

The Adversarial Stackelberg Value in Quantitative Games

Emmanuel Filiot

Université libre de Bruxelles (ULB), Belgium

Raffaella Gentilini

University of Perugia, Italy

Jean-François Raskin

Université libre de Bruxelles (ULB), Belgium

Abstract

In this paper, we study the notion of adversarial Stackelberg value for two-player non-zero sum games played on bi-weighted graphs with the mean-payoff and the discounted sum functions. The adversarial Stackelberg value of Player 0 is the largest value that Player 0 can obtain when announcing her strategy to Player 1 which in turn responds with any of his best response. For the mean-payoff function, we show that the adversarial Stackelberg value is not always achievable but ϵ -optimal strategies exist. We show how to compute this value and prove that the associated threshold problem is in NP. For the discounted sum payoff function, we draw a link with the target discounted sum problem which explains why the problem is difficult to solve for this payoff function. We also provide solutions to related gap problems.

2012 ACM Subject Classification Theory of computation \rightarrow Solution concepts in game theory; Theory of computation \rightarrow Logic and verification

Keywords and phrases Non-zero sum games, reactive synthesis, adversarial Stackelberg

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.127

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version A full version of the paper is available at <http://arxiv.org/abs/2004.12918>.

Funding This work is partially supported by the PDR project Subgame perfection in graph games (F.R.S.-FNRS), the MIS project F451019F (F.R.S.-FNRS), the ARC project Non-Zero Sum Game Graphs: Applications to Reactive Synthesis and Beyond (Fédération Wallonie-Bruxelles), the EOS project Verifying Learning Artificial Intelligence Systems (F.R.S.-FNRS & FWO), and the COST Action 16228 GAMENET (European Cooperation in Science and Technology). Emmanuel Filiot is associate researcher at F.R.S.-FNRS.

Acknowledgements We thank anonymous reviewers, Dr. Shibashis Guha and Ms. Mrudula Balachander for useful comments on a preliminary version of this paper.

1 Introduction

In this paper, we study two-player non-zero sum infinite duration quantitative games played on graph games. In non-zero sum games, the notion of worst-case value is not rich enough to reason about the (rational) behavior of players. More elaborate solution concepts have been proposed in game theory to reason about non-zero sum games: Nash equilibria, subgames perfect equilibria, admissibility, and Stackelberg equilibria are important examples of such solution concepts, see e.g. [18] and [19].

Let us first recall the abstract setting underlying the notion of Stackelberg equilibria and explain the variant that is the focus of this paper. Stackelberg games are strategic games played by two players. We note Σ_0 the set of strategies of Player 0, also called



© Emmanuel Filiot, Raffaella Gentilini, and Jean-François Raskin;
licensed under Creative Commons License CC-BY

47th International Colloquium on Automata, Languages, and Programming (ICALP 2020).

Editors: Artur Czumaj, Anuj Dawar, and Emanuela Merelli; Article No. 127; pp. 127:1–127:18

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



the *leader*, and Σ_1 the set of strategies of Player 1, also called the *follower*. Additionally, the game comes with two (usually \mathbb{R} -valued) payoff functions, Payoff_0 and Payoff_1 , that determine the payoff each player receives: if $\sigma_0 \in \Sigma_0$ and $\sigma_1 \in \Sigma_1$ are chosen then Player 0 receives the payoff $\text{Payoff}_0(\sigma_0, \sigma_1)$ while Player 1 receives the payoff $\text{Payoff}_1(\sigma_0, \sigma_1)$. Both players aim at maximizing their respective payoffs, and in a Stackelberg game, players play sequentially as follows. ① Player 0, the leader, announces her choice of strategy $\sigma_0 \in \Sigma_0$. ② Player 1, the follower, announces his choice of strategy $\sigma_1 \in \Sigma_1$. ③ Both players receive their respective payoffs: $\text{Payoff}_0(\sigma_0, \sigma_1)$ and $\text{Payoff}_1(\sigma_0, \sigma_1)$. Due to the sequential nature of the game, Player 1 knows the strategy σ_0 , and so to act rationally (s)he should choose a strategy σ_1 that maximizes the payoff $\text{Payoff}_1(\sigma_0, \sigma_1)$. If such a strategy σ_1 exists, it is called a *best-response*¹ to the strategy $\sigma_0 \in \Sigma_0$. In turn, if the leader assumes a rational response of the follower to her strategy, this should guide the leader when choosing $\sigma_0 \in \Sigma_0$. Indeed, the leader should choose a strategy $\sigma_0 \in \Sigma_0$ such that the value $\text{Payoff}_0(\sigma_0, \sigma_1)$ is as large as possible when σ_1 is a best-response of the follower.

Two different scenarios can be considered in this setting: either the best-response $\sigma_1 \in \Sigma_1$ is imposed by the leader (or equivalently chosen *cooperatively* by the two players), or the best-response is chosen *adversarially* by Player 1. In classical results from game theory and most of the close related works on games played on graphs [13, 15], with the exception of [17], only the cooperative scenario has been investigated. But, the adversarial case is interesting because it allows us to model the situation in which the leader chooses $\sigma_0 \in \Sigma_0$ only and must be prepared to face any rational response of Player 1, i.e. if Player 1 has several possible best responses then σ_0 should be designed to face all of them. In this paper, our main contribution is to investigate the second route. As already noted in [17], this route is particularly interesting for applications in *automatic synthesis*. Indeed, when designing a program, and this is especially true for reactive programs [20, 2], we aim for robust solutions that works for multiple rational usages, e.g. all the usages that respect some specification or that maximize some measure for the user.

To reflect the two scenarios above, there are two notions of *Stackelberg values*. First, the *cooperative Stackelberg value* is the largest value that Player 0 can secure by proposing a strategy σ_0 and a strategy σ_1 to the follower with the constraint that σ_1 is a best-response for the follower to σ_0 . Second, the *adversarial Stackelberg value* is the largest value that Player 0 can secure by proposing a strategy σ_0 and facing any best response σ_1 of the follower to the strategy σ_0 . In this paper, we mostly concentrate on the *adversarial* Stackelberg value, for infinite duration games played on bi-weighted game graphs for the mean-payoff function and the discounted sum function. The cooperative case has been studied in [13, 15] and we only provide some additional results when relevant for that case (see also related works below).

Main contributions. First, we consider the mean-payoff function. For this payoff function, best responses of Player 1 to a strategy $\sigma_0 \in \Sigma_0$ not always exist (Lemma 3). As a consequence, the *cooperative* (CSV) and *adversarial* (ASV) Stackelberg values are defined using ϵ -best responses. While strategies of Player 0 to achieve CSV always exist as shown in [13], we show that it is not the case for ASV (Theorem 4). The ASV can only be approached in general and memory may be necessary to play optimally or ϵ -optimally in adversarial Stackelberg games for the mean-payoff function (Theorem 4). We also provide results for related algorithmic

¹ As we will see later in the paper, sometimes, best-responses are not guaranteed to exist. In such cases, we need to resort to weaker notions such as ϵ -best-responses. We leave those technical details for later in the paper.

problems. We provide a notion of witness for proving that the ASV is (strictly) above some threshold (Theorem 5), and it is the basis for an NP algorithm to solve the threshold problem (Theorem 7). Finally, we show how the ASV can be computed effectively (Theorem 12).

Second, we consider the discounted sum function. In that case, best responses of Player 1 to strategies $\sigma_0 \in \Sigma_0$ of Player 0 always exist (Lemma 13). The CSV and ASV are directly based on best-responses in that case. Then we draw a link between the *target discounted sum problem* and the CSV threshold problem (Lemma 15). The target discounted sum problem has been studied recently in [1], left open there for the general case and shown to be related to other open problems in piecewise affine maps and the representation of numbers in nonintegral bases. As a consequence, we introduce a relaxation of the threshold problems for both CSV and ASV in the form of gap problems (or promised problems as defined in [12]). We provide algorithms to solve those gap problems (Theorem 17) both for CSV and ASV. Finally, we prove NP-hardness for the gap problems both for CSV and ASV (Theorem 18).

Closely related work. The notions of cooperative and adversarial synthesis have been introduced in [11, 17], and further studied in [7, 10]. Those two notions are closely related to our notion of cooperative and adversarial Stackelberg value respectively. The games that are considered in those papers are infinite duration games played on graphs but they consider Boolean ω -regular payoff functions or finite range ω -regular payoff functions. Neither the mean-payoff function nor the discounted sum payoff function are ω -regular, and thus they are not considered in [11, 17]. The ω -regularity of the payoff functions that they consider is central to their techniques: they show how to reduce their problems to problems on tree automata and strategy logic. Those reductions cannot be used for payoff functions that are *not* ω -regular functions and we need specific new techniques to solve our problems.

In [13, 15], the cooperative scenario for Stackelberg game is studied for mean-payoff and discounted sum respectively. Their results are sufficient to solve most of the relevant questions on the CSV but not for ASV. Indeed, the techniques that are used for CSV are closely related to the techniques that are used to reason on Nash equilibria and build on previous works [4] which in turn reduce to algorithmic solutions for zero-sum one dimensional mean-payoff (or discounted sum games). For the ASV in the context of the mean-payoff function, we have to use more elaborate multi-dim. mean-payoff games and a notion of Pareto curve adapted from [3]. Additionally, we provide new results on the CSV for the discounted sum function. First, our reduction that relates the target discounted sum problem to the CSV is new and gives additional explanations why the CSV is difficult to solve and not solved in the general case in [15]. Second, while we also leave the general problem open here, we show how to solve the gap problems related to both CSV and ASV. Finally, the authors of [14] study *incentive equilibria* for multi-player mean-payoff games. This work is an extension of their previous work [13] and again concentrates on CSV and does not consider ASV.

Structure of the paper. In Sect. 2, we introduce the necessary preliminaries for our definitions and developments. In Sect. 3, we consider the adversarial Stackelberg value for the mean-payoff function. In Sect. 4, we present our results for the discounted sum function.

2 Preliminaries and notations

Arenas. A (bi-weighted) *arena* $\mathcal{A} = (V, E, \langle V_0, V_1 \rangle, w_0, w_1)$ consists of a finite set V of vertices, a set $E \subseteq V \times V$ of edges such that for all $v \in V$ there exists $v' \in V$ such that $(v, v') \in E$, a partition $\langle V_0, V_1 \rangle$ of V , where V_0 (resp. V_1) is the set of vertices for Player 0 (resp. Player 1), and two edge weight functions $w_0 : E \mapsto \mathbb{Z}, w_1 : E \mapsto \mathbb{Z}$. In the sequel, we

denote the maximum absolute value of a weight in \mathcal{A} by W . As arenas are directed weighted graphs, we use, sometimes without recalling the details, the classical vocabulary for directed graphs. E.g., a set of vertices $S \subseteq V$ is a strongly connected component of the arena (SCC for short), if for all $s_1, s_2 \in S$, there exists a path from s_1 to s_2 and a path from s_2 to s_1 .

