One-Counter Stochastic Games

T. Brázdil[1]⋆, V. Brožek[2]⋆⋆, K. Etessami[2]

[1] Faculty of Informatics, Masaryk University
xbrazdil@fi.muni.cz
[2] School of Informatics, University of Edinburgh
{kousha,vbrozek}@inf.ed.ac.uk

Abstract. We study the computational complexity of basic decision problems for *one-counter simple stochastic games* (OC-SSGs), under various objectives. OC-SSGs are 2-player turn-based stochastic games played on the transition graph of classic one-counter automata. We study primarily the *termination* objective, where the goal of one player is to maximize the probability of reaching counter value 0, while the other player wishes to avoid this. Partly motivated by the goal of understanding termination objectives, we also study certain "limit" and "long run average" reward objectives that are closely related to some well-studied objectives for stochastic games with rewards. Examples of problems we address include: does player 1 have a strategy to ensure that the counter eventually hits 0, i.e., *terminates*, almost surely, regardless of what player 2 does? Or that the $\liminf$ (or $\limsup$) counter value equals ∞ with a desired probability? Or that the long run average reward is > 0 with desired probability? We show that the *qualitative termination problem* for OC-SSGs is in **NP** ∩ **coNP**, and is in P-time for 1-player OC-SSGs, or equivalently for *one-counter Markov Decision Processes* (OC-MDPs). Moreover, we show that *quantitative* limit problems for OC-SSGs are in **NP** ∩ **coNP**, and are in **P**-time for 1-player OC-MDPs. Both qualitative limit problems and qualitative termination problems for OC-SSGs are already at least as hard as Condon's quantitative decision problem for finite-state SSGs.

1 Introduction

There is a rich literature on the computational complexity of analyzing finite-state Markov decision processes and stochastic games. In recent years, there has also been some research done on the complexity of basic analysis problems for classes of finitely-presented but infinite-state stochastic models and games whose transition graphs arise from decidable infinite-state automata-theoretic models, including: context-free processes, one-counter processes, and pushdown processes (see, e.g., [8]). It turns out that such stochastic automata-theoretic models are intimately related to classic stochastic processes studied extensively in applied probability theory, such as (multi-type-)branching processes and (quasi-)birth-death processes (QBDs) (see [8,7,2]).

In this paper we continue this line of work by studying **one-counter simple stochastic games (OC-SSGs)**, which are turn-based 2-player zero-sum stochastic games on transition graphs of classic one-counter automata. In more detail, an OC-SSG has a finite set of control states, which are partitioned into three types: a set of *random* states, from where the next transition is chosen according to a given probability distribution, and states belonging to one of two players: *Max* or *Min*, from where the respective player chooses the next transition. Transitions can change the state and can also change the value of the (unbounded) counter by at most 1. If there are no control states belonging to *Max* (*Min*, respectively), then we call the resulting 1-player OC-SSG a *minimizing* (*maximizing*, respectively) *one-counter Markov decision process* (OC-MDP).

⋆ Supported by the Czech Science Foundation, grant No. P202/10/1469.
⋆⋆ Supported by Newton International Fellowship from the Royal Society.

Fixing strategies for the two players yields a countable state Markov chain and thus a probability space of infinite runs (trajectories). We focus in this paper on *objectives* that can be described by a (measurable) set of runs, such that player Max wants to maximize, and player Min wants to minimize, the probability of the objective. The central objective studied in this paper is *termination*: starting at a given control state and a given counter value $j > 0$, player Max (Min) wishes to maximize (minimize) the probability of eventually hitting the counter value 0 (in any control state).

Different objectives give rise to different computational problems for OC-SSGs, aimed at computing the value of the game, or optimal strategies, with respect to that objective. From general known facts about stochastic games (e.g., Martin's Blackwell determinacy theorem [13]), it follows that the games we study are *determined*, meaning they have a *value*: we can associate with each such game a *value*, ν, such that for every $\varepsilon > 0$, player Max has a strategy that ensures the objective is satisfied with probability at least $\nu - \varepsilon$ regardless of what player Min does, and likewise player Min has a strategy to ensure that the objective is satisfied with probability at most $\nu + \varepsilon$. In the case of termination objectives, the value may be *irrational* even when the input data contains only rational probabilities, and this is so even in the purely stochastic setting without any players, i.e., with only *random* control states (see [7]).

We can classify analysis problems for OC-SSGs into two kinds: *quantitative* analyses: "can the objective be achieved with probability at least/at most p" for a given $p \in [0, 1]$; or *qualitative* analyses, which ask the same question but restricted to $p \in \{0, 1\}$. We are often also interested in what kinds of strategies (e.g., memoryless, etc.) achieve these.

In a recent paper, [2], we studied *one-player* OC-SSGs, i.e., OC-MDPs, and obtained some complexity results for them under qualitative termination objectives and some quantitative limit objectives. The problems we studied included the qualitative termination problem (is the maximum probability of termination = 1?) for *maximizing* OC-MDPs. We showed that this problem is decidable in **P**-time. However, we left open the complexity of the same problem for *minimizing* OC-MDPs (is the minimum probability of termination < 1?). One of the main results of this paper is the following, which in particular resolves this open question:

Theorem 1. (Qualitative termination) *Given a OC-SSG, $\mathcal{G}$, with the objective of termination, and given an initial control state s and initial counter value $j > 0$, deciding whether the value of the game is equal to 1 is in **NP** $\cap$ **coNP**. Furthermore, the same problem is in **P**-time for 1-player OC-SSGs, i.e., for both maximizing and minimizing OC-MDPs.*

Improving on this **NP** $\cap$ **coNP** upper bound for the qualitative termination problem for OC-SSGs would require a breakthrough: we show that deciding whether the value of an OC-SSG termination game is equal to 1 is already at least as hard as Condon's [5] *quantitative* reachability problem for finite-state simple stochastic games (Corollary 1). We do not know a reduction in the other direction. We furthermore show that if the value is 1 for a OC-SSG termination game, then Max has a simple kind of optimal strategy (memoryless, counter-oblivious, and pure) that ensures termination with probability 1, regardless of Min's strategy. Similarly, if the value is less than 1, we show Min has a simple strategy (using finite memory, linearly bounded in the number of control states) that ensures the probability of termination is $< 1 - \delta$ for some positive $\delta > 0$, regardless of what Max does. We show that such strategies for both players are computable in non-deterministic polynomial time for OC-SSGs, and in deterministic P-time for (both maximizing and minimizing) 1-player OC-MDPs. We also observe that the analogous problem of deciding whether the value of a OC-SSG termination game is 0 is in **P**, which follows easily by reduction to non-probabilistic games.

OC-SSGs can be viewed as stochastic game extensions of Quasi-Birth-Death Processes (QBDs) (see [7,2]). QBDs are a heavily studied model in queuing theory and performance evaluation (the counter keeps track of the number of jobs in a queue). It is very natural to consider controlled and game extensions of such queuing theoretic models, thus allowing for adversarial modeling of queues with unknown (non-deterministic) environments or with other unknown aspects modeled non-deterministically. OC-SSGs with termination objectives also subsume "solvency games", a recently studied class of MDPs motivated by modeling of a risk-averse investment scenario [1].

Due to the presence of an unbounded counter, an OC-SSG, $\mathcal{G}$, formally describes a stochastic game with a countably-infinite state space: a "configuration" or "state" of the underlying stochastic game consists of a pair (s, j), where s is a control state of $\mathcal{G}$ and j is the current counter value. However, it is easy to see that we can equivalently view $\mathcal{G}$ as a finite-state **simple stochastic game (SSG)**, $\mathcal{H}$, with **rewards** as follows: $\mathcal{H}$ is played on the finite-state transition graph obtained from that of $\mathcal{G}$ by simply ignoring the counter values. Instead, every transition t of $\mathcal{H}$ is assigned a *reward*, $r(t) \in \{-1, 0, 1\}$, corresponding to the effect that the transition t would have on the counter in $\mathcal{G}$. Furthermore, when emulating an OC-SSG using rewards, we can easily place rewards on states rather than on transitions, by adding suitable auxiliary control states. Thus, w.l.o.g., we can assume that OC-SSGs are presented as equivalent finite-state SSGs with a reward, $r(s) \in \{-1, 0, 1\}$ labeling each state s. A *run* of $\mathcal{H}$, w, is an infinite sequence of states that is permitted by the transition structure, and we denote the i-th state along the run w by $w(i)$. The termination objective for $\mathcal{G}$, when the initial counter value is $j > 0$, can now be rephrased as the following equivalent objective for $\mathcal{H}$:

$$Term(j) := \{\, w \mid w \text{ is a run of } \mathcal{H} \text{ such that there exists } m > 0 \text{ such that } \textstyle\sum_{i=0}^{m} r(w(i)) = -j \,\} \,.$$

An important step toward our proof of Theorem 1 and related results, is to establish links between this termination objective and the following limit objectives, which are of independent interest. For $z \in \{-\infty, \infty\}$, and a comparison operator $\varDelta \in \{>, <, =\}$, consider the following objective:

$$LimInf(\varDelta\, z) := \{\, w \mid w \text{ is a run of } \mathcal{H} \text{ such that } \liminf_{n\to\infty} \textstyle\sum_{i=0}^{n} r(w(i))\, \varDelta\, z \,\} \,.$$

We will show that if j is large enough (larger than the number of control states), then the game value with respect to objective $Term(j)$ and the game value with respect to $LimInf(= -\infty)$ are either both equal to 1, or are both less than 1 (Lemma 4). We could also consider the "sup" variant of these objectives, such as $LimSup(= -\infty)$, but these are redundant. For example, by negating the sign of rewards, $LimSup(= -\infty)$ is "equivalent" to $LimInf(= +\infty)$. Indeed, the only limit objectives we need to consider for SSGs are $LimInf(= -\infty)$ and $LimInf(= +\infty)$, because the others are either the same objectives considered from the other player's points of view, or are vacuous, such as $LimInf(> +\infty)$. For both limit objectives, $LimInf(= -\infty)$ and $LimInf(= +\infty)$, we shall see that the value of the respective SSGs is always rational (Proposition 2). We shall also show that the objective $LimInf(= +\infty)$ is essentially equivalent to the following "mean payoff" objective (Lemma 2):

$$Mean(> 0) := \{\, w \mid \text{w is a run of } \mathcal{H} \text{ such that } \liminf_{n\to\infty} \textstyle\sum_{i=0}^{n-1} r(w(i))/n > 0 \,\} \,.$$

