Department of Computer Science

On Stochastic Games with Multiple Objectives

Taolue Chen, Vojtěch Forejt, Marta Kwiatkowska, Aistis Simaitis, and Clemens Wiltsche

CS-RR-13-06
(Updated 20/06/2013)

Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford, OX1 3QD

On Stochastic Games with Multiple Objectives

Taolue Chen, Vojtěch Forejt, Marta Kwiatkowska,
Aistis Simaitis, and Clemens Wiltsche

Department of Computer Science, University of Oxford, United Kingdom

Abstract. We study two-player stochastic games, where the goal of one player is to satisfy a formula given as a positive boolean combination of expected total reward objectives and the behaviour of the second player is adversarial. Such games are important for modelling, synthesis and verification of open systems with stochastic behaviour. We show that finding a winning strategy is PSPACE-hard in general and undecidable for deterministic strategies. We also prove that optimal strategies, if they exists, may require infinite memory and randomisation. However, when restricted to disjunctions of objectives only, memoryless deterministic strategies suffice, and the problem of deciding whether a winning strategy exists is NP-complete. We also present algorithms to approximate the Pareto sets of achievable objectives for the class of stopping games.

1 Introduction

Stochastic games [21] have many applications in semantics and formal verification, and have been used as abstractions for probabilistic systems [16], and more recently for quantitative verification and synthesis of competitive stochastic systems [7]. Two-player games, in particular, provide a natural representation of open systems, where one player represents the system and the other its environment, in this paper referred to as Player 1 and Player 2, respectively. Stochasticity models uncertainty or randomisation, and leads to a game where each player can select an outgoing edge in states he controls, while in stochastic states the choice is made according to a state-dependent probability distribution. A *strategy* describes which actions a player picks. A fixed pair of strategies and an initial state determines a probability space on the runs of a game, and yields expected values of given objective (payoff) functions. The problem is then to determine if Player 1 has a strategy to ensure that the expected values of the objective functions meet a given set of criteria for all strategies that Player 2 may choose.

Various objective functions have been studied, for example reachability, ω-regular, or parity [4]. We focus here on *reward functions*, which are determined by a reward structure, annotating states with rewards. A prominent example is the reward function evaluating *total reward*, which is obtained by summing up rewards for all states visited along a path. Total rewards can be conveniently used to model consumption of resources along the execution of the system, but (with a straightforward modification of the game) they can also be used to encode other objective functions, such as reachability.

Although objective functions can express various useful properties, many situations demand considering not just the value of a single objective function, but rather values of several such functions simultaneously. For example, we may wish to maximise the number of successfully provided services and, at the same time, ensure minimising resource usage. More generally, given multiple objective functions, one may ask whether an arbitrary boolean combination of upper or lower bounds on the expected values of these functions can be ensured (in this paper we restrict only to positive boolean combinations, i.e. we do not allow negations). Alternatively, one might ask to compute or approximate the *Pareto set*, i.e. the set of all bounds that can be assured by exploring trade-offs. The simultaneous optimisation of a conjunction of objectives (also known as multi-objective, multi-criteria or multi-dimensional optimisation) is actively studied in operations research [22] and used in engineering [18]. In verification it has been considered for Markov decision processes (MDPs), which can be seen as one-player stochastic games, for discounted objectives [5] and general ω-regular objectives [9]. Multiple objectives for non-stochastic games have been studied by a number of authors, including in the context of energy games [23] and strategy synthesis [6].

In this paper, we study *stochastic games* with multi-objective queries, which are expressed as positive boolean combinations of total reward functions with upper or lower bounds on the expected reward to be achieved. In that way we can, for example, give several alternatives for a valid system behaviour, such as "the expected consumption of the system is at most 10 units of energy and the probability of successfully finishing the operation is at least 70%, or the expected consumption is at most 50 units, but the probability of success is at least 99%". Another motivation for our work is assume-guarantee compositional verification [20], where the system satisfies a set of guarantees φ whenever a set of assumptions ψ is true. This can be formulated using multi-objective queries of the form $\bigwedge\psi \Rightarrow \bigwedge\varphi$. For MDPs it has been shown how to formulate assume-guarantee rules using multi-objective queries [9]. The results obtained in this paper would enable us to explore the extension to stochastic games.

Contributions. We first obtain nondeterminacy by a straightforward modification of earlier results. Then we prove the following novel results for multi-objective stochastic games:

- We prove that, even in a pure conjunction of objectives, infinite memory and randomisation are required for the winning strategy of Player 1, and that the problem of finding a *deterministic* winning strategy is undecidable.
- For the case of a pure disjunction of objectives, we show that memoryless deterministic strategies are sufficient for Player 1 to win, and we prove that determining the existence of such strategies is an NP-complete problem.
- For the general case, we show that the problem of deciding whether Player 1 has a winning strategy in a game is PSPACE-hard.
- We provide Pareto set approximation algorithms for stopping games. This result directly applies to the important class of *discounted rewards* for non-stopping games, due to an off-the-shelf reduction [8].

