Reaching Your Goal Optimally by Playing at Random with No Memory

Benjamin Monmege
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
benjamin.monmege@univ-amu.fr

Julie Parreaux
ENS Rennes, France
julie.parreaux@ens-rennes.fr

Pierre-Alain Reynier
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
pierre-alain.reynier@univ-amu.fr

Abstract

Shortest-path games are two-player zero-sum games played on a graph equipped with integer weights. One player, that we call Min, wants to reach a target set of states while minimising the total weight, and the other one has an antagonistic objective. This combination of a qualitative reachability objective and a quantitative total-payoff objective is one of the simplest settings where Min needs memory (pseudo-polynomial in the weights) to play optimally. In this article, we aim at studying a tradeoff allowing Min to play at random, but using no memory. We show that Min can achieve the same optimal value in both cases. In particular, we compute a randomised memoryless ε-optimal strategy when it exists, where probabilities are parametrised by ε. We also show that for some games, no optimal randomised strategies exist. We then characterise, and decide in polynomial time, the class of games admitting an optimal randomised memoryless strategy.

2012 ACM Subject Classification Software and its engineering → Formal software verification; Theory of computation → Algorithmic game theory

Keywords and phrases Weighted games, Algorithmic game theory, Randomisation

Digital Object Identifier 10.4230/LIPIcs.CONCUR.2020.26

Funding Benjamin Monmege and Pierre-Alain Reynier are partly funded by ANR project Ticktac (ANR-18-CE40-0015).

1 Introduction

Game theory is now an established model in the computer-aided design of correct-by-construction programs. Two players, the controller and an environment, are fighting one against the other in a zero-sum game played on a graph of all possible configurations. A winning strategy for the controller results in a correct program, while the environment is a player modelling all uncontrollable events that the program must face. Many possible objectives have been studied in such two-player zero-sum games played on graphs: reachability, safety, repeated reachability, and even all possible ω-regular objectives [10].

Apart from such *qualitative* objectives, more *quantitative* ones are useful in order to select a particular strategy among all the ones that are correct with respect to a qualitative objective. Some metrics of interest, mostly studied in the quantitative game theory literature, are mean-payoff, discounted-payoff, or total-payoff. All these objectives have in common that both players have strategies using no memory or randomness to win or play optimally [9].

Combining quantitative and qualitative objectives, enabling to select a good strategy among the valid ones for the selected metrics, often leads to the need of memory to play optimally. One of the simplest combinations showing this consists in the shortest-path

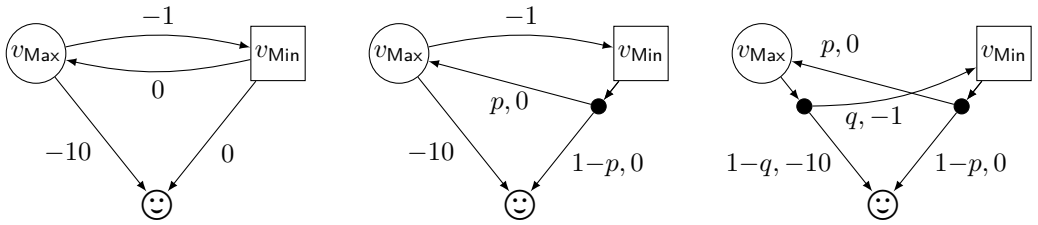

Figure 1 On the left, a shortest-path game, where Min requires memory to play optimally. In the middle, the Markov Decision Process obtained when letting Min play at random, with a parametric probability $p \in (0, 1)$. On the right, the Markov Chain obtained when Max plays along a memoryless randomised strategy, with a parametric probability $q \in [0, 1]$.

games combining a reachability objective with a total-payoff quantitative objective (studied in [11, 4] under the name of *min-cost reachability games*). Another case of interest is the combination of a parity qualitative objective (modelling every possible ω-regular condition), with a mean-payoff objective (aiming for a controller of good quality in the average long-run), where controllers need memory, and even infinite memory, to play optimally [6].

It is often crucial to enable randomisation in the strategies. For instance, Nash equilibria are only ensured to exist in matrix games (like rock-paper-scissors) when players can play at random [13]. In the context of games on graphs, a player may choose, depending on the current history, the probability distribution on the successors. In contrast, strategies that do not use randomisation are called *deterministic* (we sometimes say *pure*).

In this article, we will focus on shortest-path games, as the one depicted on the left of Figure 1. The objective of Min is to reach vertex ☺, while minimising the total weight. Let us consider the vertex v_{Min} as initial. Player Min could reach directly ☺, thus leading to a payoff of 0. But he can also choose to go to v_{Max}, in which case Max either jumps directly in ☺ (leading to a beneficial payoff -10), or comes back to v_{Min}, but having already capitalised a total payoff -1. We can continue this way *ad libitum* until Min is satisfied (at least 10 times) and jumps to ☺. This guarantees a value at most -10 for Min when starting in v_{Min}. Reciprocally, Max can guarantee a payoff at least -10 by directly jumping into ☺ when she must play for the first time. Thus, the optimal value is -10 when starting from v_{Min} or v_{Max}. However, Min cannot achieve this optimal value by playing *without memory* (we sometimes say *positionally*), since it either results in a total-payoff 0 (directly going to the target) or Max has the opportunity to keep Min in the negative cycle for ever, thus never reaching the target. Therefore, Min needs memory to play optimally. He can do so by playing a *switching strategy*, turning in the negative cycle long enough so that no matter how he reaches the target finally, the value he gets as a payoff is lower than the optimal value. This strategy uses pseudo-polynomial memory with respect to the weights of the game graph.

In this example, such a switching strategy can be *mimicked* using randomisation only (and no memory), Min deciding to go to v_{Max} with high probability $p < 1$ and to go to the target vertex with the remaining low probability $1 - p > 0$ (we enforce this probability to be positive, in order to reach the target with probability 1, no matter how the opponent is playing). The resulting *Markov Decision Process (MDP)* is depicted in the middle of Figure 1. The shortest path problem in such MDPs has been thoroughly studied in [2], where it is proved that Max does not require memory to play optimally. Denoting by q the probability that Max jumps in v_{Min} in its memoryless strategy, we obtain the *Markov chain (MC)* on the right of Figure 1. We can compute (see Example 4) the expected value in this MC, as well as the best strategy for both players: in the overall, the optimal value remains -10, even if Min no longer has an optimal strategy. He rather has an ε-optimal strategy, consisting in choosing $p = 1 - \varepsilon/10$ that ensures a value at most $-10 + \varepsilon$.

This article thus aims at studying the tradeoff between memory and randomisation in strategies for shortest-path games. The study is only interesting in the presence of both positive and negative weights, since both players have optimal memoryless deterministic strategies when the graph contains only non-negative weights [11]. The tradeoff between memory and randomisation has already been investigated in many classes of games where memory is required to win or play optimally. This is for instance the case for qualitative games like Street or Müller games thoroughly studied (with and without randomness in the arena) in [5]. The study has been extended to timed games [7] where the goal is to use as little information as possible about the precise values of real-time clocks. Memory or randomness is also crucial in multi-dimensional objectives [8]: for instance, in mean-payoff parity games, if there exists a deterministic finite-memory winning strategy, then there exists a randomised memoryless almost-sure winning strategy.

In contrast to previous work, we show that deterministic memory and memoryless randomisation provide the same power to Min. We leave the combination of memory and randomisation for future work, as explained in the discussion. After a presentation of the model of shortest-path games in Section 2, we show in Section 3 how the previous simulation of memory with randomisation can be performed for all shortest-path games. The general case is much more challenging, in particular in the presence of positive cycles in the graph, that Min cannot avoid in general. Section 4 shows reciprocally how to mimic randomised strategies with memory only. Section 5 studies the optimality of randomised strategies. Indeed, all shortest-path games admit an optimal deterministic strategy for both players, but Min may require memory to play optimally (even with randomisation allowed). We thus characterises the shortest-path games in which Min admits an optimal memoryless strategy, and decide this characterisation in polynomial time.

2 Shortest-path games: deterministic or memoryless strategies

In this section, we formally introduce the shortest-path games we consider throughout the article, as already thoroughly studied in [4] under the name of *min-cost reachability games*. We denote by $\mathbb{Z}$ the set of integers, and $\mathbb{Z}_\infty = \mathbb{Z} \cup \{-\infty, +\infty\}$. For a finite set V, we denote by $\Delta(V)$ the set of *distributions* over V, that are all mappings $\delta \colon V \to [0,1]$ such that $\sum_{v \in V} \delta(v) = 1$. The support of a distribution δ is the set $\{v \in V \mid \delta(v) > 0\}$, denoted by $\mathrm{supp}(\delta)$. A Dirac distribution is a distribution with a singleton support: the Dirac distribution of support $\{v\}$ is denoted by Dirac_v.

We consider two-player turn-based games played on weighted graphs and denote the players by Max and Min. Formally, a *shortest-path game* (SPG) is a tuple $\langle V_{\mathsf{Max}}, V_{\mathsf{Min}}, E, \omega, T \rangle$ where $V := V_{\mathsf{Max}} \uplus V_{\mathsf{Min}} \uplus T$ is a finite set of vertices partitioned into the sets V_{Max} and V_{Min} of Max and Min respectively, and a set T of target vertices, $E \subseteq V \setminus T \times V$ is a set of *directed edges*, and $\omega \colon E \to \mathbb{Z}$ is the *weight function*, associating an integer weight with each edge. In the drawings, Max vertices are depicted by circles; Min vertices by rectangles. For every vertex $v \in V$, the set of successors of v with respect to E is denoted by $E(v) = \{v' \in V \mid (v, v') \in E\}$. Without loss of generality, we assume that non-target vertices are deadlock-free, i.e. for all vertices $v \in V \setminus T$, $E(v) \neq \emptyset$. Finally, throughout this article, we let $W = \max_{(v,v') \in E} |\omega(v, v')|$ be the greatest edge weight (in absolute value) in the arena. A *finite play* is a finite sequence of vertices $\pi = v_0 v_1 \cdots v_k \in V^*$ such that for all $0 \leqslant i < k$, $(v_i, v_{i+1}) \in E$. Its *total weight* is the sum $\sum_{i=0}^{k-1} \omega(v_i, v_{i+1})$ of its weights. A *play* is either a finite play ending in a target vertex, or an infinite sequence of vertices $\pi = v_0 v_1 \cdots$ avoiding the target such that every finite prefix $v_0 \cdots v_k$, denoted by $\pi[k]$, is a finite play.

The total-payoff of a play $\pi = v_0 v_1 \dots$ is given by $\mathbf{TP}(\pi) = +\infty$ if the play is infinite (and therefore avoids T), or by the total weight $\mathbf{TP}(\pi) = \sum_{i=0}^{k-1} \omega(v_i, v_{i+1})$ if $\pi = v_0 v_1 \cdots v_k$ is a finite play ending in a vertex $v_k \in T$ (for the first time).

A *strategy* for Min over an arena $\mathcal{G} = \langle V_{\mathsf{Max}}, V_{\mathsf{Min}}, E, \omega, T\rangle$ is a mapping $\sigma\colon V^* V_{\mathsf{Min}} \to \Delta(V)$ such that for all sequences $\pi = v_0 \cdots v_k$ with $v_k \in V_{\mathsf{Min}}$, the support of the distribution $\sigma(\pi)$ is included in $E(v_k)$. A play or finite play $\pi = v_0 v_1 \cdots$ conforms to the strategy σ if for all k such that $v_k \in V_{\mathsf{Min}}$, we have that $\sigma(\pi[k])(v_{k+1}) > 0$. A similar definition allows one to define strategies $\tau\colon V^* V_{\mathsf{Max}} \to \Delta(V)$ for Max, and plays conforming to them.

