The Impatient May Use Limited Optimism to Minimize Regret

Michaël Cadilhac[1], Guillermo A. Pérez[2](✉), and Marie van den Bogaard[3]

[1] University of Oxford, Oxford, UK
michael@cadilhac.name
[2] University of Antwerp, Antwerp, Belgium
guillermoalberto.perez@uantwerpen.be
[3] Université libre de Bruxelles, Brussels, Belgium
marie.van.den.bogaard@ulb.ac.be

Abstract. Discounted-sum games provide a formal model for the study of reinforcement learning, where the agent is enticed to get rewards early since later rewards are discounted. When the agent interacts with the environment, she may realize that, with hindsight, she could have increased her reward by playing differently: this difference in outcomes constitutes her *regret value*. The agent may thus elect to follow a *regret-minimal* strategy. In this paper, it is shown that (1) there always exist regret-minimal strategies that are admissible—a strategy being inadmissible if there is another strategy that always performs better; (2) computing the minimum possible regret or checking that a strategy is regret-minimal can be done in $\mathsf{coNP}^{\mathsf{NP}}$, disregarding the computational cost of numerical analysis (otherwise, this bound becomes PSpace).

Keywords: Admissibility · Discounted-sum games · Regret minimization

1 Introduction

A pervasive model used to study the strategies of an agent in an unknown environment is *two-player infinite horizon games played on finite weighted graphs.* Therein, the set of vertices of a graph is split between two players, Adam and Eve, playing the roles of the environment and the agent, respectively. The play starts in a given vertex, and each player decides where to go next when the play reaches one of their vertices. Questions asked about these games are usually of the form: *Does there exist a strategy of Eve such that...?* For such a question to be well-formed, one should provide:

1. A valuation function: given an infinite play, what is Eve's reward?
2. Assumptions about the environment: is Adam trying to help or hinder Eve?

The valuation function can be Boolean, in which case one says that Eve *wins* or *loses* (one very classical example has Eve winning if the maximum value

M. Bojańczyk and A. Simpson (Eds.): FOSSACS 2019, LNCS 11425, pp. 133–149, 2019.
https://doi.org/10.1007/978-3-030-17127-8_8

appearing infinitely often along the edges is even). In this setting, it is often assumed that Adam is adversarial, and the question then becomes: *Can Eve always win?* (The names of the players stem from this view: *is there* a strategy of ∃ve that *always* beats ∀dam?) The literature on that subject spans more than 35 years, with newly found applications to this day (see [4] for comprehensive lecture notes, and [7] for an example of recent use in the analysis of attacks in cryptocurrencies).

The valuation function can also aggregate the numerical values along the edges into a reward value. We focus in this paper on *discounted sum*: if w is the weight of the edge taken at the n-th step, Eve's reward grows by $\lambda^n \cdot w$, where $\lambda \in (0, 1)$ is a prescribed discount factor. Discounting future rewards is a classical notion used in economics [18], Markov decision processes [9,16], systems theory [1], and is at the heart of Q-learning, a reinforcement learning technique widely used in machine learning [19]. In this setting, we consider three attitudes towards the environment:

- The adversarial environment hypothesis translates to Adam trying to minimize Eve's reward, and the question becomes: *Can Eve always achieve a reward of x?* This problem is in $\mathsf{NP} \cap \mathsf{coNP}$ [20] and showing a P upper-bound would constitute a major breakthrough (namely, it would imply the same for so-called parity games [15]). A strategy of Eve that maximizes her rewards against an adversarial environment is called *worst-case optimal.* Conversely, a strategy that maximizes her rewards assuming a *collaborative* environment is called *best-case optimal.*
- Assuming that the environment is adversarial is drastic, if not pessimistic. Eve could rather be interested in settling for a strategy σ which is not *consistently* bad: if another strategy σ' gives a better reward in one environment, there should be another environment for which σ is better than σ'. Such strategies, called *admissible* [5], can be seen as an *a priori* rational choice.
- Finally, Eve could put no assumption on the environment, but regret not having done so. Formally, the *regret value* of Eve's strategy is defined as the maximal difference, for all environments, between the best value Eve *could* have obtained and the value she actually obtained. Eve can thus be interested in following a strategy that achieves the minimal regret value, aptly called a *regret-minimal* strategy [10]. This constitutes an *a posteriori* rational choice [12]. Regret-minimal strategies were explored in several contexts, with applications including competitive online algorithm synthesis [3,11] and robot-motion planning [13,14].

In this paper, we single out a class of strategies for Eve that first follow a best-case optimal strategy, then switch to a worst-case optimal strategy after some precise time; we call these strategies *optipess.* Our main contributions are then:

1. Optipess strategies are not only regret-minimal (a fact established in [13]) but also admissible—note that there are regret-minimal strategies that are not admissible and *vice versa.* On the way, we show that for any strategy of

Eve there is an admissible strategy that performs at least as well; this is a peculiarity of discounted-sum games.

2. The regret value of a given time-switching strategy can be computed with an NP algorithm (disregarding the cost of numerical analysis). The main technical hurdle is showing that exponentially long paths can be represented succinctly, a result of independent interest.
3. The question *Can Eve's regret be bounded by x?* is decidable in $\mathsf{NP}^{\mathsf{coNP}}$ (again disregarding the cost of numerical analysis, PSpace otherwise), improving on the implicit NExp algorithm of [13]. The algorithm consists in guessing a time-switching strategy and computing its regret value; since optipess strategies are time-switching strategies that are regret-minimal, the algorithm will eventually find the minimal regret value of the input game.

Structure of the Paper. Notations and definitions are introduced in Sect. 2. The study of admissibility appears in Sect. 3, and is independent from the complexity analysis of regret. The main algorithm devised in this paper (point 2 above) is presented in Theorem 5, Sect. 6; it relies on technical lemmas that are the focus of Sects. 4 and 5. We encourage the reader to go through the statements of the lemma sections, then through the proof of Theorem 5, to get a good sense of the role each lemma plays.

In more details, in Sect. 4 we provide a crucial lemma that allows to represent long paths succinctly, and in Sect. 5, we argue that the important values of a game (regret, best-case, worst-case) have short witnesses. In Sect. 6, we use these lemmas to devise our algorithms.

2 Preliminaries

We assume familiarity with basic graph and complexity theory. Some more specific definitions and known results are recalled here.

Game, Play, History. A *(discounted-sum) game* $\mathcal{G}$ is a tuple $(V, v_0, V_\exists, E, w, \lambda)$ where V is a finite set of vertices, v_0 is the starting vertex, $V_\exists \subseteq V$ is the subset of vertices that belong to Eve, $E \subseteq V \times V$ is a set of directed edges, $w \colon E \to \mathbb{Z}$ is an (edge-)weight function, and $0 < \lambda < 1$ is a rational *discount factor*. The vertices in $V \setminus V_\exists$ are said to belong to Adam. Since we consider games played for an infinite number of turns, we will always assume that every vertex has at least one outgoing edge.