Related work. Multi-objective optimisation has been studied for various subclasses of stochastic games. For non-stochastic games, multi-dimensional objectives have been considered in [6,23]. For MDPs, multiple discounted objectives [5], long-run objectives [2], ω-regular objectives [9] and total rewards [12] have been analysed. The objectives that we study in this paper are a special case of branching time temporal logics for stochastic games [3,1]. However, already for MDPs, such logics are so powerful that it is not decidable whether there is an optimal controller [3]. A special case of the problem studied in this paper is the case where the goal of Player 1 is to achieve a *precise value* of the expectation of an objective function [8]. As regards applications, stochastic games with a single objective function have been employed and implemented for quantitative abstraction refinement for MDP models in [16]. The usefulness of techniques for verification and strategy synthesis for stochastic games with a single objective is demonstrated, e.g., for smart grid protocols [7]. Applications of multi-objective verification include assume-guarantee verification [17] and controller synthesis [13] for MDPs.

2 Preliminaries

We begin this section by introducing notations used throughout the paper. We then provide the definition of stochastic two-player games together with the concepts of strategies and paths of the game. Finally, we introduce the objectives that are studied in this paper.

2.1 Notation

Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, we use x_i to refer to its i-th component, where $1 \leq i \leq n$, and define the norm $\|\boldsymbol{x}\| \stackrel{\text{def}}{=} \sum_{i=1}^{n} |x_i|$. Given a number $y \in \mathbb{R}$, we use $\boldsymbol{x} \pm y$ to denote the vector $(x_1 \pm y, x_2 \pm y, \ldots, x_n \pm y)$. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the *dot product* of $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i \cdot y_i$, and the comparison operator $\leq$ on vectors is defined to be the componentwise ordering. The sum of two sets of vectors $X, Y \subseteq \mathbb{R}^n$ is defined by $X + Y = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in X, \boldsymbol{y} \in Y\}$. Given a set X, we define the *downward closure* of X as $\mathsf{dwc}(X) \stackrel{\text{def}}{=} \{\boldsymbol{y} \mid \exists \boldsymbol{x} \in X \,.\, \boldsymbol{y} \leq \boldsymbol{x}\}$ and the *upward closure* as $\mathsf{up}(X) \stackrel{\text{def}}{=} \{\boldsymbol{y} \mid \exists \boldsymbol{x} \in X \,.\, \boldsymbol{x} \leq \boldsymbol{y}\}$. We denote by $\mathbb{R}_{\pm\infty}$ the set $\mathbb{R} \cup \{+\infty, -\infty\}$, and we define the operations $\cdot$ and $+$ in the expected way, defining $0 \cdot x = 0$ for all $x \in \mathbb{R}_{\pm\infty}$ and leaving $-\infty + \infty$ undefined. We also define function $\mathrm{sgn}(x) : \mathbb{R}_{\pm\infty} \to \mathbb{N}$ to be 1 if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$.

A *discrete probability distribution* (or just *distribution*) over a (countable) set S is a function $\mu : S \to [0, 1]$ such that $\sum_{s \in S} \mu(s) = 1$. We write $\mathcal{D}(S)$ for the set of all distributions over S. Let $\mathsf{supp}(\mu) = \{s \in S \mid \mu(s) > 0\}$ be the *support set* of $\mu \in \mathcal{D}(S)$. We say that a distribution $\mu \in \mathcal{D}(S)$ is a *Dirac distribution* if $\mu(s) = 1$ for some $s \in S$. We represent a distribution $\mu \in \mathcal{D}(S)$ on a set $S = \{s_1, \ldots, s_n\}$ as a map $[s_1 \mapsto \mu(s_1), \ldots, s_n \mapsto \mu(s_n)]$ and omit the elements of S outside $\mathsf{supp}(\mu)$ to simplify the presentation. If the context is clear we sometimes identify a Dirac distribution μ with the unique element in $\mathsf{supp}(\mu)$.

2.2 Stochastic games

In this section we introduce turn-based stochastic two-player games.

Stochastic two-player games. A *stochastic two-player game* is a tuple $\mathcal{G} = \langle S, (S_\Box, S_\Diamond, S_\bigcirc), \Delta\rangle$ where S is a finite set of states partitioned into sets $S_\Box$, $S_\Diamond$, and $S_\bigcirc$; $\Delta : S \times S \to [0,1]$ is a probabilistic transition function such that $\Delta(\langle s,t\rangle) \in \{0,1\}$ if $s \in S_\Box \cup S_\Diamond$ and $\sum_{t\in S} \Delta(\langle s,t\rangle) = 1$ if $s \in S_\bigcirc$.

$S_\Box$ and $S_\Diamond$ represent the sets of states controlled by players $\mathsf{Player\ 1}$ and $\mathsf{Player\ 2}$, respectively, while $S_\bigcirc$ is the set of stochastic states. For a state $s \in S$, the set of successor states is denoted by $\Delta(s) \stackrel{\text{def}}{=} \{t \in S \mid \Delta(\langle s,t\rangle){>}0\}$. We assume that $\Delta(s) \neq \emptyset$ for all $s \in S$. A state from which no other states except for itself are reachable is called *terminal*, and the set of terminal states is denoted by $\mathsf{Term} \stackrel{\text{def}}{=} \{s \in S \mid \Delta(\langle s,t\rangle){=}1 \text{ iff } s = t\}$.