Plays and histories. A *play* in \mathcal{A} is an infinite sequence of vertices $\pi = \pi_0\pi_1 \cdots \in V^\omega$ such that for all $k \in \mathbb{N}$, $(\pi_k, \pi_{k+1}) \in E$. We denote by $\text{Plays}_{\mathcal{A}}$ the set of plays in \mathcal{A} , omitting the subscript \mathcal{A} when the underlying arena is clear from the context. Given $\pi = \pi_0\pi_1 \cdots \in \text{Plays}_{\mathcal{A}}$ and $k \in \mathbb{N}$, the prefix $\pi_0\pi_1 \cdots \pi_k$ of π (resp. suffix $\pi_k\pi_{k+1} \cdots$ of π) is denoted by $\pi_{\leq k}$ (resp. $\pi_{\geq k}$). An *history* in \mathcal{A} is a (non-empty) prefix of a play in \mathcal{A} . The length $|h|$ of an history $h = \pi_{\leq k}$ is the number $|h| = k$ of its edges. We denote by $\text{Hist}_{\mathcal{A}}$ the set of histories in \mathcal{A} , \mathcal{A} is omitted when clear from the context. Given $i \in \{0, 1\}$, the set $\text{Hist}_{\mathcal{A}}^i$ denotes the set of histories such that their last vertex belongs to V_i . We denote the last vertex of a history h by $\text{last}(h)$. We write $h \leq \pi$ whenever h is a prefix of π . A play π is called a *lasso* if it is obtained as the concatenation of a history h concatenated with the infinite repetition of another history l , i.e. $\pi = h \cdot l^\omega$ with $h, l \in \text{Hist}_{\mathcal{A}}$ (notice that l is not necessary a simple cycle). The *size* of a lasso $h \cdot l^\omega$ is defined as $|h \cdot l|$. Given a vertex $v \in V$ in the arena \mathcal{A} , we denote by $\text{Succ}(v) = \{v' \mid (v, v') \in E\}$ the set of successors of v and by Succ^* its transitive closure.

Games. A *game* $\mathcal{G} = (\mathcal{A}, \langle \text{Val}_0, \text{Val}_1 \rangle)$ consists of a bi-weighted arena \mathcal{A} , a value function $\text{Val}_0 : \text{Plays}_{\mathcal{A}} \mapsto \mathbb{R}$ for Player 0 and a value function $\text{Val}_1 : \text{Plays}_{\mathcal{A}} \mapsto \mathbb{R}$ for Player 1. In this paper, we consider the classical *mean-payoff* and *discounted-sum* value functions. Both are played in bi-weighted arenas.

In a *mean-payoff* game $\mathcal{G} = (\mathcal{A}, \langle \text{MP}_0, \text{MP}_1 \rangle)$ the payoff functions MP_0, MP_1 are defined as follows. Given a play $\pi \in \text{Plays}_{\mathcal{A}}$ and $i \in \{0, 1\}$, the payoff $\text{MP}_i(\pi)$ is given by $\text{MP}_i(\pi) = \liminf_{k \rightarrow \infty} \frac{1}{k} w_i(\pi_{\leq k})$, where the weight $w_i(h)$ of an history $h \in \text{Hist}$ is the sum of the weights assigned by w_i to its edges. In our definition of the mean-payoff, we have used \liminf , we will also need the \limsup case for technical reasons. Here is the formal definition together with its notation: $\overline{\text{MP}}_i(\pi) = \limsup_{k \rightarrow \infty} \frac{1}{k} w_i(\pi_{\leq k})$

For a given discount factor $0 < \lambda < 1$, a *discounted sum* game is a game $\mathcal{G} = (\mathcal{A}, \langle \text{DS}_0^\lambda, \text{DS}_1^\lambda \rangle)$ where the payoff functions $\text{DS}_0^\lambda, \text{DS}_1^\lambda$ are defined as follows. Given a play $\pi \in \text{Plays}_{\mathcal{A}}$ and $i \in \{0, 1\}$, the payoff $\text{DS}_i^\lambda(\pi)$ is defined as $\text{DS}_i^\lambda(\pi) = \sum_{k=0}^{\infty} \lambda^k w_i(\pi_k, \pi_{k+1})$.

Strategies and payoffs. A strategy for Player $i \in \{0, 1\}$ in a game $\mathcal{G} = (\mathcal{A}, \langle \text{Val}_0, \text{Val}_1 \rangle)$ is a function $\sigma : \text{Hist}_{\mathcal{A}}^i \mapsto V$ that maps histories ending with a vertex $v \in V_i$ to a successor of v . The set of all strategies of Player $i \in \{0, 1\}$ in the game \mathcal{G} is denoted $\Sigma_i(\mathcal{G})$, or Σ_i when \mathcal{G} is clear from the context.

A strategy has memory M if it can be realized as the output of a finite state machine with M states. A memoryless (or positional) strategy is a strategy with memory 1, that is, a function that only depends on the last element of the given partial play. We note Σ_i^{ML} the set of memoryless strategies of Player i , and Σ_i^{FM} its set of finite memory strategies. A *profile* is a pair of strategies $\bar{\sigma} = (\sigma_0, \sigma_1)$, where $\sigma_0 \in \Sigma_0(\mathcal{G})$ and $\sigma_1 \in \Sigma_1(\mathcal{G})$. As we consider games with perfect information and deterministic transitions, any profile $\bar{\sigma}$ yields, from any history h , a unique *play* or *outcome*, denoted $\text{Out}_h(\mathcal{G}, \bar{\sigma})$. Formally, $\text{Out}_h(\mathcal{G}, \bar{\sigma})$ is the play π such that $\pi_{\leq |h|-1} = h$ and $\forall k \geq |h| - 1$ it holds that $\pi_{k+1} = \sigma_i(\pi_{\leq k})$ if $\pi_k \in V_i$. The set of outcomes (resp. histories) compatible with a strategy $\sigma \in \Sigma_{i \in \{0,1\}}(\mathcal{G})$ after a history h is $\text{Out}_h(\mathcal{G}, \sigma) = \{\pi \mid \exists \sigma' \in \Sigma_{1-i}(\mathcal{G}) \text{ such that } \pi = \text{Out}_h(\mathcal{G}, (\sigma, \sigma'))\}$ (resp. $\text{Hist}_h(\sigma) = \{h' \in \text{Hist}(\mathcal{G}) \mid \exists \pi \in \text{Out}_h(\mathcal{G}, \sigma), n \in \mathbb{N} : h' = \pi_{\leq n}\}$).

Each outcome π in $\mathcal{G} = (\mathcal{A}, \langle \text{Val}_0, \text{Val}_1 \rangle)$ yields a payoff $\text{Val}(\pi) = (\text{Val}_0(\pi), \text{Val}_1(\pi))$, where $\text{Val}_0(\pi)$ is the payoff for Player 0 and $\text{Val}_1(\pi)$ is the payoff for Player 1. We denote by $\text{Val}(h, \bar{\sigma}) = \text{Val}(\text{Out}_h(\mathcal{G}, \bar{\sigma}))$ the payoff of a profile of strategies $\bar{\sigma}$ after a history h .

Usually, we consider game instances such that players start to play at a fixed vertex v_0 . Thus, we call an initialized game a pair (\mathcal{G}, v_0) , where \mathcal{G} is a game and $v_0 \in V$ is the initial vertex. When the initial vertex v_0 is clear from context, we speak directly of $\mathcal{G}, \text{Out}(\mathcal{G}, \bar{\sigma}), \text{Out}(\mathcal{G}, \sigma), \text{Val}(\bar{\sigma})$ instead of $(\mathcal{G}, v_0), \text{Out}_{v_0}(\mathcal{G}, \bar{\sigma}), \text{Out}_{v_0}(\mathcal{G}, \sigma), \text{Val}(v_0, \bar{\sigma})$. We sometimes simplify further the notation omitting also \mathcal{G} , when the latter is clear from the context.

Best-responses and adversarial value in zero-sum games. Let $\mathcal{G} = (\mathcal{A}, \langle \text{Val}_0, \text{Val}_1 \rangle)$ be a $(\text{Val}_0, \text{Val}_1)$ -game on the bi-weighted arena \mathcal{A} . Given a strategy σ_0 for Player 0, we define two sets of strategies for Player 1. His *best-responses* to σ_0 , noted $\text{BR}_1(\sigma_0)$, and defined as:

$$\{\sigma_1 \in \Sigma_1 \mid \forall v \in V \cdot \forall \sigma'_1 \in \Sigma_1 : \text{Val}_1(\text{Out}_v(\sigma_0, \sigma_1)) \geq \text{Val}_1(\text{Out}_v(\sigma_0, \sigma'_1))\}.$$

And his ϵ -*best-responses* to σ_0 , for $\epsilon > 0$, noted $\text{BR}_1^\epsilon(\sigma_0)$, and defined as:

$$\{\sigma_1 \in \Sigma_1 \mid \forall v \in V \cdot \forall \sigma'_1 \in \Sigma_1 : \text{Val}_1(\text{Out}_v(\sigma_0, \sigma_1)) \geq \text{Val}_1(\text{Out}_v(\sigma_0, \sigma'_1)) - \epsilon\}.$$

We also introduce notations for zero-sum games (that are needed as intermediary steps in our algorithms). The adversarial value that Player 1 can enforce in the game \mathcal{G} from vertex v as: $\text{WCV}_1(v) = \sup_{\sigma_1 \in \Sigma_1} \inf_{\sigma_0 \in \Sigma_0} \text{Val}_1(\text{Out}_v(\sigma_0, \sigma_1))$. Let \mathcal{A} be an arena, $v \in V$ one of its states, and $\mathcal{O} \subseteq \text{Plays}_{\mathcal{A}}$ be a set of plays (called objective), then we write $\mathcal{A}, v \models \ll i \gg \mathcal{O}$, if $\exists \sigma_i \in \Sigma_i \cdot \forall \sigma_{1-i} \in \Sigma_{1-i} : \text{Out}_v(\mathcal{A}, (\sigma, \sigma')) \in \mathcal{O}$, for $i \in \{0, 1\}$. Here the underlying interpretation is zero-sum: Player i wants to force an outcome in \mathcal{O} and Player $1 - i$ has the opposite goal. All the zero-sum games we consider in this paper are *determined* meaning that for all \mathcal{A} , for all objectives $\mathcal{O} \subseteq \text{Plays}_{\mathcal{A}}$ we have that: $\mathcal{A}, v \models \ll i \gg \mathcal{O}$ iff $\mathcal{A}, v \not\models \ll 1 - i \gg \text{Plays}_{\mathcal{A}} \setminus \mathcal{O}$.