A *play* is an infinite path $v_1 v_2 \cdots \in V^\omega$ in the digraph (V, E). A *history* $h = v_1 \cdots v_n$ is a finite path. The *length of* h, written $|h|$, is the number of *edges* it contains: $|h| \overset{\text{def}}{=} n - 1$. The set **Hist** consists of all histories that start in v_0 and end in a vertex from $V_\exists$.

Strategies. A *strategy of Eve* in $\mathcal{G}$ is a function σ that maps histories ending in some vertex $v \in V_\exists$ to a neighbouring vertex v' (i.e., $(v, v') \in E$). The strategy

σ is *positional* if for all histories h, h' ending in the same vertex, $\sigma(h) = \sigma(h')$. *Strategies of Adam* are defined similarly.

A history $h = v_1 \cdots v_n$ is said to be *consistent with a strategy* σ of Eve if for all $i \geq 2$ such that $v_i \in V_\exists$, we have that $\sigma(v_1 \cdots v_{i-1}) = v_i$. Consistency with strategies of Adam is defined similarly. We write $\mathbf{Hist}(\sigma)$ for the set of histories in **Hist** that are consistent with σ. A play is consistent with a strategy (of either player) if all its prefixes are consistent with it.

Given a vertex v and both Adam and Eve's strategies, τ and σ respectively, there is a unique play starting in v that is consistent with both, called the *outcome* of τ and σ on v. This play is denoted $\mathbf{out}^v(\sigma, \tau)$.

For a strategy σ of Eve and a history $h \in \mathbf{Hist}(\sigma)$, we let σ_h be the strategy of Eve that assumes h has already been played. Formally, $\sigma_h(h') = \sigma(h \cdot h')$ for any history h' (we will use this notation only on histories h' that start with the ending vertex of h).

Values. The *value of a history* $h = v_1 \cdots v_n$ is the discounted sum of the weights on the edges:

$$\mathbf{Val}(h) \stackrel{\text{def}}{=} \sum_{i=0}^{|h|-1} \lambda^i w(v_i, v_{i+1}) \ .$$

The *value of a play* is simply the limit of the values of its prefixes.

The *antagonistic value* of a strategy σ of Eve with history $h = v_1 \cdots v_n$ is the value Eve achieves when Adam tries to hinder her, after h:

$$\mathbf{aVal}^h(\sigma) \stackrel{\text{def}}{=} \mathbf{Val}(h) + \lambda^{|h|} \cdot \inf_\tau \mathbf{Val}(\mathbf{out}^{v_n}(\sigma_h, \tau)) \ ,$$

where τ ranges over all strategies of Adam. The *collaborative value* $\mathbf{cVal}^h(\sigma)$ is defined in a similar way, by substituting "sup" for "inf." We write $\mathbf{aVal}^h$ (resp. $\mathbf{cVal}^h$) for the best antagonistic (resp. collaborative) value achievable by Eve with any strategy.

Types of Strategies. A strategy σ of Eve is *strongly worst-case optimal* (SWO) if for every history h we have $\mathbf{aVal}^h(\sigma) = \mathbf{aVal}^h$; it is *strongly best-case optimal* (SBO) if for every history h we have $\mathbf{cVal}^h(\sigma) = \mathbf{cVal}^h$.

We single out a class of SWO strategies that perform well if Adam turns out to be helping. A SWO strategy σ of Eve is *strongly best worst-case optimal* (SBWO) if for every history h we have $\mathbf{cVal}^h(\sigma) = \mathbf{acVal}^h$, where:

$$\mathbf{acVal}^h \stackrel{\text{def}}{=} \sup\{\mathbf{cVal}^h(\sigma') \mid \sigma' \text{ is a SWO strategy of Eve}\} \ .$$

In the context of discounted-sum games, strategies that are positional and strongly optimal always exist. Furthermore, the set of all such strategies can be characterized by local conditions.

Lemma 1 (Follows from [20, Theorem 5.1]). *There exist positional SWO, SBO, and SBWO strategies in every game. For any positional strategy σ of Eve:*

- $(\forall v \in V)\,[\mathbf{aVal}^v(\sigma) = \mathbf{aVal}^v]$ *iff σ is SWO;*
- $(\forall v \in V)\,[\mathbf{cVal}^v(\sigma) = \mathbf{cVal}^v]$ *iff σ is SBO;*
- $(\forall v \in V)\,[\mathbf{aVal}^v(\sigma) = \mathbf{aVal}^v \wedge \mathbf{cVal}^v(\sigma) = \mathbf{acVal}^v]$ *iff σ is SBWO.*

Regret. The *regret* of a strategy σ of Eve is the maximal difference between the value obtained by using σ and the value obtained by using an alternative strategy:

$$\mathbf{Reg}\,(\sigma) \overset{\text{def}}{=} \sup_{\tau} \left(\left(\sup_{\sigma'} \mathbf{Val}(\mathbf{out}^{v_0}(\sigma', \tau)) \right) - \mathbf{Val}(\mathbf{out}^{v_0}(\sigma, \tau)) \right) ,$$

where τ and σ' range over all strategies of Adam and Eve, respectively. The *(minimal) regret of $\mathcal{G}$* is then $\mathbf{Reg} \overset{\text{def}}{=} \inf_\sigma \mathbf{Reg}\,(\sigma)$.

Regret can also be characterized by considering the point in history when Eve should have done things differently. Formally, for any vertices u and v let $\mathbf{cVal}^u_{\neg v}$ be the maximal $\mathbf{cVal}^u(\sigma)$ for strategies σ verifying $\sigma(u) \neq v$. Then:

Lemma 2 ([13, Lemma 13]). *For all strategies σ of Eve:*

$$\mathbf{Reg}\,(\sigma) = \sup \left\{ \lambda^n \left(\mathbf{cVal}^{v_n}_{\neg\sigma(h)} - \mathbf{aVal}^{v_n}(\sigma_h) \right) \,\middle|\, h = v_0 \cdots v_n \in \mathbf{Hist}(\sigma) \right\} .$$

Switching and Optipess Strategies. Given strategies σ_1, σ_2 of Eve and a *threshold function* $t \colon V_\exists \to \mathbb{N} \cup \{\infty\}$, we define the *switching strategy* $\sigma_1 \xrightarrow{t} \sigma_2$ for any history $h = v_1 \cdots v_n$ ending in $V_\exists$ as:

$$\sigma_1 \xrightarrow{t} \sigma_2(h) = \begin{cases} \sigma_2(h) & \text{if } (\exists i)[i \geq t(v_i)], \\ \sigma_1(h) & \text{otherwise.} \end{cases}$$

We refer to histories for which the first condition above holds as *switched histories*, to all others as *unswitched histories*. The strategy $\sigma = \sigma_1 \xrightarrow{t} \sigma_2$ is said to be *bipositional* if both σ_1 and σ_2 are positional. Note that in that case, for all histories h, if h is switched then $\sigma_h = \sigma_2$, and otherwise σ_h is the same as σ but with $t(v)$ changed to $\max\{0, t(v) - |h|\}$ for all $v \in V_\exists$. In particular, if $|h|$ is greater than $\max\{t(v) < \infty\}$, then σ_h is nearly positional: it switches to σ_2 as soon as it sees a vertex with $t(v) \neq \infty$.