Paths. An infinite *path* λ of a stochastic game $\mathcal{G}$ is an infinite sequence $s_0 s_1 \dots$ of states such that $s_{i+1} \in \Delta(s_i)$ for all $i \geq 0$. A finite path is a finite such sequence. For a finite or infinite path λ we write $\mathsf{len}(\lambda)$ for the number of states in the path. For $i < \mathsf{len}(\lambda)$ we write λ_i to refer to the i-th state s_i of λ. For a finite path λ we write $\mathsf{last}(\lambda)$ for the last state of the path. For a game $\mathcal{G}$ we write $\Omega^+_{\mathcal{G}}$ for the set of all finite paths, and $\Omega_{\mathcal{G}}$ for the set of all infinite paths, and $\Omega_{\mathcal{G},s}$ for the set of infinite paths starting in state s. We denote the set of paths that reach a state in $T \subseteq S$ by $\Diamond T \stackrel{\text{def}}{=} \{\omega \in \Omega_{\mathcal{G}} \mid \exists i\,.\,\omega_i \in T\}$.

Strategies. A *strategy* of $\mathsf{Player\ 1}$ is a (partial) function $\pi : \Omega^+_{\mathcal{G}} {\to} \mathcal{D}(S)$, which is defined for $\lambda \in \Omega^+_{\mathcal{G}}$ only if $\mathsf{last}(\lambda) \in S_\Box$, such that $s \in \mathsf{supp}(\pi(\lambda))$ only if $\Delta(\langle \mathsf{last}(\lambda), s\rangle) = 1$. A strategy π is a *finite-memory* strategy if there is a finite automaton $\mathcal{A}$ over the alphabet S such that $\pi(\lambda)$ is determined by $\mathsf{last}(\lambda)$ and the state of $\mathcal{A}$ in which it ends after reading the word λ. We say that π is *memoryless* if $\mathsf{last}(\lambda){=}\mathsf{last}(\lambda')$ implies $\pi(\lambda){=}\pi(\lambda')$, and *deterministic* if $\pi(\lambda)$ is Dirac for all $\lambda \in \Omega^+_{\mathcal{G}}$. If π is a memoryless strategy for $\mathsf{Player\ 1}$ then we identify it with the mapping $\pi\colon S_\Box \to \mathcal{D}(S)$. A strategy σ for $\mathsf{Player\ 2}$ is defined similarly. We denote by Π and Σ the sets of all strategies for $\mathsf{Player\ 1}$ and $\mathsf{Player\ 2}$, respectively.

Probability measures. A stochastic game $\mathcal{G}$, together with a strategy pair $(\pi,\sigma) \in \Pi \times \Sigma$ and a starting state s, induces an infinite Markov chain on the game (see e.g. [8]). We define the probability measure of this Markov chain by $\Pr^{\pi,\sigma}_{\mathcal{G},s}$. The expected value of a measurable function $f\colon S^\omega {\to} \mathbb{R}_{\pm\infty}$ is defined as $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \stackrel{\text{def}}{=} \int_{\Omega_{\mathcal{G},s}} f \, d\Pr^{\pi,\sigma}_{\mathcal{G},s}$. We say that a game $\mathcal{G}$ is a *stopping game* if, for every pair of strategies π and σ, a terminal state is reached with probability 1.

Rewards. A reward function $\boldsymbol{r} : S \to \mathbb{Q}^n$ assigns a reward vector $\boldsymbol{r}(s) \in \mathbb{Q}^n$ to each state s of the game $\mathcal{G}$. We use r_i for the function defined by $r_i(t) = \boldsymbol{r}(t)_i$ for all t. We assume that for each i the reward assigned by r_i is either non-negative or non-positive for all states (we adopt this approach in order to express minimisation problems via maximisation, as explained in the next subsection). The analysis of more general reward functions is left for future work. We define the vector of *total reward* random variables $rew(\boldsymbol{r})$ such that, given a path λ, $rew(\boldsymbol{r})(\lambda) = \sum_{j \geq 0} \boldsymbol{r}(\lambda_j)$.

2.3 Multi-objective queries

A *multi-objective query* (MQ) φ is a positive boolean combination (i.e. disjunctions and conjunctions) of predicates (or *objectives*) of the form $r \bowtie v$, where r is a reward function, $v \in \mathbb{Q}$ is a bound and $\bowtie \in \{\geq, \leq\}$ is a comparison operator. The validity of an MQ is defined inductively on the structure of the query: an objective $r \bowtie v$ is true in a state s of $\mathcal{G}$ under a pair of strategies (π, σ) if and only if $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(r)] \bowtie v$, and the truth value of disjunctions and conjunctions of queries is defined straightforwardly. Using the definition of the reward function above, we can express the operator $\leq$ by using $\geq$, applying the equivalence $r \leq v \equiv (-r \geq -v)$. Thus, throughout the paper we often assume that MQs only contain the operator $\geq$.