A strategy σ is *deterministic* (or *pure*) if for all finite plays π, $\sigma(\pi)$ is a Dirac distribution: in this case, we let $\sigma(\pi)$ denote the unique vertex in the support of this Dirac distribution. We let $\mathsf{d}\Sigma_{\mathsf{Min}}$ and $\mathsf{d}\Sigma_{\mathsf{Max}}$ be the deterministic strategies of players Min and Max, respectively. A strategy σ is *memoryless* if for all finite plays π, π', and all vertices $v \in V$, we have that $\sigma(\pi v) = \sigma(\pi' v)$ for all $v \in V$. We let $\mathsf{m}\Sigma_{\mathsf{Min}}$ and $\mathsf{m}\Sigma_{\mathsf{Max}}$ be the memoryless strategies of players Min and Max, respectively. To distinguish them easily from deterministic strategies, we will denote a memoryless strategy of Min using letter ρ (for *r*andom).

In this article, we focus on deterministic strategies on the one hand, and memoryless strategies on the other hand. Even if the notion of values that we will now introduce could be defined in a more general setting, we prefer to give two simpler definitions in the two separate cases, for the sake of clarity.

2.1 Deterministic strategies

In case of deterministic strategies, for all vertices v, we let $\mathsf{Play}(v, \sigma, \tau)$ be the unique play conforming to strategies σ and τ of Min and Max, respectively, and starting in v. This unique play has a payoff $\mathbf{TP}(\mathsf{Play}(v, \sigma, \tau))$. Then, we define the value of strategies σ and τ by letting for all v,

$$\mathsf{dVal}^{\sigma}(v) = \sup_{\tau' \in \mathsf{d}\Sigma_{\mathsf{Max}}} \mathbf{TP}(\mathsf{Play}(v, \tau', \sigma)) \qquad \text{and} \qquad \mathsf{dVal}^{\tau}(v) = \inf_{\sigma' \in \mathsf{d}\Sigma_{\mathsf{Min}}} \mathbf{TP}(\mathsf{Play}(v, \tau, \sigma'))$$

Finally, the game itself has two possible values, an *upper value* describing the best Min can hope for, and a *lower value* describing the best Max can hope for: for all vertices v,

$$\overline{\mathsf{dVal}}(v) = \inf_{\sigma \in \mathsf{d}\Sigma_{\mathsf{Min}}} \mathsf{dVal}^{\sigma}(v) \qquad \text{and} \qquad \underline{\mathsf{dVal}}(v) = \sup_{\tau \in \mathsf{d}\Sigma_{\mathsf{Max}}} \mathsf{dVal}^{\tau}(v)$$

We may easily show that $\underline{\mathsf{dVal}}(v) \leqslant \overline{\mathsf{dVal}}(v)$ for all initial vertices v. In [3, Theorem 1], shortest-path games are shown to be determined when both players use deterministic strategies, i.e. $\underline{\mathsf{dVal}}(v) = \overline{\mathsf{dVal}}(v)$. We thus denote $\mathsf{dVal}(v)$ this common value. We say that deterministic strategies $\sigma^\star$ of Min and $\tau^\star$ of Max are optimal (respectively, ε-optimal for a positive real number ε) if, for all vertices v: $\mathsf{dVal}^{\sigma^\star}(v) = \mathsf{dVal}(v)$ and $\mathsf{dVal}^{\tau^\star}(v) = \mathsf{dVal}(v)$ (respectively, $\mathsf{dVal}^{\sigma^\star}(v) \leqslant \mathsf{dVal}(v) + \varepsilon$ and $\mathsf{dVal}^{\tau^\star}(v) \geqslant \mathsf{dVal}(v) - \varepsilon$).

▶ **Example 1.** The deterministic value of the game on the left of Figure 1 is described in the introduction: $\mathsf{dVal}(v_{\mathsf{Min}}) = \mathsf{dVal}(v_{\mathsf{Max}}) = -10$. An optimal strategy for player Min consists in going to v_{Max} the first 10 times, and switching to the target vertex afterwards. An optimal strategy for player Max consists in directly going towards the target vertex.
If we remove the edge from v_{Max} to the target (of weight -10), we obtain another game in which $\mathsf{dVal}(v_{\mathsf{Min}}) = \mathsf{dVal}(v_{\mathsf{Max}}) = -\infty$ since Min can decide to turn as long as he wants in the negative cycle, before switching to the target. There is no optimal strategy for Min but a sequence of strategies guaranteeing a value as low as we want.

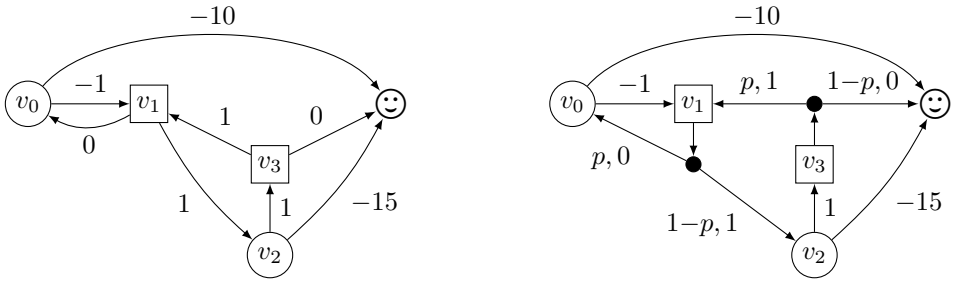

Figure 2 On the left, a more complex example of shortest-path game. On the right, the MDP associated with a randomised strategy of Min with a parametric probability $p \in (0, 1)$.

2.2 Memoryless strategies

Definitions above can be adapted for memoryless (randomised) strategies. In order to keep the explanations simple, we only define the upper value above, without relying on hypothetical determinacy results in this context. Once we fix a memoryless (randomised) strategy $\rho \in \mathsf{m}\Sigma_{\mathsf{Min}}$, we obtain a *Markov decision process* (MDP) where the other player must still choose how to react. An MDP is a tuple $\langle V, A, P \rangle$ where V is a set of vertices, A is a set of actions, and $P \colon V \times A \to \Delta(V)$ is a partial function mapping to some pair of vertices and actions a distribution of probabilities over the successor vertices. In our context, we let $\mathcal{G}^\rho$ be the MDP with the same set V of vertices as $\mathcal{G}$, actions $A = V \cup \{\bot\}$ being either successor vertices of the game or an additional action $\bot$ denoting the random choice of ρ, and a probability distribution P defined by:

- if $v \in V_{\mathsf{Max}}$, $P(v, v')$ is only defined if $(v, v') \in E$ in which case $P(v, v') = \mathsf{Dirac}_{v'}$, and $P(v, \bot)$ is also undefined;
- if $v \in V_{\mathsf{Min}}$, $P(v, \bot) = \rho(v)$, and $P(v, v')$ is undefined for all $v' \in V$.

In drawings of MDPs (and also of Markov chains, later), we show weights as trivially transferred from the game graph.

▶ **Example 2.** In Figure 1, a shortest-path game is presented on the left, with the MDP in the middle obtained by picking as a memoryless strategy for Min the one choosing to go to v_{Max} with probability $p \in (0, 1)$ and to the target vertex with probability $1 - p$. Another more complex example is given in Figure 2 where the memoryless strategy for Min consists, in vertex v_1, to choose successor v_0 with probability $p \in (0, 1)$ and successor v_2 with probability $1 - p$, and in vertex v_3, to choose successor v_1 with the same probability p and the target vertex with probability $1 - p$.

In such an MDP, when player Max has chosen her strategy, there will remain no "choices" to make, and we will thus end up in a *Markov chain*. A Markov chain (*MC*) is a tuple $\mathcal{M} = \langle V, P \rangle$ where V is a set of vertices, and $P \colon V \to \Delta(V)$ associates to each vertex a distribution of probabilities over the successor vertices. In our context, for all memoryless strategies $\chi \in \mathsf{m}\Sigma_{\mathsf{Max}}$, we let $\mathcal{G}^{\rho,\chi}$ the MC obtained from the MDP $\mathcal{G}^\rho$ by following strategy χ and action $\bot$. Formally, it consists of the same set V of vertices as $\mathcal{G}$, and mapping P associating to a vertex $v \in V_{\mathsf{Min}}$, $P(v) = \rho(v)$ and to a vertex $v \in V_{\mathsf{Max}}$, $P(v) = \chi(v)$.

▶ **Example 3.** On the right of Figure 1 is depicted the MC obtained when Max decides to go to v_{Min} with probability $q \in [0, 1]$ and to the target vertex with probability $1 - q$.

When starting in a given initial vertex v, we let $\mathbb{P}_v^{\rho,\chi}$ denote the induced probability measure over the sets of paths in the MC $\mathcal{G}^{\rho,\chi}$ (as before, $\mathcal{G}$ is made implicit in the notation). A *property* is any measurable subset of finite or infinite paths in the MC with respect to the

standard cylindrical sigma-algebra. For instance, we denote by $\mathbb{P}_v^{\rho,\chi}(\diamond T)$ the probability of the set of plays that reach the target set $T \subseteq V$ of vertices. Given a random variable X over the infinite paths in the MC, we let $\mathbb{E}_v^{\rho,\chi}(X)$ be the expectation of X with respect to the probability measure $\mathbb{P}_v^{\rho,\chi}$. Therefore, $\mathbb{E}_v^{\rho,\chi}(\mathbf{TP})$ is the expected weight of a path in the MC, weights being the ones taken from $\mathcal{G}$.

The objective of Max is to maximise the payoff in the MDP $\mathcal{G}^\rho$. We therefore define the value of strategy ρ of Min as the best case scenario for Max:

$$\mathsf{mVal}^\rho(v) = \sup_{\chi \in \mathsf{m}\Sigma_{\mathsf{Max}}} \mathbb{E}_v^{\rho,\chi}(\mathbf{TP})$$

By [1, Section 10.5.1], the value $\mathsf{mVal}^\rho(v)$ is finite if and only if $\mathbb{P}_v^{\rho,\chi}(\diamond T) = 1$ for all χ, i.e. if strategy ρ ensures the reachability of a target vertex with probability 1, no matter how the opponent plays. In this case, letting P be the probability mapping defining the MC $\mathcal{G}^{\rho,\chi}$, the vector $(\mathbb{E}_v^{\rho,\chi}(\mathbf{TP}))_{v\in V}$ is the only solution of the system of equations

$$\mathbb{E}_v^{\rho,\chi}(\mathbf{TP}) = \begin{cases} 0 & \text{if } v \in T \\ \sum_{v' \in E(v)} P(v,v') \times (\omega(v,v') + \mathbb{E}_{v'}^{\rho,\chi}(\mathbf{TP})) & \text{if } v \notin T \end{cases} \tag{1}$$

Since Min wants to minimise the shortest-path payoff, we finally define the memoryless upper value as

$$\overline{\mathsf{mVal}}(v) = \inf_{\rho \in \mathsf{m}\Sigma_{\mathsf{Min}}} \mathsf{mVal}^\rho(v)$$

Once again, we say that a memoryless strategy ρ is optimal (respectively, ε-optimal for a positive real number ε) if $\mathsf{mVal}^\rho(v) = \overline{\mathsf{mVal}}(v)$ (respectively, $\mathsf{mVal}^\rho(v) \leqslant \overline{\mathsf{mVal}}(v) + \varepsilon$). With respect to player Max, we only consider optimality and ε-optimality in the MDP $\mathcal{G}^\rho$.