This "intuitively obvious equivalence" is not so easy to prove. (Note also that $LimInf(= -\infty)$ is certainly not equivalent to $Mean(\leq 0)$.) We establish the equivalence by a combination of new methods and by using recent results by Gimbert, Horn and Zielonka [11,12]. Mean payoff objectives are of course very heavily studied for stochastic games and for MDPs (see [15]). However, there is a subtle but important difference here: mean payoff objectives are typically formulated via *expected payoffs*: the Max player wishes to maximize the *expected* mean payoff, while the Min player wishes to minimize this. Instead, in the above $Mean(> 0)$ objective we wish to maximize (minimize) the *probability* that the mean payoff is > 0. These require new algorithms. Our main result about such limit objectives is the following:

Theorem 2. *For both limit objectives, $O \in \{LimInf(= -\infty), LimInf(= +\infty)\}$, given a finite-state SSG, $\mathcal{G}$, with rewards, and given a rational probability threshold, p, $0 \leq p \leq 1$, deciding whether the value of $\mathcal{G}$ with objective O is $>p$ (or $\geq p$) is in **NP** $\cap$ **coNP**. If $\mathcal{G}$ is a 1-player SSG (i.e., a maximizing or minimizing MDP), then the game value can be computed in P-time.*

Although our upper bounds for both these objectives look the same, their proofs are quite different. We show that both players have pure and memoryless optimal strategies in these games (Proposition 1), which can be computed in P-time for 1-player (Max or Min) MDPs. Furthermore, we show that even deciding whether the value of these games is either 1 or 0, given input for which one of these two is promised to be the case, is already at least as hard as Condon's [5] *quantitative* reachability problem for finite-state simple stochastic games (Proposition 4). Thus, even any non-trivial *approximation* of the value of SSGs with such limit objectives is not easier than Condon's problem.

We already considered in [2] the problem of maximizing the probability of $\mathit{LimInf}(= -\infty)$ in a OC-MDP. There we showed that the maximum probability can be computed in P-time. However, again, we did not resolve the complementary problem of minimizing the probability of $\mathit{LimInf}(= -\infty)$ in a OC-MDP. Thus we could not address two-player OC-SSGs with either of these objectives, and we left these as key open problems, which we resolve here. An important distinction between *maximizing* and *minimizing* the probability of objective $\mathit{LimInf}(= -\infty)$ is that maximizing this objective satisfies a *submixing* property defined by Gimbert [10], which he showed implies the existence of optimal memoryless strategies, whereas minimizing the objective is not submixing, and thus we require new methods to tackle it, which we develop in this paper.

Finally, we mention that one can also consider OC-SSGs with the objective of terminating in a *selected* subset of states, F. Such objectives were considered for OC-MDPs in [2]. Using our termination results in this paper, we can also show that given an OC-SSG it is decidable (in double exponential time) whether Max can achieve a termination probability 1 in a selected subset of states, F. The computational complexity of selective termination is higher than for non-selective termination: PSPACE-hardness holds already for OC-MDPs without Min ([2]). Due to space limitations, we omit results about selective termination from this conference paper, and will include them in the journal version of this paper.

Related work. As mentioned earlier, we initiated the study of some classes of 1-player OC-SSGs (i.e., OC-MDPs) in a recent paper [2]. The reader will find extensive references to earlier related literature in [2]. No earlier work considered OC-SSGs explicitly, but as we have highlighted already there are close connections between OC-SSGs and finite-state stochastic games with certain interesting limiting average reward objectives. One-counter automata with a non-negative counter are equivalent to pushdown automata restricted to a 1-letter stack alphabet (see [7]), and thus OC-SSGs with the termination objective form a subclass of pushdown stochastic games, or equivalently, Recursive simple stochastic games (RSSGs). These more general stochastic games were introduced and studied in [8], where it is shown that many interesting computational problems for the general RSSG and RMDP models are undecidable, including generalizations of qualitative termination problems for RMDPs. It was also established in [8] that for stochastic context-free games (1-exit RSSGs), which correspond to pushdown stochastic games with only one state, both qualitative and quantitative termination problems are decidable, and in fact qualitative termination problems are decidable in $\mathbf{NP} \cap \mathbf{coNP}$ ([9]). Solving termination objectives is a key ingredient for many more general analyses and model checking problems for stochastic games. OC-SSGs form another natural subclass of RSSGs, which is incompatible with stochastic context-free games. Specifically, for OC-SSGs with the termination objective, the number of stack symbols, rather than the number of control states, of a pushdown stochastic game is being restricted to 1. As we show in this paper, this restriction again yields decidability of the qualitative termination problem. However, the decidability of the quantitative termination problem for OC-SSGs remains an open problem (see below).

Open problems. Our results complete part of the picture for decidability and complexity of several problems for OC-SSGs. However, our results also leave many open questions. The most important open question for OC-SSGs is whether the *quantitative* termination problem, even for OC-MDPs, is decidable. Specifically, we do not know whether the following is decidable: given a OC-MDP, and a rational probability $p \in (0, 1)$, decide whether the maximum probability of termination is $>p$ (or $\geq p$). Substantial new obstacles arise for deciding this. In particular, we know that an optimal strategy may in general need to use different actions at the same control state for arbitrarily large counter values (so strategies cannot ignore the value of the counter, even for arbitrarily large values), and this holds already for the extremely simple case of solvency games [1, Theorem 3.7].

Outline of paper. We fix notation and key definitions in Section 2. In Section 3, we prove Theorem 2. Building on Section 3, we prove Theorem 1 in Section 4. Many proofs are in the appendix.

2 Preliminaries

Definition 1. *A* **simple stochastic game (SSG)** *is given by a finite, or countably infinite directed graph, $(V, \hookrightarrow)$, where V is the set of* vertices *(which we also call* states*), and $\hookrightarrow$ is the edge (also called* transition*) relation, together with a partition $(V_\top, V_\bot, V_P)$ of V, as well as a* probability assignment, *Prob, which to each $v \in V_P$ assigns a rational probability distribution on its set of outgoing edges. States in V_P are called* random, *states in $V_\top$ belong to player Max, and states in $V_\bot$ belong to player Min. We assume that for every $v \in V$ there is at least one $u \in V$ such that $v \hookrightarrow u$. We often write $v \overset{x}{\hookrightarrow} u$ instead of $Prob(v \hookrightarrow u) = x$. If $V_\bot = \emptyset$ we call $\mathcal{G}$ a* maximizing **Markov decision process (MDP)***. If $V_\top = \emptyset$ we call it a* minimizing *MDP. If $V_\bot = V_\top = \emptyset$ then we call $\mathcal{G}$ a* **Markov chain***. A SSG (or a MDP or a Markov chain) can be equipped with a reward function, r, which assigns to each state, $v \in V$, a number $r(v) \in \{-1, 0, 1\}$.*[3] *Similarly, rewards can be assigned to transitions.*

For a *path*, $w = w(0)w(1)\cdots w(n-1)$, of states in a graph, we use $len(w) = n$ to denote the length of w. A *run* in a SSG, $\mathcal{G}$, is an infinite path in the underlying directed graph. The set of all runs in $\mathcal{G}$ is denoted by $Run_{\mathcal{G}}$, and the set of all runs starting with a finite path w is $Run_{\mathcal{G}}(w)$. These sets generate the standard Borel algebra on $Run_{\mathcal{G}}$.

A *strategy* for player Max is a function, σ, which to each *history* $w \in V^+$ ending in some $v \in V_\top$, assigns a probability distribution on the set of outgoing transitions of v. We say that a strategy σ is *memoryless* if $\sigma(w)$ depends only on the last state, v, and *pure* if $\sigma(w)$ assigns probability 1 to some transition, for each history w. When σ is pure, we write $\sigma(w) = v'$ instead of $\sigma(w)(v, v') = 1$. Strategies for player Min are defined similarly, just by substituting $V_\top$ with $V_\bot$.

Assume we fix a starting state s, and a pair of strategies: σ for player Max, and π for Min in a SSG, $\mathcal{G}$. There is a unique probabilistic measure, $\mathbb{P}_s^{\sigma,\pi}$, on the Borel space of runs $Run_{\mathcal{G}}$, satisfying for all finite paths w starting in s: $\mathbb{P}_s^{\sigma,\pi}(Run_{\mathcal{G}}(w)) = \prod_{i=1}^{len(w)-1} x_i$ where x_i, $1 \le i < len(w)$ are defined by requiring that (a) if $w(i-1) \in V_P$ then $w(i-1) \overset{x_i}{\hookrightarrow} w(i)$; and (b) if $w(i-1) \in V_\top$ then $\sigma(w(0)\cdots w(i-1))$ assigns x_i to the transition $w(i-1) \hookrightarrow w(i)$; and (c) if $w(i-1) \in V_\bot$ then $\pi(w(0)\cdots w(i-1))$ assigns x_i to the transition $w(i-1) \hookrightarrow w(i)$. In particular, $\mathbb{P}_s^{\sigma,\pi}(Run_{\mathcal{G}}(s)) = 1$. In cases where $\mathcal{G}$ is a maximizing MDP, a minimizing MDP, or a Markov chain, we denote this probability measure by $\mathbb{P}_s^{\sigma}$, $\mathbb{P}_s^{\pi}$, or $\mathbb{P}_s$, respectively. See, e.g., [15, p. 30], for the existence and uniqueness of the measure $\mathbb{P}_s^{\sigma}$ in the case of MDPs. It is straightforward then to establish existence and uniqueness of $\mathbb{P}_s^{\sigma,\pi}$ for SSGs, by considering pairs of strategies to be one strategy.

In this paper, an *objective* for a stochastic game is given by a measurable set of runs. An objective, O, is called a *tail* objective if for all runs w and all suffixes w' of w, we have $w' \in O \iff w \in O$.

Assume we have fixed a SSG, an objective, O, and a starting state, s. We define the *value of $\mathcal{G}$ in s* as $Val^O(s) := \sup_\sigma \inf_\pi \mathbb{P}_s^{\sigma,\pi}(O)$. It follows from Martin's Blackwell determinacy theorem [13] that these games are *determined*, meaning $Val^O(s) = \inf_\pi \sup_\sigma \mathbb{P}_s^{\sigma,\pi}(O)$. A strategy σ for Max is *optimal in s* if $\mathbb{P}_s^{\sigma,\pi}(O) \ge Val^O(s)$ for every π. Similarly a strategy π for Min is *optimal in s* if $\mathbb{P}_s^{\sigma,\pi}(O) \le Val^O(s)$ for every σ. A strategy is called *optimal* if it is optimal in every state.