Convex hull and F_{\min} . First, we need some additional notations and vocabulary related to linear algebra. Given a finite set of d -dim. vectors $X \subset \mathbb{R}^d$, we note the set of all their convex combinations as $\text{CH}(X) = \{v \mid v = \sum_{x \in X} \alpha_x \cdot x \wedge \forall x \in X : \alpha_x \in [0, 1] \wedge \sum_{x \in X} \alpha_x = 1\}$, this set is called the *convex hull* of X . We also need the following additional, and less standard notions, introduced in [5]. Given a finite set of d -dim. vectors $X \subset \mathbb{R}^d$, let $f_{\min}(X)$ be the vector $v = (v_1, v_2, \dots, v_d)$ where $v_i = \min \{c \mid \exists x \in X : x_i = c\}$, i.e. the vector v is the pointwise minimum of the vectors in X . Let $S \subseteq \mathbb{R}^d$, then $F_{\min}(S) = \{f_{\min}(P) \mid P \text{ is a finite subset of } S\}$. The following proposition expresses properties of the $F_{\min}(S)$ operator that are useful for us in the sequel. The interested reader will find more results about the F_{\min} operator in [5].

► **Proposition 1.** *For all sets $S \subseteq \mathbb{R}^d$, for all $x \in F_{\min}(S)$, there exists $y \in S$ such that $x \leq y$. If S is a closed bounded set then $F_{\min}(S)$ is also a closed bounded set.*

In the sequel, we also use formulas of the theory of the reals with addition and order, noted $\langle \mathbb{R}, +, \leq \rangle$, in order to define subsets of \mathbb{R}^n . This theory is decidable and admits effective quantifier elimination [8].

3 Adversarial Stackelberg value for mean-payoff games

Mean-payoffs induced by simple cycles. Given a play $\pi \in \text{Plays}_{\mathcal{A}}$, we note $\text{inf}(\pi)$ the set of vertices v that appear infinitely many times along π , i.e. $\text{inf}(\pi) = \{v \mid \forall i \in \mathbb{N} \cdot \exists j \geq i : v = \pi_j\}$. It is easy to see that $\text{inf}(\pi)$ is an SCC in the underlying graph of the arena \mathcal{A} . A *cycle* c is a sequence of edges that starts and stops in a given vertex v , it is *simple* if it does not contain any other repetition of vertices. Given an SCC S , we write $\mathbb{C}(S)$ for the set of simple cycles inside S . Given a simple cycle c , for $i \in \{0, 1\}$, let $\text{MP}_i(c) = \frac{w_i(c)}{|c|}$ be the mean of w_i weights along edges in the simple cycle c , and we call the pair $(\text{MP}_0(c), \text{MP}_1(c))$ the *mean-payoff coordinate* of the cycle c . We write $\text{CH}(\mathbb{C}(S))$ for the convex-hull of the set of mean-payoff coordinates of simple cycles of S . The following theorem relates the d -dim. mean-payoff values of infinite plays and the d -dim. mean-payoff of simple cycles in the arena.

► **Theorem 2** ([5]). *Let S be an SCC in the arena \mathcal{A} , the following two properties hold: (i) for all $\pi \in \text{Plays}_{\mathcal{A}}$, if $\text{inf}(\pi) \subseteq S$ then $(\text{MP}_0(\pi), \text{MP}_1(\pi)) \in F_{\min}(\text{CH}(\mathbb{C}(S)))$ (ii) for all $(x, y) \in F_{\min}(\text{CH}(\mathbb{C}(S)))$, there exists $\pi \in \text{Plays}_{\mathcal{A}}$ such that $\text{inf}(\pi) = S$ and $(\text{MP}_0(\pi), \text{MP}_1(\pi)) = (x, y)$. Furthermore, the set $F_{\min}(\text{CH}(\mathbb{C}(S)))$ is effectively expressible in $\langle \mathbb{R}, +, \leq \rangle$.*

In the sequel, we denote by $\Phi_S(x, y)$ the formula with two free variables in $\langle \mathbb{R}, +, \leq \rangle$ such that for all $(u, v) \in \mathbb{R}^2$, $(u, v) \in F_{\min}(\text{CH}(\mathbb{C}(S)))$ if and only if $\Phi_S(x, y)[x/u, y/v]$ is true.

On the existence of best-responses for MP. We start the study of mean-payoff games with some considerations about the existence of best-responses and ϵ -best-responses for Player 1 to strategies of Player 0.

► **Lemma 3.** *There is a mean-payoff game \mathcal{G} and a strategy $\sigma_0 \in \Sigma_0(\mathcal{G})$ such that $\text{BR}_1(\sigma_0) = \emptyset$. For all mean-payoff games \mathcal{G} and finite memory strategies $\sigma_0 \in \Sigma_0^{\text{FM}}(\mathcal{G})$, $\text{BR}_1(\sigma_0) \neq \emptyset$. For all mean-payoff games \mathcal{G} , for all strategies $\sigma_0 \in \Sigma_0(\mathcal{G})$, for all $\epsilon > 0$, $\text{BR}_1^\epsilon(\sigma_0) \neq \emptyset$.*

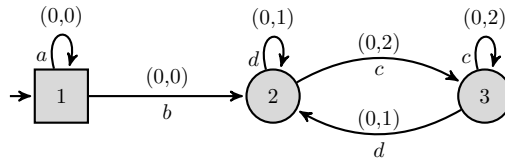
Proof sketch - full proof in the full version [9]. First, in the arena of Fig. 1, we consider the strategy of Player 0 that plays the actions c and d with a frequency that is equal to $1 - \frac{1}{k}$ for c and $\frac{1}{k}$ for d where k is the number of times that Player 1 has played a in state 1 before sending the game to state 2. We claim that there is no best response of Player 1 to this strategy of Player 0. Indeed, taking a one more time before going to state 2 is better for Player 1.

Second, if Player 0 plays a finite memory strategy, then a best response for Player 1 is an optimal path for the mean-payoff of Player 1 in the finite graph obtained as the product of the original game arena with the finite state strategy of Player 0. Optimal mean-payoff paths are guaranteed to exist [16].

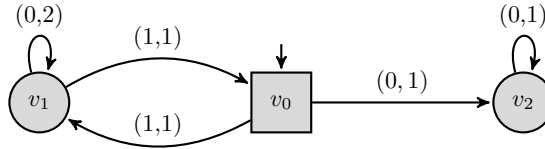
Finally, the existence of ϵ -best responses for $\epsilon > 0$, is guaranteed by an analysis of the infinite tree obtained as the unfolding of the game arena with the (potentially infinite memory) strategy of Player 1. Branches of this tree witness responses of Player 1 to the strategy of Player 0. The supremum of the values of those branches for Player 1 is always approachable to any $\epsilon > 0$. ◀

According to Lemma 3, the set of best-responses of Player 1 to a strategy of Player 0 can be empty. As a consequence, we need to use the notion of ϵ -best-responses (which are always guaranteed to exist) when we define the adversarial Stackelberg value:

$$\text{ASV}(\sigma_0)(v) = \sup_{\epsilon \geq 0 \mid \text{BR}_1^\epsilon(\sigma_0) \neq \emptyset} \inf_{\sigma_1 \in \text{BR}_1^\epsilon(\sigma_0)} \text{MP}_0(\text{Out}_v(\sigma_0, \sigma_1)) \text{ and } \text{ASV}(v) = \sup_{\sigma_0 \in \Sigma_0} \text{ASV}(\sigma_0)(v)$$



■ **Figure 1** A mean-payoff game in which there exists a Player 0's strategy σ_0 such that $BR_1(\sigma_0) = \emptyset$.



■ **Figure 2** In this game, $ASV(v_0) = 1$ but there is no Player 0 strategy to achieve this value.

We note that when best-responses to a strategy σ_0 exist, then as expected the following equality holds, because $BR_1(\sigma_0) = BR_1^0(\sigma_0)$ and $BR_1^{\epsilon_1}(\sigma_0) \subseteq BR_1^{\epsilon_2}(\sigma_0)$ for all $\epsilon_1 \leq \epsilon_2$, ϵ should be taken equal to 0:

$$ASV(\sigma_0)(v) = \sup_{\epsilon \geq 0 \mid BR_1^\epsilon(\sigma_0) \neq \emptyset} \inf_{\sigma_1 \in BR_1^\epsilon(\sigma_0)} \underline{MP}_0(\text{Out}_v(\sigma_0, \sigma_1)) = \inf_{\sigma_1 \in BR_1(\sigma_0)} \underline{MP}_0(\text{Out}_v(\sigma_0, \sigma_1))$$

Finally, we note that changing the sup over ϵ into an inf in our definition, we get the classical notion of worst-case value in which the rationality of Player 1 and his payoff are ignored. We also recall the definition of CSV, the cooperative Stackelberg value:

$$CSV(\sigma_0)(v) = \sup_{\epsilon \geq 0 \mid BR_1^\epsilon(\sigma_0) \neq \emptyset} \sup_{\sigma_1 \in BR_1^\epsilon(\sigma_0)} \underline{MP}_0(\text{Out}_v(\sigma_0, \sigma_1)) \text{ and } CSV(v) = \sup_{\sigma_0 \in \Sigma_0} CSV(\sigma_0)(v)$$

The interest reader is referred to [13] for an in-depth treatment of this value.

The adversarial Stackelberg value may not be achievable. In contrast with results in [13] that show that CSV can always be achieved, the following statement expresses the fact that the adversarial Stackelberg value may not be achievable but it can always be approximated by a strategy of Player 0.

► **Theorem 4.** *There exists a mean-payoff game \mathcal{G} in which Player 0 has no strategy which enforces the adversarial Stackelberg value. Furthermore, for all mean-payoff games \mathcal{G} , for all vertices $v \in V$, for all $\epsilon > 0$, there exists a strategy $\sigma_0 \in \Sigma_0$ such that $ASV(\sigma_0)(v) > ASV(v) - \epsilon$. Memory is needed to achieve high ASV.*

Proof sketch - full proof in the full version [9]. First, consider the game depicted in Fig 2. In this game, $ASV(v_0) = 1$ and it is not achievable. Player 0 needs to ensure that Player 1 does not take the transition from v_0 to v_2 otherwise she gets a payoff of 0. To ensure this, Player 0 needs to choose a strategy (that cycles within $\{v_0, v_1\}$) and that gives to Player 1 at least $1 + \epsilon$ with $\epsilon > 0$. Such strategies gives $1 - \epsilon$ to Player 0, and the value 1 cannot be reached.