▶ **Example 4.** For the game of Figure 1, we let σ and τ the memoryless strategies that result in the MC on the right. Letting $x = \mathbb{E}_{v_{\mathsf{Min}}}^{\rho,\chi}(\mathbf{TP})$ and $y = \mathbb{E}_{v_{\mathsf{Max}}}^{\rho,\chi}(\mathbf{TP})$, the system (1) rewrites as $x = (1-p) \times 0 + p \times y$ and $y = q \times (-1 + x) + (1-q) \times (-10)$. We thus have $x = p(9q-10)/(1-pq)$. Two cases happen, depending on the value of p: if $p < 9/10$, then Max maximises x by choosing $q = 1$, while she chooses $q = 0$ when $p \geqslant 9/10$. In all cases, player Max will therefore play deterministically: if $p < 9/10$, the expected payoff from v_{Min} will then be $\mathsf{mVal}^\rho(v_{\mathsf{Min}}) = -p/(1-p)$; if $p \geqslant 9/10$, it will be $\mathsf{mVal}^\rho(v_{\mathsf{Min}}) = -10p$. This value is always greater than the optimum -10 that Min were able to achieve with memory, since we must keep $1 - p > 0$ to ensure reaching the target with probability 1. We thus obtain $\overline{\mathsf{mVal}}(v_{\mathsf{Min}}) = \overline{\mathsf{mVal}}(v_{\mathsf{Max}}) = -10$ as before. There are no optimal strategies for Min, but an ε-optimal one consisting in choosing probability $p \geqslant 1 - \varepsilon/10$.

The fact that Max can play optimally with a deterministic strategy in the MDP $\mathcal{G}^\rho$ is not specific to this example. Indeed, in an MDP $\mathcal{G}^\rho$ such that $\mathbb{P}_v^{\rho,\chi}(\diamond T) = 1$ for all χ, Max cannot avoid reaching the target: she must then ensure the most expensive play possible. Considering the MDP $\tilde{\mathcal{G}}^\rho$ obtained by multiplying all the weights in the graph by -1, the objective of Max becomes a shortest-path objective. We can then deduce from [2] that she has an optimal deterministic memoryless strategy: the same applies in the original MDP $\mathcal{G}^\rho$.

▶ **Proposition 5.** *In the MDP $\mathcal{G}^\rho$ such that $\mathbb{P}_v^{\rho,\chi}(\diamond T) = 1$ for all χ,* Max *has an optimal deterministic memoryless strategy.*

2.3 Contribution

Our contribution consists in showing that optimal values are the same when restricting both players to memoryless or deterministic strategies:

▶ **Theorem 6.** *For all games $\mathcal{G}$ with a shortest-path objective, for all vertices v, we have* $\mathsf{dVal}(v) = \overline{\mathsf{mVal}}(v)$.

We show this theorem in the two next sections by a simulation of deterministic strategies with memoryless ones, and vice versa. We start here by ruling out the case of values $+\infty$. Indeed, $\mathsf{dVal}(v) = +\infty$ signifies that Min is not able to reach a target vertex from v with deterministic strategies. This also implies that Min has no memoryless randomised strategies to ensure reaching the target with probability 1, and thus $\overline{\mathsf{mVal}}(v) = +\infty$. Reciprocally, if $\overline{\mathsf{mVal}}(v) = +\infty$, then Min has no memoryless strategies to reach the target with probability 1 (since this is the only reason for having a value $+\infty$). Since reachability is a purely qualitative objective, and the game graph does not contain probabilities, Min cannot use memory in order to guarantee reaching the target: therefore, this also means that $\mathsf{dVal}(v) = +\infty$. In the end, we have shown that $\mathsf{dVal}(v) = +\infty$ if and only if $\overline{\mathsf{mVal}}(v) = +\infty$. We thus remove every such vertex from now on, which does not change the values of other vertices in the game.

▶ **Assumption.** From now on, all games $\mathcal{G}$ with a shortest-path objective are such that $\mathsf{dVal}(v)$ and $\overline{\mathsf{mVal}}(v)$ are different from $+\infty$, for all vertices v.

3 Simulating deterministic strategies with memoryless strategies

Towards proving Theorem 6, we show in this section that, for all shortest-path games $\mathcal{G} = \langle V, E, \omega, \mathbf{P}\rangle$ (where no values are $+\infty$) and vertices $v \in V$, $\overline{\mathsf{mVal}}(v) \leqslant \mathsf{dVal}(v)$. This is done by considering the *switching strategies* originated from [3], which are a particular kind of deterministic strategies: they are optimal from vertices of finite value, and they can get a value as low as wanted from vertices of value $-\infty$. A switching strategy $\sigma = \langle \sigma_1, \sigma_2, \alpha\rangle$ is described by two deterministic memoryless strategies σ_1 and σ_2, as well as a switching parameter α. The strategy σ consists in playing along σ_1, until eventually switching to σ_2 when the length of the current finite play is greater than α. Strategy σ_2 is thus any *attractor strategy* ensuring that plays reach the target set of vertices: it can be computed via a classical attractor computation. Strategy σ_1 is chosen so that every cyclic finite play $v_0v_1\dots v_kv_0$ conforming to σ_1 has a negative total weight: this is called an *NC-strategy* (for *negative-cycle-strategy*) in [3]. The *fake-value* of σ_1 from a vertex v_0 is defined by $\mathsf{fake}^{\sigma_1}(v_0) = \sup\{\mathbf{TP}(v_0v_1\cdots v_k) \mid v_k \in T, v_0v_1\cdots v_k \text{ conforming to } \sigma_1\}$, letting $\sup\emptyset = -\infty$: it consists of only considering plays conforming σ_1 that reach the target. Strategy σ_1 is said to be *fake-optimal* if $\mathsf{fake}^{\sigma_1}(v) \leqslant \mathsf{dVal}(v)$ for all vertices v: in this case, if a play from v conforms to σ_1 (or σ before the switch happens) and reaches the target set of vertices, it has a weight at most $\mathsf{dVal}(v)$.

▶ **Proposition 7** ([3]). *There exists a fake-optimal NC-strategy σ_1. Moreover, for all such fake-optimal NC-strategies σ_1, for all attractor strategies σ_2, and for all $n \in \mathbb{N}$, the switching parameter $\alpha = (2W(|V|-1)+n)|V|+1$ defines a switching strategy $\sigma = \langle\sigma_1, \sigma_2, \alpha\rangle$ with a value $\mathsf{dVal}^\sigma(v) \leqslant \max(-n, \mathsf{dVal}(v))$, from all initial vertices $v \in V$.*

In particular, if $\mathsf{dVal}(v)$ is finite, for n large enough, the switching strategy is optimal. If $\mathsf{dVal}(v) = -\infty$ however, the sequence $(\sigma^n)_{n\in\mathbb{N}}$ of strategies, each with a different parameter n, has a value that tends to $-\infty$.

▶ **Example 8.** For all $n \in \mathbb{N}$, let $\sigma = (\sigma_1, \sigma_2, \alpha)$ the switching strategy described above. In Figure 1, we have $\sigma_1(v_{\mathsf{Min}}) = v_{\mathsf{Max}}$, $\sigma_2(v_{\mathsf{Min}}) = \☺$ and $\alpha = 3(40+n)+1$. In Figure 2, σ_1 chooses v_0 from v_1 and v_1 from v_3, σ_2 chooses v_2 from v_1 and ☺ from v_3 and $\alpha = 5(60+n)+1$, for all $n \in \mathbb{N}$.

Definition of a memoryless (randomised) strategy. Let $n \in \mathbb{N}$, we consider the switching strategy $\sigma = \langle \sigma_1, \sigma_2, \alpha \rangle$ described before, of value $\mathsf{dVal}^\sigma(v) \leqslant \max(-n, \mathsf{dVal}(v))$, and simulate it with a memoryless (randomised) strategy for Min, denoted ρ_p, with a parametrised probability $p \in (0,1)$. This new strategy is a probabilistic superposition of the two memoryless deterministic strategies σ_1 and σ_2.

Formally, we define ρ_p on each strongly connected components (SCC) of the graph according to the presence of a negative cycle. In an SCC that does not contain negative cycles, for each vertex $v \in V_{\mathsf{Min}}$ of the SCC, we let $\rho_p(v) = \mathsf{Dirac}_{\sigma_1(v)}$: player Min chooses to play the first strategy σ_1 of the switching strategy, thus looking for a negative cycle in the next SCCs (in topological order) if any. In an SCC that contains a negative cycle, for each vertex $v \in V_{\mathsf{Min}}$ of the SCC, we let $\rho_p(v)$ be the distribution of support $\{\sigma_1(v), \sigma_2(v)\}$ that chooses $\sigma_1(v)$ with probability p and $\sigma_2(v)$ with probability $1-p$, except if $\sigma_1(v) = \sigma_2(v)$ in which case we choose it with probability 1. Note that MDPs in Figures 1 and 2 are obtained by applying this strategy ρ_p.

We fix some vertex $v_0 \in V$. In the rest of this section, we prove the following result:

▶ **Proposition 9.** *For ε small enough and p close enough to* 1, $\mathsf{mVal}^{\rho_p,\tau}(v_0) \leqslant \mathsf{dVal}^\sigma(v_0) + \varepsilon$.

This entails the expected result. Indeed, if $\mathsf{dVal}(v_0) \in \mathbb{Z}$, we get (with $n = |\mathsf{dVal}(v_0)|$) that $\mathsf{mVal}^{\rho_p}(v_0) \leqslant \mathsf{dVal}(v_0) + \varepsilon$, and thus $\overline{\mathsf{mVal}}(v_0) \leqslant \mathsf{dVal}(v_0)$ since this holds for all $\varepsilon > 0$. Otherwise, $\mathsf{dVal}(v_0) = -\infty$, and letting n tend towards $+\infty$, we also get $\overline{\mathsf{mVal}}(v_0) = -\infty$.

We first prove that ρ_p is one of the strategies of Min that guarantee to reach the target with probability 1 in the MDP $\mathcal{G}^{\rho_p}$ no matter how Max reacts.

▶ **Proposition 10.** *For all strategies $\chi \in \mathsf{m}\Sigma_{\mathsf{Max}}$, $\mathbb{P}_{v_0}^{\rho_p,\chi}(\Diamond\, T) = 1$.*

Proof. Recall that we designed our graph games so that target vertices are the only deadlocks. Thus, by using the characterisation of [1, Lemma 10.111], $\min_{\chi \in \mathsf{m}\Sigma_{\mathsf{Max}}} \mathbb{P}_{v_0}^{\rho_p,\chi}(\Diamond\, T) = 1$ if and only if for all $\chi \in \mathsf{m}\Sigma_{\mathsf{Max}}$, all bottom SCCs of the MC $\mathcal{G}^{\rho_p,\chi}$ (the ones from which we cannot exit) consist in a unique target vertex. Suppose in the contrary that Max has a memoryless strategy χ such that the MC $\mathcal{G}^{\rho_p,\chi}$ contains a bottom SCC $\mathcal{C}$ with no target vertices.

If all vertices of $\mathcal{C}$ belong to Max, then they all have a successor in $\mathcal{C}$ and therefore there also exists a deterministic memoryless strategy τ' for which all vertices $v \in \mathcal{C}$ are such that $\mathsf{dVal}^{\tau'}(v) = +\infty$, and thus $\mathsf{dVal}(v) = +\infty$: this contradicts our hypothesis that all vertices have a deterministic value different from $+\infty$.