[3] Rewards can generally be arbitrary rational values, but for this paper we confine ourselves to rewards in $\{-1, 0, 1\}$.

An important objective for us is *reachability*. Given a set $T \subseteq V$, we define the objective $Reach(T) := \{w \in Run_{\mathcal{G}} \mid \exists i \geq 0 : w(i) \in T\}$. The following fact is well known:

Fact 3 *(See, e.g., [15,5,6].) For both maximizing and minimizing finite-state MDPs with reachability objectives, pure memoryless optimal strategies exist and can be computed, together with the optimal value, in polynomial time.*

3 Limit objectives

All MDPs and SSGs in this section have finitely many states. Rewards are assigned to states, not to transitions. The main goal of this section is to prove Theorem 2. We start by proving that both players have optimal pure and memoryless strategies for objectives $\mathit{LimInf}(= -\infty)$, $\mathit{LimInf}(= +\infty)$, and $\mathit{Mean}(> 0)$. The following is a corollary of a result by Gimbert and Zielonka, which allows us to concentrate on MDPs instead of SSGs:

Fact 4 *(See [12, Theorem 2].) Fix any objective, O, and suppose that in every maximizing and minimizing MDP with objective O, the unique player has a pure memoryless optimal strategy. Then in all SSGs with objective O, both players have optimal pure and memoryless strategies.*

Note that the probability of $\mathit{LimInf}(= -\infty)$ is minimized iff the probability of $\mathit{LimInf}(> -\infty)$ is maximized, similarly with $\mathit{LimInf}(= +\infty)$ vs. $\mathit{LimInf}(< +\infty)$, and $\mathit{Mean}(> 0)$ vs. $\mathit{Mean}(\leq 0)$.

Fact 5 *(See [11, Theorem 4.5].) Let O be a tail objective. Assume that for every maximizing MDP and for every state, s, with* $Val^{O}(s) = 1$, *there is an optimal pure memoryless strategy starting in s. Then for all s there is an optimal pure memoryless strategy starting in s, without restricting* $Val^{O}(s)$.

Proposition 1. *For every SSG, considered with any of the objectives* $\mathit{LimInf}(= -\infty)$, $\mathit{LimInf}(= +\infty)$, *or* $\mathit{Mean}(> 0)$, *both players Max and Min have optimal pure memoryless strategies.*

Proof. (Sketch.) Using Fact 4 we consider only maximizing MDPs, and prove the proposition for the objectives listed and their complements. Note that since all these objectives are tail, a play under an optimal strategy, starting from a state with value 1, cannot visit a state with value < 1. By Fact 5 we may thus safely assume that the value is 1 in all states. We discuss different groups of objectives:

$\mathit{LimInf}(= -\infty)$, $\mathit{LimInf}(< +\infty)$, $\mathit{Mean}(\leq 0)$, $\mathit{Mean}(> 0)$: The first three (with $\mathit{LimInf}(= -\infty)$ also handled explicitly in [2]) are *tail* objectives and are also *submixing* (see [10]). Therefore, Theorem 1 of [10] immediately yields the desired result. $\mathit{Mean}(> 0)$ can be equivalently rephrased via a submixing lim sup variant. See Section A.1 in the appendix for details.

$\mathit{LimInf}(= +\infty)$: is a tail objective, so there is always a pure optimal strategy, τ, by [11, Theorem 3.1]. Note that $\mathit{LimInf}(= +\infty)$ is *not submixing*, so Theorem 1 of [10] does not apply. In the following we proceed in two steps: we start with τ and convert it to a finite-memory strategy[4], σ. Finally, we reduce the use of memory to get a memoryless strategy.

First, we obtain a finite-memory optimal strategy, starting in some state, s. For a run $w \in Run_{\mathcal{G}}(s)$ and $i \geq 0$, we denote by $r[i](w)$ the accumulated reward $\sum_{j=0}^{i} r(w(j))$ up to step i. Observe that because τ is optimal there is some $m > 0$ and a (measurable) set of runs $A \subseteq Run_{\mathcal{G}}(s)$, such that $\mathbb{P}_s^{\tau}(A) \geq \frac{1}{2}$, and for all

[4] A finite-memory strategy is specified by a finite state automaton, $\mathcal{A}$, over the alphabet V. Given $w \in V^{+}$, the value $\sigma(w)$ is determined by the state of $\mathcal{A}$ after reading w.

$w \in A$ we have that the accumulated reward along w never reaches $-m$ (i.e. $\inf_{i\geq 0} r[i](w) > -m$). Since for almost all runs of A we have $\lim_{i\to\infty} r[i](w) = \infty$, there is some $n > 0$ and a set $B \subseteq A$ such that $\mathbb{P}^{\tau}_{s}(B \mid A) \geq \frac{1}{2}$ (and hence, $\mathbb{P}^{\tau}_{s}(B) \geq \frac{1}{4}$), and for all $w \in B$ we have that the accumulated reward along w reaches $4m$ before the n-th step. Thus with probability at least $\frac{1}{4}$, a run $w \in Run_{\mathcal{G}}(s)$ satisfies $\inf_{i\geq 0} r[i](w) > -m$ and $\max_{0\leq i\leq n} r[i](w) \geq 4m$.

We denote by $T_s(w)$ the *stopping time* over $Run_{\mathcal{G}}(s)$ which for every $w \in Run_{\mathcal{G}}(s)$ returns the least number $i \geq 0$ such that either $r[i](w) \notin (-m, 4m)$, or $i = n$. Observe that the expected accumulated reward at the stopping time T_s is at least $\frac{1}{4} \cdot 4m + \frac{3}{4}(-m) = \frac{m}{4} > 0$. Let us define a new strategy σ as follows. Starting in a state $s \in V$, the strategy σ chooses the same transitions as τ started in s, up to the stopping time T_s. Once the stopping time is reached, say in a state v, the strategy σ erases its memory and behaves like τ started anew in v. Subsequently, σ follows the behavior of τ up to the stopping time T_v. Once the stopping time T_v is reached, say in a state u, σ erases its memory and starts to behave as τ started anew in u, and so on. Observe that the strategy σ uses only finite memory because each stopping time T_s is bounded for every state s. Because τ is pure, so is σ.

Now we argue that σ is optimal. Intuitively, this is because, on average, the accumulated reward strictly increases between resets of the memory of σ. To formally argue that this implies that the accumulated reward increases indefinitely, we employ the theory of random walks on $\mathbb{Z}$ and sums of i.i.d. random variables (see, e.g., Chapter 8 of [4]). Essentially, we define a set of random walks, one for each state s, capturing the sequence of changes to the accumulated reward between each reset in s and the next reset (in any state). We can then apply random walk results, e.g., from [4, Chapter 8], to conclude that these walks diverge to ∞ almost surely. For details see Lemmas 11 and 10 in the appendix.

Taking the product of the finite-memory strategy σ and $\mathcal{G}$ yields a finite-state Markov chain. By analyzing its bottom strongly connected components we can eliminate the use of memory, and obtain a pure and memoryless optimal strategy, see Lemma 12 in the appendix.

$LimInf(> -\infty)$: Like $LimInf(= +\infty)$, the objective $LimInf(> -\infty)$ is tail, but not submixing. Thus there is always a pure optimal strategy, τ, for $LimInf(> -\infty)$, by [11, Theorem 3.1], but Theorem 1 of [10] does not apply. We will prove Proposition 1 for $LimInf(> -\infty)$ using the results for $LimInf(= +\infty)$, and also a new objective, $All(\geq 0) := \{w \in Run_{\mathcal{G}} \mid \forall n \geq 0 : \sum_{j=0}^{n} r(w(j)) \geq 0\}$. Let W_∞ and W_+ denote the sets of states s such that $Val^{LimInf(=+\infty)}(s) = 1$, and $Val^{All(\geq 0)}(s) = 1$, respectively. Then, as we prove in the appendix, Lemma 13, for every state, s, with $Val^{LimInf(>-\infty)}(s) = 1$:

$$\exists\sigma : \mathbb{P}^{\sigma}_{s}(Reach(W_\infty \cup W_+)) = 1 \tag{1}$$

Moreover, we prove that whenever $Val^{All(\geq 0)}(s) = 1$ then Max has a pure and memoryless strategy σ_+ which is optimal in s for $All(\geq 0)$. Indeed, observe that player Max achieves $All(\geq 0)$ with probability 1 iff *all* runs satisfy it. So we may consider the MDP $\mathcal{G}$ as a 2-player non-stochastic game, where random nodes are now treated as player Min's. In this case, Theorem 12 of [3] guarantees the existence of the promised strategy σ_+. The proof is now finished by observing that, by Fact 3, there is a pure and memoryless strategy σ maximizing the probability of reaching $W_\infty \cup W_+$. The resulting pure and memoryless strategy, optimal for $LimInf(> -\infty)$, can be obtained by "stitching" σ together with the respective optimal strategies for $LimInf(= +\infty)$ and $All(\geq 0)$. □

The following simple lemma is proved in the appendix.

Lemma 1. *Let $\mathcal{M}$ be a finite, strongly connected (irreducible) Markov chain, and O be a tail objective. Then there is $x \in \{0, 1\}$ such that $\mathbb{P}_s(O) = x$ for all states s.*

A corollary of the previous proposition and lemma is the following:

Proposition 2. *Let $O \in \{LimInf(= -\infty), LimInf(= +\infty), Mean(> 0)\}$. Then in every SSG, and for all states, s, $Val^O(s)$ is rational, with a polynomial length binary encoding.*

Proof. By Proposition 1, there are memoryless optimal strategies: σ for Max, and π for Min. Fixing them induces a Markov chain on the states of $\mathcal{G}$. By Lemma 1, in every fixed bottom strongly connected component (BSCC), C, of this finite-state Markov chain, all states $v \in C$ have the same value, x_C, which is either 0 or 1. Denote by W the union of all BSCCs, C, with $x_C = 1$. By optimality of σ and π, $Val^O(s) = \mathbb{P}_s^{\sigma,\pi}(Reach(W))$ for every $s \in V$. By, e.g., [6, Section 3], this probability is rational, with polynomial length bit encoding, since reaching W is a regular event, and every Markov chain is a special case of a MDP. $\square$

Proof of Theorem 2. We will need a couple of preliminary lemmas:

Lemma 2. *Let $\mathcal{G}$ be a MDP with rewards, and s a state of $\mathcal{G}$. Then for every memoryless strategy σ:*

$$\mathbb{P}_s^\sigma(Mean(> 0)) = \mathbb{P}_s^\sigma(LimInf(= +\infty))$$

In particular, both objectives are equivalent with respect to both the value and optimal strategies.