Second, by definition of the ASV, the value is obtained as the sup over all strategies of Player 0. As a consequence, ϵ -optimal strategies (for $\epsilon > 0$) exist. ◀

Witnesses for the ASV. Given a mean-payoff game \mathcal{G} , we associate with each vertex v , the following set of pairs of real numbers: $\Lambda(v) = \{(c, d) \in \mathbb{R}^2 \mid v \models \ll 1 \gg \underline{\text{MP}}_0 \leq c \wedge \underline{\text{MP}}_1 \geq d\}$. We say that v is (c, d) -bad if $(c, d) \in \Lambda(v)$. Let $c' \in \mathbb{R}$. A play π in \mathcal{G} is called a (c', d) -witness of $\text{ASV}(v) > c$ if it starts from v , $(\underline{\text{MP}}_0(\pi), \underline{\text{MP}}_1(\pi)) = (c', d)$, $c' > c$ and π does not contain any (c, d) -bad vertex. A play π is called a witness of $\text{ASV}(v) > c$ if it is a (c', d) -witness of $\text{ASV}(v) > c$ for some c', d . The following theorem justifies the name witness.

► **Theorem 5.** *Let \mathcal{G} be a mean-payoff game and v be one of its vertices. $\text{ASV}(v) > c$ iff there exists a play π in \mathcal{G} such that π is a witness of $\text{ASV}(v) > c$.*

Proof. *From right to left.* Assume the existence of a (c', d) -witness π and let us show that there exists a strategy σ_0 which forces $\text{ASV}(\sigma_0)(v) > c$. We define σ_0 as follows:

1. for all histories $h \leq \pi$ such that *last* (h) belongs to Player 0, $\sigma_0(h)$ follows π .
2. for all histories $h \not\leq \pi$ where there has been a deviation from π by Player 1, we assume that Player 0 switches to a strategy that we call *punishing*. This strategy is defined as follows. In the subgame after history h' where h' is a first deviation by Player 1 from π , we know that Player 0 has a strategy to enforce the objective: $\underline{\text{MP}}_0 > c \vee \underline{\text{MP}}_1 < d$. This is true because π does not cross any (c, d) -bad vertex. So, we know that $h' \not\models \ll 1 \gg \underline{\text{MP}}_0 \leq c \wedge \underline{\text{MP}}_1 \geq d$ which entails the previous statement by determinacy of n -dimension mean-payoff games [21] (here $n = 2$).
3. for all other histories h , Player 0 can behave arbitrarily as those histories are never reached when Player 0 plays as defined in point 1 and 2 above.

Let us now establish that the strategy σ_0 satisfies $\text{ASV}(\sigma_0)(v) > c$. We have to show the existence of some $\epsilon \geq 0$ such that $\text{BR}_1^\epsilon(\sigma_0) \neq \emptyset$ and for all $\sigma_1 \in \text{BR}_1^\epsilon(\sigma_0)$, $\underline{\text{MP}}_0(\text{Out}_v(\sigma_0, \sigma_1)) > c$ holds. For that, we consider two subcases:

1. $\sup_{\sigma_1} \underline{\text{MP}}_1(\text{Out}_v(\sigma_0, \sigma_1)) = d = \underline{\text{MP}}_1(\pi)$. This means that any strategy σ_1 of Player 1 that follows π is for $\epsilon = 0$ a best-response to σ_0 . Now let us consider any strategy $\sigma_1 \in \text{BR}_1^0(\sigma_0)$. Clearly, $\pi' = \text{Out}_v(\sigma_0, \sigma_1)$ is such that $\underline{\text{MP}}_1(\pi') \geq d$. If $\pi' = \pi$, we have that $\underline{\text{MP}}_0(\pi') = c' > c$. If $\pi' \neq \pi$, then when π' deviates from π , we know that σ_0 behaves as the punishing strategy and so we have that $\underline{\text{MP}}_0(\pi') > c \vee \underline{\text{MP}}_1(\pi') < d$. But as $\sigma_1 \in \text{BR}_1^0(\sigma_0)$, we conclude that $\underline{\text{MP}}_1(\pi') \geq d$, and so in turn, we obtain that $\underline{\text{MP}}_0(\pi') > c$.
2. $\sup_{\sigma_1} \underline{\text{MP}}_1(\text{Out}_v(\sigma_0, \sigma_1)) = d' > d$. Let $\epsilon > 0$ be such that $d' - \epsilon > d$. By Lemma 3, $\text{BR}_1^\epsilon(\sigma_0) \neq \emptyset$. Let us now characterize the value that Player 0 receives against any strategy $\sigma_1 \in \text{BR}_1^\epsilon(\sigma_0)$. First, if σ_1 follows π then Player 0 receives $c' > c$. Second, if σ_1 deviates from π , Player 1 receives at least $d' - \epsilon > d$. But by definition of σ_0 , we know that if the play deviates from π then Player 0 applies her punishing strategy. Then we know that the outcome satisfies $\underline{\text{MP}}_0 > c \vee \underline{\text{MP}}_1 < d$. But as $d' - \epsilon > d$, we must conclude that the outcome π' is such that $\underline{\text{MP}}_0(\pi') > c$.

From left to right. Let σ_0 such that $\text{ASV}(\sigma_0)(v) > c$. Then by the equivalence shown in the proof of Theorem 4, we know that

$$\exists \epsilon \geq 0 : \text{BR}_1^\epsilon(\sigma_0) \neq \emptyset \wedge \forall \sigma_1 \in \text{BR}_1^\epsilon(\sigma_0) : \text{Out}_v(\sigma_0, \sigma_1) > c \quad (1)$$

Let ϵ^* be a value for ϵ that makes eq. (1) true. Take any $\sigma_1 \in \text{BR}_1^{\epsilon^*}(\sigma_0)$ and consider $\pi = \text{Out}_v(\sigma_0, \sigma_1)$. We will show that π is a witness for $\text{ASV}(v) > c$.

We have that $\underline{\text{MP}}_0(\pi) > c$. Let $d_1 = \underline{\text{MP}}_1(\pi)$ and consider any $\pi' \in \text{Out}_v(\sigma_0)$. Clearly if $\underline{\text{MP}}_1(\pi') \geq d_1$ then there exists $\sigma'_1 \in \text{BR}_1^{\epsilon^*}(\sigma_0)$ such that $\pi' = \text{Out}_v(\sigma_0, \sigma'_1)$ and we conclude that $\underline{\text{MP}}_0(\pi') > c$. So all deviations of Player 1 w.r.t. π against σ_0 are either giving him a $\underline{\text{MP}}_1$ which is less than d_1 or it gives to Player 0 a $\underline{\text{MP}}_0$ which is larger than c . So π is a $(\underline{\text{MP}}_0(\pi), \underline{\text{MP}}_1(\pi))$ -witness for $\text{ASV}(v) > c$ as we have shown that π never crosses an $(c, \underline{\text{MP}}_1(\pi))$ -bad vertex, and we are done. ◀

The following statement is a direct consequence of the proof of the previous theorem.

► **Corollary 6.** *If π is a witness for $ASV(v) > c$ then all π' such that: $\pi'(0) = v$, the set of vertices visited along π and π' are the same, and $\underline{MP}_0(\pi') \geq \underline{MP}_0(\pi)$ and $\underline{MP}_1(\pi') \geq \underline{MP}_1(\pi)$, are also witnesses for $ASV(v) > c$.*

Small witnesses and NP membership. Here, we refine Theorem 5 to establish membership of the threshold problem to NP.

► **Theorem 7.** *Given a mean-payoff game \mathcal{G} , a vertex v and a rational value $c \in \mathbb{Q}$, it can be decided in nondeterministic polynomial time if $ASV(v) > c$.*

Proof of Thm. 7 relies on the existence of small witnesses established in the following lemma:

► **Lemma 8.** *Given a mean-payoff game \mathcal{G} , a vertex v and $c \in \mathbb{Q}$, $ASV(v) > c$ if and only if there exists an SCC reachable from v that contains two simple cycles ℓ_1, ℓ_2 such that: (i) there exist $\alpha, \beta \in \mathbb{Q}$ such that $\alpha \cdot w_0(\ell_1) + \beta \cdot w_0(\ell_2) = c' > c$, and $\alpha \cdot w_1(\ell_1) + \beta \cdot w_1(\ell_2) = d$ (ii) there is no (c, d) -bad vertex v' along the path from v to ℓ_1 , the path from ℓ_1 to ℓ_2 , and the path from ℓ_2 to ℓ_1 .*

Proof sketch - full proof in the full version [9]. Theorem 5 establishes the existence of a witness π for $ASV(v) > c$. In turn, we show here that the existence of such a π can be established by a polynomially checkable witness composed of the following elements. First, a simple path from v to the SCC in which π gets trapped in the long run, (ii) two simple cycles (that can produce the value (c', d) of π) by looping at the right frequencies along the two cycles. Indeed, $(\underline{MP}_0(\pi), \underline{MP}_1(\pi))$ only depends on the suffix in the SCC in which it gets trapped. Furthermore, by Theorem 2, Proposition 1 and Corollary 6, we know that the mean-payoff of witnesses can be obtained as the convex combination of the mean-payoff coordinates of simple cycles, and 3 such simple cycles are sufficient by the Carathéodory baricenter theorem. A finer analysis of the geometry of the sets allows us to go to 2 cycles only (see the full proof in [9]). ◀

Proof of Theorem 7. According to Lemma 8, the nondeterministic algorithm that establishes the membership to NP guesses a reachable SCC together with the two simple cycles ℓ_1 and ℓ_2 , and parameters α and β . Additionally, for each vertex v' that appears along the paths to reach the SCC, on the simple cycles ℓ_1 and ℓ_2 , and to connect those simple cycles, the algorithm guesses a memoryless strategy $\sigma_0^{v'}$ for Player 0 that establishes $v' \not\models \ll 1 \gg \underline{MP}_0 \leq c \wedge \underline{MP}_1 \geq d$ which means by determinacy of multi-dimensional mean-payoff games, that $v' \models \ll 0 \gg \underline{MP}_0 > c \vee \underline{MP}_1 < d$. The existence of those memoryless strategy is established in Propositions 20 and 21 in the full version [9] (in turn those propositions rely on results from [21]). Those memoryless strategies are checkable in PTime [16]. ◀

Computing the ASV in mean-payoff games. The previous theorems establish the existence of a notion of witness for the adversarial Stackelberg value in non zero-sum two-player mean-payoff games. This notion of witness can be used to decide the threshold problem in NPTime. We now show how to use this notion to effectively compute the ASV. This algorithm is also based on the computation of an effective representation, for each vertex v of the game graph, of the infinite set of pairs $\Lambda(v)$. The following lemma expresses that a symbolic representation of this set of pairs can be constructed effectively. This result is using techniques that have been introduced in [3].