Otherwise, for all vertices $v \in V_{\mathsf{Min}} \cap \mathcal{C}$, since $\mathcal{C}$ is a bottom SCC of $\mathcal{G}^{\rho_p,\chi}$, the distribution $\rho_p(v)$ has its support included in $\mathcal{C}$. If $\mathcal{C}$ is included in a SCC of $\mathcal{G}$ with no negative cycles, $\mathrm{supp}(\rho_p(v)) = \{\sigma_1(v)\}$: playing $\sigma_1(v)$ in $\mathcal{C}$ will end up in a cycle (since there are no deadlocks) that must be negative, by the hypothesis on σ_1, which is impossible. Thus, $\mathcal{C}$ must be included in an SCC of $\mathcal{G}$ with a negative cycle. Then, $\mathrm{supp}(\rho_p(v)) = \{\sigma_1(v), \sigma_2(v)\} \subseteq \mathcal{C}$, and in particular the attractor strategy is not able to reach a target vertex: playing the deterministic switching strategy σ will result in not reaching a target vertex either, so that $\mathsf{dVal}(v) = +\infty$ for $v \in V_{\mathsf{Min}} \cap \mathcal{C}$, which also contradicts our hypothesis. ◀

We can therefore apply Proposition 5. This result is very helpful since it allows us to only consider deterministic memoryless strategies τ to compute $\mathsf{mVal}^{\rho_p}(v_0) = \sup_\tau \mathsf{mVal}^{\rho_p,\tau}(v_0)$, for all initial vertices v_0. We thus consider such a strategy τ and we now show that

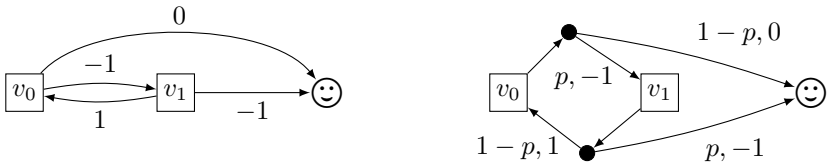

Figure 3 On the left, a game graph with no negative cycles where ρ_p is optimal. The MC obtained when playing a different randomised memoryless strategy.

$\mathsf{mVal}^{\rho_p,\tau}(v_0) \leqslant \mathsf{dVal}^{\sigma}(v) + \varepsilon$ whenever $p < 1$ is close enough to 1 (in function of $\varepsilon > 0$). By gathering the finite number of lower bounds about p, for all deterministic memoryless strategies of Max (there are a finite number of such), we obtain a lower bound for p such that $\mathsf{mVal}^{\rho_p}(v_0) \leqslant \mathsf{dVal}^{\sigma}(v_0) + \varepsilon$, as expected to prove Proposition 9.

The case where the whole game graph does not contain any negative cycles is easy. In this case, ρ_p chooses the strategy σ_1 with probability 1, by definition since no SCC contain a negative cycle (this is the only reason why we defined ρ_p as it is, for such SCCs): a play from initial vertex v_0 conforming to ρ_p is thus conforming to σ_1. Since the graph contains no negative cycles and all cycles conforming to σ_1 must be negative, all plays from v_0 conforming to σ_1 reach the target set of vertices, with a total payoff at most $\mathsf{dVal}^{\sigma}(v_0)$. This single play has probability 1 in the MC $\mathcal{G}^{\rho_p,\tau}$, thus $\mathbb{E}_{v_0}^{\rho_p,\tau}(\mathbf{TP}) \leqslant \mathsf{dVal}^{\sigma}(v_0)$, which proves that $\mathsf{mVal}^{\rho_p}(v) \leqslant \mathsf{dVal}^{\sigma}(v_0)$ as expected.

▶ **Example 11.** If the definition of ρ_p would not distinguish the SCCs with no negative cycles from the other SCCs, we would not have the optimality of ρ_p as shown before. Indeed, consider the game graph on the left of Figure 3, which has no negative cycles. We have $\mathsf{dVal}(v_0) = -2$ and $\mathsf{dVal}(v_1) = -1$. As a switching strategy, we can choose $\sigma_1(v_0) = v_1$, $\sigma_1(v_1) = ☺$, $\sigma_2(v_0) = ☺$ and $\sigma_2(v_1) = v_0$. Then, ρ_p is equal to σ_1 (and thus independent of p), and $\mathsf{mVal}^{\rho_p}(v_0) = -2$ and $\mathsf{mVal}^{\rho_p}(v_1) = -1$. However, if we would have chosen to still mix σ_1 and σ_2, we would obtain a strategy ρ'_p, and the MC on the right of Figure 3. Then, we get $\mathsf{mVal}^{\rho'_p}(v_0) = -2p^2/(1-p(1-p))$ and $\mathsf{mVal}^{\rho'_p}(v_1) = (p^2-3p+1)/(1-p(1-p))$ whose limits are -2 and -1 respectively, when p tends to 1. This strategy ρ'_p would then still be ε-optimal for p close enough to 1.

Now, suppose that the graph game contains negative cycles. We let $c > 0$ be the maximal size of an elementary cycle (that visits a vertex at most once) in $\mathcal{G}$, $w^- > 0$ be the opposite of the maximal weight of an elementary negative cycle in $\mathcal{G}$, and $w^+ \geqslant 0$ be the maximal weight of an elementary non-negative cycle in $\mathcal{G}$ (or 0 if such cycle does not exist).

▶ **Example 12.** In the graph of Figure 1, we have $c = 2$, $w^- = 1$, and $w^+ = 0$ (since there is no non-negative cycles). In the game graph of Figure 2, we have $c = 3$, $w^- = 1$, and $w^+ = 3$.

The difficulty initiates from the possible presence of non-negative cycles too. Indeed, when applying the switching strategy σ, all cycles conforming to σ_1 have a negative weight. This is no longer true with the probabilistic superposition ρ_p, as can be seen in the example of Figure 2. Finding an adequate lower-bound for p requires to estimate $\mathbb{E}_{v_0}^{\rho_p,\tau}(\mathbf{TP})$, by controlling the weight and probability of non-negative cycles, balancing them with the ones of negative cycles. The crucial argument comes from the definition of the superposition ρ_p:

▶ **Lemma 13.** *All cycles in $\mathcal{G}^{\rho_p,\tau}$ of non-negative total weight contain at least one edge of probability $1-p$.*

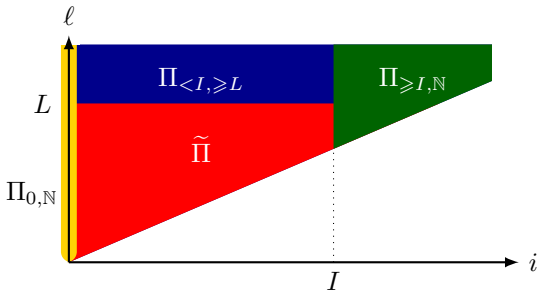

Figure 4 Partition of plays Π.

Proof. Suppose on the contrary that all edges have probability p or 1, then the cycle is conforming to strategy σ_1, and has therefore a negative weight. ◀

Proof of Proposition 9. The proof that $\mathsf{mVal}^{\rho_p,\tau}(v_0) \leqslant \mathsf{dVal}^{\sigma}(v_0)+\varepsilon$ is done by partitioning the set Π of plays starting in v_0, conforming to ρ_p and τ, and reaching the target set of vertices, into subsets $\Pi_{i,\ell}$ according to the number i of edges of probability $1-p$ they go through, and their length ℓ (we always have $i \leqslant \ell$). The partition is depicted in Figure 4:

- $\Pi_{0,\mathbb{N}}$, depicted in yellow, contains all plays with no edges of probability $1-p$;
- $\Pi_{\geqslant I,\mathbb{N}}$, depicted in green, contains all plays having at least

$$I = \left\lceil \frac{2w^+}{\gamma W} + \frac{8(w^+ + |V|W)}{\varepsilon} \right\rceil$$

  edges[1] of probability $1-p$ where $\gamma = c\left(1+\frac{w^+}{w^-}\right) \geqslant 1$;
- $\Pi_{<I,\geqslant L}$, depicted in blue, contains all plays with at most I edges of probability $1-p$, and of length at least $L = I\gamma + \frac{2|\mathsf{dVal}^{\sigma}(v_0)|+|V|W}{w^-}c + |V|$;
- $\widetilde{\Pi}$, depicted in red, is the rest of the plays.

We let $\gamma_{0,\mathbb{N}}$ (respectively, $\gamma_{<I,\geqslant L}$, $\gamma_{\geqslant I,\mathbb{N}}$, and $\widetilde{\gamma}$) be the expectation $\mathbb{E}_{v_0}^{\rho_p,\tau}(\mathbf{TP})$ restricted to plays in $\Pi_{0,\mathbb{N}}$ (respectively, $\Pi_{<I,\geqslant L}$, $\Pi_{\geqslant I,\mathbb{N}}$, $\widetilde{\Pi}$). By linearity of expectation,

$$\mathsf{mVal}^{\rho_p,\tau}(v_0) = \mathbb{E}_{v_0}^{\rho_p,\tau}(\mathbf{TP}) = \gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} + \gamma_{\geqslant I,\mathbb{N}} + \widetilde{\gamma} \tag{2}$$

Partitioning the plays allows us to carefully control non-negative cycles: plays with a large number of non-negative cycles contain a large number of edges of probability $1-p$, by Lemma 13; thus if p is made close enough to 1, the probability of this set of plays will be small enough. We thus control separately the four terms of (2) to obtain $\mathsf{mVal}^{\rho_p,\tau}(v_0) \leqslant \mathsf{dVal}^{\sigma}(v_0) + \varepsilon$.

Yellow and blue zones are such that $\gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} \leqslant \mathsf{dVal}^{\sigma}(v_0) + \varepsilon/2$

All plays of $\Pi_{0,\mathbb{N}}$ reach the target without edges of probability $1-p$, i.e. by conforming to σ_1. By fake-optimality of σ_1, their total payoff is upper-bounded by $\mathsf{dVal}^{\sigma}(v_0)$. Notice that, in case $\mathsf{dVal}(v_0) = -\infty$, no plays conforming to σ_1 starting in v_0 reach the target, since Min has the opportunity to stay as long as he wants in negative cycles: thus $\Pi_{0,\mathbb{N}} = \emptyset$ in this case, and $\gamma_{0,\mathbb{N}} = 0$.

[1] This intricate definition of I, as well as L in the next item, is justified by the computations that will follow in the proof.

All plays of $\Pi_{i,\ell}$, with $1 \leqslant i < I$ and $\ell \geqslant L$, go through i edges of probability $1-p$. By Lemma 13, they contain at most i elementary cycles of non-negative total weight (each of weight at most w^+). The total length of these cycles is at most ic. Once we have removed these cycles from the play, it remains a play of length at least $\ell - ic$. By a repeated pumping argument, it still contains at least $\left\lfloor \frac{\ell - ic - |V|}{c} \right\rfloor$ elementary cycles, that have all a negative total weight (each has a weight at most $-w^-$). The remaining part, once removed the last negative cycles it contains, has length at most $|V|$, and thus a total payoff at most $|V|W$. In summary, the total payoff of a play in $\Pi_{i,\ell}$ is at most

$$\begin{aligned} iw^+ + \left\lfloor \frac{\ell - ic - |V|}{c} \right\rfloor (-w^-) + |V|W &\leqslant Iw^+ + \frac{L - Ic - |V|}{c}(-w^-) + |V|W \\ &= -2|\mathsf{dVal}^\sigma(v_0)| \leqslant 0 \end{aligned} \tag{3}$$

Let us then consider three cases.