Proof. (Sketch.) The inequality $\leq$ is true for all strategies, since $Mean(> 0) \subseteq LimInf(= +\infty)$. In the other direction, the property that σ is memoryless is needed, so that fixing σ yields a Markov chain on the states of $\mathcal{G}$. In this Markov chain, by Lemma 1, for every BSCC, C, there are $x_C \leq y_C \in \{0, 1\}$, such that $\mathbb{P}_s^\sigma(Mean(> 0) \mid Reach(C)) = x_C$, and $\mathbb{P}_s^\sigma(LimInf(= +\infty) \mid Reach(C)) = y_C$. By random walk arguments, considering the rewards accumulated between subsequent visits to a fixed state in C, we can prove that $y_C = 1 \implies x_C = 1$, see Lemma 14 in the appendix. Proposition 1 finishes the proof. $\square$

Lemma 3. *For an objective $O \in \{LimInf(= -\infty), LimInf(> -\infty), LimInf(= +\infty), LimInf(< +\infty)\}$, and a maximizing MDP, $\mathcal{G}$, denote by W the set of all $s \in V$ satisfying $Val^O(s) = 1$. Then $Val^O(s) = Val^{Reach(W)}(s)$ for every state s.*

Proof. Proposition 1 gives us a memoryless optimal strategy, σ. By fixing it, we obtain a Markov chain on states of $\mathcal{G}$. We denote by W' the union of all BSCCs of this Markov chain, in which at least one state has a positive value. By Lemma 1, all states from W' have, in fact, value 1. Since $W' \subseteq W$, and σ is optimal, we get

$$Val^O(s) = \mathbb{P}_s^\sigma(O) = \mathbb{P}_s^\sigma(Reach(W')) \leq \mathbb{P}_s^\sigma(Reach(W)) \leq Val^{Reach(W)}(s)$$

for every state s. Because O is a tail objective, we easily obtain $Val^O(s) \geq Val^{Reach(W)}(s)$. $\square$

To prove Theorem 2, we start with the MDP case. By Proposition 1, pure memoryless strategies are sufficient for optimizing the probability of all the objectives considered in this theorem, so we can restrict ourselves to such strategies for this proof. Given an objective O, we will write W^O to denote the set of states s with $Val^O(s) = 1$. As $\mathcal{G}$ is a MDP, optimal strategies for *reaching* any state in W^O can be computed in polynomial time, by Fact 3. If O is any of the objectives mentioned in the statement of Lemma 3, then by that Lemma, in order to compute optimal strategies and values for objective O, it suffices to compute the set W^O and optimal strategies for the objective O in states in W^O. The resulting optimal strategy "stitches" these and the optimal strategy for reaching W^O.

Proposition 3. *For every MDP, $\mathcal{G}$, and an objective $O = LimInf(= -\infty)$, $LimInf(= +\infty)$, or $Mean(> 0)$, the problem whether $s \in W^O$ is decidable in **P**-time. If $s \in W^O$, then a strategy optimal in s is computable in **P**-time.*

Proof. (Sketch.) From Lemma 2 we know that $\mathit{LimInf}(=+\infty)$ is equivalent to $\mathit{Mean}(>0)$, and thus we only have to consider $O = \mathit{LimInf}(=-\infty)$ and $O = \mathit{Mean}(>0)$. For a uniform presentation, we assume that $\mathcal{G}$ is a maximizing MDP, and consider two cases: $O = \mathit{Mean}(>0)$, and $\mathit{LimInf}(>-\infty)$. The remaining cases were solved in [2] – Theorem 3.1 there solves the case $O = \mathit{LimInf}(=-\infty)$, and Section 3.3 solves $O = \mathit{Mean}(\leq 0)$.

$O = \mathit{Mean}(>0)$: We design an algorithm to decide whether $\max_\sigma \mathbb{P}^\sigma_s(\mathit{Mean}(>0)) = 1$, using the existing polynomial time algorithm, based on linear programming, for maximizing the *expected* mean payoff and computing optimal strategies for it (see, e.g., [15]). Note that, as shown in the appendix (Lemma 7), it does not matter whether lim inf or lim sup is used in the definition of $\mathit{Mean}(>0)$. Under a memoryless strategy σ, almost all runs in $\mathcal{G}$ reach one of the bottom strongly connected components (BSCCs). Almost all runs initiated in some BSCC, C, visit all states of C infinitely often, and it follows from standard Markov chain theory (e.g., [14]) that almost all runs in C have the same mean payoff, which equals the expected mean payoff for the Markov chain induced by C.

Procedure `MP`(s)

Data: A state s.
Result: Decide $\mathit{Val}^{\mathit{Mean}(>0)}(s) \stackrel{?}{=} 1$. If yes, return a strategy σ with $\mathbb{P}^\sigma_s(\mathit{Mean}(>0)) = 1$.

1 **repeat**
2   Compute a strategy σ_{mp} maximizing the expected mean payoff.
3   **if** $\mathbb{E}^{\sigma_{mp}}_s(\mathit{mean\ payoff}) \leq 0$ **then return** No
4   Fix σ_{mp} to get a Markov chain on $\mathcal{G}$. Find a BSCC, C, with mean payoff almost surely positive.
5   Compute a strategy σ_C maximizing the probability of $\mathit{Reach}(C)$.
6   **foreach** v *with* $\mathbb{P}^{\sigma_C}_v(\mathit{Reach}(C)) = 1$ **do**
7   Remove state v.
8   **if** $v \in C$ **then** $\sigma(v) \leftarrow \sigma_{mp}(v)$ **else** $\sigma(v) \leftarrow \sigma_C(v)$
9 **until** *s is cut off*
10 **return** (Yes, σ)

The algorithm is given here as Procedure `MP`(s). Both step 2, as well as verifying the condition from step 4, can be done in P-time, because, as observed above, this is equivalent to verifying that the expected mean payoff in C is positive, which can be done in P-time (see [15, Theorem 9.3.8]). Step 5 can be done in P-time by Fact 3. To obtain a formally correct MDP, we introduce a new state z with a self-loop, and after the removal of any state v in step 7 of the for loop, we redirect all stochastic transitions leading to v to this new state z, and eliminate all other transitions into v. The reward of the new state z is set to 0. This will not affect the sign of subsequent optimal expected mean payoffs starting from s, unless s has been already removed. Thus, the algorithm can be implemented so that each iteration of the repeat-loop takes P-time, and so the algorithm terminates in P-time, since in each iteration at least one state must be removed. If the algorithm outputs (Yes, σ) then clearly $\mathbb{P}^\sigma_s(\mathit{Mean}(>0)) = 1$. On the other hand, by an easy induction on the number of iterations of the repeat-loop one can prove that if $\mathit{Val}^{\mathit{Mean}(>0)}(s) = 1$ then the following is an invariant of line 9: either s has been removed, or the maximal expected mean payoff starting in s is positive. In particular, the algorithm cannot output No. Thus we have completed the case when $O = \mathit{Mean}(>0)$.

$O = \mathit{LimInf}(>-\infty)$: Recall first the auxiliary objective $\mathit{All}(\geq 0) := \{w \in \mathit{Run}_\mathcal{G} \mid \forall n \geq 0 : \sum_{j=0}^{n} r(w(j)) \geq 0\}$ from the proof of Proposition 1, and also the sets $W_\infty = \{v \mid \mathit{Val}^{\mathit{LimInf}(=+\infty)}(v) = 1\}$, and $W_+ = \{v \mid \mathit{Val}^{\mathit{All}(\geq 0)}(v) = 1\}$. Note that $W_\infty = W^{\mathit{Mean}(>0)}$, by Lemma 2. Finally, recall from the equation (1) in the proof of Proposition 1, that the probability of $\mathit{LimInf}(>-\infty)$ is maximized by almost surely reaching $W_\infty \cup W_+$

and then satisfying $All(\geq 0)$ or $LimInf(=+\infty)$. We note that the strategy σ_+, optimal for $All(\geq 0)$, from the proof of Proposition 1, can be computed in polynomial time by [3, Theorem 12]. The results on $Mean(>0)$ and Fact 3 conclude the proof. $\square$

Now we finish the proof of Theorem 2. Proposition 3 and Fact 3 together establish the MDP case. Establishing the NP $\cap$ coNP upper bound for SSGs proceeds in a standard way: guess a strategy for one player, fix it to get a MDP, and verify in polynomial time (Proposition 3) that the other player cannot do better than the given value p. To decide whether, e.g., $Val^{O}(s) \geq p$, guess a strategy σ for Max, fix it to get an MDP, and verify that Min has no strategy π so that $\mathbb{P}_s^{\sigma,\pi}(O) < p$. Other cases are similar. $\square$

Finally, we show that the upper bound from Theorem 2 is hard to improve upon:

Proposition 4. *Assume that a SSG, $\mathcal{G}$, a state s, and a reward function r are given, and let O be an objective from $\{LimInf(=-\infty), LimInf(=+\infty), Mean(>0)\}$. Moreover, assume the property (promise) that either $Val^{O}(s) = 1$ or $Val^{O}(s) = 0$. Then deciding which is the case is at least as hard as Condon's [5]* quantitative *reachability problem w.r.t. polynomial time reductions.*

Proof. The problem studied by Condon [5] is: given a SSG, $\mathcal{H}$, an initial state s, and a target state t, decide whether $Val^{Reach(t)}(s) \geq 1/2$. Deciding whether $Val^{Reach(()t)}(s) > 1/2$ is P-time equivalent. Moreover, we may safely assume there is a state $t' \neq t$, such that whatever strategies are employed, we reach t or t', with probability 1. Consider the following reduction: given a SSG, $\mathcal{H}$, with distinguished states s, t, and t' as above, produce a new SSG, $\mathcal{G}$, with rewards as follows: remove all outgoing transitions from t and t', add transitions $t \hookrightarrow s$ and $t' \hookrightarrow s$, and make both t and t' belong to Max. Let r be the reward function over states of $\mathcal{G}$, defined as follows: $r(t) := -1$, $r(t') := +1$ and $r(z) := 0$ for all other $z \notin \{t, t'\}$. It follows from basic random walk theory that in $\mathcal{G}$, $Val^{LimInf(=-\infty)}(s) = 1$ if $Val^{Reach(t)}(s) \geq 1/2$, and $Val^{LimInf(=-\infty)}(s) = 0$ otherwise. Likewise, $Val^{LimInf(=+\infty)}(s) = 1$ if $Val^{Reach(t')}(s) > 1/2$, and $Val^{LimInf(=+\infty)}(s) = 0$ otherwise, and identically for the objective $Mean(>0)$ which we already showed to be equivalent to $LimInf(=+\infty)$. $\square$

4 Termination

In this section we prove Theorem 1. We continue viewing OC-SSGs as finite-state SSGs with rewards, as discussed in the introduction. However, for notational convenience this time we consider rewards on *transitions* rather than on states. It is easy to observe that Theorem 2 remains valid even if we sum rewards on transitions instead of rewards on states in the definition of $LimInf(=-\infty)$. We fix a SSG, $\mathcal{G}$, with state set V, and a reward function r.