► **Lemma 9.** *Given a bi-weighted game graph \mathcal{G} and a vertex $v \in V$, we can effectively construct a formula $\Psi_v(x, y)$ of $\langle \mathbb{R}, +, \leq \rangle$ with two free variables such that $(c, d) \in \Lambda(v)$ if and only if the formula $\Psi_v(x, y)[x/c, y/d]$ is true.*

Extended graph game. From the graph game $\mathcal{G} = (V, E, w_0, w_1)$, we construct the extended graph game $\mathcal{G}^{\text{ext}} = (V^{\text{ext}}, E^{\text{ext}}, w_0^{\text{ext}}, w_1^{\text{ext}})$, whose vertices and edges are defined as follows. The set of vertices is $V^{\text{ext}} = V \times 2^V$. With an history h in \mathcal{G} , we associate a vertex in \mathcal{G}^{ext} which is a pair (v, P) , where $v = \text{last}(h)$ and P is the set of the vertices traversed along h . Accordingly the set of edges and the weight functions are defined as $E^{\text{ext}} = \{((v, P), (v', P')) \mid (v, v') \in E \wedge P' = P \cup \{v'\}\}$ and $w_i^{\text{ext}}((v, P), (v', P')) = w_i((v, v'))$, for $i \in \{0, 1\}$. Clearly, there exists a bijection between the plays π in \mathcal{G} and the plays π^{ext} in \mathcal{G}^{ext} which start in vertices of the form $(v, \{v\})$, i.e. π^{ext} is mapped to the play π in \mathcal{G} that is obtained by erasing the second dimension of its vertices.

► **Proposition 10.** *For all game graph \mathcal{G} , the following holds:*

1. *Let π^{ext} be an infinite play in the extended graph and π be its projection into the original graph \mathcal{G} (over the first component of each vertex), the following properties hold: (i) For all $i < j$: if $\pi^{\text{ext}}(i) = (v_i, P_i)$ and $\pi^{\text{ext}}(j) = (v_j, P_j)$ then $P_i \subseteq P_j$. (ii) $\underline{MP}_i(\pi^{\text{ext}}) = \underline{MP}_i(\pi)$, for $i \in \{0, 1\}$.*
2. *The unfolding of \mathcal{G} from v and the unfolding of \mathcal{G}^{ext} from $(v, \{v\})$ are isomorphic, and so $\text{ASV}(v) = \text{ASV}(v, \{v\})$.*

By the first point of the latter proposition and since the set of vertices of the graph is finite, the second component of any play π^{ext} stabilises into a set of vertices of \mathcal{G} which we denote by $V^*(\pi^{\text{ext}})$.

We now show how to characterize $\text{ASV}(v)$ with the notion of witness introduced above and the decomposition of \mathcal{G}^{ext} into SCC. This is formalized in the following lemma:

► **Lemma 11.** *For all mean-payoff games \mathcal{G} , for all vertices $v \in V$, let $\text{SCC}^{\text{ext}}(v)$ be the set of strongly-connected components in \mathcal{G}^{ext} which are reachable from $(v, \{v\})$, then we have*

$$\text{ASV}(v) = \max_{S \in \text{SCC}^{\text{ext}}(v)} \sup\{c \in \mathbb{R} \mid \exists \pi^{\text{ext}} : \pi^{\text{ext}} \text{ is a witness for } \text{ASV}(v, \{v\}) > c \text{ and } V^*(\pi^{\text{ext}}) = S\}$$

Proof. First, we note the following sequence of equalities:

$$\begin{aligned} & \text{ASV}(v) \\ &= \sup\{c \in \mathbb{R} \mid \text{ASV}(v) \geq c\} \\ &= \sup\{c \in \mathbb{R} \mid \text{ASV}(v) > c\} \\ &= \sup\{c \in \mathbb{R} \mid \exists \pi : \pi \text{ is a witness for } \text{ASV}(v) > c\} \\ &= \sup\{c \in \mathbb{R} \mid \exists \pi^{\text{ext}} : \pi^{\text{ext}} \text{ is a witness for } \text{ASV}(v, \{v\}) > c\} \\ &= \max_{S \in \text{SCC}^{\text{ext}}(v)} \sup\{c \in \mathbb{R} \mid \exists \pi^{\text{ext}} : \pi^{\text{ext}} \text{ is a witness for } \text{ASV}(v, \{v\}) > c \text{ and } V^*(\pi^{\text{ext}}) = S\} \end{aligned}$$

The first two equalities are direct consequences of the definition of the supremum and that $\text{ASV}(v) \in \mathbb{R}$. The third is a consequence of Theorem 5 that guarantees the existence of witnesses for strict inequalities. The fourth equality is a consequence of point 2 in Proposition 10. The last equality is the consequence of point 1 in Proposition 10. ◀

By definition of \mathcal{G}^{ext} , for all SCC S of \mathcal{G}^{ext} , there exists a set of vertices of \mathcal{G} which we also denote by $V^*(S)$ such that any vertex of S is of the form $(v, V^*(S))$. The set of bad thresholds for S is then defined as $\Lambda^{\text{ext}}(S) = \bigcup_{v \in V^*(S)} \Lambda(v)$. Applying Lemma 9, we can construct a formula $\Psi_S(x, y)$ which symbolic encodes the set $\Lambda^{\text{ext}}(S)$.

Now, we are equipped to prove that $\text{ASV}(v)$ is effectively computable. This is expressed by the following theorem and established in its proof.

► **Theorem 12.** *For all mean-payoff games \mathcal{G} , for all vertices $v \in V$, the value $ASV(v)$ can be effectively expressed by a formula ρ_v in $\langle \mathbb{R}, +, \leq \rangle$ and explicitly computed from this formula.*

Proof. To establish this theorem, we show how to build the formula $\rho_v(z)$ that is true iff $ASV(v) = z$. We use Lemma 11, to reduce this to the construction of a formula that expresses the existence of witnesses for $ASV(v)$ from $(v, \{v\})$:

$$ASV(v) = \max_{S \in \text{SCC}^{\text{ext}}(v)} \sup\{c \in \mathbb{R} \mid \exists \pi^{\text{ext}} : \pi^{\text{ext}} \text{ is a witness for } ASV(v, \{v\}) > c \text{ and } V^*(\pi^{\text{ext}}) = S\}$$

As $\max_{S \in \text{SCC}^{\text{ext}}(v)}$ is easily expressed in $\langle \mathbb{R}, +, \leq \rangle$, we concentrate on one SCC S reachable from $(v, \{v\})$ and we show how to express

$$\sup\{c \in \mathbb{R} \mid \exists \pi^{\text{ext}} : \pi^{\text{ext}} \text{ is a witness for } ASV(v, \{v\}) > c \text{ and } V^*(\pi^{\text{ext}}) = S\}$$

First, we define a formula that express the existence of a witness for $ASV(v) > c$. This is done by the following formula:

$$\rho_{v_0}^S(c) \equiv \exists x, y \cdot x > c \wedge \Phi_S(x, y) \wedge \neg \Psi_S(c, y)$$

Where $\Phi_S(x, y)$ is the symbolic encoding of $F_{\min}(\text{CH}(\mathbb{C}(S)))$ as defined in Theorem 2. This ensures that the values (x, y) are the mean-payoff values realizable by some path in S . By Lemma 9, $\neg \Psi_S(c, y)$ expresses that the path does not cross a (c, y) -bad vertex. So the conjunction $\exists x, y \cdot x > c \wedge \Phi_S(x, y) \wedge \neg \Psi_S(c, y)$ establishes the existence of a witness with mean-payoff values (x, y) for the threshold c . From this formula, we can compute the ASV by quantifier elimination in:

$$\exists z \cdot \forall e > 0 \cdot (\rho_{v_0}^S(z - e) \wedge (\forall y \cdot \rho_{v_0}^S(y) \implies y \leq z))$$

and obtain the unique value of z that makes the formula true. ◀

4 Stackelberg values for discounted-sum games

In this section, we study the notion of Stackelberg value in the case of discounted sum measures. Beside the adversarial setting considered so far, we also refer to a *cooperative* framework for discounted sum-games, since we add some results to [15], where the cooperative Stackelberg value for discounted-sum measures has been previously introduced and studied.

On the existence of best-responses for DS. First, we show that the set of best-responses for Player 1 to strategies of Player 0 is guaranteed to be nonempty for discounted sum games, while this was not the case in mean-payoff games.

► **Lemma 13.** *For all discounted sum games \mathcal{G} and strategies $\sigma_0 \in \Sigma_0(\mathcal{G})$, $BR_1(\sigma_0) \neq \emptyset$.*

Proof. Given $\sigma \in \Sigma_0(\mathcal{G})$, consider $S = \{\text{DS}_1(\text{Out}(\sigma, \tau)) \mid \tau \in \Sigma_1(\mathcal{G})\}$. S is a non empty limited subset of \mathbb{R} , since for each $\tau \in \Sigma_1(\mathcal{G})$ it holds $\text{DS}_1(\text{Out}(\sigma, \tau)) \leq \frac{W}{1 - \lambda}$, where W is the maximum absolute value of a weight in \mathcal{G} . Hence, S admits a unique superior extreme $s = \text{sup}(S)$. By definition of superior extreme, for each $\epsilon > 0$, there exists $v_\epsilon \in S$ such that $s \geq v_\epsilon > s - \epsilon$. Therefore, for each $\epsilon > 0$ there exists $\tau_\epsilon \in \Sigma_1(\mathcal{G})$ such that $s \geq \text{DS}_1(\text{Out}(\sigma, \tau_\epsilon)) > s - \epsilon$, i.e.:

$$0 \leq s - \text{DS}_1(\text{Out}(\sigma, \tau_\epsilon)) < \epsilon \tag{2}$$

127:12 The Adversarial Stackelberg Value in Quantitative Games

We show that this implies that $\text{Out}(\sigma)$ contains a play π^* such that $\text{DS}_1(\pi^*) = s$, which leads to $\text{BR}_1(\sigma) \neq \emptyset$, since Player 1 has a strategy to achieve $s = \sup(\{\text{DS}_1(\text{Out}(\sigma, \tau)) \mid \tau \in \Sigma_1(\mathcal{G})\})$. By contradiction, suppose that $\text{Out}(\sigma)$ does not contain any play π such that $\text{DS}_1(\pi) = s$. Hence, for each $\pi \in \text{Out}(\sigma)$, it holds that $\text{DS}_1(\pi) < s$ and π admits a prefix $\pi_{\leq k}$ such that:

$$\text{DS}_1(\pi_{\leq k}) + W \frac{\lambda^k}{1-\lambda} < s \quad (3)$$

Hence, we can cut each play in $\text{Out}(\sigma)$ as soon as Equation 3 is accomplished, leading to a finite tree T (by Konig lemma, since $\text{Out}(\sigma)$ is finitely branching). Let $\pi_T^* = \pi_0 \dots \pi_k$ be a branch in the finite tree T such that the value $v(\pi_T^*) = s - (\text{DS}_1(\pi_T^*) + W \frac{\lambda^k}{1-\lambda})$ is minimal. Note that, by Equation 3, $v(\pi_T^*) > 0$ since $v(\pi_T^*) = s - (\text{DS}_1(\pi_T^*) + W \frac{\lambda^k}{1-\lambda}) > s - s = 0$.