We say that Player 1 *achieves* the MQ φ (i.e., *wins* the game) in a state s if it has a strategy π such that for all strategies σ of Player 2 the query φ evaluates to true under (π, σ). An MQ φ is a *conjunctive query* (CQ) if it is a conjunction of objectives, and a *disjunctive query* (DQ) if it is a disjunction of objectives.

For a MQ φ containing n objectives $r_i \bowtie_i v_i$ for $1 \leq i \leq n$ and for $\boldsymbol{x} \in \mathbb{R}^n$ we use $\varphi[\boldsymbol{x}]$ to denote φ in which each $r_i \bowtie_i v_i$ is replaced with $r_i \bowtie_i x_i$.

Reachability. We can enrich multi-objective queries with *reachability objectives*, i.e. objectives $\Diamond T \geq p$ for a set of target states $T \subseteq S$, where $p \in [0, 1]$ is a bound. The objective $\Diamond T \geq p$ is true under a pair of strategies (π, σ) if $\Pr^{\pi,\sigma}_{\mathcal{G},s}(\Diamond T) \geq p$, and notions such as achieving a query are defined straightforwardly. Note that queries containing reachability objectives can be reduced to queries with total expected reward only (see Appendix A for a reduction). It also follows from the construction that if all target sets contain only terminal states, the reduction works in polynomial time.

Pareto sets. Let φ be an MQ containing n objectives. The vector $\boldsymbol{v} \in \mathbb{R}^n$ is a *Pareto vector* if and only if (a) $\varphi[\boldsymbol{v} - \varepsilon]$ is achievable for all $\varepsilon > 0$, and (b) $\varphi[\boldsymbol{v} + \varepsilon]$ is not achievable for any $\varepsilon > 0$. The set P of all such vectors is called a *Pareto set*. Given $\varepsilon > 0$, an *ε-approximation of a Pareto set* is a set of vectors Q satisfying that, for any $\boldsymbol{w} \in Q$, there is a vector $\boldsymbol{v}$ in the Pareto set such that $\|\boldsymbol{v} - \boldsymbol{w}\| \leq \varepsilon$, and for every $\boldsymbol{v}$ in the Pareto set there is a vector $\boldsymbol{w} \in Q$ such that $\|\boldsymbol{v} - \boldsymbol{w}\| \leq \varepsilon$.

Example. Consider the game $\mathcal{G}$ from Figure 1 (left). It consists of one Player 1 state s_0, one Player 2 state s_1, six stochastic states s_2, s_3, s_4, s_5, t_1 and t_2, as well as two terminal states t'_1 and t'_2. Outgoing edges of stochastic states are assigned uniform distributions by convention. For the MQ $\varphi_1 = r_1 \geq \frac{2}{3} \wedge r_2 \geq \frac{1}{6}$, where the reward functions are defined by $r_1(t_1)$=$r_2(t_2)$=1 and all other values are zero, the Pareto set for the initial state s_0 is shown in Figure 1 (centre). Hence, φ_1 is satisfied at s_0, as $(\frac{2}{3}, \frac{1}{6})$ is in the Pareto set. For the MQ $\varphi_2 = r_1 \geq \frac{2}{3} \wedge -r_2 \geq -\frac{1}{6}$, Figure 1 (right) illustrates the Pareto set for s_0, showing that φ_2 is not satisfied

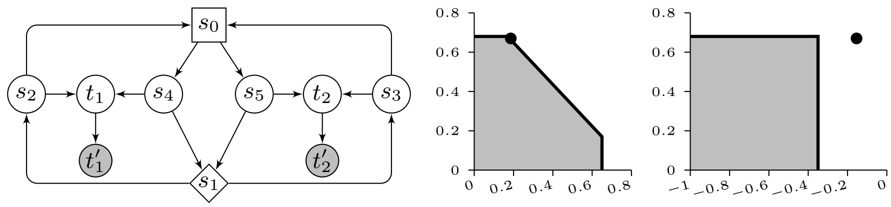


Fig. 1: An example game (left), Pareto set for φ_1 at s_0 (centre), and Pareto set for φ_2 at s_0 (right), with bounds indicated by a dot. Note that the sets are unbounded towards $-\infty$.

at s_0. Note that φ_1 and φ_2 correspond to the combination of reachability and safety objectives, i.e., $\Diamond\{t'_1\} \geq \frac{2}{3} \wedge \Diamond\{t'_2\} \geq \frac{1}{6}$ and $\Diamond\{t'_1\} \geq \frac{2}{3} \wedge \Diamond\{t'_2\} \leq \frac{1}{6}$.

3 Conjunctions of Objectives

In this section we present the results for CQs. We first recall that the games are not determined, and then show that Player 1 may require an infinite-memory randomised strategy to win, while it is not decidable whether deterministic winning strategies exist. We also provide fixpoint equations characterising the Pareto sets of achievable vectors and their successive approximations.