Lemma 4. *Assume that $j \geq |V|$. Then for all states s: $Val^{Term(j)}(s) = 1$ iff $Val^{LimInf(=-\infty)}(s) = 1$.*

Proof. If $\mathcal{G}$ is a maximizing MDP, the proposition is true by results of [2, Section 4]. Consider now the general case, when $\mathcal{G}$ is a SSG. If $Val^{LimInf(=-\infty)}(s) = 1$ then clearly $Val^{Term(j)}(s) = 1$. Now assume that $Val^{Term(j)}(s) = 1$ and consider the memoryless strategy of player Min, optimal for $LimInf(=-\infty)$, which exists by Proposition 1. Fixing it, we get a maximizing MDP, in which the value of $Term(j)$ in s is, of course, still 1. We already know from the above discussion that the value of $LimInf(=-\infty)$ in s is thus also 1 in this MDP. Since the fixed strategy for Min was optimal, we get that $Val^{LimInf(=-\infty)}(s) = 1$ in $\mathcal{G}$. Thus, if $Val^{Term(j)}(s) = 1$ then $Val^{LimInf(=-\infty)}(s) = 1$. $\square$

Proof of Theorem 1. For cases where $j \geq |V|$, the theorem follows directly from Lemma 4 and Theorem 2. If $j < |V|$ then we have to perform a simple reachability analysis, similar to the one presented in [2]. The following SSG, $\mathcal{G}'$, keeps track of the accumulated rewards as long as they are between $-j$ and $|V| - j$: its set of states is $V' := \{(u, i) \mid u \in V, -j \leq i \leq |V| - j\}$.

States (u, i) with $i \in \{-j, |V| - j\}$ are absorbing, and for $i \notin \{-j, |V| - j\}$ we have $(u, i) \rightarrow (t, k)$ iff $u \rightarrow t$ and $k = i + r(u \rightarrow t)$. Every (u, i) belongs to the player who owned u. The probability of every transition $(u, i) \rightarrow (t, k)$, $u \in V_P$, is the same as that of $u \rightarrow t$. There is no reward function for $\mathcal{G}'$, we consider a reachability objective instead, given by the target set $R := \{(u, -j) \mid u \in V\} \cup \{(u, i) \mid -j \leq i \leq |V| - j, Val^{LimInf(=-\infty)}(u) = 1\}$. Finally, let us observe that, by Lemma 4, $Val^{Reach(R)}((s, 0)) = 1$ iff $Val^{Term(j)}(s) = 1$. Since the size of $\mathcal{G}'$ is polynomial in the size of $\mathcal{G}$, Theorem 1 is proved. $\square$

Proposition 5. *For all $j > 0$, $s \in V$, there are pure strategies, σ for Max, and π for Min, such that*

1. *If $Val^{Term(j)}(s) = 1$ then σ is optimal in s for $Term(j)$.*
2. *If $Val^{Term(j)}(s) < 1$ then $\sup_\tau \mathbb{P}_s^{\tau,\pi}(Term(j)) < 1$.*

Moreover, σ is memoryless, and π only uses memory of size $|V|$. Such strategies can be computed in P-time for MDPs.

The proof goes along the lines of the proof of Theorem 1. It can be found in the appendix, Section A.6, together with an example that shows the memory use in π is necessary.

Similarly, both $Val^{Term(j)}(s) = 0$ and $Val^{Term(j)}(s) > 0$ are witnessed by pure and memoryless strategies for the respective players. Deciding which is the case is in P-time, by assigning the random states to player Max, obtaining a non-stochastic 2-player one-counter game, and using, e.g., [3, Theorem 12]. Finally, we note that from Proposition 4 and Lemma 4, it follows that:

Corollary 1. *Given an SSG, $\mathcal{G}$, and reward function r, deciding whether the value of the termination objective $Term(j)$ equals 1 is at least as hard as Condon's [5]* quantitative *reachability problem, w.r.t. P-time many-one reductions.*

References

1. N. Berger, N. Kapur, L. J. Schulman, and V. Vazirani. Solvency Games. In *Proc. of FSTTCS'08*, 2008.
2. T. Brázdil, V. Brožek, K. Etessami, A. Kučera, and D. Wojtczak. One-Counter Markov Decision Processes. In *ACM-SIAM SODA*, pages 863–874, 2010. Full tech report: CoRR, abs/0904.2511, 2009. http://arxiv.org/abs/0904.2511.
3. T. Brázdil, P. Jančar, and A. Kučera. Reachability Games on Extended Vector Addition Systems with States. In *Proc. 37th Int. Coll. on Automata, Languages, and Programming (ICALP)*, 2010. To appear. Full tech report: FIMU-RS-2010-02, Faculty of Informatics, Masaryk University.
4. K. L. Chung. *A Course in Probability Theory*. Academic Press, 3rd edition, 2001.
5. A. Condon. The Complexity of Stochastic Games. *Inform. and Comput.*, 96:203–224, 1992.
6. C. Courcoubetis and M. Yannakakis. Markov decision processes and regular events. *IEEE Trans. Automat. Control*, 43(10):1399–1418, 1998.
7. K. Etessami, D. Wojtczak, and M. Yannakakis. Quasi-birth-death processes, tree-like QBDs, probabilistic 1-counter automata, and pushdown systems. In *Proc. 5th Int. Symp. on Quantitative Evaluation of Systems (QEST)*, pages 243–253, 2008.
8. K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *Proc. 32nd ICALP*, pages 891–903, 2005.
9. K. Etessami and M. Yannakakis. Efficient qualitative analysis of classes of recursive Markov decision processes and simple stochastic games. In *Proc. of 23rd STACS'06*. Springer, 2006.

10. H. Gimbert. Pure stationary optimal strategies in markov decision processes. In *STACS*, pages 200–211, 2007.
11. H. Gimbert and F. Horn. Solving Simple Stochastic Tail Games. In *ACM-SIAM Symposium on Discrete Algorithms (SODA10)*, pages 847–862, 2010.
12. H. Gimbert and W. Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games. Technical Report hal-00438359, HAL, http://hal.archives-ouvertes.fr/hal-00438359/, 2009.
13. D. A. Martin. The Determinacy of Blackwell Games. *The Journal of Symbolic Logic*, 63(4):1565–1581, December 1998.
14. J. R. Norris. *Markov chains*. Cambridge University Press, 1998.
15. M. L. Puterman. *Markov Decision Processes*. J. Wiley and Sons, 1994.

A Appendix

In the entire appendix, when referring to MDPs and SSGs, we mean *finite-state* MDPs and SSGs.

A.1 Proof of Proposition 1, objectives $\mathit{LimInf}(< +\infty)$, $\mathit{Mean}(> 0)$ and $\mathit{Mean}(\leq 0)$

An objective O is *submixing* if for every run $w = u_1 v_1 u_2 v_2 \cdots u_k v_k \cdots$, such that u_i and v_i are finite paths for every i, and such that both $u = u_1 u_2 \cdots u_k \cdots$ and $v = v_1 v_2 \cdots v_k \cdots$ are also runs, we have $w \in O \implies (u \in O \vee v \in O)$. This notion is taken directly from [10], where it has been defined in a more general setting. (See also [2, Section 3] for more details.) By [10, Theorem 1], for every maximizing MDP and every tail submixing objective, O, player Max has a pure and memoryless optimal strategy.

Lemma 5. *The objective* $\mathit{LimInf}(< +\infty)$ *is a submixing and tail objective.*

Proof. Obviously it is tail. As for the submixing property, let $\{a_i\}_{i=1}^{\infty}$ be a sequence of numbers, and consider an arbitrary splitting of this sequence into two infinite subsequences $\{b_i\}_{i=1}^{\infty}$, $\{c_i\}_{i=1}^{\infty}$. For $x \in \{a, b, c\}$ we define

$$L_x \coloneqq \liminf_{n\to\infty} \sum_{i=1}^{n} x_i \ .$$

It is easy to verify that if at least one of L_b, L_c is finite, or if they are infinite with the same sign, then $L_a \geq L_b + L_c$. In particular, if $L_a < \infty$ then $\min\{L_b, L_c\} < \infty$. Applying this to the sequences of rewards finishes the proof. ◻

By changing the lim inf to lim sup in the definition of $\mathit{Mean}(> 0)$ we obtain a new objective:

$$\mathit{Mean}(> 0)_+ \coloneqq \{w \in \mathit{Run}_{\mathcal{G}} \mid \limsup_{n\to\infty} \sum_{i=0}^{n-1} r(w(i))/n > 0\} \ .$$

Lemma 6. *Both* $\mathit{Mean}(> 0)_+$ *and* $\mathit{Mean}(\leq 0)$ *are tail and submixing.*

Proof. Both are clearly tail. For the submixing property, let us start with $\mathit{Mean}(> 0)_+$. Let $A = \{a_i\}_{i=1}^{\infty}$ be a sequence of numbers, and consider an arbitrary splitting of this sequence into two infinite subsequences $B = \{b_i\}_{i=1}^{\infty}$, $C = \{c_i\}_{i=1}^{\infty}$. For a fixed $n \geq 1$ denote by $n_b \leq n$ the number of elements of B among the first n elements of A. Then, assuming $n_b < n$

$$\frac{\sum_{i=1}^{n} a_i}{n} = \frac{\sum_{i=1}^{n_b} b_i + \sum_{i=1}^{n-n_b} c_i}{n} = \frac{\sum_{i=1}^{n_b} b_i}{n_b} \cdot \frac{n_b}{n} + \frac{\sum_{i=1}^{n-n_b} c_i}{n-n_b} \cdot \left(1 - \frac{n_b}{n}\right) \ .$$

Consequently,

$$\frac{\sum_{i=1}^{n} a_i}{n} \leq \frac{\sum_{i=1}^{n_b} b_i}{n_b} \qquad \text{or} \qquad \frac{\sum_{i=1}^{n} a_i}{n} \leq \frac{\sum_{i=1}^{n-n_b} c_i}{n-n_b} \ ,$$

and thus there is $x \in \{b, c\}$ such that

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^{n} a_i}{n} \leq \limsup_{n\to\infty} \frac{\sum_{i=1}^{n} x_i}{n} \ .$$

The proof for $\mathit{Mean}(\leq 0)$ proceeds similarly, only with reversed signs. ◻

Now we show that $Mean(>0)$ is equivalent to $Mean(>0)_+$ for memoryless strategies.

