it suffices to prove that for every directed set D and every vertex v we have $L(\sup_{\boldsymbol{d} \in D} \boldsymbol{d})_v \leq \sup_{\boldsymbol{d} \in D} L(\boldsymbol{d})_v$. Note that $(\sup_{\boldsymbol{d} \in D} \boldsymbol{d})_v = \sup_{\boldsymbol{d} \in D} \boldsymbol{d}_v$. We consider three cases:

(1.) $v \in V_\Box$. Then we trivially have

$$L(\sup_{\boldsymbol{d} \in D} \boldsymbol{d})_v = \sup_{v \to v'} \sup_{\boldsymbol{d} \in D} \boldsymbol{d}_{v'} = \sup_{\boldsymbol{d} \in D} \sup_{v \to v'} \boldsymbol{d}_{v'} = \sup_{\boldsymbol{d} \in D} L(\boldsymbol{d})_v.$$

(2.) $v \in V_\Diamond$. Assume, for the sake of contradiction, that $\inf_{v \to v'} \sup_{\boldsymbol{d} \in D} \boldsymbol{d}_{v'} > \sup_{\boldsymbol{d} \in D} \inf_{v \to v'} \boldsymbol{d}_{v'}$. Then for each of the finitely many transitions $v \to v'$ there is a vector $\boldsymbol{d}(v') \in D$ such that $\boldsymbol{d}(v')_{v'} > \sup_{\boldsymbol{d} \in D} \inf_{v \to v'} \boldsymbol{d}_{v'}$. But since the set D is directed and there are only finitely many $v \to v'$, there is a vector $\boldsymbol{d}^* \in D$ such that $\boldsymbol{d}(v') \sqsubseteq \boldsymbol{d}^*$ for every successor v' of v. We thus have

$$\sup_{\boldsymbol{d} \in D} \inf_{v \to v'} \boldsymbol{d}_{v'} \geq \inf_{v \to v'} \boldsymbol{d}^*_{v'} \geq \inf_{v \to v'} \boldsymbol{d}(v')_{v'} > \inf_{v \to v'} \sup_{\boldsymbol{d} \in D} \inf_{v \to v'} \boldsymbol{d}_{v'} = \sup_{\boldsymbol{d} \in D} \inf_{v \to v'} \boldsymbol{d}_{v'},$$

a contradiction. (Above, the second inequality follows from the fact that $\boldsymbol{d}(v') \sqsubseteq \boldsymbol{d}^*$ for every v' and the first inequality and the last equality are trivial. The third inequality is strict because there are only finitely many successors of v.)

(3.) $v \in V_\bigcirc$. Then we again trivially have

$$L(\sup_{\boldsymbol{d} \in D} \boldsymbol{d})_v = \sum_{v \to v'} \mathit{Prob}(v)(v, v') \cdot \sup_{\boldsymbol{d} \in D} \boldsymbol{d}_{v'} = \sup_{\boldsymbol{d} \in D} \sum_{v \to v'} \mathit{Prob}(v)(v, v') \cdot \boldsymbol{d}_{v'} = \sup_{\boldsymbol{d} \in D} L(\boldsymbol{d})_v.$$

□

From the Kleene fixed-point theorem it follows that $L^{\omega}(\mathbf{0}) = \mathbf{K}$, i.e. that the ordinal number α from Lemmas 13 and 15 can be assumed to be equal to ω. Fix a labeling d of finite paths starting in v that satisfies the conditions (a)–(c) in Lemma 13 (or Lemma 15, if there are some vertices with infinite HR-value). Then v is labeled by ω and all other elements of $Fpath(v)$ are labeled with nonnegative integers. Recall that $\tau(\omega)$ denotes the least k such that $d(\omega(0), \ldots, \omega(k)) = 0$.

Now let u be the unique successor of v. We set $n = d(vu) + 1$. To see that this n satisfies (8), consider the deterministic $(\varepsilon/8)$-HR-optimal strategy $\sigma_{\varepsilon/8}$ constructed in the proof of Lemma 6. From Lemma 13 (or Lemma 15) it follows that

$$\inf_{\pi \in \mathrm{HR}_{\diamond}} \mathbb{E}_v^{\sigma_{\varepsilon/8},\pi}[\sum_{i=0}^{\tau(\omega)} r(\omega(i)) \mid Run(v)] \geq \mathrm{Val}_{\mathrm{HR}} \ominus \frac{\varepsilon}{8}.$$

But now we clearly have $\tau(\omega) \leq n = d(vu) + 1$ for all runs ω starting in v. Thus, we have

$$\inf_{\pi \in \mathrm{HR}_{\diamond}} \mathbb{E}_v^{\sigma_{\varepsilon/8},\pi}(Acc_n) \geq \inf_{\pi \in \mathrm{HR}_{\diamond}} \mathbb{E}_v^{\sigma_{\varepsilon/8},\pi}[\sum_{i=0}^{\tau(\omega)} r(\omega(i)) \mid Run(v)] \geq \mathrm{Val}_{\mathrm{HR}} \ominus \frac{\varepsilon}{8} > \mathrm{Val}_{\mathrm{HR}} \ominus \frac{\varepsilon}{4}.$$

This finishes the proof of Lemma 11.

D MD-optimal strategies for player $\diamond$

We prove Item 2 of Proposition 8, i.e. the fact that for every $\diamond$-finitely-branching game G there is $\pi \in \mathrm{MD}_{\diamond}$ such that π is HR-optimal in every vertex. We have already defined π 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. But the HR-optimality of this strategy immediately follows from Lemma 12 (note that this lemma works for $\varepsilon = 0$) and Lemma 5 (which says that the least fixed-point $\boldsymbol{K}$ of L is equal to the vector of HR-values).