A strategy σ is *perfectly optimistic-then-pessimistic* (optipess, for short) if there are positional SBO and SBWO strategies σ^{sbo} and σ^{sbwo} such that $\sigma = \sigma^{\text{sbo}} \xrightarrow{t} \sigma^{\text{sbwo}}$ where $t(v) = \inf \left\{ i \in \mathbb{N} \,\middle|\, \lambda^i \left(\mathbf{cVal}^v - \mathbf{aVal}^v \right) \leq \mathbf{Reg} \right\}$.

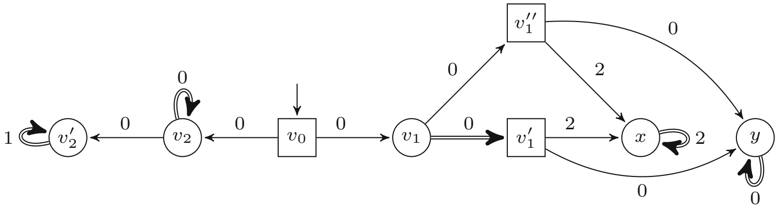
Theorem 1 **([13]).** *For all optipess strategies σ of Eve,* $\mathbf{Reg}\,(\sigma) = \mathbf{Reg}$.

Conventions. As we have done so far, we will assume throughout the paper that a game $\mathcal{G}$ is fixed—with the notable exception of the results on complexity, in which we assume that the game is given with all numbers in binary. Regarding strategies, we assume that bipositional strategies are given as two positional strategies and a threshold function encoded as a table with binary-encoded entries.

⋆

⋆ ⋆

Example 1. Consider the following game, where round vertices are owned by Eve, and square ones by Adam. The double edges represent Eve's positional strategy σ:



Eve's strategy has a regret value of $2\lambda^2/(1-\lambda)$. This is realized when Adam plays from v_0 to v_1, from v_1'' to x, and from v_1' to y. Against that strategy, Eve ensures a discounted-sum value of 0 by playing according to σ while regretting not having played to v_1'' to obtain $2\lambda^2/(1-\lambda)$. ■

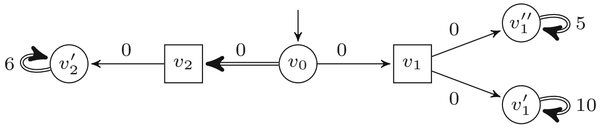
3 Admissible Strategies and Regret

There is no reason for Eve to choose a strategy that is consistently worse than another one. This classical idea is formalized using the notions of *strategy domination* and *admissible strategies.* In this section, which is independent from the rest of the paper, we study the relation between admissible and regret-minimal strategies. Let us start by formally introducing the relevant notions:

Definition 1. *Let σ_1, σ_2 be two strategies of Eve. We say that σ_1 is* weakly dominated *by σ_2 if* $\mathbf{Val}(\mathbf{out}^{v_0}(\sigma_1, \tau)) \le \mathbf{Val}(\mathbf{out}^{v_0}(\sigma_2, \tau))$ *for every strategy τ of Adam. We say that σ_1 is* dominated *by σ_2 if σ_1 is* weakly dominated *by σ_2 but not conversely. A strategy σ of Eve is* admissible *if it is not dominated by any other strategy.*

In other words, admissible strategies are maximal elements for the weak-domination pre-order.

Example 2. Consider the following game, where the strategy σ of Eve is shown by the double edges:



This strategy guarantees a discounted-sum value of $6\lambda^2(1-\lambda)$ against any strategy of Adam. Furthermore, it is worst-case optimal since playing to v_1 instead of v_2 would allow Adam the opportunity to ensure a strictly smaller value by playing to v_1''. The latter also implies that σ is admissible. Interestingly, playing to v_1 is also an admissible behavior of Eve since, against a strategy of Adam that plays from v_1 to v_1', it obtains $10\lambda^2(1-\lambda) > 6\lambda^2(1-\lambda)$. ■

The two examples above can be used to argue that the sets of strategies that are regret minimal and admissible, respectively, are in fact incomparable.

Proposition 1. *There are regret-optimal strategies that are not admissible and admissible strategies that have suboptimal regret.*

Proof (Sketch). Consider once more the game depicted in Example 1 and recall that the strategy σ of Eve corresponding to the double edges has minimal regret. This strategy is *not* admissible: it is dominated by the alternative strategy σ' of Eve that behaves like σ from v_1 but plays to v_2' from v_2. Indeed, if Adam plays to v_1 from v_0 then the outcomes of σ and σ' are the same. However, if Adam plays to v_2 then the value of the outcome of σ is 0 while the value of the outcome of σ' is strictly greater than 0.

Similarly, the strategy σ depicted by double edges in the game from Example 2 is admissible but *not* regret-minimizing. In fact, her strategy σ' that consists in playing v_1 from v_0 has a smaller regret. □

In the rest of this section, we show that (1) any strategy is weakly dominated by an admissible strategy; (2) being dominated entails more regret; (3) optipess strategies are both regret-minimal and admissible. We will need the following:

Lemma 3 ([6]). *A strategy σ of Eve is admissible if and only if for every history $h \in \mathbf{Hist}(\sigma)$ the following holds: either $\mathbf{cVal}^h(\sigma) > \mathbf{aVal}^h$ or $\mathbf{aVal}^h(\sigma) = \mathbf{cVal}^h(\sigma) = \mathbf{aVal}^h = \mathbf{acVal}^h$.*

The above characterization of admissible strategies in so-called *well-formed games* was proved in [6, Theorem 11]. Lemma 3 follows from the fact that discounted-sum games are well-formed.

3.1 Any Strategy Is Weakly Dominated by an Admissible Strategy

We show that discounted-sum games have the distinctive property that every strategy is weakly dominated by an admissible strategy. This is in stark contrast with most cases where admissibility has been studied previously [6].