Lemma 7. *Under a memoryless strategy, σ, for a MDP, $\mathcal{G}$, with a reward function, r, for almost all runs, w:*

$$\liminf_{n\to\infty} \sum_{i=0}^{n-1} r(w(i))/n = \limsup_{n\to\infty} \sum_{i=0}^{n-1} r(w(i))/n \ .$$

Proof. Fix σ to get a Markov chain on the states of $\mathcal{G}$. Almost all runs visit some bottom strongly connected component (BSCC), and the above equality establishes a prefix independent property. We thus safely assume that w starts in a BSCC, C. On C, σ induces an irreducible Markov chain, and applying the Ergodic theorem (see Theorem 1.10.2 from [14]) finishes the proof. □

Lemma 8. *For every maximizing MDP, there is always a pure and memoryless strategy, σ, optimal for $Mean(>0)$.*

Proof. Choose σ to be optimal for $Mean(>0)_+$. This is possible, because $Mean(>0)_+$ is a submixing and tail objective. Observe that since $Mean(>0) \subseteq Mean(>0)_+$, we have $Val^{Mean(>0)}(s) \leq Val^{Mean(>0)_+}(s)$ for all states s. Finally, due to Lemma 7, for all states s:

$$Val^{Mean(>0)_+}(s) = \mathbb{P}^{\sigma}_{s}(Mean(>0)_+) = \mathbb{P}^{\sigma}_{s}(Mean(>0)) \leq Val^{Mean(>0)}(s) \ .$$

□

One may be tempted to believe that all of the objectives we study are submixing. This is, however, not true for $LimInf(=+\infty)$ and $LimInf(>-\infty)$, where we have to employ other methods for proving the existence of pure and memoryless optimal strategies.

Lemma 9. *The objectives $LimInf(=+\infty)$ and $LimInf(>-\infty)$ are not submixing.*

Proof. Consider the following finite sequences A_k over $\{\pm 1\}$, parametrized by $k \geq 1$, and defined inductively by $A_1 := +1, -1$, and $A_{k+1} := +1, A_k, -1$. We build an infinite sequence $A = \{a_i\}_{i=1}^{\infty}$ by concatenating them, $A := A_1, A_2, A_3, \ldots$. Obviously $\liminf \sum_{i=1}^{n} a_i = 0$. Now we define two particular subsequences of A, denoted by $B := \{b_i\}_{i=1}^{\infty}$, $C := \{c_i\}_{i=1}^{\infty}$, so that

$$\liminf_{n\to\infty} \sum_{i=1}^{n} b_i = \liminf_{n\to\infty} \sum_{i=1}^{n} c_i = -\infty \ . \tag{2}$$

We do it inductively by saying for every $k \geq 1$, whether the k-th element, a_k, of A belongs to B, or C. Assume we have already decided for each of the first k elements of A whether it belongs to B or C, so that we have already defined the finite prefixes $b_1, \ldots, b_M$ of B, and $c_1, \ldots, c_N$ of C. Set $s_B^i := \sum_{j=1}^{i} b_j$, and similarly $s_C^i := \sum_{j=1}^{i} c_j$. If either $a_{k+1} = -1$ and $\min_{i=1}^{M} s_B^i \geq \min_{i=1}^{N} s_C^i$, or $a_{k+1} = 1$ and $\min_{i=1}^{M} s_B^i < \min_{i=1}^{N} s_C^i$, then a_{k+1} belongs to B, otherwise it belongs to C. It is easy to verify that for every number m we have some n such that $s_B^n < m$, and some n' such that $s_C^{n'} < m$. (In fact, this is the idea behind the construction – the sequences B and C take turns in achieving lower and lower partial sums.) Thus (2) is true. As the sequence A can be easily obtained as a sequence of rewards associated to a run of a very simple MDP with rewards, this proves that $LimInf(>-\infty)$ is not submixing.

Similarly goes the proof that $LimInf(=+\infty)$ is not submixing. Along the lines of the previous proof, just consider the following modifications: Take the sequence $A = \{a_i\}_{i=1}^{\infty}$ to be defined by $a_i = -1$ iff $i \equiv 0 \pmod 3$ and $a_i = +1$ otherwise. Further, in the inductive process of building the sequences B and C, denote by $z_B := |\{i \le M \mid s^i_B = 0\}|$, and by $z_C := |\{i \le N \mid s^i_C = 0\}|$. Finally, apply the rule of assigning a_{k+1} to B iff either $a_{k+1} = -1$, $z_C \ge z_B$, and $s^M_B > 0$, or $a_{k+1} = +1$ and $z_C < z_B$ or $s^M_B = 0$. (Here the intuition is that B and C take turns in revisiting 0 from above.) It is easy to show that for every $m \ge 0$ there is some $n \ge m$ such that $s^n_B = 0$, and some $n' \ge m$ such that $s^{n'}_C = 0$. This shows that $\liminf B = \liminf C = 0$, while $\liminf A = \infty$. As a consequence, $\mathit{LimInf}(= +\infty)$ is not submixing. $\square$

A.2 Proof of Proposition 1, objective $\mathit{LimInf}(= +\infty)$

First we set up a tool to analyze finite-state Markov chains with respect to the objective $\mathit{LimInf}(= +\infty)$. Consider a finite-state Markov chain, $\mathcal{M}$, with the underlying transition graph $(S, \mapsto)$, and with a reward function, $r : S \to \{-1, 0, +1\}$. Assume, moreover, that $\mathcal{M}$ is irreducible. Also assume that some initial state, s, is fixed. We derive here one condition sufficient for $\mathbb{P}_s(\mathit{LimInf}(= +\infty)) = 1$, and another one sufficient for $\mathbb{P}_s(\mathit{LimInf}(= +\infty)) = 0$ in $\mathcal{M}$. The conditions are parametrized by a choice of a subset $R \subseteq S$ of the states of $\mathcal{M}$. To formulate them we need the following random variables.

- V^t_k, $k \ge 0$, $t \in R$ returns the time of the k-th visit (thus "V") to t.
- G^t_k, $k \ge 0$, $t \in R$ is the reward gained ("G") between time V^t_k (inclusive) and the next visit to R (exclusive).

By standard facts from probability theory, almost all runs in $\mathcal{M}$ visit all states infinitely often. Thus these random variables are almost surely defined. For a fixed $t \in R$, all the variables G^t_k are i.i.d., and, as the expected time to visit R from t is finite, their common mean, μ_t, is well defined and finite. Observe also that the values μ_t do not depend on the choice of the initial state.

Lemma 10. *For every finite-state irreducible Markov chain, $\mathcal{M}$, and every subset, R, of states, and every $t \in R$, considering the numbers μ_s, $s \in R$, derived as above, the following is true:*

- *If $\mu_s > 0$ for all $s \in R$ then $\mathbb{P}_t(\mathit{LimInf}(= +\infty)) = 1$.*
- *If $\mu_s \le 0$ for all $s \in R$ then $\mathbb{P}_t(\mathit{LimInf}(= +\infty)) = 0$.*

Proof. We use the following random variables on runs from $\mathit{Run}_{\mathcal{M}}(t)$:

- V_k, $k \ge 1$, the time of the k-th visit to t. (Note: $V_1 \equiv 0$.)
- A_k, $k \ge 1$, the reward accumulated ("A") between time V_k (inclusive) and V_{k+1} (exclusive).
- $S_m := \sum_{k=1}^m A_k$, $m \ge 0$. ("S" for "sum". Note: $S_0 \equiv 0$.)

Since $\mathcal{M}$ is a Markov chain, we get that the variables A_k are i.i.d., in particular there is some μ such that $\mu = \mathbb{E}_t(A_k)$ for all $k \ge 1$. [5]

Claim. If all $\mu_s > 0$ then $\mu > 0$. If all $\mu_s \le 0$ then $\mu \le 0$.

Proof. For every $s \in R$ and $w \in \mathit{Run}_{\mathcal{M}}(t)$ denote by $v_s(w)$ the number of visits to s before the first revisit to t: $v_s(w) = \mathit{card}\,(\{k \ge 0 \mid w(k) = s \wedge \forall l < k : w(l) = t \implies l = 0\})$. Then, writing $R = \{t_1, \ldots, t_\ell\}$,

$$\mu = \sum_{c_1,\ldots,c_\ell \ge 0} \mathbb{P}_t\left(\bigwedge_{j=1}^{\ell} v_{t_j} = c_j\right) \cdot \sum_{j=1}^{\ell} c_j \cdot \mu_{t_j} \ .$$

Since all the coefficients of μ_s, $s \in R$ are non-negative, the claim is proved. $\square$

[5] By $\mathbb{E}$ we denote the expectation.

For $\mu \leq 0$ standard results on random walks (see [4, Theorem 8.3.4]) yield $\liminf_{n\to\infty} S_n < \infty$ almost surely. Immediately, $\liminf_{n\to\infty} \sum_{i=0}^{n} r(w(i)) < \infty$ almost surely, thus $\mathbb{P}_t(\mathit{LimInf}(=+\infty)) = 0$.
The case when $\mu > 0$ is more subtle, and we need to introduce two more random variables:

- M, the least m such that $S_m > 0$. ("M" for "maximum".)
- M', $M' := V_M$ (the actual number of steps to M).

Claim. **(cf. [4, Theorem 8.4.4])** $\mathbb{E}_t(M) < \infty$.

Claim. $\mathbb{E}_t(V_{k+1} - V_k) = \mathbb{E}_t(V_2 - V_1) < \infty$ for all $k \geq 1$.

Proof. Since $\mathcal{M}$ is a Markov chain, we get the equality. By standard results on Markov chains (see [14, Theorem 1.7.7]) we obtain $\mathbb{E}_t(V_2 - V_1) = \mathbb{E}_t(V_2) = (\pi(t))^{-1}$ where π is an invariant (and positive) distribution over the states of $\mathcal{M}$. Thus the inequality follows. □

Claim. $\mathbb{E}_t(M') < \infty$.