- If $\mathsf{dVal}^\sigma(v_0) \geqslant 0$, we note that all plays in $\Pi_{<I,\geqslant L}$ have a non-positive total payoff, therefore at most $\mathsf{dVal}^\sigma(v_0)$. Thus,

$$\begin{aligned} \gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} &\leqslant \mathsf{dVal}^\sigma(v_0)\mathbb{P}(\Pi_{0,\mathbb{N}}) + \mathsf{dVal}^\sigma(v_0)\mathbb{P}(\Pi_{<I,\geqslant L}) \\ &= \mathsf{dVal}^\sigma(v_0)\big(\mathbb{P}(\Pi_{0,\mathbb{N}}) + \mathbb{P}(\Pi_{<I,\geqslant L})\big) \leqslant \mathsf{dVal}^\sigma(v_0) \end{aligned}$$

- If $\mathsf{dVal}^\sigma(v_0) < 0$ and $\Pi_{<I,\geqslant L} \neq \emptyset$, we have $\gamma_{0,\mathbb{N}} \leqslant 0$ (whatever $\mathsf{dVal}(v_0) = -\infty$ or not). Moreover, a play in $\Pi_{i,\ell}$ goes through i edges of probability $1-p$ and at most ℓ edges of probability p, other edges having probability 1. So, it has probability at least $(1-p)^i p^\ell$. We can deduce that

$$\gamma_{<I,\geqslant L} \leqslant \sum_{i=1}^{I-1}\sum_{\ell=L}^{\infty}(1-p)^i p^\ell \underbrace{\left(iw^+ + \left\lfloor \frac{\ell - ic - |V|}{c} \right\rfloor (-w^-) + |V|W\right)}_{\leqslant 0 \text{ by } (3)} \leqslant \mathsf{dVal}^\sigma(v_0)$$

  the last inequality being true when p is close enough to 1, as shown in Appendix A.
- If $\mathsf{dVal}^\sigma(v_0) < 0$ and $\Pi_{<I,\geqslant L} = \emptyset$, then $\mathsf{dVal}(v_0) \neq -\infty$, since otherwise a play conforming to strategy σ_1 for L rounds, and then switching to σ_2 for at most $|V| \leqslant I$ rounds, would be in $\Pi_{<I,\geqslant L}$. Thus, $\gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} = \gamma_{0,\mathbb{N}} \leqslant \mathsf{dVal}^\sigma(v_0)\mathbb{P}(\Pi_{0,\mathbb{N}})$. Moreover, by the same argument, all plays in $\Pi_{0,\mathbb{N}}$ are acyclic and their length is at most $|V|$: they go through no edges of probability $1-p$, and thus at most $|V|$ edges of probability p. Therefore, $\mathbb{P}(\Pi_{0,\mathbb{N}}) \geqslant p^{|V|}$, and thus, once again because $\mathsf{dVal}^\sigma(v_0) < 0$, when $p \geqslant (1 - \varepsilon/2|\mathsf{dVal}^\sigma(v_0)|)^{1/|V|}$ which is less than 1 for ε small enough,

$$\gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} \leqslant \mathsf{dVal}^\sigma(v_0)p^{|V|} \leqslant \mathsf{dVal}^\sigma(v_0) + \varepsilon/2$$

In all cases, we have $\gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} \leqslant \mathsf{dVal}^\sigma(v_0) + \varepsilon/2$.

Red and green zones are such that $\gamma_{\geqslant I,\mathbb{N}} + \widetilde{\gamma} \leqslant \varepsilon/2$

First, a play of $\Pi_{\geqslant I,\mathbb{N}}$ has a large total payoff, but a low probability to happen, which enables us to control its expected payoff. Indeed, consider a play of $\Pi_{i,\mathbb{N}}$, with $i \geqslant I$. By Lemma 13, it contains at most i elementary cycles of non-negative total weight. The remaining of the play may contain negative cycles, as well as an acyclic part reaching the target in at most $|V|$ steps. The total payoff of the whole play is thus at most $iw^+ + |V|W$. Moreover, $\mathbb{P}(\Pi_{i,\mathbb{N}}) \leqslant (1-p)^i$ since all the plays contain i edges of probability $1-p$. Overall,

$$\gamma_{\geqslant I,\mathbb{N}} \leqslant \sum_{i=I}^{\infty}(iw^+ + |V|W)(1-p)^i = (1-p)^I\left(\frac{w^+}{p}I + \frac{w^+(1-p)}{p^2} + \frac{|V|W}{p}\right) \leqslant \frac{\varepsilon}{4}$$

where the last inequality holds for p close enough to 1, as shown in Appendix A.

Finally, all plays of $\widetilde{\Pi}$ have a length less than L (and thus a total payoff at most LW) and a number i of edges of probability $1-p$ such that $0 < i < I$. By a similar argument as before, if $p \geqslant LW/(LW+\varepsilon/4)$, we have

$$\widetilde{\gamma} \leqslant \sum_{i=1}^{I} LW(1-p)^i = LW\frac{(1-p)(1-(1-p)^I)}{p} \leqslant LW\frac{1-p}{p} \leqslant \frac{\varepsilon}{4}$$

since $p \mapsto (1-p)/p$ is decreasing on $(0,1)$. ◀

This ends the proof that for all vertices v, $\overline{\mathsf{mVal}}(v) \leqslant \mathsf{dVal}(v)$. Let us illustrate the computation of the lower-bound on probability p of the memoryless strategy ρ_p in the previously studied examples.

▶ **Example 14.** For the game in Figure 1, with initial vertex v_{Min}, we have $\gamma = 2$. For $\varepsilon = 0.1$, we then have $I = 2400$, and $L = 4903$. The lower-bound on p is then $q = 0.9999995$, which gives a value $\mathsf{mVal}^{\rho_p}(v_{\mathsf{Min}}) = -10p = -9.999995$. For the game in Figure 2, with initial vertex v_2, we have $\gamma = 12$. For $\varepsilon = 0.1$, we then have $I = 3121$, and $L = 37730$. The lower-bound on p is then $q = 0.99999998$, which gives a value $\mathsf{mVal}^{\rho_p}(v_2) \approx -7.9999996$. We see that the lower-bound are correct, even if they could certainly be made coarser.

4 Simulating memoryless strategies with deterministic strategies

To finish the proof of Theorem 6, we will show that $\mathsf{dVal}(v) \leqslant \overline{\mathsf{mVal}}(v)$, for all vertices v. For a given memoryless strategy ρ ensuring that Min reaches the target set T with probability 1, we build a deterministic strategy σ which guarantees a value $\mathsf{dVal}^\sigma(v) \leqslant \mathsf{mVal}^\rho(v)$ from vertex v. Then, as in the previous section, if $\overline{\mathsf{mVal}}(v)$ is finite, for an ε-optimal memoryless strategy ρ, we get a deterministic strategy such that $\mathsf{dVal}^\sigma(v) \leqslant \overline{\mathsf{mVal}}(v) + \varepsilon$, and thus $\mathsf{dVal}(v) \leqslant \overline{\mathsf{mVal}}(v) + \varepsilon$. We can conclude since this holds for all $\varepsilon > 0$. In case $\overline{\mathsf{mVal}}(v) = -\infty$, if ρ guarantees a value at most $-n$ with $n \in \mathbb{N}$, then so does the deterministic strategy σ, which also ensures that $\mathsf{dVal}(v) = -\infty$.

We fix a memoryless strategy ρ, and an initial vertex v_0. The first attempt to build a deterministic strategy σ such that $\mathsf{dVal}^\sigma(v) \leqslant \overline{\mathsf{mVal}}(v) + \varepsilon$ would be to use classical techniques of finite-memory strategies, for instance in Street or Müller games: for instance, to ensure the visit of two vertices v_1 and v_2 infinitely often during an infinite play (to win a Müller game with winning objective $\{v_1, v_2\}$), we would try to reach v_1 with a first memoryless strategy, and then reach v_2 with another memoryless strategy, before switching again to reach v_1 again, etc.

▶ **Example 15.** Let us try this technique on the shortest-path game of Figure 1. We consider as a starting point the memoryless strategy ρ such that $\rho(v_{\mathsf{Min}}) = \delta$ with $\delta(☺) = 2/3$ and $\delta(v_{\mathsf{Max}}) = 1/3$ (this is the case $p = 1/3$ in the MDP on the middle of Figure 1). As seen in Example 4, this strategy has value $\mathsf{mVal}^\rho(v_{\mathsf{Min}}) = -1/2$ et $\mathsf{mVal}^\rho(v_{\mathsf{Max}}) = -3/2$. Naively, we could try to mimic the distribution δ by using memory as follows: when in v_{Min}, go to ☺ two thirds of the time and to v_{Min} one third of the time. Moreover, we would naively try to follow first the choice with greatest probability. In this case, the strategy σ would first choose to go to ☹, thus stopping immediately the play. We thus get $\mathsf{dVal}^\sigma(v_{\mathsf{Min}}) = 0 > -1/2 + \varepsilon$ as soon as $\varepsilon < 1/2$.

The main reason why this naive approach fails is that the plays are essentially finite in shortest-path games. We thus cannot delay the choices and must carefully play as soon as the play starts. Instead, our solution is to define a switching strategy $\sigma = \langle \sigma_1, \sigma_2, \alpha \rangle$, with σ_2 any attractor strategy, and $\alpha = \max(0, |V|W - \mathsf{mVal}^\rho(v_0)) \times |V| + 1$.

▶ **Example 16** (Example 15 continued). In the game of Figure 1, the attractor strategy is $\sigma_2(v_{\mathsf{Min}}) = ☺$. We then choose $\sigma_1(v_{\mathsf{Min}})$ so as to minimise the immediate reward obtained by playing one turn and then getting the value ensured by ρ:

$$\sigma_1(v_{\mathsf{Min}}) = \operatorname{argmin}_{v' \in \{v_{\mathsf{Max}}, ☺\}} [w(v, v') + \mathsf{mVal}^\rho(v')] = v_{\mathsf{Max}}$$

For an appropriate choice of α, we thus recover the optimal switching strategy for this game.

In the rest of this section, we will detail how to define strategy σ_1 in general so as to obtain the following property:

▶ **Proposition 17.** *The switching strategy $\sigma = \langle \sigma_1, \sigma_2, \alpha \rangle$ built from the memoryless (randomised) strategy ρ satisfies* $\mathsf{dVal}^\sigma(v_0) \leqslant \mathsf{mVal}^\rho(v_0)$.

The construction of σ_1 is split in two parts. First, we restrict the possibilities for $\sigma_1(v)$ to a subset $\widetilde{E}(v)$ of $\mathrm{supp}(\rho(v))$ in (4): with respect to Example 15, this will forbid the use of edge $(v_{\mathsf{Min}}, ☺)$ in particular. The definition of $\sigma_1(v)$ is then given later in (7).

We restrict our attention to edges present in the MDP $\mathcal{G}^\rho$, and for each vertex $v \in V_{\mathsf{Min}}$, we let

$$\widetilde{E}(v) = \operatorname*{argmin}_{v' \in \mathrm{supp}(\rho(v))} [w(v, v') + \mathsf{mVal}^\rho(v')] \tag{4}$$

be the successors of v that minimise the expected value at horizon 1. We let $\widetilde{\mathcal{G}}$ be the game obtained from $\mathcal{G}$ by removing all edges (v, v') from a vertex $v \in V_{\mathsf{Min}}$ such that $v' \notin \widetilde{E}(v)$.