Theorem 2. *Any strategy of Eve is weakly dominated by an admissible strategy.*

Proof (*Sketch*). The main idea is to construct, based on σ, a strategy σ' that will switch to a SBWO strategy as soon as σ does not satisfy the characterization of Lemma 3. The first part of the argument consists in showing that σ is indeed weakly dominated by σ'. This is easily done by comparing, against each strategy τ of Adam, the values of σ and σ'. The second part consists in verifying that σ' is indeed admissible. This is done by checking that each history h consistent with σ' satisfies the characterization of Lemma 3, that is $\mathbf{cVal}^h(\sigma') > \mathbf{aVal}^h$ or $\mathbf{aVal}^h(\sigma') = \mathbf{cVal}^h(\sigma') = \mathbf{aVal}^h = \mathbf{acVal}^h$. □

3.2 Being Dominated Is Regretful

Theorem 3. *For all strategies σ, σ' of Eve such that σ is weakly dominated by σ', it holds that $\mathbf{Reg}(\sigma') \leq \mathbf{Reg}(\sigma)$.*

Proof. Let σ, σ' be such that σ is weakly dominated by σ'. This means that for every strategy τ of Adam, we have that $\mathbf{Val}(\pi) \leq \mathbf{Val}(\pi')$ where $\pi = \mathbf{out}^{v_0}(\sigma, \tau)$ and $\pi' = \mathbf{out}^{v_0}(\sigma', \tau)$. Consequently: we obtain

$$\left(\sup_{\sigma''} \mathbf{Val}(\mathbf{out}^{v_0}(\sigma'', \tau))\right) - \mathbf{Val}(\pi') \leq \left(\sup_{\sigma''} \mathbf{Val}(\mathbf{out}^{v_0}(\sigma'', \tau))\right) - \mathbf{Val}(\pi) \ .$$

As this holds for any τ, we can conclude that $\sup_\tau \sup_{\sigma''}(\mathbf{Val}(\mathbf{out}^{v_0}(\sigma'', \tau)) - \mathbf{Val}(\mathbf{out}^{v_0}(\sigma', \tau))) \leq \sup_\tau \sup_{\sigma''}(\mathbf{Val}(\mathbf{out}^{v_0}(\sigma'', \tau)) - \mathbf{Val}(\mathbf{out}^{v_0}(\sigma, \tau)))$, that is $\mathbf{Reg}(\sigma') \leq \mathbf{Reg}(\sigma)$. □

It follows from Proposition 1, however, that the converse of the theorem is false.

3.3 Optipess Strategies Are both Regret-Minimal and Admissible

Recall that there are admissible strategies that are not regret-minimal and *vice versa* (Proposition 1). However, as a direct consequence of Theorems 2 and 3, there always exist regret-minimal admissible strategies. It turns out that optipess strategies, which are regret-minimal (Theorem 1), are also admissible:

Theorem 4. *All optipess strategies of Eve are admissible.*

Proof. Let $\sigma = \sigma^{\text{sbo}} \xrightarrow{t} \sigma^{\text{sbwo}}$ be an optipess strategy; we show it is admissible. To this end, let $h = v_0 \ldots v_n \in \mathbf{Hist}(\sigma)$; we show that one of the properties of Lemma 3 holds. There are two cases:

(h is switched.) In that case, $\sigma_h = \sigma^{\text{sbwo}}$. Since σ^{sbwo} is an SBWO strategy, $\mathbf{cVal}^h(\sigma^{\text{sbwo}}) = \mathbf{acVal}^h$. Now if $\mathbf{acVal}^h > \mathbf{aVal}^h$, then:

$$\mathbf{cVal}^h(\sigma) = \mathbf{cVal}^h(\sigma^{\text{sbwo}}) = \mathbf{acVal}^h > \mathbf{aVal}^h \ ,$$

and σ satisfies the first property of Lemma 3. Otherwise $\mathbf{acVal}^h = \mathbf{aVal}^h$ and the second property holds: we have that $\mathbf{cVal}^h(\sigma) = \mathbf{acVal}^h$, and as σ^{sbwo} is an SWO and $\mathbf{aVal}^h(\sigma) = \mathbf{aVal}^h(\sigma^{\text{sbwo}})$, we also have that $\mathbf{aVal}^h(\sigma) = \mathbf{aVal}^h$.

(h is unswitched.) We show that $\mathbf{cVal}^h(\sigma) > \mathbf{aVal}^h$. Since h is unswitched, we have in particular that:

$$\mathbf{Reg}\,(\sigma) = \mathbf{Reg} < \lambda^n \left(\mathbf{cVal}^{v_n} - \mathbf{aVal}^{v_n}\right) \ . \tag{1}$$

Furthermore:

$$\begin{aligned}\lambda^n \left(\mathbf{cVal}^{v_n} - \mathbf{aVal}^{v_n}\right) &= \left(\mathbf{Val}(h) + \lambda^n \mathbf{cVal}^{v_n}\right) - \left(\mathbf{Val}(h) + \lambda^n \mathbf{aVal}^{v_n}\right) \\ &= \mathbf{cVal}^h - \mathbf{aVal}^h \ ,\end{aligned}$$

and combining the previous equation with Eq. 1, we obtain:

$$\mathbf{cVal}^h - \mathbf{Reg}\,(\sigma) > \mathbf{aVal}^h \ .$$

To conclude, we show that $\mathbf{Reg}\,(\sigma) \geq \mathbf{cVal}^h - \mathbf{cVal}^h(\sigma)$. Consider a strategy τ of Adam such that h is consistent with both σ^{sbo} and τ and satisfying $\mathbf{Val}(\mathbf{out}^{v_0}(\sigma^{\text{sbo}}, \tau)) = \mathbf{cVal}^h$. (That such a τ exists is intuitively clear since σ has been following the SBO strategy σ^{sbo} along h.) It holds immediately that $\mathbf{cVal}^h(\sigma) \geq \mathbf{Val}(\mathbf{out}^{v_0}(\sigma, \tau))$. Now by definition of the regret:

$$\begin{aligned}\mathbf{Reg}\,(\sigma) &\geq \mathbf{Val}(\mathbf{out}^{v_0}(\sigma^{\text{sbo}}, \tau)) - \mathbf{Val}(\mathbf{out}^{v_0}(\sigma, \tau)) \\ &\geq \mathbf{cVal}^h - \mathbf{cVal}^h(\sigma) \ .\end{aligned}$$

□

4 Minimal Values Are Witnessed by a Single Iterated Cycle

We start our technical work towards a better algorithm to compute the regret value of a game. Here, we show that there are succinctly presentable histories that witness small values in the game. Our intention is to later use this result to apply a modified version of Lemma 2 to bipositional strategies to argue there are small witnesses of a strategy having too much regret.

More specifically, we show that for any history h, there is another history h' of the same length that has smaller value and such that $h' = \alpha \cdot \beta^k \cdot \gamma$ where $|\alpha\beta\gamma|$ is small. This will allow us to find the smallest possible value among exponentially long histories by guessing α, β, γ, and k, which will all be small. This property holds for a wealth of different valuation functions, hinting at possible further applications. For discounted-sum games, the following suffices to prove the desired property holds.