Then, for each play π , let $\pi_{\leq p}$ be the longest prefix of π which is also a branch in the finite tree T . By definition of π_T^* , we have:

$$s - \text{DS}_1(\pi) \geq s - (\text{DS}_1(\pi_{\leq p}) + W \frac{\lambda^p}{1-\lambda}) \geq v(\pi_T^*) > 0 \quad (4)$$

This leads to a contradiction to the fact that for all $\epsilon > 0$ there exists $\tau \in \Sigma_1(\mathcal{G})$ such that $s - \text{DS}_1(\text{Out}(\sigma, \tau_\epsilon)) < \epsilon$, established within Equation 2. \blacktriangleleft

Stackelberg values for DS in the adversarial and cooperative settings. The existence of best-responses allows us to simplify the notion of Stackelberg value for discounted sum measures, avoiding the parameter ϵ used for mean-payoff games. In particular, the adversarial Stackelberg value $\text{ASV}(v)$ for discounted sum games is defined for all $\sigma_0 \in \Sigma_0(\mathcal{G})$ as:

$$\text{ASV}(\sigma_0)(v) = \inf_{\sigma_1 \in \text{BR}_1(\sigma_0)} \text{DS}_0^\lambda(\text{Out}_v(\sigma_0, \sigma_1)) \text{ and } \text{ASV}(v) = \sup_{\sigma_0 \in \Sigma_0} \text{ASV}(\sigma_0)(v)$$

As previously announced, we also consider the notion of Stackelberg value for discounted sum measures in the cooperative setting, where Player 0 suggests a profile of strategies (σ_0, σ_1) and Player 1 agrees to play σ_1 if the latter strategy is a best response to σ_0 . Formally, the cooperative Stackelberg value $\text{CSV}(v)$ for discounted sum games is defined as:

$$\text{CSV}(\sigma_0)(v) = \sup_{\sigma_1 \in \text{BR}_1(\sigma_0)} \text{DS}_0^\lambda(\text{Out}_v(\sigma_0, \sigma_1)) \text{ and } \text{CSV}(v) = \sup_{\sigma_0 \in \Sigma_0} \text{CSV}(\sigma_0)(v)$$

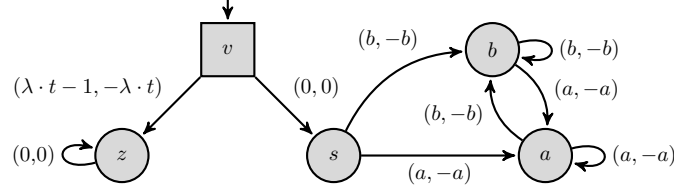
Lemma 15 below links the cooperative Stackelberg value for discounted-sum measures to the *target discounted-sum problem* [1] (cfr. Definition 14), whose decidability is notoriously hard to solve and relates to several open questions in mathematics and computer science [1].

► **Definition 14** (Target Discount Sum Problem [1] (TDS)). *Given a rational discount factor $0 < \lambda < 1$ and three rationals a, b, t does there exist an infinite sequence $w \in \{a, b\}^\omega$ such that $\sum_{i=0}^\infty w(i)\lambda^i = t$?*

In particular, given an instance $I = (a, b, t, \lambda)$ of the TDS problem, Figure 3 depicts a discounted sum game \mathcal{G}^I such that I admits a solution iff $\text{CSV}(v) \geq \lambda \cdot t$.

► **Lemma 15.** *The target discounted-sum problem reduces to the problem of deciding if $\text{CSV}(v) \geq c$ in discounted-sum games.*

Proof. Let $I = (a, b, t, \lambda)$ be an instance of the target discounted sum problem and consider the game \mathcal{G}^I depicted in Figure 3. We prove that I admits a solution iff $\text{CSV}(v) \geq \lambda \cdot t$.



■ **Figure 3** The instance of TDS $I = (a, b, \lambda, t)$ admits a solution iff $\text{CSV}(v) \geq \lambda \cdot t$.

Suppose that I admits a solution and let $w \in \{a, b\}^\omega$ such that $\sum_{i=0}^{\infty} w(i)\lambda^i = t$. Consider the following strategy σ for Player 0: for all $\alpha \in \{a, b\}^*$, $\sigma(vs\alpha) = x$ if $w(|\alpha|) = x$, where $x \in \{a, b\}$. We prove that if τ is a best response to σ , then $\text{DS}_0(\text{Out}(\sigma, \tau)) = \lambda \cdot t$. In fact, Player 1 has two choices from v . Let us denote τ_s (resp. τ_z) the strategy that prescribes to Player 1 to proceed to vertex s (resp. z) out from v . We have that $\text{DS}_1(\text{Out}(\sigma, \tau_s)) = \text{DS}_1(\text{Out}(\sigma, \tau_z)) = -\lambda \cdot t$, by definition of σ and \mathcal{G}^I . Hence, τ_s is a best response to σ which guarantees to Player 0 a payoff $\text{DS}_0(\text{Out}(\sigma, \tau_s)) = \lambda \cdot t$.

In the other direction, suppose that I does not admit any solution, i.e. there does not exist an infinite sequence $w \in \{a, b\}^\omega$ such that $\sum_{i=0}^{\infty} w(i)\lambda^i = t$. We prove that for any strategy σ for Player 0, if τ is a best response of Player 1 to σ then $\text{DS}_0(\text{Out}(\sigma, \tau)) < \lambda \cdot t$. Let σ be an arbitrary strategy for Player 0 and consider the strategy τ_z for Player 1.

We have two cases to consider depending on whether τ_z is a best response to σ or not. In the first case, we have that $\text{DS}(\text{Out}(\sigma, \tau_z)) = (\lambda \cdot t - 1, -\lambda \cdot t)$ and, since τ_z is a best response to σ , we need to have $\text{DS}_1(\text{Out}(\sigma, \tau_s)) \leq -\lambda \cdot t$. We can not have that $\text{DS}_1(\text{Out}(\sigma, \tau_s)) = -\lambda \cdot t$, since this would imply $\text{DS}_0(\text{Out}(\sigma, \tau_s)) = -\text{DS}_1(\text{Out}(\sigma, \tau_s)) = \lambda \cdot t$ contradicting our hypothesis that I does not admit any solution. Therefore, $\text{DS}_1(\text{Out}(\sigma, \tau_s)) < -\lambda \cdot t$, meaning that τ_s is not a best response to σ and $\text{CSV}(v) = \lambda \cdot t - 1 < \lambda \cdot t$.

In the second case, where τ_z is not a best response to σ , we have that $\text{DS}_1(\text{Out}(\sigma, \tau_s)) > -\lambda \cdot t$ which implies that $\text{CSV}(v) = \text{DS}_0(\text{Out}(\sigma, \tau_s)) = -\text{DS}_1(\text{Out}(\sigma, \tau_s)) < \lambda \cdot t$. ◀

The construction used to link the cooperative Stackelberg value to the target discounted sum problem can be properly modified² to prove that infinite memory may be necessary to allow Player 0 to achieve her CSV, recovering a result originally proved in [15]. In the same paper, the authors show that in 3-player discounted sum games the cooperative Stackelberg value cannot be approximated by considering strategies with bounded memory only. In the next section, we show that this is not the case for 2-player discounted sum games.

Gap problems and their algorithmic solutions. We consider a gap approximation of the Stackelberg value problem in both the cooperative and the adversarial settings. Given $\epsilon > 0$ and $c \in \mathbb{Q}$, and $\text{VAL} \in \{\text{CSV}, \text{ASV}\}$, let us define the sets of games:

- $\text{Yes}_{\text{VAL}}^{\epsilon, c} = \{(\mathcal{G}, v) \mid \mathcal{G} \text{ is a game with } \text{VAL}(v) > c + \epsilon\}$
- $\text{No}_{\text{VAL}}^{\epsilon, c} = \{(\mathcal{G}, v) \mid \mathcal{G} \text{ is a game with } \text{VAL}(v) < c - \epsilon\}$

The (ϵ, c) -CSV-gap (resp. (ϵ, c) -ASV-gap) problem (also referred to as (ϵ, c) -gap problems or just gap problems when ϵ and c are clear from the context) consists in determining if a given game \mathcal{G} and an initial vertex v belong to $\text{Yes}_{\text{CSV}}^{\epsilon, c}$ or $\text{No}_{\text{CSV}}^{\epsilon, c}$ (resp. $\text{Yes}_{\text{ASV}}^{\epsilon, c}$ or $\text{No}_{\text{ASV}}^{\epsilon, c}$). More

² Consider the game \mathcal{G}^I depicted in Figure 3 for $a = 0, b = 1, \lambda = \frac{2}{3}, t = \frac{3}{2}$. By Proposition 1 in [6], Player 0 can achieve $\frac{3}{2}$ from s – and therefore $\text{CSV}(v) = 1$ – only with infinite memory.

precisely, solving the Stackelberg value gap problem in e.g. the cooperative setting amounts to answer **Yes** if the instance of the game belongs to $\text{Yes}_{\text{CSV}}^{\epsilon, c}$, answer **No** if the instance belongs to $\text{No}_{\text{CSV}}^{\epsilon, c}$, never answer or answer arbitrarily otherwise.

Theorem 17 below uses the results in Lemma 16 to provide an algorithm that solves the Stackelberg value gap problem in the cooperative and adversarial settings, for games with discounted sum objectives. In particular, Lemma 16, shows that finite memory strategies are sufficient to witness Stackelberg values strictly greater than a threshold $c \in \mathbb{Q}$.