▶ **Lemma 18.** *(i) Each finite play of $\widetilde{\mathcal{G}}$ from a vertex v has a total payoff at most* $\mathsf{mVal}^\rho(v)$. *(ii) Each cycle in the game $\widetilde{\mathcal{G}}$ has a non-positive total weight.*

Proof. We prove the property (i) on finite plays π of $\widetilde{\mathcal{G}}$ by induction on the length of π, for all initial vertices v. If π has length 0, this means that $v \in T$, in which case $\mathbf{TP}(\pi) = 0 = \mathsf{mVal}^\rho(v)$. Consider then a play $\pi = v\pi'$ of length at least 1, with π' starting from v', so that $\mathbf{TP}(\pi) = \omega(v, v') + \mathbf{TP}(\pi')$. By induction hypothesis, $\mathbf{TP}(\pi') \leqslant \mathsf{mVal}^\rho(v')$, so that $\mathbf{TP}(\pi) \leqslant \omega(v, v') + \mathsf{mVal}^\rho(v')$.

Suppose first that $v \in V_{\mathsf{Max}}$. By Proposition 5, we know that Max can play optimally in the MDP $\mathcal{G}^\rho$ with a deterministic and memoryless strategy. For each possible deterministic and memoryless strategy τ of Max, we have $\mathsf{mVal}^\rho(u) \geqslant \mathbb{E}_u^{\rho,\tau}(\mathbf{TP})$ for all $u \in V_{\mathsf{Max}}$, and by the system (1) of equations, letting $u' = \tau(u)$, $\mathbb{E}_u^{\rho,\tau}(\mathbf{TP}) = \omega(u, u') + \mathbb{E}_{u'}^{\rho,\tau}(\mathbf{TP})$. We thus know that $\mathsf{mVal}^\rho(u) \geqslant \omega(u, u') + \mathbb{E}_{u'}^{\rho,\tau}(\mathbf{TP})$. By taking a maximum over all deterministic and memoryless strategies τ of Max, Proposition 5 ensures that

$$\forall u \in V_{\mathsf{Max}} \quad \forall u' \in E(u) \qquad \mathsf{mVal}^\rho(u) \geqslant \omega(u, u') + \mathsf{mVal}^\rho(u') \tag{5}$$

In particular, $\mathsf{mVal}^\rho(v) \geqslant \omega(v, v') + \mathsf{mVal}^\rho(v') \geqslant \mathbf{TP}(\pi)$.

If $v \in V_{\mathsf{Min}}$, then $v' \in \widetilde{E}(v)$ so that $\omega(v, v') + \mathsf{mVal}^\rho(v')$ is minimum over all possible successors $v' \in \mathrm{supp}(\rho(v))$. The system (1) of equations implies that, for an optimal strategy χ of Max,

$$\begin{aligned} \mathsf{mVal}^\rho(v) = \mathbb{E}_v^{\rho,\chi}(\mathbf{TP}) &= \sum_{v'' \in E(v)} P(v, v'') \times (\omega(v, v'') + \mathbb{E}_{v''}^{\rho,\chi}(\mathbf{TP})) \\ &= \sum_{v'' \in \mathrm{supp}(\rho(v))} P(v, v'') \times (\omega(v, v'') + \mathsf{mVal}^\rho(v'')) \geqslant \omega(v, v') + \mathsf{mVal}^\rho(v') \end{aligned} \tag{6}$$

so that we also get $\mathsf{mVal}^\rho(v) \geqslant \mathbf{TP}(\pi)$.

We then prove the property (ii) on cycles. Consider thus a cycle $v_1v_2\cdots v_kv_1$ of $\widetilde{\mathcal{G}}$, and let $\omega_1 = \omega(v_1,v_2), \omega_2 = \omega(v_2,v_3), \dots, \omega_k = \omega(v_k,v_1)$ be the sequence of weights of edges. We also let $v_{k+1} = v_1$. We show that $\omega_1+\omega_2+\cdots+\omega_k \leqslant 0$. Let $i \in \{1,2,\dots,k\}$. If $v_i \in V_{\mathsf{Max}}$, by (5), $\mathsf{mVal}^\rho(v_i) \geqslant \omega_i + \mathsf{mVal}^\rho(v_{i+1})$. If $v_i \in V_{\mathsf{Min}}$, by the reasoning applied above in (6), we also know that $\mathsf{mVal}^\rho(v_i) \geqslant \omega_i + \mathsf{mVal}^\rho(v_{i+1})$. By summing all the inequalities above, we get

$$\sum_{i=1}^{k} \mathsf{mVal}^\rho(v_i) \geqslant \sum_{i=1}^{k} \omega_i + \sum_{i=1}^{k} \mathsf{mVal}^\rho(v_i) \qquad \text{i.e.} \quad \omega_1+\omega_2+\cdots+\omega_k \leqslant 0$$ ◀

▶ **Example 19.** Consider again the game graph on the left of Figure 3, and the memoryless strategy ρ'_p giving rise to the MDP/MC on the right of Figure 3. Recall that $\mathsf{mVal}^{\rho'_p}(v_0) = -2p^2/(1-p(1-p))$ and $\mathsf{mVal}^{\rho'_p}(v_1) = (p^2-3p+1)/(1-p(1-p))$. Consider p close enough to 1 so that $\mathsf{mVal}^{\rho'_p}(v_0) \leqslant -3/2$ and $\mathsf{mVal}^{\rho'_p}(v_1) \leqslant -1/2$. Then, we have $\widetilde{E}(v_0) = \{v_1\}$ and $\widetilde{E}(v_1) = \{\text{☺}\}$. The corresponding game graph $\widetilde{\mathcal{G}}$ contains only edges (v_0,v_1) and $(v_1,\text{☺})$, and thus no cycles. The unique finite play from vertex v_0 has total-payoff $-2 \leqslant \mathsf{mVal}^{\rho'_p}(v_0)$. In particular, the only possible memoryless deterministic strategy σ_1 in $\widetilde{\mathcal{G}}$ is optimal in $\mathcal{G}$.

For each vertex v in the game, we let $d(v)$ be the distance (number of steps) of v to the target given by an attractor computation to the target in $\mathcal{G}^\rho$ (notice that this may be different from the distance given in the whole game graph, since some edges are taken with probability 0 in ρ, but still $d(v) < +\infty$ since ρ ensures to reach T with probability 1). We then let, for all vertices $v \in V_{\mathsf{Min}}$,

$$\sigma_1(v) = \operatorname*{argmin}_{v' \in \widetilde{E}(v)} d(v') \tag{7}$$

▶ **Example 20.** Consider once again the game graph of Figure 3, but with a new memoryless strategy ρ''_p defined by $\rho''_p(v_0) = \mathsf{Dirac}_{v_1}$ and $\rho''_p(v_1) = \delta$ such that $\delta(v_0) = 1-p$ and $\delta(\text{☺}) = p$, where $p \in (0,1)$. Then, we can check that $\mathsf{mVal}^{\rho''_p}(v_0) = -2$ and $\mathsf{mVal}^{\rho''_p}(v_1) = -1$. Thus, $\widetilde{E}(v_0) = \{v_1\}$ and $\widetilde{E}(v_1) = \{v_0, \text{☺}\}$. Not all memoryless deterministic strategies taken in $\widetilde{\mathcal{G}}$ are NC-strategies, since it contains the cycle $v_0v_1v_0$ of total weight 0. We thus apply the construction before, using the fact that $d(\text{☺}) = 0$, $d(v_1) = 1$ and $d(v_0) = 2$ (since the edge $(v_0,\text{☺})$ is not present in $\widetilde{\mathcal{G}}$). Thus, σ_1 is defined by $\sigma_1(v_0) = v_1$ and $\sigma_1(v_1) = \text{☺}$, and is indeed an NC-strategy.

▶ **Lemma 21.** *Strategy σ_1 is an NC-strategy, i.e. all cycles of $\widetilde{\mathcal{G}}$ conforming with σ_1 have a negative total weight.*

Proof. Let $v_1v_2\cdots v_kv_1$ be a cycle of $\widetilde{\mathcal{G}}$ that conforms to σ_1, with v_1 a vertex of minimal distance $d(v_1)$ among the ones of the cycle. We can choose v_1 such that it belongs to Min: otherwise, this would contradict the attractor computation in $\widetilde{\mathcal{G}}$. By Lemma 18(ii), its total weight is non-positive. Suppose that it is 0. Then, in the proof of Lemma 18(ii), all inequalities $\mathsf{mVal}^\rho(v_i) \geqslant \omega_i + \mathsf{mVal}^\rho(v_{i+1})$ are indeed equalities. In particular, $\mathsf{mVal}^\rho(v_1) = \omega_1 + \mathsf{mVal}^\rho(v_2)$. Since $v_2 \in \widetilde{E}(v_1)$, (6) ensures that all successors $v' \in \mathrm{supp}(\rho(v_1))$, $\mathsf{mVal}^\rho(v_1) = \omega(v_1,v') + \mathsf{mVal}^\rho(v')$. Since v_1 has minimal distance among all vertices of the cycle, it exists $v' \in \widetilde{E}(v_1)$ such that $d(v') = d(v_1) - 1$. But $d(v_2) \geqslant d(v_1) > d(v')$, which contradicts the choice of v_2 for $\sigma_1(v_1)$ in (7). ◀

Proof of Proposition 17. Let π be a play conforming to σ, from vertex v_0. Since σ is a switching strategy, it necessarily reaches T. If σ conforms with σ_1, by Lemma 18(i), it has a total-payoff $\mathbf{TP}(\pi) \leqslant \mathsf{mVal}^\rho(v_0)$. Otherwise, it is obtained by a switch, and is thus longer than

$\alpha = \max(0, |V|W - \mathsf{mVal}^\rho(v_0)) \times |V| + 1$. Then, it contains at least $\max(0, |V|W - \mathsf{mVal}^\rho(v_0))$ elementary cycles, before it switches to the attractor strategy σ_2. Once we remove the cycles, it remains a play of length at most $|V|$, and thus of total payoff at most $|V|W$. Since all cycles conforming to σ_1 have a total weight at most -1, by Lemma 21, $\mathbf{TP}(\pi)$ is at most $(-1) \times \max(0, |V|W - \mathsf{mVal}^\rho(v_0)) + |V|W \leqslant \mathsf{mVal}^\rho(v_0)$. ◀

This concludes the proof of Theorem 6.

5 Characterisation of optimality

All shortest-path games admit an optimal deterministic strategy for both players: however, as we have seen in Example 1, Min may require memory to play optimally. In this case, we also have seen in Example 4 that Min does not have an optimal memoryless (randomised) strategy: he only has ε-optimal ones, for all $\varepsilon > 0$. But some shortest-path games indeed admit optimal memoryless strategies for Min: the strategy ρ_p described in Section 3 is indeed optimal in graph games not containing negative cycles, for instance. In this final section, we characterise the shortest-path games in which Min admits an optimal memoryless strategy. For sure, Min does not have an optimal strategy if there is some vertex v of value $\mathsf{dVal}(v) = -\infty$.

▶ **Assumption.** In this last section, we therefore suppose that all shortest-path games are such that $\mathsf{dVal}(v) \neq -\infty$ for all vertices v.