Proof.

$$\begin{aligned}
\mathbb{E}_t(M') &= \sum_{m=1}^{\infty} \mathbb{P}_t(M = m) \cdot \mathbb{E}_t(V_m) \\
&= \sum_{m=1}^{\infty} \mathbb{P}_t(M = m) \cdot \mathbb{E}_t((V_m - V_{m-1}) + (V_{m-1} - V_{m-2}) + \cdots + (V_2 - V_1)) \\
&= \sum_{m=1}^{\infty} \mathbb{P}_t(M = m) \cdot (m-1) \cdot \mathbb{E}_t(V_2 - V_1) \\
&= (\mathbb{E}_t(M) - 1) \cdot \mathbb{E}_t(V_2 - V_1)
\end{aligned}$$

□

As a generalization of the variable M, we define, inductively and for almost all runs from $\mathit{Run}_{\mathcal{M}}(t)$, yet another sequence M_k, $k \geq 0$ of random variables by setting $M_0 \equiv 0$, and M_{k+1} to be the least m such that $S_m > S_{M_k}$. (We get $M = M_1$.) In other words, M_k are the times when maximal rewards were achieved on revisit to t. We also define a sequence of events, Z_k, $k \geq 1$: A run $w \in \mathit{Run}_{\mathcal{M}}(t)$ is in Z_k iff there is some j, $V_{M_k} \leq j < V_{M_{k+1}}$ such that the reward accumulated on $w(0) \cdots w(j)$ is 0. ("Z" for "zero".)

Claim. $\sum_{k\geq 1}^{\infty} \mathbb{P}_t(Z_k) < \infty$.

Proof. It takes at least $S_{M_k} \geq k$ steps to gain reward 0 starting at time V_{M_k}. Since $V_{M_{k+1}} - V_{M_k}$ has the same distribution as M', we get $\mathbb{P}_t(Z_k) \leq \mathbb{P}_t(M' \geq k)$. Now

$$\sum_{k\geq 1}^{\infty} \mathbb{P}_t(Z_k) \leq \sum_{k\geq 1}^{\infty} \mathbb{P}_t(M' \geq k) = \sum_{k\geq 1}^{\infty} \sum_{l\geq k}^{\infty} \mathbb{P}_t(M' = k) = \sum_{k\geq 1}^{\infty} k \cdot \mathbb{P}_t(M' = k) = \mathbb{E}_t(M') < \infty\,.$$

□

Thus by the Borel-Cantelli lemma, the probability that Z_k occurs for infinitely many k is 0. Consequently $\liminf_{n\to\infty} \sum_{i=0}^{n} r(w(i)) > 0$ for almost all w. Similarly we can prove for all $h > 0$ that $\liminf_{n\to\infty} \sum_{i=0}^{n} r(w(i)) > h$ for almost all w. Hence, $\liminf_{n\to\infty} \sum_{i=0}^{n} r(w(i)) = \infty$ almost surely, because a countable intersection of sets of probability 1 has probability 1. Thus $\mathbb{P}_t(\mathit{LimInf}(=+\infty)) = 1$. □

Lemma 11. *The finite-memory strategy σ from the proof of Proposition 1 is optimal for $LimInf(=+\infty)$.*

Proof. Observe that fixing σ yields a finite-state Markov chain, $\mathcal{G}(\sigma)$, on the parallel composition of $\mathcal{G}$ and the finite automaton used for updating the memory of σ. Let us fix an arbitrary bottom strongly connected component (BSCC), C, of $\mathcal{G}(\sigma)$, and denote by R the states of C in which the memory of σ is being reset. We are now going to analyze, using Lemma 10, the irreducible MC, $\mathcal{M}$, induced by restricting $\mathcal{G}(\sigma)$ to C. Fix an arbitrary $s \in R$. Recall, that the variable G_k^u, defined before stating Lemma 10, returns the reward accumulated between the k-th visit to s and the next visit to R. It is easy to verify that the common mean, μ_u, of G_k^u is equal to the mean of the stopping time T_s introduced in the main text of the proof, ant thus positive. Therefore Lemma 10 guarantees that for every state $s \in R$ lying in some BSCC we have $\mathbb{P}_s(LimInf(=+\infty)) = 1$. Since $\mathcal{G}(\sigma)$ is finite, almost every run in it reaches some BSCC and every state in it. Because $LimInf(=+\infty)$ is a tail objective we get $\mathbb{P}_s(LimInf(=+\infty)) = 1$ for every state s. □

Lemma 12. *In a maximizing MDP, $\mathcal{G}$, with value* 1 *in all states, given a pure finite-memory strategy σ optimal for $LimInf(=+\infty)$, a pure and memoryless optimal strategy τ can be constructed.*

Proof. As in the proof of Lemma 11, given a finite-memory strategy, ϱ, we denote by $\mathcal{G}(\varrho)$ the finite-state Markov chain, states of which are pairs (s, q) where s is a state of $\mathcal{G}$, and q is a state of the finite automaton representing the memory of ϱ. Probabilities are obtained in the natural way from ϱ and $\mathcal{G}$. Consider now the Markov chain $G(\sigma)$. The initial state is (s_0, q_0) where s_0, q_0 are initial states of $\mathcal{G}$, and the automaton for σ, respectively. For technical reasons we assume that for each q there is at most one s so that (s, q) is reachable from (s_0, q_0).

If there are two states, $q \neq p$, of the automaton for σ, and a state s of $\mathcal{G}$ such that both (s, q) and (s, p) are reachable from (s_0, q_0), we call both q and p *ambiguous*. If there is no ambiguous state, σ is already memoryless. If there are ambiguous states, we show how to modify σ to get another pure and finite-memory optimal strategy σ', such that the associated Markov chain, $\mathcal{G}(\sigma')$, has fewer ambiguous states. As there are only finitely many ambiguous states in the beginning, repeating this process inevitably leads to the optimal pure and memoryless strategy τ.

We thus assume that there is a state s of $\mathcal{G}$ such that $A := \{(s, q) \mid (s, q) \text{ is reachable from } (s_0, q_0)\}$ has at least two elements. For every fixed choice of $(s, q) \in A$ we now define a new finite-memory strategy σ_q. This is derived by modifying the finite automaton for σ so that all transitions leading to some p, where $(s, p) \in A$, are redirected to q. From this, due to our technical assumption, already follows that σ' has fewer ambiguous states. It remains to prove that there is some q such that σ_q is optimal.

There are two cases to consider. First, consider the situation where there is $(s, q) \in A$ such that with some positive probability states from $A \setminus \{(s, q)\}$ are visited only finitely often in $\mathcal{G}(\sigma)$. This implies that there is a BSCC, S, of $\mathcal{G}(\sigma)$, such that $|S \cap A| \leq 1$. We choose (s, q) so that it minimizes the distance (in the transition graph of $\mathcal{G}(\sigma)$) to S among the states from A. This implies that, starting in (s, q), states from $A \setminus \{(s, q)\}$ are avoided with some positive probability, δ. We now prove that σ_q is optimal. Indeed, let $\neg A$ be the event of not visiting A, and let E be an arbitrary event. Then $\mathbb{P}^{\sigma}_{(s_0,q_0)}(E \mid \neg A) = \mathbb{P}^{\sigma_q}_{(s_0,q_0)}(E \mid \neg A)$. On the other hand, every run in $\mathcal{G}(\sigma)$ visiting A projects to $\mathcal{G}(\sigma_q)$, as a run w visiting (s, q). Here we have two possibilities. Either $\delta = 1$, and we set w_q to be the suffix of w starting with the first occurrence of (s, q). Or $\delta < 1$, implying that $S \cap A = \emptyset$ and thus (s, q) is not in a BSCC. Thus on almost all runs (s, q) is visited only finitely many times, and we may define w_q to be the suffix starting with the last occurrence of (s, q) in w. For every event E we define the set $E' := \{w \in Run_{\mathcal{G}} \mid w_q \in E\}$. Denoting simply by A the event of visiting A, it is easy to verify for all E that $\mathbb{P}^{\sigma}_{(s_0,q_0)}(E \mid A) = \mathbb{P}^{\sigma_q}_{(s_0,q_0)}(E' \mid A)$. Since $LimInf(=+\infty)$ is tail, we have $LimInf(=+\infty)' \subseteq LimInf(=+\infty)$. Thus almost all runs in $\mathcal{G}(\sigma_q)$ satisfy $LimInf(=+\infty)$.

If the first case does not apply then there must be a BSCC, S, such that $|S \cap A| \geq 2$. Using Lemma 10 for $\mathcal{G}(\sigma)$ restricted to S, with $R = A$, we obtain that there must be at least one $(s, q) \in A$ such that the expected accumulated reward until revisiting A, μ_q, is positive. Observe that $(S \setminus A) \cup \{(s,q)\}$ forms a BSCC, S_q, in $\mathcal{G}(\sigma_q)$. Using Lemma 10 on $\mathcal{G}(\sigma_q)$ restricted to S_q, with $R = \{(s,q)\}$, we obtain that all runs in $\mathcal{G}(\sigma_q)$ started in S_q satisfy $\mathit{LimInf}(=+\infty)$. Because, similarly to the previous case, almost all runs in $\mathcal{G}(\sigma_q)$ remain either unaffected or visit S_q, we obtain again that almost all runs in $\mathcal{G}(\sigma_q)$ satisfy $\mathit{LimInf}(=+\infty)$. ◻

A.3 Proof of Proposition 1, objective $\mathit{LimInf}(>-\infty)$

Recall that $W_\infty = \{s \mid \mathit{Val}^{\mathit{LimInf}(=+\infty)}(s) = 1\}$, and $W_+ = \{s \mid \mathit{Val}^{\mathit{All}(\geq 0)}(s) = 1\}$.

Lemma 13. *For every maximizing MDP, $\mathcal{G}$, and its state, s, if $\mathit{Val}^{\mathit{LimInf}(>-\infty)}(s) = 1$ then there is a strategy, τ, such that $\mathbb{P}^\tau_s(\mathit{Reach}(W_\infty \cup W_+)) = 1$.*

Proof. If $\mathit{Val}^{\mathit{Reach}(W_\infty)}(s) = 1$ then we are already done for this state. Assume $\mathit{Val}^{\mathit{Reach}(W_\infty)}(s) < 1$.