► **Lemma 16.** *Let \mathcal{G} be a discounted-sum game and consider $c \in \mathbb{Q}$ and $\epsilon > 0$. If Player 0 has a strategy σ_0 such that $\text{CSV}(\sigma_0)(v) > c + \epsilon$ (resp. $\text{ASV}(\sigma_0)(v) > c + \epsilon$), then Player 0 has a strategy σ_0^* with finite memory $M(\epsilon)$ such that $\text{CSV}(\sigma_0^*)(v) > c$ (resp. $\text{ASV}(\sigma_0^*)(v) > c$). Moreover, $M(\epsilon)$ is computable given ϵ .*

Proof. Let $\sigma_{\min}^{\text{DS}_1} \in \Sigma_0$ a memoryless strategy for Player 0 minimizing $\sup_{\tau \in \Sigma_1} \text{DS}_1(\text{Out}(\sigma, \tau))$. Let $\sigma_{\max}^{\text{DS}_1} \in \Sigma_0$ be a memoryless strategy for Player 0 that maximizes $\sup_{\tau \in \Sigma_1} \text{DS}_1(\text{Out}(\sigma, \tau))$. Such memoryless strategies exist since 2-player (single-valued) discounted-sum games are memoryless determined. In particular, $\sigma_{\min}^{\text{DS}_1} \in \Sigma_0$ can be obtained by using standard algorithms for two players (single-valued) discounted-sum games. In turn, $\sigma_{\max}^{\text{DS}_1} \in \Sigma_0$ can be computed by solving a single player (single valued) discounted-sum game, in which all the nodes are controlled by the maximizer who aims at maximizing DS_1 .

Cooperative Setting: Let $\sigma^* \in \Sigma_0(\mathcal{G})$ be a strategy for Player 0 s.t. $\text{DS}_0(\text{Out}(\sigma^*, \tau)) > c + \epsilon$ for some strategy $\tau \in \text{BR}_1(\sigma^*)$. Denote by π^* the play $\pi^* = \text{Out}(\sigma^*, \tau)$ and let N such that $\lambda^N \frac{W}{1-\lambda} < \frac{\epsilon}{2}$. Given the above premises, consider the finite memory strategy $\sigma' \in \Sigma_0$ for Player 0 that follows σ^* for the first N steps and then either apply the memoryless strategy $\sigma_{\min}^{\text{DS}_1} \in \Sigma_0$ or the memoryless strategy $\sigma_{\max}^{\text{DS}_1} \in \Sigma_0$, depending on the history h followed up to N . In particular, if $h = \pi_{\leq N}^*$, then the strategy σ' prescribes to Player 0 to follow $\sigma_{\max}^{\text{DS}_1} \in \Sigma_0$, cooperating with Player 1 at maximizing DS_1 . Otherwise ($h \neq \pi_{\leq N}^*$), the strategy σ' prescribes to Player 0 to follow $\sigma_{\min}^{\text{DS}_1} \in \Sigma_0$, minimizing the payoff of the adversary. We show that a best response τ' for Player 1 to σ' consists in following π^* up to N and then applying the memoryless strategy $\tau_{\max}^{\text{DS}_1} \in \Sigma_1$, i.e. maximizing $\sup_{\sigma \in \Sigma_0} \text{DS}_1(\text{Out}(\sigma, \tau))$. In fact, by definition of σ' and τ' we have that:

- $\text{DS}_1(\text{Out}(\sigma', \tau')) \geq \text{DS}_1(\pi^*)$
- for any other strategy $\tau'' \neq \tau'$ for Player 1:
 - if $\text{Out}(\sigma', \tau'')_{\leq N} = x \neq \pi_{\leq N}^*$, then:

$$\begin{aligned} \text{DS}_1(\text{Out}(\sigma', \tau'')) &= \text{DS}_1(x) + \lambda^N \text{DS}_1(\text{Out}_x(\sigma_{\min}^{\text{DS}_1}, \tau'')) \leq \\ &\leq \text{DS}_1(x) + \lambda^N (\sup_{\tau \in \Sigma_1} (\text{DS}_1(\text{Out}_x(\sigma_{\min}^{\text{DS}_1}, \tau)))) \leq \\ &\leq \text{DS}_1(x) + \lambda^N (\sup_{\tau \in \Sigma_1} (\text{DS}_1(\text{Out}_x(\sigma^*, \tau)))) = \text{DS}_1(\pi^*) \leq \text{DS}_1(\text{Out}(\sigma', \tau')) \end{aligned}$$

since $\text{DS}_1(\pi^*)$ is the payoff (for player 1) of a best response of Player 1 to σ^* .

- if $\text{Out}(\sigma', \tau'')_{\leq N} = x = \pi_{\leq N}^*$, then:

$$\begin{aligned} \text{DS}_1(\text{Out}(\sigma', \tau'')) &\leq \text{DS}_1(x) + \lambda^N \cdot \sup\{\text{DS}_1(\pi) \mid \pi \in \text{Plays}(\mathcal{G}) \wedge \pi \text{ starts at } \text{last}(x)\} = \\ &= \text{DS}_1(x) + \lambda^N \cdot \text{DS}_1(\text{Out}_x(\sigma_{\max}^{\text{DS}_1}, \tau_{\max}^{\text{DS}_1})) = \text{DS}_1(\text{Out}(\sigma', \tau')) \end{aligned}$$

Finally, we show that the best response π' of Player 1 to σ' guarantees to Player 0 a payoff greater than c . In fact, $\text{DS}_0(\text{Out}(\sigma', \tau')) > \text{DS}_0(\pi_{\leq N}^*) - \frac{\epsilon}{2} > c + \frac{\epsilon}{2} - \frac{\epsilon}{2} = c$, since $\text{DS}_0(\pi_{\leq N}^*) > c + \frac{\epsilon}{2}$. Due to the choice of N , having $\text{DS}_0(\pi_{\leq N}^*) \leq c + \frac{\epsilon}{2}$ would lead in fact to the following contradiction: $\text{DS}_0(\pi^*) \leq \text{DS}_0(\pi_{\leq N}^*) + \lambda^N \frac{W}{1-\lambda} < \text{DS}_0(\pi_{\leq N}^*) + \frac{\epsilon}{2} \leq c + \frac{\epsilon}{2} + \frac{\epsilon}{2} = c + \epsilon$, i.e. $\text{DS}_0(\pi^*) \leq c + \epsilon$.

Adversarial Setting: Let $\sigma \in \Sigma_0$ be a strategy for Player 0 such that for all $\tau \in BR_1(\sigma)$ it holds $DS_0(\text{Out}(\sigma, \tau)) > c + \epsilon$. Let N such that $\lambda^N \frac{W}{1-\lambda} < \frac{\epsilon}{2}$ and consider the unfolding T of $\text{Out}(\sigma)$ up to N . For each maximal root-to-leaf branch b of T , color its leaf $\text{last}(b)$ green if b is the prefix $\pi_{\leq N}$ of some play $\pi = \text{Out}(\sigma, \tau)$ such that $\tau \in BR_1(\sigma)$. Otherwise, let the leaf $\text{last}(b)$ of b be colored by red. We show that the finite memory strategy $\sigma^* \in \Sigma_0$ that prescribes to Player 0 to follow σ up to N and then:

- from each green node apply the memoryless strategy $\sigma_{max}^{DS_1} \in \Sigma_0$ (i.e. cooperate with Player 1 to maximize the payoff DS_1)
- from each red node apply the memoryless strategy $\sigma_{min}^{DS_1} \in \Sigma_0$ (i.e. minimize the payoff DS_1 of the adversary)

is such that $ASV(\sigma^*) > c$. Let $d = \sup\{DS_1(\text{Out}(\sigma, \tau)) \mid \tau \in \Sigma_1(\mathcal{G})\}$ and consider $\pi \in \text{Out}(\sigma^*)$.

First, we show that if π contains a green node then $DS_0(\pi_{\leq N}) > c$. In fact, $DS_0(\pi_{\leq N}) > DS_0(\pi_{\leq N}) - \lambda^N \frac{W}{1-\lambda} > c + \frac{\epsilon}{2} - \frac{\epsilon}{2} = c$, since $\lambda^N \frac{W}{1-\lambda} < \frac{\epsilon}{2}$ by definition of N and since $DS_0(\pi_{\leq N}) > c + \frac{\epsilon}{2}$ being $\text{last}(\pi_{\leq N})$ a green node (witnessing that $\pi_{\leq N}$ is the prefix of a play π' compatible with a best response of Player 1 to σ^* , for which $DS_0(\pi') > c + \epsilon$).

Moreover, there is a play $\pi \in \text{Out}(\sigma^*)$ containing a green node for which $DS_1(\pi) \geq d$. This is because of two reasons. First, a play in $\text{Out}(\sigma)$ compatible with a best response to σ by Player 1 is of the form $hv\pi'$, where hv is a maximal root-to-leaf branch b of T with $\text{last}(b) = v$ green (by definition of green nodes). Second, for each history hv such that hv is a maximal root-to-leaf branch b of T with $\text{last}(b) = v$ green, $\text{Out}(\sigma^*)$ contains a play $hv\bar{\pi}$, where $\bar{\pi}$ is a play starting in v maximizing DS_1 . Therefore $DS_1(hv\bar{\pi}) = DS_1(hv) + \lambda^N DS_1(\bar{\pi}) \geq DS_1(hv) + \lambda^N DS_1(\pi') = d$, where $hv\pi'$ is a play compatible with a best response of Player 1 to σ . To conclude our proof, we need just to show that each play $\pi \in \text{Out}(\sigma^*)$ containing a red node is such that $DS_1(\pi) < d$. In fact, being $\text{last}(\pi_{\leq N})$ red, the history $\pi_{\leq N}$ can not be a prefix of any play in $\text{Out}(\sigma)$ compatible with a best response of Player 1 to σ . In other words, by playing σ Player 0 allows the adversary to gain a payoff that is at most $r < d$ on each play $\pi = hv\pi'$ with v red. Therefore, switching her strategy from σ to σ^* (i.e. playing σ for the first N turns and then switching to the memoryless strategy $\sigma_{min}^{DS_1} \in \Sigma_0$) Player 0 is sure to let Player 1 gain a payoff that is at most $r' \leq r < d$ on each play $\pi = hv\pi'$ with v red.