We first recall the computations performed in [3] to compute values $\mathsf{dVal}(v)$. It consists of an iterated computation, called *value iteration* based on the operator $\mathcal{F} \colon (\mathbb{Z} \cup \{+\infty\})^V \to (\mathbb{Z} \cup \{+\infty\})^V$ defined for all $x = (x_v)_{v \in V} \in (\mathbb{Z} \cup \{+\infty\})^V$ and all vertices $v \in V$ by

$$\mathcal{F}(x)_v = \begin{cases} 0 & \text{if } v \in T \\ \min_{v' \in E(v)} (\omega(v, v') + x_{v'}) & \text{if } v \in V_{\mathsf{Min}} \\ \max_{v' \in E(v)} (\omega(v, v') + x_{v'}) & \text{if } v \in V_{\mathsf{Max}} \end{cases}$$

We let $f_v^{(0)} = 0$ if $v \in T$ and $+\infty$ otherwise. By monotony of $\mathcal{F}$, the sequence $(f^{(i)} = \mathcal{F}^i(f^{(0)}))_{i \in \mathbb{N}}$ is non-increasing. It is proved to be stationary, and convergent towards $(\mathsf{dVal}(v))_{v \in V}$, the smallest fixed-point of $\mathcal{F}$. The pseudo-polynomial complexity of solving shortest-path games comes from the fact that this sequence may becomes stationary after a pseudo-polynomial (and not polynomial) number of steps: the game of Figure 1 is one of the typical examples.

We introduce a new notion, being the most permissive strategy of Min at each step $i \geqslant 0$ of the computation. It maps each vertex $v \in V_{\mathsf{Min}}$ to the set

$$\widetilde{E}^{(i)}(v) = \{v' \in E(v) \mid \omega(v, v') + f_{v'}^{(i-1)} = f_v^{(i)}\}$$

of vertices that Min can choose. For each such most permissive strategy $\widetilde{E}^{(i)}$, we let $\widetilde{\mathcal{G}}^{(i)}$ be the game graph where we remove all edges (v, v') with $v \in V_{\mathsf{Min}}$ and $v' \notin \widetilde{E}^{(i)}(v)$. This allows us to state the following result, whose proof is in Appendix B.

▶ **Proposition 22.** *The following assertions are equivalent:*

1. Min *has an optimal memoryless deterministic strategy in* $\mathcal{G}$ *(for* dVal*);*
2. Min *has an optimal memoryless (randomised) strategy in* $\mathcal{G}$ *(for* $\overline{\mathsf{mVal}}$*);*

3. $f_v^{(|V|-1)} = f_v^{(|V|)} = \mathsf{dVal}(v)$ *for all vertices* v *(this means that the sequence* $(f^{(i)})$ *is stationary as soon as step* $|V|-1$*), and* Min *can guarantee to reach* T *from all vertices in the game graph* $\widetilde{\mathcal{G}}^{(|V|-1)}$.

This characterisation of the existence of optimal memoryless strategy is testable in polynomial time since it is enough to compute vectors $f^{(|V|-1)}$ and $f^{(|V|)}$, check their equality, compute the sets $\widetilde{E}^{(|V|-1)}(v)$ (this can be done while computing $f^{(|V|)}$) and check whether Min can guarantee reaching the target in $\widetilde{\mathcal{G}}^{(|V|-1)}$ by an attractor computation. The proof of implication $3 \Rightarrow 1$ is constructive and actually allows one to build an optimal memoryless deterministic strategy when it exists.

6 Discussion

This article studies the tradeoff between memoryless and deterministic strategies, showing that Min guarantees the same value when restricted to these two kinds of strategies. We also studied the existence of optimal memoryless strategies, which turns out to be equivalent to the existence of optimal memoryless deterministic strategies, and testable in polynomial time.

We could also define a more general lower and upper values $\underline{\mathsf{Val}}(v)/\overline{\mathsf{Val}}(v)$ when we let Min and Max play unrestricted strategies (randomised and with memory). The Blackwell determinacy results [12] implies that, for such unrestricted strategies, shortest-path games are still determined so that $\overline{\mathsf{Val}}(v) = \underline{\mathsf{Val}}(v) = \mathsf{Val}(v)$. The reasoning of Section 4 only used the vector of values $(\mathsf{mVal}^{\rho}(v))_{v\in V}$ to define the deterministic switching strategy σ, without using anywhere that ρ is memoryless. We thus indeed showed that $\mathsf{dVal}(v) \leqslant \mathsf{Val}(v)$. However, the proof of Section 3 is not directly translatable if we allow Min to use memory and randomisation. In particular, we know nothing anymore about how Max can react, which may break the result of Proposition 10. We leave this further study for future work.

References

1 Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT Press, 2008.

2 Dimitri P. Bertsekas and John N. Tsitsiklis. An analysis of stochastic shortest path problems. *Math. Oper. Res.*, 16(3):580–595, 1991.

3 Thomas Brihaye, Gilles Geeraerts, Axel Haddad, and Benjamin Monmege. Pseudopolynomial iterative algorithm to solve total-payoff games and min-cost reachability games. *Acta Informatica*, 54, July 2016.

4 Thomas Brihaye, Gilles Geeraerts, Axel Haddad, and Benjamin Monmege. Pseudopolynomial iterative algorithm to solve total-payoff games and min-cost reachability games. *Acta Informatica*, 54(1):85–125, February 2017. doi:10.1007/s00236-016-0276-z.

5 Krishnendu Chatterjee, Luca de Alfaro, and Thomas A. Henzinger. Trading memory for randomness. In *Proceedings of the The Quantitative Evaluation of Systems, First International Conference*, QEST '04, pages 206–217, Washington, DC, USA, 2004. IEEE Computer Society. URL: http://dl.acm.org/citation.cfm?id=1025129.1026090.

6 Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdziński. Mean-payoff parity games. In *Proceedings of the 20th Annual Symposium on Logic in Computer Science (LICS'05)*, pages 178–187. IEEE Computer Society Press, 2005.

7 Krishnendu Chatterjee, Thomas A. Henzinger, and Vinayak S. Prabhu. Trading infinite memory for uniform randomness in timed games. In *Hybrid Systems: Computation and Control, 11th International Workshop, HSCC 2008, St. Louis, MO, USA, April 22-24, 2008. Proceedings*, pages 87–100, 2008. doi:10.1007/978-3-540-78929-1_7.

8 Krishnendu Chatterjee, Mickael Randour, and Jean-François Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Informatica*, 51:129–163, 2014. doi:10.1007/s00236-013-0182-6.

9 Hugo Gimbert and Wiesław Zielonka. When can you play positionally? In *Proceedings of the 29th International Conference on Mathematical Foundations of Computer Science (MFCS'04)*, volume 3153 of *Lecture Notes in Computer Science*, pages 686–698. Springer, 2004.

10 Erich Grädel, Wolfgang Thomas, and Thomas Wilke. *Automata, Logics, and Infinite Games: A Guide to Current Research*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002.

11 Leonid Khachiyan, Endre Boros, Konrad Borys, Khaled Elbassioni, Vladimir Gurvich, Gabor Rudolf, and Jihui Zhao. On short paths interdiction problems: Total and node-wise limited interdiction. *Theory of Computing Systems*, 43:204–233, 2008.

12 Donald A. Martin. The determinacy of Blackwell games. *The Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

13 John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950.

A Computations for proof of Proposition 9

A.1 Computations for $\gamma_{0,\mathbb{N}} + \gamma_{<I,\geqslant L} \leqslant \mathsf{dVal}^\sigma(v_0)$

When $\mathsf{dVal}^\sigma(v_0) < 0$ and $\Pi_{<I,\geqslant L} \neq \emptyset$, it remains to show under which conditions over p,

$$S = \sum_{i=1}^{I-1} \sum_{\ell=L}^{\infty} (1-p)^i p^\ell \left(i w^+ + \left\lfloor \frac{\ell - ic - |V|}{c} \right\rfloor (-w^-) + |V| W \right) \leqslant \mathsf{dVal}^\sigma(v_0)$$

Upper-bounding $\left\lfloor \frac{\ell-ic-|V|}{c} \right\rfloor (-w^-)$ by $\left(\frac{\ell-ic-|V|}{c} - 1 \right)(-w^-) = \frac{\ell-ic-|V|-c}{c}(-w^-)$, we can split the double sum S in three parts:

$$S = (w^+ + w^-) \underbrace{\sum_{i=1}^{I-1} \sum_{\ell=L}^{\infty} (1-p)^i p^\ell i}_{S_1} - \frac{w^-}{c} \underbrace{\sum_{i=1}^{I-1} \sum_{\ell=L}^{\infty} (1-p)^i p^\ell \ell}_{S_2} + \left(\frac{-|V|-c}{c}(-w^-) + |V| W \right) \underbrace{\sum_{i=1}^{I-1} \sum_{\ell=L}^{\infty} (1-p)^i p^\ell}_{S_3}$$

Using the fact that $L \geqslant 2$ ($L = I\gamma + \frac{2|\mathsf{dVal}^\sigma(v_0)|+|V|W}{w^-} c + |V| > |V| > 1$ otherwise, for the unique $v \in V$, $\mathsf{dVal}(v) = 0$ or $+\infty$ regarding $v \in T$ or not), we have

$$S_1 \leqslant \sum_{i=1}^{\infty} i(1-p)^i \sum_{\ell=2}^{\infty} p^\ell = \frac{1-p}{p^2} \times \frac{p^2}{1-p} = 1$$

$$S_3 \leqslant \sum_{i=1}^{\infty} (1-p)^i \sum_{\ell=1}^{\infty} p^\ell = \frac{1-p}{p} \times \frac{p}{1-p} = 1$$

and

$$\begin{aligned}
S_2 &= \sum_{i=1}^{I-1}(1-p)^i \sum_{\ell=L}^{\infty} p^\ell \ell \\
&= (1-p)\frac{1-(1-p)^{I-1}}{p} \times \frac{p^L(-Lp+L+p)}{(1-p)^2} \\
&= \frac{(1-(1-p)^{I-1})p^{L-1}(-Lp+L+p)}{1-p} \\
&\geqslant \frac{(1-(1-p)^{I-1})p^L}{1-p} && \text{(since } -Lp+L \geqslant 0\text{)} \\
&\geqslant \frac{1}{4(1-p)} && \text{(since } p \geqslant \tfrac{1}{2^{1/L}} \geqslant \tfrac{1}{2} \text{ and } 1-(1-p)^{I-1} \geqslant \tfrac{1}{2} \text{ by } 0 \leqslant I\text{)}
\end{aligned}$$

Therefore, we obtain

$$S \leqslant (w^+ + w^-) - \frac{w^-}{c}\frac{1}{4(1-p)} + \left(\frac{-|V|-c}{c}(-w^-) + |V|W\right)$$

The right term goes towards $-\infty$ when $p \to 1$. In particular, when

$$p \geqslant 1 - \frac{w^-}{4(cw^+ + 2cw^- + |V|w^- + c|V|W - \mathsf{dVal}^\sigma(v_0)c)}$$

we obtain

$$S \leqslant \mathsf{dVal}^\sigma(v_0)$$

A.2 Computations for $\gamma_{\geqslant I,\mathbb{N}} \leqslant \varepsilon/4$

It remains to show that

$$(1-p)^I\left(\frac{w^+}{p}I + \frac{w^+(1-p)}{p^2} + \frac{|V|W}{p}\right) \leqslant \frac{\varepsilon}{4}$$

We let here $\delta = \frac{2|\mathsf{dVal}^\sigma(v_0)|+|V|W}{w^-}c + |V|$ so that $L = I\gamma + \delta$. Since, $p \geqslant LW/(LW+\varepsilon/4) = (I\gamma W + \delta W)/(I\gamma W + \delta W + \varepsilon/4)$,

$$1-p \leqslant \frac{\varepsilon/4}{I\gamma W + \delta W + \varepsilon/4} = \frac{1}{4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1}$$

By also using that $p \geqslant 1/2 \geqslant 1/4$, thus $1/p \leqslant 4$ and $1/p^2 \leqslant 4$, we obtain

$$\gamma_{\geqslant I,\mathbb{N}} \leqslant \left(\frac{1}{4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1}\right)^I \left(4w^+ I + 4(w^+ + |V|W)\right)$$