Theorem 1 (Non-determinacy, optimal strategies [8]). *Stochastic games with multiple objectives are, in general, not determined, and optimal strategies might not exist, already for CQs with two objectives.*

Theorem 1 carries over from the results for precise value games, because the problem of reaching a set of terminal states $T \subseteq \mathsf{Term}$ with probability precisely p is a special case of multi-objective stochastic games and can be expressed as a CQ $\varphi = \Diamond T \geq p \wedge \Diamond T \leq p$.

Theorem 2 (Infinite memory). *An infinite-memory randomised strategy may be required for Player 1 to win a multi-objective stochastic game with a CQ even for stopping games with reachability objectives.*

Proof. To prove the theorem we will use the example game from Figure 2. We only explain the intuition behind the need of infinite memory here; the formal proof is presented in Appendix B.1. First, we note that it is sufficient to consider deterministic counter-strategies for Player 2, since, after Player 1 has proposed his strategy, the resulting model is an MDP with finite branching [19]. Consider the game starting in the initial state s_0 and a CQ $\varphi = \bigwedge_{i=1}^{3} \Diamond T_i \geq \frac{1}{3}$, where the target sets T_1, T_2 and T_3 contain states labelled 1, 2 and 3, respectively. We note that target sets are terminal and disjoint, and for any π and σ we have that $\sum_{i=1}^{3} \mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_i) = 1$, and hence for any winning Player 1 strategy π it must be the case that, for any σ, $\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_i) = \frac{1}{3}$ for $1 \leq i \leq 3$.

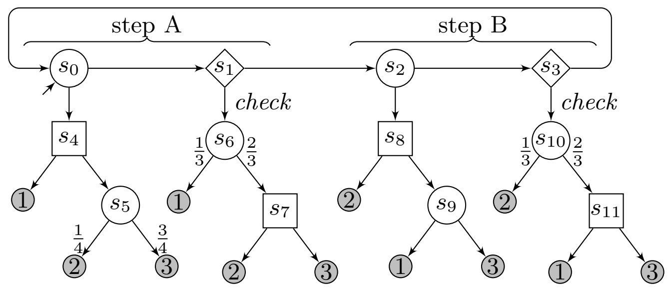


Fig. 2: Game where Player 1 requires infinite memory to win.

Let E be the set of runs which never take any transition *check*. The game proceeds by alternating between the two steps A and B as indicated in Figure 2. In step A, Player 1 chooses a probability to go to T_1 from state s_4, and then Player 2 gets an opportunity to "verify" that the probability $\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_1|E)$ of runs reaching T_1 conditional on the event that no *check* action was taken is $\frac{1}{3}$. She can do this by taking the action *check* and so ensuring that $\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_1|\Omega_{\mathcal{G}} \setminus E) = \frac{1}{3}$. If Player 2 again does not choose to take *check*, the game continues in step B, where the same happens for T_2, and so on.

When first performing step A, Player 1 has to pick probability $\frac{1}{3}$ to go to T_1. But since the probability of going from s_4 to T_2 is $< \frac{1}{3}$, when step B is performed for the first time, Player 1 must go to T_2 with probability $y_0 > \frac{1}{3}$ to compensate for the "loss" of the probability in step A. However, this decreases the probability of reaching T_1 at step B, and so Player 1 must compensate for it in the subsequent step A by taking probability $> \frac{1}{3}$ of going to T_1. This decreases the probability of reaching T_2 in the second step B even more (compared to first execution of step A), for which Player 1 must compensate by picking $y_1 > y_0 > \frac{1}{3}$ in the second execution of step B, and so on. So, in order to win, Player 1 has to play infinitely many different probability distributions in states s_4 and s_8. Note that, if Player 2 takes action "check", Player 1 can always randomise in states s_7 and s_{11} to achieve expectations exactly $\frac{1}{3}$ for all objectives. $\square$

In fact, the above idea allows us to encode natural numbers together with operations of increment and decrement, and obtain a reduction of the location reachability problem in the two-counter machine (which is known to be undecidable [15]) to the problem of deciding whether there exists a *deterministic* winning strategy for Player 1 in a multi-objective stochastic game.

Theorem 3 (Undecidability). *The problem whether there exists a* deterministic *winning strategy for Player 1 in a multi-objective stochastic game is undecidable already for stopping games and conjunctions of reachability objectives.*

Our proof is inspired by the proof of [3] which shows that the problem of existence of a winning strategy in an MDP for a PCTL formula is undecidable. However, the proof of [3] relies on branching time features of PCTL to ensure the counter

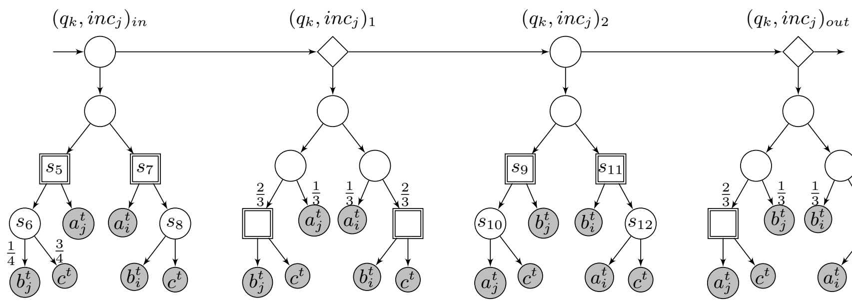


Fig. 3: Increment gadget for counter j.