Lemma 4. *For any history $h = \alpha \cdot \beta \cdot \gamma$ with α and γ same-length cycles:*

$$\min\{\mathbf{Val}(\alpha^2 \cdot \beta), \mathbf{Val}(\beta \cdot \gamma^2)\} \leq \mathbf{Val}(h) \ .$$

Within the proof of the key lemma of this section, and later on when we use it (Lemma 9), we will rely on the following notion of cycle decomposition:

Definition 2. *A* simple-cycle decomposition *(SCD) is a pair consisting of paths and iterated simple cycles. Formally, an SCD is a pair $D = \langle (\alpha_i)_{i=0}^n, (\beta_j, k_j)_{j=1}^n \rangle$, where each α_i is a path, each β_j is a simple cycle, and each k_j is a positive integer. We write $D(j) = \beta_j^{k_j} \cdot \alpha_j$ and $D(\star) = \alpha_0 \cdot D(1)D(2) \cdots D(n)$.*

By carefully iterating Lemma 4, we have:

Lemma 5. *For any history h there exists an history $h' = \alpha \cdot \beta^k \cdot \gamma$ with:*

- *h and h' have the same starting and ending vertices, and the same length;*
- *$\mathbf{Val}(h') \leq \mathbf{Val}(h)$;*
- *$|\alpha\beta\gamma| \leq 4|V|^3$ and β is a simple cycle.*

Proof. In this proof, we focus on SCDs for which each path α_i is simple; we call them ßCDs. We define a wellfounded partial order on ßCDs. Let $D = \langle (\alpha_i)_{i=0}^n, (\beta_j, k_j)_{j=1}^n \rangle$ and $D' = \langle (\alpha'_i)_{i=0}^{n'}, (\beta'_j, k'_j)_{j=1}^{n'} \rangle$ be two ßCDs; we write $D' < D$ iff all the following holds:

- $D(\star)$ and $D'(\star)$ have the same starting and ending vertices, the same length, and satisfy $\mathbf{Val}(D'(\star)) \leq \mathbf{Val}(D(\star))$ and $n' \leq n$;
- Either $n' < n$, or $|\alpha'_0 \cdots \alpha'_{n'}| < |\alpha_0 \cdots \alpha_n|$, or $|\{k'_i \geq |V|\}| < |\{k_i \geq |V|\}|$.

That this order has no infinite descending chain is clear. We show two claims:

1. Any ßCD with n greater than $|V|$ has a smaller ßCD;
2. Any ßCD with two $k_j, k_{j'} > |V|$ has a smaller ßCD.

Together they imply that for a smallest ßCD D, $D(\star)$ is of the required form. Indeed let j be the unique value for which $k_j > |V|$, then the statement of the Lemma is satisfied by letting $\alpha = \alpha_0 \cdot D(1) \cdots D(j-1)$, $\beta = \beta_j$, $k = k_j$, and $\gamma = \alpha_j \cdot D(j+1) \cdots D(n)$.

Claim 1. Suppose D has $n > |V|$. Since all cycles are simple, there are two cycles $\beta_j, \beta_{j'}$, $j < j'$, of same length. We can apply Lemma 4 on the path $\beta_j \cdot (\alpha_j D(j+1) \cdots D(j'-1)) \cdot \beta_{j'}$, and remove one of the two cycles while duplicating the other; we thus obtain a similar path of smaller value. This can be done repeatedly until we obtain a path with only one of the two cycles, say $\beta_{j'}$, the other case being similar. Substituting this path in $D(\star)$ results in:

$$\alpha_0 \cdot D(1) \cdots D(j) \cdot \left(\alpha_j \cdot D(j+1) \cdots D(j'-1) \cdot \beta_{j'}^{k_j + k_{j'}}\right) \cdot \alpha_{j'} \cdot D(j'+1) \cdots D(n) \enspace .$$

This gives rise to a smaller ßCD as follows. If $\alpha_{j-1}\alpha_j$ is still a simple path, then the above history is expressible as an ßCD with a smaller number of cycles. Otherwise, we rewrite $\alpha_{j-1}\alpha_j = \alpha'_{j-1}\beta'_j\alpha'_j$ where α'_{j-1} and α'_j are simple paths and β'_j is a simple cycle; since $|\alpha'_{j-1}\alpha'_j| < |\alpha_{j-1}\alpha_j|$, the resulting ßCD is smaller.

Claim 2. Suppose D has two $k_j, k_{j'} > |V|$, $j < j'$. Since each cycle in the ßCD is simple, k_j and $k_{j'}$ are greater than both $|\beta_j|$ and $|\beta_{j'}|$; let us write $k_j = b|\beta_{j'}| + r$ with $0 \le r < |\beta_{j'}|$, and similarly, $k_{j'} = b'|\beta_j| + r'$. We have:

$$D(j) \cdots D(j') = \beta_j^r \cdot \left((\beta_j^{|\beta_{j'}|})^b \cdot \alpha_j \cdot D(j+1) \cdots D(j'-1) \cdot (\beta_{j'}^{|\beta_j|})^{b'} \right) \cdot \beta_{j'}^{r'} \cdot \alpha_{j'} \ .$$

Noting that $\beta_{j'}^{|\beta_j|}$ and $\beta_j^{|\beta_{j'}|}$ are cycles of the same length, we can transfer all the occurrences of one to the other, as in Claim 1. Similarly, if two simple paths get merged and give rise to a cycle, a smaller ßCD can be constructed; if not, then there are now at most $r < |V|$ occurrences of $\beta_{j'}$ (or conversely, r' of β_j), again resulting in a smaller ßCD. □

5 Short Witnesses for Regret, Antagonistic, and Collaborative Values

We continue our technical work towards our algorithm for computing the regret value. In this section, the overarching theme is that of *short witnesses*. We show that (1) the regret value of a strategy is witnessed by histories of bounded length; (2) the collaborative value of a game is witnessed by a simple path and an iterated cycle; (3) the antagonistic value of a strategy is witnessed by an SCD and an iterated cycle.