The value $4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1$ being greater than 1, we can write

$$\gamma_{\geqslant I,\mathbb{N}} \leqslant \left(\frac{1}{4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1}\right)^{I-1} \left(4w^+ \frac{I}{4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1} + 4(w^+ + |V|W)\right)$$

Since $x/(ax+b) \leqslant 1/a$ whenever $a, x, b \geqslant 0$, we have $\frac{I}{4I\gamma W/\varepsilon + 4\delta W/\varepsilon + 1} \leqslant \frac{\varepsilon}{4\gamma W}$. Moreover, $\frac{4I\gamma W}{\varepsilon} + \frac{4\delta W}{\varepsilon} + 1 > \frac{I\gamma W}{2\varepsilon}$ and thus

$$\gamma_{\geqslant I,\mathbb{N}} \leqslant \left(\frac{2\varepsilon}{I\gamma W}\right)^{I-1} \left(\frac{\varepsilon w^+}{\gamma W} + 4(w^+ + |V|W)\right)$$

But

$$\left(\frac{2\varepsilon}{I\gamma W}\right)^{I-1}\left(\frac{\varepsilon w^+}{\gamma W}+4(w^++|V|W)\right)\leqslant\frac{\varepsilon}{4}$$

if and only if

$$\left(\frac{I\gamma W}{2\varepsilon}\right)^{I-1}\geqslant\frac{4w^+}{\gamma W}+\frac{16(w^++|V|W)}{\varepsilon}\geqslant\frac{2w^+}{\gamma W}+\frac{8(w^++|V|W)}{\varepsilon}$$

if and only if

$$(I-1)\ln\left(\frac{I\gamma W}{2\varepsilon}\right)\geqslant\ln\left(\frac{2w^+}{\gamma W}+\frac{8(w^++|V|W)}{\varepsilon}\right)=\ln\left(\frac{\xi\gamma W}{2\varepsilon}\right)$$

where $\xi=\frac{4\varepsilon w^+}{\gamma^2W^2}+\frac{16(w^++|V|W)}{\gamma W}$. Consider ε small enough so that $\gamma W/2\varepsilon\geqslant 1$ and $\xi\gamma W/2\varepsilon\geqslant 2$ (the two terms tend to $+\infty$ when ε tends to 0). Then, $(I-1)\ln\left(\frac{I\gamma W}{2\varepsilon}\right)\geqslant(I-1)\ln(I)$, and it is sufficient to prove that

$$(I-1)\ln(I)\geqslant\ln\left(\frac{\xi\gamma W}{2\varepsilon}\right)$$

Since the mapping $I\mapsto(I-1)\ln(I)$ is increasing, and $I\geqslant\frac{\xi\gamma W}{2\varepsilon}$ (by definition),

$$(I-1)\ln(I)\geqslant\left(\frac{\xi\gamma W}{2\varepsilon}-1\right)\ln\left(\frac{\xi\gamma W}{2\varepsilon}\right)\geqslant\ln\left(\frac{\xi\gamma W}{2\varepsilon}\right)$$

A.3 Lower bound over p

If we gather all the lower bounds over p that we need in the proof, we get that:

- if $\mathsf{dVal}^\sigma(v_0)\geqslant 0$, we must have

$$p\geqslant\max\left(\frac{LW}{LW+\varepsilon/4},\frac{1}{2}\right)$$

- if $\mathsf{dVal}^\sigma(v_0)<0$, we must have

$$\max\left(\frac{LW}{LW+\varepsilon/4},\frac{1}{2^{1/L}},\left(1-\frac{\varepsilon}{2|\mathsf{dVal}^\sigma(v_0)|}\right)^{\frac{1}{|V|}},\right.$$
$$\left.1-\frac{w^-}{4(cw^++2cw^-+|V|w^-+c|V|W+|\mathsf{dVal}^\sigma(v_0)|c)}\right)$$

  with ε small enough so that this bound is less than 1.

B Proof of Proposition 22

Implication $1\Rightarrow 2$ is trivial by the result of Theorem 6.

For implication $3\Rightarrow 1$, consider any memoryless deterministic strategy σ^* that guarantees Min to reach T from all vertices in the game graph $\widetilde{\mathcal{G}}^{(|V|-1)}$. Then, for all vertices v, we show by induction on n, that each play π from v that reaches the target in at most n steps, and conforming to σ^*, has a total-payoff $\mathbf{TP}(\pi)\leqslant\mathsf{dVal}(v)$. This is trivial for $n=0$. If $\pi=v\pi'$ with π' starting in v, then

$$\mathbf{TP}(\pi)=\omega(v,v')+\mathbf{TP}(\pi')\leqslant\omega(v,v')+\mathsf{dVal}(v')=\omega(v,v')+f_v^{(|V|-1)}$$

If $v \in V_{\mathsf{Max}}$, we have

$$\mathbf{TP}(\pi) \leqslant \omega(v,v') + f_v^{(|V|-1)} \leqslant f_v^{(|V|)} = \mathsf{dVal}(v)$$

If $v \in V_{\mathsf{Min}}$, since $v' \in \widetilde{E}^{(|V|-1)}(v)$,

$$\mathbf{TP}(\pi) = f_v^{(|V|)} = \mathsf{dVal}(v)$$

This ends the proof by induction. To conclude that 1 holds, since σ^* guarantees to reach the target, all plays conforming to it reach the target in less than $|V|$ steps, which proves that $\mathsf{dVal}^{\sigma^*}(v) \leqslant \mathsf{dVal}(v)$, showing that σ^* is optimal.

For implication $1 \Rightarrow 3$, consider an optimal deterministic memoryless strategy σ^*, such that for all v, $\mathsf{dVal}^{\sigma^*}(v) = \mathsf{dVal}(v)$.

First, we show that $f_v^{(|V|-1)} = \mathsf{dVal}(v)$ for all vertices v. For that, consider the deterministic strategy τ of Max defined for all finite plays π having $n \leqslant |V|$ vertices, ending in a vertex $v \in V_{\mathsf{Max}}$, by $\tau(\pi) = v'$ such that $\omega(v,v') + f_{v'}^{(|V|-1-n)} = f_v^{(|V|-n)}$. For longer finite plays, we define τ arbitrarily. Then, let π be the play from v conforming to σ^* and τ. Since σ^* ensures reaching the target and is memoryless deterministic, π reaches the target in at most $|V|-1$ steps. Let $\pi = v_0 v_1 v_2 \cdots v_{k-1} v_k$ with $k \leqslant |V|$. Let us show that $\mathbf{TP}(\pi) \geqslant f_v^{(|V|-1)}$. We prove by induction on $0 \leqslant j \leqslant k$ that

$$\sum_{i=j}^{k-1} \omega(v_i, v_{i+1}) \geqslant f_{v_j}^{(|V|-1-j)}$$

When $j = k-1$, the result is trivial since the sum is

$$0 = f_{v_k}^{(0)} \geqslant f_{v_k}^{(|V|-1-(k-1))}$$

Otherwise, by induction hypothesis

$$\sum_{i=j}^{k-1} \omega(v_i, v_{i+1}) \geqslant \omega(v_j, v_{j+1}) + f_{v_{j+1}}^{(|V|-1-(j+1))}$$

If $v_j \in V_{\mathsf{Max}}$, v_{j+1} is chosen by τ so that

$$\omega(v_j, v_{j+1}) + f_{v_{j+1}}^{(|V|-1-(j+1))} = f_{v_j}^{(|V|-1-j)}$$

If $v \in V_{\mathsf{Min}}$, by definition of $\mathcal{F}$,

$$\omega(v_j, v_{j+1}) + f_{v_{j+1}}^{(|V|-1-(j+1))} \geqslant f_{v_j}^{(|V|-1-j)}$$

We can conclude in all cases, so that $f_v^{(|V|-1)} = \mathsf{dVal}(v)$ for all vertices v.

Then, we show that Min can guarantee to reach T from all vertices in the game graph $\widetilde{\mathcal{G}}^{(|V|-1)}$. Let us suppose that this is not the case. Then, there exists a set V' of vertices in which Max can guarantee to keep Min for ever, in the game $\widetilde{\mathcal{G}}^{(|V|-1)}$: for all $v' \in V' \cap V_{\mathsf{Min}}$, $\widetilde{E}^{(|V|-1)}(v') \subseteq V'$, and for all $v' \in V' \cap V_{\mathsf{Max}}$, $E(v) \cap V' \neq \emptyset$. Since σ^* guarantees to reach the target, there exists $v \in V' \cap V_{\mathsf{Min}}$ such that $\sigma^*(v) = v' \notin V'$: then $\omega(v,v') + \mathsf{dVal}(v') > \mathsf{dVal}(v)$ (here we use that $\mathsf{dVal}(v) = f_v^{(|V|-1)} = f_v^{(|V|)}$). Consider an optimal deterministic memoryless strategy τ^* of Max in $\mathcal{G}$. Then, the play π from v conforming to σ^* and τ^* starts by taking the edge (v,v') and continues with a play π'. By optimality, we know that $\mathbf{TP}(\pi) = \mathsf{dVal}(v)$ and $\mathbf{TP}(\pi') = \mathsf{dVal}(v')$. However,

$$\mathbf{TP}(\pi) = \omega(v,v') + \mathbf{TP}(\pi') = \omega(v,v') + \mathsf{dVal}(v') > \mathsf{dVal}(v)$$

which raises a contradiction.

We finish the proof by showing $2 \Rightarrow 1$. For that, consider an optimal memoryless strategy ρ^* for $\overline{\mathsf{mVal}}$. By following the construction of Section 4, we build a memoryless deterministic strategy σ_1. Lemma 21 ensures that σ_1 is an NC-strategy so that every cycle conforming to σ_1 has a negative total weight. Let us show that such a negative cycle cannot exist, which will ensure that all plays conforming to σ_1 reach the target, and thus the optimality of σ_1. Suppose that a cycle $v_1 v_2 \cdots v_k v_1$ conforms to σ_1. By following the notations of the proof of Lemma 18(ii), we suppose that v_1 is a vertex of minimal distance $d(v_1)$ to the target, and that it is owned by V_{Min}. Note that such a vertex exists, otherwise only Max has the minimal distance vertices on the cycle and that contradicts the attractor computation. By minimality of $d(v_1)$ among the vertices of the cycle, $d(v_2) \geqslant d_{v_1}$. Moreover, by the attractor computation, there exists $u \in E(v_1)$ such that $d(u) = d(v_1) - 1 < d(v_1)$. By definition of σ_1, we know for sure that $u \notin \widetilde{E}(v_1)$, so that

$$\omega(v_1, u) + \mathsf{mVal}^{\rho^*}(u) > \omega(v_1, v_2) + \mathsf{mVal}^{\rho^*}(v_2)$$

By (6), we know that in this case

$$\mathsf{mVal}^{\rho^*}(v_1) > \omega(v_1, v_2) + \mathsf{mVal}^{\rho^*}(v_2)$$

By optimality of ρ^*, this rewrites in

$$\overline{\mathsf{mVal}}(v_1) > \omega(v_1, v_2) + \overline{\mathsf{mVal}}(v_2)$$

By Theorem 6, this also rewrites in

$$\mathsf{dVal}(v_1) > \omega(v_1, v_2) + \mathsf{dVal}(v_2) \geqslant \mathcal{F}\big((\mathsf{dVal}(v))_{v \in V}\big)(v_1)$$

(since $v_1 \in V_{\mathsf{Min}}$): this contradicts the fact that the vector $(\mathsf{dVal}(v))_{v \in V}$ is a fixed-point of $\mathcal{F}$.