values of the two-counter machine are encoded correctly. Since MQs only allow us to express combinations of linear-time properties, we need to take a different approach, utilising ideas of Theorem 2. We present the proof idea here; for the full proof see Appendix B.2. We encode the counter machine instructions in gadgets similar to the ones used for the proof of Theorem 2, where Player 1 has to change the probabilities with which he goes to the target states based on the current value of the counter. For example, the gadget in Figure 3 encodes the instruction to *increment* the counter j. The basic idea is that, if the counter value is c_j when entering the increment gadget, then in state s_5 Player 1 has to assign probability exactly $\frac{2}{3 \cdot 2^{c_j}}$ to the edge $\langle s_5, s_6 \rangle$, and then probability $\frac{2}{3 \cdot 2^{c_j+1}}$ to the edge $\langle s_9, s_{10} \rangle$ in s_9, resulting in the counter being incremented. The gadgets for counter decrement and zero-check can be found in the appendix. The resulting query contains six target sets. In particular, there is a conjunct $\Diamond T_t \geq 1$, where the set T_t is not reached with probability 1 only if the gadget representing the target counter machine location is reached. The remaining five objectives ensure that Player 1 updates the counter values correctly (by picking corresponding probability distributions) and so the strategy encodes a valid computation of the two-counter machine. Hence, the counter machine terminates if and only if there does not exist a winning strategy for Player 1.

We note that the problem of deciding whether there is a *randomised* winning strategy for Player 1 remains open, since the gadgets modelling decrement instructions in our construction rely on the strategy being deterministic. Nevertheless, for stopping games, in Theorem 4 below we provide a functional that, given a CQ φ, computes $\varepsilon-$approximations of the Pareto sets, i.e. the sets containing the bounds $\boldsymbol{x}$ so that Player 1 has a winning strategy for $\varphi[\boldsymbol{x} - \varepsilon]$. As a corollary of the theorem, using a simple reduction (see e.g. [8]) we get an approximation algorithm for the Pareto sets in non-stopping games with (multiple) *discounted reward objectives.*

Theorem 4 (Pareto set approximation). *For a stopping game $\mathcal{G}$ and a CQ $\varphi = \bigwedge_{i=1}^{n} r_i \geq v_i$, an $\varepsilon-$approximation of the Pareto sets for all states can be computed in $k = |S| + \lceil |S| \cdot \frac{\ln(\varepsilon \cdot (n \cdot M)^{-1})}{\ln(1-\delta)} \rceil$ iterations of the operator $F : (S \to$*

$\mathcal{P}(\mathbb{R}^n)) \to (S \to \mathcal{P}(\mathbb{R}^n))$ *defined by*

$$F(X)(s) \stackrel{\text{def}}{=} \begin{cases} \mathit{dwc}(\mathit{conv}(\bigcup_{t\in\Delta(s)} X_t) + \boldsymbol{r}(s)) & \text{if } s \in S_\Box \\ \mathit{dwc}(\bigcap_{t\in\Delta(s)} X_t + \boldsymbol{r}(s)) & \text{if } s \in S_\Diamond \\ \mathit{dwc}(\sum_{t\in\Delta(s)} \Delta(\langle s,t\rangle)\cdot X_t + \boldsymbol{r}(s)) & \text{if } s \in S_\bigcirc, \end{cases}$$

where the initial sets are $X_s^0 \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \le \boldsymbol{r}(s)\}$ *for all* $s \in S$*, and* $M = |S| \cdot \frac{\max_{s\in S,i} |r_i(s)|}{\delta}$ *for* $\delta = p_{\min}^{|S|}$ *and* $p_{\min}$ *being the smallest positive probability in* $\mathcal{G}$*.*

We first explain the intuition behind the operations when $\boldsymbol{r}(s) = \boldsymbol{0}$. For $s \in S_\Box$, Player 1 can randomise between successor states, so any convex combination of achievable points in X_t^{k-1} for the successors $t \in \Delta(s)$ is achievable in X_s^k, and so we take the convex closure of the union. For $s \in S_\Diamond$, a value in X_s^k is achievable if it is achievable in X_s^{k-1} for all successors $t \in \Delta(s)$, and hence we take the intersection. Finally, stochastic states $s \in S_\bigcirc$ are like Player 1 states with a fixed probability distribution, and hence the operation performed is the weighted Minkowski sum. When $\boldsymbol{r}(s) \neq \boldsymbol{0}$, the reward is added as a contribution to what is achievable at s.

Proof (Outline). The proof, presented in Appendix B.3, consists of two parts. First, we prove in Proposition 2 that the result of the k-th iteration of F contains exactly the points achievable by some strategy in k steps; this is done by applying induction on k. As the next step, we observe that, since the game is *stopping*, after $|S|$ steps the game has terminated with probability at least $\delta = p_{\min}^{|S|}$. Hence, the maximum change to any dimension to any vector in X_s^k after k steps of the iteration is less than $M \cdot (1-\delta)^{\lfloor \frac{k}{|S|} \rfloor}$. It follows that $k = |S| + \lceil |S| \cdot \frac{\ln(\varepsilon\cdot(n\cdot M)^{-1})}{\ln(1-\delta)} \rceil$ iterations of F suffice to yield all points which are within ε from the Pareto points for $\boldsymbol{r}$.