5.1 Regret Is Witnessed by Histories of Bounded Length

Lemma 6. *Let* $\sigma = \sigma_1 \xrightarrow{t} \sigma_2$ *be an arbitrary bipositional switching strategy of Eve and let* $C = 2|V| + \max\{t(v) < \infty\}$. *We have that:*

$$\mathbf{Reg}\,(\sigma) = \max \Big\{ \lambda^n \left(\mathbf{cVal}^{v_n}_{\neg\sigma(h)} - \mathbf{aVal}^{v_n}(\sigma_h) \right) \Big| \; h = v_0 \ldots v_n \in \mathbf{Hist}(\sigma), n \le C \Big\} \ .$$

Proof. Consider a history h of length greater than C, and write $h = h_1 \cdot h_2$ with $|h_1| = \max\{t(v) < \infty\}$. Let $h_2 = p \cdot p'$ where p is the maximal prefix of h_2 such that $h_1 \cdot p$ is unswitched—we set $p = \epsilon$ if h is switched. Note that one of p or p' is longer than $|V|$—say p, the other case being similar. This implies that there is a cycle in p, i.e., $p = \alpha \cdot \beta \cdot \gamma$ with β a cycle. Let $h' = h_1 \cdot \alpha \cdot \gamma \cdot p'$; this history has the same starting and ending vertex as h. Moreover, since $|h_1|$ is larger than any value of the threshold function, $\sigma_h = \sigma_{h'}$. Lastly, h' is still in $\mathbf{Hist}(\sigma)$, since the removed cycle did not play a role in switching strategy. This shows:

$$\mathbf{cVal}^{v_n}_{\neg\sigma(h)} - \mathbf{aVal}^{v_n}(\sigma_h) = \mathbf{cVal}^{v_n}_{\neg\sigma(h')} - \mathbf{aVal}^{v_n}(\sigma_{h'}) \ .$$

Since the length of h is greater than the length of h', the discounted value for h' will be greater than that of h, resulting in a higher regret value. There is thus no need to consider histories of size greater than C. □

It may seem from this lemma and the fact that $t(v)$ may be very large that we will need to guess histories of important length. However, since we will be considering bipositional switching strategies, we will only be interested in guessing *some* properties of the histories that are not hard to verify:

Lemma 7. *The following problem is decidable in* NP*:*

> ***Given:*** *A game, a bipositional switching strategy σ,*
> *a number n in binary, a Boolean b, and two vertices v, v'*
>
> ***Question:*** *Is there a $h \in \mathbf{Hist}(\sigma)$ of length n, switched if b,*
> *ending in v, with $\sigma(h) = v'$?*

Proof. This is done by guessing multiple flows within the graph (V, E). Here, we call *flow* a valuation of the edges E by integers, that describes the number of times a path crosses each edge. Given a vector in $\mathbb{N}^E$, it is not hard to check whether there is a path that it represents, and to extract the initial and final vertices of that path [17].

We first order the different thresholds from the strategy $\sigma = \sigma_1 \xrightarrow{t} \sigma_2$: let $V_\exists = \{v_1, v_2, \ldots, v_k\}$ with $t(v_i) \leq t(v_{i+1})$ for all i. We analyze the structure of histories consistent with σ. Let $h \in \mathbf{Hist}(\sigma)$, and write $h = h' \cdot h''$ where h' is the maximal unswitched prefix of h. Naturally, h' is consistent with σ_1 and h'' is consistent with σ_2. Then $h' = h_0 h_1 \cdots h_i$, for some $i < |V_\exists|$, with:

- $|h_0| = t(v_1)$ and for all $1 \leq j < i$, $|h_j| = t(v_{j+1}) - t(v_j)$;
- For all $0 \leq j \leq i$, h_j does not contain a vertex v_k with $k \leq j$.

To confirm the existence of a history with the given parameters, it is thus sufficient to guess the value $i \leq |V_\exists|$, and to guess i connected flows (rather than paths) with the above properties that are consistent with σ_1. Finally, we guess a flow for h'' consistent with σ_2 if we need a switched history, and verify that it is starting at a switching vertex. The flows must sum to $n + 1$, with the last vertex being v', and the previous v. □

5.2 Short Witnesses for the Collaborative and Antagonistic Values

Lemma 8. *There is a set P of pairs (α, β) with α a simple path and β a simple cycle such that:*

- $\mathbf{cVal}^{v_0} = \max\{\mathbf{Val}(\alpha \cdot \beta^\omega) \mid (\alpha, \beta) \in P\}$ *and*
- *membership in P is decidable in polynomial time w.r.t. the game.*

Proof. We argue that the set P of all pairs (α, β) with α a simple path, β a simple cycle, and such that $\alpha \cdot \beta$ is a path, gives us the result.

The first part of the claim is a consequence of Lemma 1: Consider positional SBO strategies τ and σ of Adam and Eve, respectively. Since they are positional, the path $\mathbf{out}^{v_0}(\sigma, \tau)$ is of the form $\alpha \cdot \beta^\omega$, as required, and its value is $\mathbf{cVal}^{v_0}$. We can thus let P be the set of all pairs obtained from such SBO strategies.

Moreover, it can be easily checked that for all pairs (α, β) such that $\alpha \cdot \beta$ is a path in the game there exists a pair of strategies with outcome $\alpha \cdot \beta^\omega$. (Note that verifying whether $\alpha \cdot \beta$ is a path can indeed be done in polynomial time given α and β.) Finally, the value $\mathbf{Val}(\alpha \cdot \beta^\omega)$ will, by definition, be at most $\mathbf{cVal}^{v_0}$. □

Lemma 9. *Let σ be a bipositional switching strategy of Eve. There is a set K of pairs (D, β) with D an SCD and β a simple cycle such that:*

- $\mathbf{aVal}^{v_0}(\sigma) = \min\{\mathbf{Val}(D(\star) \cdot \beta^\omega) \mid (D, \beta) \in K\}$ *and*
- *the size of each pair is polynomially bounded, and membership in K is decidable in polynomial time w.r.t. σ and the game.*

Proof. We will prove that the set K of all pairs (D, β) with D an SCD of polynomial length (which will be specified below), β a simple cycle, and such that $D(\star) \cdot \beta$ is a path, satisfies our claims.

Let $C = \max\{t(v) < \infty\}$, and consider a play π consistent with σ that achieves the value $\mathbf{aVal}^{v_0}(\sigma)$. Write $\pi = h \cdot \pi'$ with $|h| = C$, and let v be the final vertex of h. Naturally:

$$\mathbf{aVal}^{v_0}(\sigma) = \mathbf{Val}(\pi) = \mathbf{Val}(h) + \lambda^{|h|}\mathbf{Val}(\pi') \ .$$

We first show how to replace π' by some $\alpha \cdot \beta^\omega$, with α a simple path and β a simple cycle. First, since π witnesses $\mathbf{aVal}^{v_0}(\sigma)$, we have that $\mathbf{Val}(\pi') = \mathbf{aVal}^{v}(\sigma_h)$. Now σ_h is positional, because $|h| \geq C$.[1] It is known that there are optimal positional antagonistic strategies τ for Adam, that is, that satisfy $\mathbf{aVal}^{v}(\sigma_h) = \mathbf{out}^{v}(\sigma_h, \tau)$. As in the proof of Lemma 8, this implies that $\mathbf{aVal}^{v}(\sigma_h) = \mathbf{Val}(\alpha \cdot \beta^\omega) = \mathbf{Val}(\pi')$ for some α and β; additionally, any (α, β) that are consistent with σ_h and a potential strategy for Adam will give rise to a larger value.

We now argue that $\mathbf{Val}(h)$ is witnessed by an SCD of polynomial size. This bears similarity to the proof of Lemma 7. Specifically, we will reuse the fact that histories consistent with σ can be split into histories played "between thresholds."