Claim. There is a strategy, τ, such that

1. $\mathbb{P}^\tau_s(\mathit{LimInf}(>-\infty)) = 1$;
2. τ restricted to W_∞ is memoryless and $\mathbb{P}^\tau_v(\mathit{LimInf}(=+\infty)) = 1$ for all $v \in W_\infty$;
3. almost all $w \in \mathit{Run}_{\mathcal{G}(\tau)}(s)$ which do not visit W_∞ satisfy

$$\infty > \liminf_{n\to\infty} \sum_{i=0}^{n} r(w(i)) > -\infty \ . \tag{3}$$

Proof. Choose a strategy satisfying 1, it must exist by [11, Theorem 3.1]. By Proposition 1 for $\mathit{LimInf}(=+\infty)$ we obtain 2, because $\mathit{LimInf}(=+\infty) \subseteq \mathit{LimInf}(>-\infty)$. On the other hand, runs avoiding W_∞ belong to $\mathit{LimInf}(=+\infty)$ with probability 0, as a consequence of Lemma 3 for the objective $\mathit{LimInf}(=+\infty)$, [6] proving 3. ◻

We now define an event $\mathit{Inf}(v)$ for all states v. Consider a run w, satisfying (3). There must be some integer ℓ such that $\sum_{i=0}^{n} r(w(i)) \geq \ell$ for all $n \geq 0$. Choosing the greatest such ℓ, there is some index, j, such that $\sum_{i=0}^{j} r(w(i)) = \ell$. We call the smallest such j to be the *minimum* of w, and ℓ is said to be the *minimal value* of w. According to this, we define inductively the following functions: $M_1(w)$ is the minimum of w, and, given $n := M_k(w)$, and the suffix $w' = w(n{+}1)\, w(n{+}2) \cdots$ of w, we set $M_{k+1}(w) := M_1(w') + n + 1$ for $k \geq 1$. Further, $m_k := \sum_{i=0}^{M_k(w)} r(w(i))$. (See also Figure 1 for an example.) The sequence $\{m_k\}_{k=1}^{\infty}$ is non-decreasing and, due to the first inequality in (3) also bounded, hence it has a well defined finite limit, $\bar{m}$. Given some state v, we define an event, $\mathit{Inf}(v)$, by the condition that there are infinitely many k such that the state visited at time M_k is v and $m_k = \bar{m}$.

Claim. For all states v, if $\mathbb{P}^\tau_v(\mathit{Inf}(v)) > 0$ then $v \in W_+$.

Proof. Fix a state v satisfying the assumption of the claim. Note that due to our choice of τ, for all such v, W_∞ is not reached on a run from $\mathit{Inf}(v)$. Observe that $\mathit{Inf}(v)$ is tail, so by [11, Theorem 3.1] there is a state v' and a strategy π such that $\mathbb{P}^\pi_{v'}(\mathit{Inf}(v)) = 1$. In particular, this must be true for $v' = v$ since $\mathit{Inf}(v) \subseteq \mathit{Reach}(v)$, and the objective is tail. Finally, the strategy can be chosen so that almost surely $m_k = 0$ for all $k \geq 1$. In other words, $\mathbb{P}^\pi_v(\mathit{All}(\geq 0)) = 1$. In particular, $v \in W_+$. ◻

Since there are only finitely many states, the union $\bigcup_v \mathit{Inf}(v)$ has probability 1 on the condition of not reaching W_∞. The last claim showed that $\mathbb{P}^\tau_s(\mathit{Reach}(W_+) \mid \mathit{Inf}(v)) = 1$ for all $v \in V$ with $\mathbb{P}^\tau_s(\mathit{Inf}(v)) > 0$. This proves $\mathbb{P}^\tau_s(\mathit{Reach}(W_\infty \cup W_+)) = 1$. ◻

[6] A careful reader may suspect a circular dependency since Lemma 3 uses Proposition 1. This is, however, a correct use, since it only uses the proposition for $\mathit{LimInf}(=+\infty)$, which has already been proved.

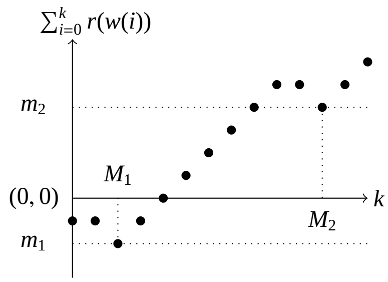

Fig. 1. An example of a run and its minima.

A.4 Proof of Lemma 1

Lemma 1. *Let $\mathcal{M}$ be a finite, strongly connected (irreducible) Markov chain, and O be a tail objective. Then there is $x \in \{0, 1\}$ such that $\mathbb{P}_s(O) = x$ for all states s.*

Proof. From every state, s, every other state, t, is visited almost surely. O is tail, thus $\mathbb{P}_s(O) = \mathbb{P}_t(O)$. Assume that $\mathbb{P}_s(O) > 0$ for some s, and thus for all s. Since a Markov chain is a special case of a SSG, we directly apply [11, Theorem 3.2] and get that $\mathbb{P}_s(O) = 1$. ☐

A.5 Proof of Lemma 2

Lemma 14. *Let $\mathcal{M}$ be an irreducible Markov chain with rewards on states, and s a fixed state of $\mathcal{M}$. If $\mathbb{P}_s(LimInf(= +\infty)) = 1$ then $\mathbb{P}_s(Mean(> 0)) = 1$.*

Proof. We fix s as a starting state. Denote by X_k, $k \geq 1$ the reward accumulated between the k-th (inclusive) and k+1-st (exclusive) visit to s. Since $\mathcal{M}$ is a Markov chain, these variables are i.i.d.; we denote by μ their common mean. Choosing $R = \{s\}$ in Lemma 10 yields that $\mu > 0$. Thus the sums, $S_\ell := \sum_{k=1}^{\ell} X_k$ define a homogeneous random walk with a positive drift. Define:

- V_k, $k \geq 1$ to be the time of the k-th visit to s (note: $V_1 \equiv 0$),
- M to be the least k such that $S_k > 0$, and
- $M' := V_M$.

By Claim A.2 from the proof of Lemma 10 we know that $\mathbb{E}_s(M') < \infty$. Further we define:

- $M_0 \equiv 0$, and M_k, $k > 0$ to be the least m such that $S_m > S_{M_{k-1}}$, and
- $Y_k := V_{M_k+1} - V_{M_{k-1}+1}$, $k > 0$. (Note: $\sum_{k=1}^{n} Y_n = V_{M_n+1}$.)

The variables in the sequence Y_k are independent and distributed identically with M', thus we may apply the strong law of large numbers (see, e.g., Theorem 1.10.1 in [14]) and obtain that almost surely

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} Y_n}{n} = \lim_{n\to\infty} \frac{V_{M_n+1}}{n} = \mathbb{E}_s(M') < \infty \,.$$

Because $S_{M_n+1} \geq n$, we have almost surely

$$\lim_{n\to\infty} \frac{S_{M_n+1}}{V_{M_n+1}} \geq \lim_{n\to\infty} \frac{n}{V_{M_n+1}} = \frac{1}{\mathbb{E}_s(M')} > 0 \,.$$

Because the leftmost term is equal to the mean payoff, we conclude that $\mathbb{P}_s(Mean(> 0)) = 1$. ☐

A.6 Proof of Proposition 5

Recall the SSG $\mathcal{G}'$ with the reachability objective R from the proof of Theorem 1. This game emulates playing $\mathcal{G}$ until (1) the accumulated reward exceeds $|V| - j$ or a state u with $\mathit{Val}^{\mathit{LimInf}(=-\infty)}(u) = 1$ is visited – then, by Lemma 4 the players may switch to optimizing the probability of $\mathit{LimInf}(= -\infty)$ instead – or until (2) the accumulated reward is $-j$. Memoryless strategies for $\mathcal{G}'$ induce strategies for $\mathcal{G}$ which use memory of size $|V|$ to store the accumulated reward until it exceeds $|V| - j$ or hits $-j$. From this and from the analysis in the proof of Theorem 1 we can see that the strategies σ and π from the statement of the proposition, are easy to construct, with the promised time complexity, to be pure and using only a finite memory of size $|V|$.

The last thing to show is how to transform σ to some memoryless σ', preserving the optimality for $\mathit{Term}(j)$ in s. Restricted to states u with $\mathit{Val}^{\mathit{LimInf}(=-\infty)}(u) = 1$, σ is already memoryless. Call these u *safe*. We set $\sigma'(u) = \sigma(u)$ for every safe u. We further call *unsafe* those states u which are not safe, but there is some strategy τ such that $\mathbb{P}_s^{\sigma,\tau}(\mathit{Reach}(u)) > 0$. Unsafe states may have been visited with various accumulated rewards so far, but from what we already proved it follows that all these accumulated rewards lie between $-j$ and $|V| - j$ (excl.). For an unsafe u, denote by i_u the maximal such accumulated reward, and by w_u some history along which this was accumulated. It remains to define σ' for unsafe u. We simply set $\sigma'(u) = \sigma(w_u)$. Since, under σ', no unsafe state is reached from a safe state, σ' is still optimal for $\mathit{Term}(j)$ in all safe states. Consider an unsafe u, and some i, $-j < i \leq i_u$, and an arbitrary strategy π' for Min. Then in $\mathcal{G}$, under the strategies (σ', π'), on condition that u was visited with an accumulated reward i, almost all runs from u either visit a safe state, or the accumulated reward reaches $-j$ at some point, or an unsafe state t is visited, and at the same time the accumulated reward is at most $i_t + i - i_u$. Thus by double induction, first on $|V| - j - i_u$ then on i, for all unsafe u and $i \leq i_u$ we have that σ is optimal for $\mathit{Term}(i)$ in u. Thus σ is pure, memoryless and optimal for $\mathit{Term}(j)$ in s. ◻

An example where memory for Min is needed. This example shows that the strategy π of player Min from Proposition 5 may indeed have to use memory. Consider this minimizing MDP:

- States: $v, \mathit{low}, \mathit{up}, \mathit{back}, \mathit{down}$; Min owns $V_\perp = \{v\}$.
- Transitions (and their rewards): $v \to \mathit{back}$ (reward 0), $v \to \mathit{low}$ (-1), $\mathit{low} \to \mathit{up}$ $(+1)$, $\mathit{up} \to \mathit{up}$ $(+1)$, $\mathit{back} \to \mathit{down}$ (0), $\mathit{back} \to v$ $(+1)$, $\mathit{down} \to \mathit{down}$ (-1).
- From probabilistic states the successor is chosen uniformly among available transitions.

Then for all $j > 0$: $\mathit{Val}^{\mathit{Term}(j)}(v) < 1$, as witnessed by the strategy choosing *back* as a successor of v whenever the reward accumulated so far is 1, and *low* in all other cases. However, there are only two pure and memoryless strategies for Min:

- choosing the transition $v \to \mathit{back}$ makes the probability of $\mathit{Term}(j)$ to be 1 for all $i > 0$;
- choosing $v \to \mathit{low}$ makes the probability of $\mathit{Term}(1)$ to be 1.