4 General Multi-Objective Queries

In this section we consider the general case where the objective is expressed as an arbitrary MQ. The nondeterminacy result from Theorem 1 carries over to the more general MQs, and, even if we restrict to DQs, the games stay nondetermined (see Appendix C.1 for a proof). The following theorem establishes lower complexity bounds for the problem of deciding the existence of the winning strategy for Player 1.

Theorem 5. *The problem of deciding whether there is a winning strategy for Player 1 for an MQ* φ *is PSPACE-hard in general, and NP-hard if* φ *is a DQ.*

The above theorem is proved by reductions from QBF and 3SAT, respectively (see Appendix C.2 and Appendix C.3). The reduction from QBF is similar to the one in [10], the major differences being that our results apply even when the target states are terminal, and that we need to deal with possible randomisation of the strategies.

We now establish conditions under which a winning strategy for Player 1 exists. Before we proceed, we note that it suffices to consider MQs in conjunctive normal form (CNF) that contain no negations, since any MQ can be converted to CNF using standard methods of propositional logic. Before presenting the proof of Theorem 6, we give the following reformulation of the separating hyperplane theorem, proved in Appendix C.6.

Lemma 1. *Let $W \subseteq \mathbb{R}^m_{\pm\infty}$ be a convex set satisfying the following. For all j, whenever there is $\boldsymbol{x} \in W$ such that $sgn(x_j) \geq 0$ (resp. $sgn(x_j) \leq 0$), then $sgn(y_j) \geq 0$ (resp. $sgn(y_j) \leq 0$) for all $\boldsymbol{y} \in W$. Let $\boldsymbol{z} \in \mathbb{R}^m$ be a point which does not lie in the closure of $\mathsf{up}(W)$. Then there is a non-zero vector $\boldsymbol{x} \in \mathbb{R}^m$ such that the following conditions hold:*

1. *for all $1 \leq j \leq m$ we have $x_j \geq 0$;*
2. *for all $1 \leq j \leq m$, if there is $\boldsymbol{w} \in W$ satisfying $w_j = -\infty$, then $x_j = 0$; and*
3. *for all $\boldsymbol{w} \in W$, the product $\boldsymbol{w} \cdot \boldsymbol{x}$ is defined and satisfies $\boldsymbol{w} \cdot \boldsymbol{x} \geq \boldsymbol{z} \cdot \boldsymbol{x}$.*

Theorem 6. *Let $\psi = \bigwedge_{i=1}^n \bigvee_{j=1}^m q_{i,j} \geq u_{i,j}$ be an MQ in CNF, and let π be a strategy of Player 1. The following two conditions are equivalent.*

- *The strategy π achieves ψ.*
- *For all $\varepsilon > 0$ there are nonzero vectors $\boldsymbol{x}_1, \ldots \boldsymbol{x}_n \in \mathbb{R}^m_{\geq 0}$, such that π achieves the conjunctive query $\varphi = \bigwedge_{i=1}^n r_i \geq v_i$, where $r_i(s) = \boldsymbol{x}_i \cdot (q_{i,1}(s), \ldots, q_{i,m}(s))$ and $v_i = \boldsymbol{x}_i \cdot (u_{i,1} - \varepsilon, \ldots, u_{i,m} - \varepsilon)$ for all $1 \leq i \leq n$.*

Proof (Sketch). We only present high-level intuition here, see Appendix C.4 for the full proof. Using the separating hyperplane theorem we show that if there exists a winning strategy for Player 1, then there exist separating hyperplanes, one per conjunct, separating the objective vectors within each conjunct from the set of points that Player 2 can enforce, and vice versa. This allows us to reduce the MQ expressed in CNF into a CQ, by obtaining one reward function per conjuct, which is constructed by weighthing the original reward function by the characteristic vector of the hyperplane.

When we restrict to DQs only, it follows from Theorem 6 that there exists a strategy achieving a DQ if and only if there is a strategy achieving a certain single-objective expected total reward, and hence we obtain the following theorem.

Theorem 7 (Memoryless deterministic strategies). *Memoryless deterministic strategies are sufficient for Player 1 to achieve a DQ.*

Since memoryless deterministic strategies suffice for optimising single total reward, to determine whether a DQ is achievable we can guess such a strategy for Player 1, which uniquely determines an MDP. We can then use the polynomial time algorithm of [9] to verify that there exists no winning Player 2 strategy. This NP algorithm, together with Theorem 5, gives us the following corollary.

Corollary 1. *The problem whether a DQ is achievable is NP-complete.*

Using Theorem 6 we can construct an approximation algorithm computing Pareto sets for disjunctive objectives for stopping games, which performs multiple calls to the algorithm for computing optimal value for the single-objective reward.