Let us write $\sigma = \sigma_1 \xrightarrow{t} \sigma_2$. Again, we let $V_\exists = \{v_1, v_2, \ldots, v_k\}$ with $t(v_i) \leq t(v_{i+1})$ for all i and write $h = h' \cdot h''$ where h' is the maximal unswitched prefix of h. We note that h' is consistent with σ_1 and h'' is consistent with σ_2. Then $h' = h_0 h_1 \cdots h_i$, for some $i < |V_\exists|$, with:

- $|h_0| = t(v_1)$ and for all $1 \leq j < i$, $|h_j| = t(v_{j+1}) - t(v_j)$;
- For all $0 \leq j \leq i$, h_j does not contain a vertex v_k with $k \leq j$.

We now diverge from the proof of Lemma 7. We apply Lemma 5 on each h_j in the game where the strategy σ_1 is hardcoded (that is, we first remove every edge $(u, v) \in V_\exists \times V$ that does not satisfy $\sigma_1(u) = v$). We obtain a history $h'_0 h'_1 \cdots h'_i$ that is still in $\mathbf{Hist}(\sigma)$, thanks to the previous splitting of h. We also apply Lemma 5 to h', this time in the game where σ_2 is hardcoded, obtaining h''. Since each h'_j and h'' are expressed as $\alpha \cdot \beta^k \cdot \gamma$, there is an SCD D with no more

[1] Technically, σ_h is positional in the game that records whether the switch was made.

than $|V_\exists|$ elements that satisfies $\mathbf{Val}(D(\star)) \leq \mathbf{Val}(h)$—naturally, since $\mathbf{Val}(h)$ is minimal and $D(\star) \in \mathbf{Hist}(\sigma)$, this means that the two values are equal. Note that it is not hard, given an SCD D, to check whether $D(\star) \in \mathbf{Hist}(\sigma)$, and that SCDs that are not valued $\mathbf{Val}(h)$ have a larger value. □

6 The Complexity of Regret

We are finally equipped to present our algorithms. To account for the cost of numerical analysis, we rely on the problem PosSLP [2]. This problem consists in determining whether an arithmetic circuit with addition, subtraction, and multiplication gates, together with input values, evaluates to a positive integer. PosSLP is known to be decidable in the so-called counting hierarchy, itself contained in the set of problems decidable using polynomial space.

Theorem 5. *The following problem is decidable in* $\mathsf{NP}^{\mathsf{PosSLP}}$*:*

Given: *A game, a bipositional switching strategy* σ*,*
a value $r \in \mathbb{Q}$ *in binary*
Question: *Is* $\mathbf{Reg}\,(\sigma) > r$*?*

Proof. Let us write $\sigma = \sigma_1 \xrightarrow{t} \sigma_2$. Lemma 6 indicates that $\mathbf{Reg}\,(\sigma) > r$ holds if there is a history h of some length $n \leq C = 2|V| + \max\{t(v) < \infty\}$, ending in some v_n such that:

$$\lambda^n \left(\mathbf{cVal}^{v_n}_{\neg\sigma(h)} - \mathbf{aVal}^{v_n}(\sigma_h) \right) > r \ . \tag{2}$$

Note that since σ is bipositional, we do not need to know everything about h. Indeed, the following properties suffice: its length n, final vertex v_n, $v' = \sigma(h)$, and whether it is switched. Rather than guessing h, we can thus rely on Lemma 7 to get the required information. We start by simulating the NP machine that this lemma provides, and verify that n, v_n, and v are consistent with a potential history.

Let us now concentrate on the collaborative value that we need to evaluate in Eq. 2. To compute $\mathbf{cVal}$, we rely on Lemma 8, which we apply in the game where v_n is set initial, and its successor forced not to be v. We guess a pair $(\alpha_{\mathrm{c}}, \beta_{\mathrm{c}}) \in P$; we thus have $\mathbf{Val}(\alpha_{\mathrm{c}} \cdot \beta_{\mathrm{c}}^\omega) \leq \mathbf{cVal}^{v_n}_{\neg\sigma(h)}$, with at least one guessed pair $(\alpha_{\mathrm{c}}, \beta_{\mathrm{c}})$ reaching that latter value.

Let us now focus on computing $\mathbf{aVal}^{v_n}(\sigma_h)$. Since σ is a bipositional switching strategy, σ_h is simply σ where $t(v)$ is changed to $\max\{0, t(v) - n\}$. Lemma 9 can thus be used to compute our value. To do so, we guess a pair $(D, \beta_{\mathrm{a}}) \in K$; we thus have $\mathbf{Val}(D(\star) \cdot \beta_{\mathrm{a}}^\omega) \geq \mathbf{aVal}^{v_n}(\sigma_h)$, and at least one pair (D, β_{a}) reaches that latter value.

Our guesses satisfy:

$$\mathbf{cVal}^{v_n}_{\neg\sigma(h)} - \mathbf{aVal}^{v_n}(\sigma_h) \geq \mathbf{Val}(\alpha_{\mathrm{c}} \cdot \beta_{\mathrm{c}}^\omega) - \mathbf{Val}(D(\star) \cdot \beta_{\mathrm{a}}^\omega) \ ,$$

and there is a choice of our guessed paths and SCD that gives exactly the left-hand side. Comparing the left-hand side with r can be done using an oracle to PosSLP, concluding the proof. □

Theorem 6. *The following problem is decidable in* $\mathsf{coNP}^{\mathsf{NP}^{\mathsf{PosSLP}}}$*:*

Given: *A game, a value* $r \in \mathbb{Q}$ *in binary*

Question: *Is* $\mathbf{Reg} > r$*?*

Proof. To decide the problem at hand, we ought to check that *every* strategy has a regret value greater than r. However, optipess strategies being regret-minimal, we need only check this for a class of strategies that contains optipess strategies: bipositional switching strategies form one such class.

What is left to show is that optipess strategies can be encoded in *polynomial space*. Naturally, the two positional strategies contained in an optipess strategy can be encoded succinctly. We thus only need to show that, with t as in the definition of optipess strategies (page 5), $t(v)$ is at most exponential for every $v \in V_\exists$ with $t(v) \in \mathbb{N}$. This is shown in the long version of this paper. □

Theorem 7. *The following problem is decidable in* $\mathsf{coNP}^{\mathsf{NP}^{\mathsf{PosSLP}}}$*:*

Given: *A game, a bipositional switching strategy* σ

Question: *Is* σ *regret optimal?*

Proof. A consequence of the proof of Theorem 5 and the existence of optipess strategies is that the value $\mathbf{Reg}$ of a game can be computed by a polynomial size arithmetic circuit. Moreover, our reliance on PosSLP allows the input r in Theorem 5 to be represented as an arithmetic circuit without impacting the complexity. We can thus verify that for all bipositional switching strategies σ' (with sufficiently large threshold functions) and all possible polynomial size arithmetic circuits, $\mathbf{Reg}(\sigma) > r$ implies that $\mathbf{Reg}(\sigma') > r$. The latter holds if and only if σ is regret optimal since, as we have argued in the proof of Theorem 6, such strategies σ' include optipess strategies and thus regret-minimal strategies. □

7 Conclusion

We studied *regret*, a notion of interest for an agent that does not want to assume that the environment she plays in is simply adversarial. We showed that there are strategies that both minimize regret, and are not consistently worse than any other strategies. The problem of computing the minimum regret value of a game was then explored, and a better algorithm was provided for it.