As a conclusion, against σ^* Player 1 can achieve at least a value d . Hence, each best response to σ^* visits a green node (if it does not, then $DS_1 < d$ which is a contradiction). This guarantees that $DS_0 > c$. ◀

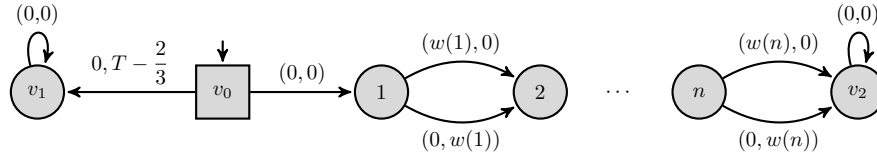
The approximation algorithm for solving the Stackelberg values gap problems introduced in Theorem 17 roughly works as follows. Given a discounted sum game \mathcal{G} , a rational threshold $c \in \mathbb{Q}$ and a tolerance rational value $\epsilon > 0$, the procedure checks whether there exists a strategy $\sigma_0 \in \Sigma_0(\mathcal{G})$ with finite memory $M(\epsilon)$ such that $ASV(\sigma_0) > c$ (resp. $CSV(\sigma_0) > c$). If such a strategy exists, the procedure answers **Yes**, otherwise it answers **No**. The correctness of the outlined procedure follows directly from Lemma 16.

► **Theorem 17.** *The (ϵ, c) -gap problems for both the CSV and ASV are decidable for games with discounted-sum objectives.*

We conclude this subsection by providing a reduction from the partition problem to our gap problems (for both CSV and ASV), showing NP-hardness for the corresponding problems.

► **Theorem 18.** *The (ϵ, c) -gap problems for both the CSV and ASV are NP-hard for games with discounted-sum objectives, where ϵ and c are given as input.*

127:16 The Adversarial Stackelberg Value in Quantitative Games



■ **Figure 4** Arena for hardness proof of the gap problem.

Proof. We do a reduction from the Partition problem to our gap problems, working for both CSV and ASV. Let us consider an instance of the partition problem defined by a set $A = \{1, 2, \dots, n\}$, a function $w : A \rightarrow \mathbb{N}_0$. The partition problem asks if there exists $B \subset A$ such that $\sum_{a \in B} w(a) = \sum_{a \in A \setminus B} w(a)$. W.l.o.g., we assume $\sum_{a \in A} w(a) = 2 \cdot T$ for some T .

To define our reduction, we first fix c to $T - \frac{1}{2}$, and the two parameters $\lambda \in (0, 1)$ and $\epsilon > 0$ by choosing values that respect the following two constraints:

$$T \cdot \lambda^{n+1} > T - \frac{1}{2} + \epsilon \quad (T - 1) \cdot \lambda^{n+1} < T - \frac{1}{2} - \epsilon \quad (5)$$

It is not difficult to see that such values always exist and they can be computed in polynomial time from the description of the partition problem. Then, we construct the bi-weighted arena \mathcal{A} depicted in Fig. 4. In this arena, Player 1 has only two choices in the starting state of the game v_0 . There, he can either send the game to the state v_1 , and get a payoff of $T - \frac{2}{3}$, or he can go to state 1.

From state 1, Player 0 can simulate a partition of the elements of A by choosing edges: left edges simulate the choice of putting the object corresponding to the state in the left class and right edges simulate the choice of putting the corresponding object in the right class. Let D_0 and D_1 be the discounted sum obtained by Player 0 and Player 1 when arriving in v_2 . Because λ and ϵ have been chosen according to eq. (5), we have that: $D_0 > T - \frac{1}{2} + \epsilon \wedge D_1 > T - \frac{1}{2} + \epsilon$ if and only if the choices of edges of Player 0 correspond to a valid partition of A .

Indeed, assume that $B \subseteq A$ is a solution to the partition problem. Assume that Player 0 follows the choices defined by B . Then when the game reaches state v_2 , the discounted sum of rewards for both players is larger than $T \cdot \lambda^{n+1}$. This is because along the way to v_2 , the discounted factor applied on the rewards obtained by both players has always been smaller than λ^{n+1} as they were equal to λ^{i+1} for all $i \leq n$. Additionally, we know that sum of (non-discounted) rewards for both players is equal to T as B is a correct partition. Now, it should be clear that both $\text{ASV}(v_0)$ and $\text{CSV}(v_0)$ are greater than $T - \frac{1}{2} + \epsilon$ as in the two cases, Player 1 has no incentive to deviate from the play that goes to v_1 as Player 1 would only get $T - \frac{2}{3}$ which is strictly smaller than D_1 .

Now, assume that there is no solution to the partition problem. In that case, Player 0 cannot avoid to give less than $T - 1$ to herself or to Player 1 when going from v_0 to v_2 . In the first case, its reward is less than $T - 1$ and in the second case, the reward of Player 1 is less than $T - 1$ and Player 1 has an incentive to deviate to state v_1 . In the two cases, we have that both $\text{ASV}(v_0)$ and $\text{CSV}(v_0)$ are less than $T - \frac{1}{2} - \epsilon$. So, we have established that the answer to the gap problem is yes if the partition instance is positive, and the answer is no if the partition instance is negative. ◀

References

- 1 Udi Boker, Thomas A. Henzinger, and Jan Otop. The target discounted-sum problem. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, Kyoto, Japan, July 6-10, 2015*, pages 750–761, 2015. URL: <https://ieeexplore.ieee.org/xpl/conhome/7174833/proceeding>.
- 2 Romain Brenguier, Lorenzo Clemente, Paul Hunter, Guillermo A. Pérez, Mickael Randour, Jean-François Raskin, Ocan Sankur, and Mathieu Sassolas. Non-zero sum games for reactive synthesis. In *Language and Automata Theory and Applications - 10th International Conference, LATA 2016, Prague, Czech Republic, March 14-18, 2016, Proceedings*, volume 9618 of *Lecture Notes in Computer Science*, pages 3–23. Springer, 2016. doi:10.1007/978-3-319-30000-9.
- 3 Romain Brenguier and Jean-François Raskin. Pareto curves of multidimensional mean-payoff games. In *Computer Aided Verification - 27th International Conference, CAV 2015, San Francisco, CA, USA, July 18-24, 2015, Proceedings, Part II*, volume 9207 of *Lecture Notes in Computer Science*, pages 251–267. Springer, 2015. doi:10.1007/978-3-319-21668-3.
- 4 Thomas Brihaye, Julie De Pril, and Sven Schewe. Multiplayer cost games with simple nash equilibria. In *Logical Foundations of Computer Science, International Symposium, LFCS 2013, San Diego, CA, USA, January 6-8, 2013. Proceedings*, volume 7734 of *Lecture Notes in Computer Science*, pages 59–73. Springer, 2013. doi:10.1007/978-3-642-35722-0.
- 5 Krishnendu Chatterjee, Laurent Doyen, Herbert Edelsbrunner, Thomas A. Henzinger, and Philippe Rannou. Mean-payoff automaton expressions. In *CONCUR 2010 - Concurrency Theory, 21th International Conference, CONCUR 2010, Paris, France, August 31-September 3, 2010. Proceedings*, volume 6269, pages 269–283. Springer, 2010. doi:10.1007/978-3-642-15375-4.
- 6 Krishnendu Chatterjee, Vojtěch Forejt, and Dominik Wojtczak. Multi-objective discounted reward verification in graphs and mdps. In Ken McMillan, Aart Middeldorp, and Andrei Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning*, pages 228–242. Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- 7 Rodica Condurache, Emmanuel Filiot, Raffaella Gentilini, and Jean-François Raskin. The complexity of rational synthesis. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPICs*, pages 121:1–121:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016. URL: <http://www.dagstuhl.de/dagpub/978-3-95977-013-2>.
- 8 Jeanne Ferrante and Charles Rackoff. A decision procedure for the first order theory of real addition with order. *SIAM J. Comput.*, 4(1):69–76, 1975. doi:10.1137/0204006.
- 9 Emmanuel Filiot, Raffaella Gentilini, and Jean-François Raskin. The adversarial stackelberg value in quantitative games, 2020. arXiv:2004.12918.
- 10 Emmanuel Filiot, Raffaella Gentilini, and Jean-François Raskin. Rational synthesis under imperfect information. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, pages 422–431. ACM, 2018. doi:10.1145/3209108.
- 11 Dana Fisman, Orna Kupferman, and Yoav Lustig. Rational synthesis. In *Tools and Algorithms for the Construction and Analysis of Systems, 16th International Conference, TACAS, volume 6015 of Lecture Notes in Computer Science*, pages 190–204. Springer, 2010. doi:10.1007/978-3-642-12002-2.
- 12 Oded Goldreich. On promise problems: A survey. In *Theoretical Computer Science, Essays in Memory of Shimon Even*, volume 3895 of *Lecture Notes in Computer Science*, pages 254–290. Springer, 2006. doi:10.1007/11685654.
- 13 Anshul Gupta and Sven Schewe. Quantitative verification in rational environments. In *21st International Symposium on Temporal Representation and Reasoning, TIME 2014, Verona, Italy, September 8-10, 2014*, pages 123–131. IEEE Computer Society, 2014. URL: <https://ieeexplore.ieee.org/xpl/conhome/6933234/proceeding>.

- 14 Anshul Gupta, Sven Schewe, Ashutosh Trivedi, Maram Sai Krishna Deepak, and Bharath Kumar Padarthy. Incentive stackelberg mean-payoff games. In *Software Engineering and Formal Methods - 14th International Conference, SEFM 2016, Held as Part of STAF 2016, Vienna, Austria, July 4-8, 2016, Proceedings*, volume 9763 of *Lecture Notes in Computer Science*, pages 304–320. Springer, 2016.
- 15 Anshul Gupta, Sven Schewe, and Dominik Wojtczak. Making the best of limited memory in multi-player discounted sum games. In *Proceedings Sixth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2015, Genoa, Italy, 21-22nd September 2015*, volume 193 of *EPTCS*, pages 16–30, 2015. doi:10.4204/EPTCS.193.
- 16 Richard M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Mathematics*, 23(3):309–311, 1978. doi:10.1016/0012-365X(78)90011-0.
- 17 Orna Kupferman, Giuseppe Perelli, and Moshe Y. Vardi. Synthesis with rational environments. *Ann. Math. Artif. Intell.*, 78(1):3–20, 2016. doi:10.1007/s10472-016-9508-8.
- 18 J. F. Nash. Equilibrium points in n -person games. In *PNAS*, volume 36, pages 48–49. National Academy of Sciences, 1950.
- 19 Martin J. Osborne. *An introduction to game theory*. Oxford Univ. Press, 2004.
- 20 Amir Pnueli and Roni Rosner. On the synthesis of a reactive module. In *Conference Record of the Sixteenth Annual ACM Symposium on Principles of Programming Languages, Austin, Texas, USA, January 11-13, 1989*, pages 179–190, 1989. URL: <http://dl.acm.org/citation.cfm?id=75277>.
- 21 Yaron Velner, Krishnendu Chatterjee, Laurent Doyen, Thomas A. Henzinger, Alexander Moshe Rabinovich, and Jean-François Raskin. The complexity of multi-mean-payoff and multi-energy games. *Inf. Comput.*, 241:177–196, 2015. doi:10.1016/j.ic.2015.03.001.