The exact complexity of this problem remains however open. The only known lower bound, a straightforward adaptation of [14, Lemma 3] for discounted-sum games, shows that it is at least as hard as solving parity games [15].

Our upper bound could be significantly improved if we could efficiently solve the following problem:

PosRatBase

Given: $(a_i)_{i=1}^n \in \mathbb{Z}^n$, $(b_i)_{i=1}^n \in \mathbb{N}^n$, and $r \in \mathbb{Q}$ all in binary,

Question: Is $\sum_{i=1}^n a_i \cdot r^{b_i} > 0$?

This can be seen as the problem of comparing succinctly represented numbers in a rational base. The PosSLP oracle in Theorem 5 can be replaced by an oracle for this seemingly simpler arithmetic problem. The variant of PosRatBase in which r is an integer was shown to be in P by Cucker, Koiran, and Smale [8], and they mention that the complexity is open for rational values. To the best of our knowledge, the exact complexity of PosRatBase is open even for $n = 3$.

Acknowledgements. We thank Raphaël Berthon and Ismaël Jecker for helpful conversations on the length of maximal (and minimal) histories in discounted-sum games, James Worrell and Joël Ouaknine for pointers on the complexity of comparing succinctly represented integers, and George Kenison for his writing help.

References

1. de Alfaro, L., Henzinger, T.A., Majumdar, R.: Discounting the future in systems theory. In: Baeten, J.C.M., Lenstra, J.K., Parrow, J., Woeginger, G.J. (eds.) ICALP 2003. LNCS, vol. 2719, pp. 1022–1037. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45061-0_79
2. Allender, E., Bürgisser, P., Kjeldgaard-Pedersen, J., Miltersen, P.B.: On the complexity of numerical analysis. SIAM J. Comput. **38**(5), 1987–2006 (2009). https://doi.org/10.1137/070697926
3. Aminof, B., Kupferman, O., Lampert, R.: Reasoning about online algorithms with weighted automata. ACM Trans. Algorithms **6**(2), 28:1–28:36 (2010). https://doi.org/10.1145/1721837.1721844
4. Apt, K.R., Grädel, E.: Lectures in Game Theory for Computer Scientists. Cambridge University Press, New York (2011)
5. Brenguier, R., et al.: Non-zero sum games for reactive synthesis. In: Dediu, A.-H., Janoušek, J., Martín-Vide, C., Truthe, B. (eds.) LATA 2016. LNCS, vol. 9618, pp. 3–23. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30000-9_1
6. Brenguier, R., Pérez, G.A., Raskin, J.F., Sankur, O.: Admissibility in quantitative graph games. In: Lal, A., Akshay, S., Saurabh, S., Sen, S. (eds.) 36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2016. LIPIcs, Chennai, India, 13–15 December 2016, vol. 65, pp. 42:1–42:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2016). https://doi.org/10.4230/LIPIcs.FSTTCS.2016.42
7. Chatterjee, K., Goharshady, A.K., Ibsen-Jensen, R., Velner, Y.: Ergodic mean-payoff games for the analysis of attacks in crypto-currencies. In: Schewe, S., Zhang, L. (eds.) 29th International Conference on Concurrency Theory, CONCUR 2018. LIPIcs, Beijing, China, 4–7 September 2018, vol. 118, pp. 11:1–11:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2018). https://doi.org/10.4230/LIPIcs.CONCUR.2018.11

8. Cucker, F., Koiran, P., Smale, S.: A polynomial time algorithm for diophantine equations in one variable. J. Symb. Comput. **27**(1), 21–29 (1999). https://doi.org/10.1006/jsco.1998.0242
9. Filar, J., Vrieze, K.: Competitive Markov Decision Processes. Springer, Heidelberg (2012). https://doi.org/10.1007/978-1-4612-4054-9
10. Filiot, E., Le Gall, T., Raskin, J.-F.: Iterated regret minimization in game graphs. In: Hliněný, P., Kučera, A. (eds.) MFCS 2010. LNCS, vol. 6281, pp. 342–354. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15155-2_31
11. Filiot, E., Jecker, I., Lhote, N., Pérez, G.A., Raskin, J.F.: On delay and regret determinization of max-plus automata. In: 32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, 20–23 June 2017, pp. 1–12. IEEE Computer Society (2017). https://doi.org/10.1109/LICS.2017.8005096
12. Halpern, J.Y., Pass, R.: Iterated regret minimization: a new solution concept. Games Econ. Behav. **74**(1), 184–207 (2012). https://doi.org/10.1016/j.geb.2011.05.012
13. Hunter, P., Pérez, G.A., Raskin, J.F.: Minimizing regret in discounted-sum games. In: Talbot, J.M., Regnier, L. (eds.) 25th EACSL Annual Conference on Computer Science Logic, CSL 2016. LIPIcs, Marseille, France, 29 August–1 September 2016, vol. 62, pp. 30:1–30:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2016). https://doi.org/10.4230/LIPIcs.CSL.2016.30
14. Hunter, P., Pérez, G.A., Raskin, J.F.: Reactive synthesis without regret. Acta Inf. **54**(1), 3–39 (2017). https://doi.org/10.1007/s00236-016-0268-z
15. Jurdzinski, M.: Deciding the winner in parity games is in UP ∩ co-UP. Inf. Process. Lett. **68**(3), 119–124 (1998). https://doi.org/10.1016/S0020-0190(98)00150-1
16. Puterman, M.L.: Markov Decision Processes. Wiley-Interscience, New York (2005)
17. Reutenauer, C.: The Mathematics of Petri Nets. Prentice-Hall Inc., Upper Saddle River (1990)
18. Shapley, L.S.: Stochastic games. Proc. Natl. Acad. Sci. **39**(10), 1095–1100 (1953)
19. Watkins, C.J.C.H., Dayan, P.: Technical note Q-learning. Mach. Learn. **8**, 279–292 (1992). https://doi.org/10.1007/BF00992698
20. Zwick, U., Paterson, M.: The complexity of mean payoff games on graphs. Theor. Comput. Sci. **158**(1&2), 343–359 (1996). https://doi.org/10.1016/0304-3975(95)00188-3