Theorem 8 (Pareto sets). *For stopping games, given a vector $\boldsymbol{r} = (r_1, \ldots, r_m)$ of reward functions, an ε-approximation of the Pareto sets for disjunction of objectives for $\boldsymbol{r}$ can be computed by $(\frac{2\cdot m^2\cdot(M+1)}{\varepsilon})^{m-1}$ calls to a NP$\cap$coNP algorithm computing single-objective total reward, where M is as in Theorem 4.*

Proof (Sketch). By Theorem 6 and Lemma 3 (see Appendix C.6), we have that a DQ $\varphi = \bigvee_{j=1}^{m} r_j \geq v_j$ is achievable if and only if there exists π and $\boldsymbol{x} \in \mathbb{R}^m_{\geq 0}$ such that $\forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{x} \cdot rew(\boldsymbol{r})] \geq \boldsymbol{x} \cdot \boldsymbol{v}$, which is a single-objective query decidable by an NP$\cap$coNP oracle. Given a finite set $X \subseteq \mathbb{R}^m$, we can compute values $d_{\boldsymbol{x}} = \sup_\pi \inf_\sigma \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{x} \cdot rew(\boldsymbol{r})]$ for all $\boldsymbol{x} \in X$, and define $U_X = \bigcup_{\boldsymbol{x}\in X}\{\boldsymbol{p} \mid \boldsymbol{x} \cdot \boldsymbol{p} \leq d_{\boldsymbol{x}}\}$. It is not difficult to see that U_X yields an under-approximation of achievable points. Let $\tau = \frac{\varepsilon}{2\cdot m^2\cdot(M+1)}$. We argue that when we let X be the set of all non-zero vectors $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 1$, and where all x_i are of the form $\tau \cdot k_i$ for some $k_i \in \mathbb{N}$, we obtain an ε-approximation of the Pareto set by taking all Pareto points on U_X (see Appendix C.7 for a proof).

The above approach, together with the algorithm for Pareto set approximations for CQs from Theorem 4, can be used to compute ε-approximations of the Pareto sets for MQs expressed in CNF. The set U_X would then contain tuples of vectors, one per conjunct.

5 Conclusions

We studied stochastic games with multiple expected total reward objectives, and analysed the complexity of the related algorithmic problems. There are several interesting directions for future research. Probably the most obvious is settling the question whether the problem of existence of a strategy achieving a MQ is decidable. Further, it is natural to extend the algorithms to handle long-run objectives containing mean-payoff or ω-regular goals, or to lift the restriction on reward functions to allow both negative and positive rewards at the same time. Another direction is to investigate practical algorithms for the solution for the problems studied here, such as more sophisticated methods for the approximation of Pareto sets.

Acknowledgements. The authors would like to thank Klaus Draeger, Ashutosh Trivedi and Michael Ummels for the discussions about the problem. The authors are partially supported by ERC Advanced Grant VERIWARE, the Institute for the Future of Computing at the Oxford Martin School, EPSRC grant EP/F001096, and the German Academic Exchange Service (DAAD). V. Forejt was suppoted by the Newton Fellowship of Royal Society and is also affiliated with the Faculty of Informatics, Masaryk University, Czech Republic.

References

1. C. Baier, T. Brázdil, M. Größer, and A. Kucera. Stochastic game logic. *Acta Inf.*, 49(4):203–224, 2012.
2. T. Brázdil, V. Brožek, K. Chatterjee, V. Forejt, and A. Kučera. Two views on multiple mean-payoff objectives in Markov decision processes. In *LICS*, 2011.
3. T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Stochastic games with branching-time winning objectives. In *LICS*, pages 349–358, 2006.
4. K. Chatterjee. *Stochastic Omega-Regular Games*. PhD thesis, EECS Department, University of California, Berkeley, October 2007.
5. K. Chatterjee, R. Majumdar, and T. Henzinger. Markov decision processes with multiple objectives. In *STACS*, pages 325–336. Springer, 2006.
6. K. Chatterjee, M. Randour, and J.-F. Raskin. Strategy synthesis for multi-dimensional quantitative objectives. In *CONCUR*, pages 115–131, 2012.
7. T. Chen, V. Forejt, M. Z. Kwiatkowska, D. Parker, and A. Simaitis. Automatic verification of competitive stochastic systems. In *TACAS*, pages 315–330, 2012.
8. T. Chen, V. Forejt, M. Z. Kwiatkowska, A. Simaitis, A. Trivedi, and M. Ummels. Playing stochastic games precisely. In *CONCUR*, pages 348–363, 2012.
9. K. Etessami, M. Kwiatkowska, M. Vardi, and M. Yannakakis. Multi-objective model checking of Markov decision processes. *LMCS*, 4(4):1–21, 2008.
10. N. Fijalkow and F. Horn. The surprizing complexity of reachability games. *CoRR*, abs/1010.2420, 2010.
11. J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.
12. V. Forejt, M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Quantitative multi-objective verification for probabilistic systems. In *TACAS*, 2011.
13. V. Forejt, M. Kwiatkowska, and D. Parker. Pareto curves for probabilistic model checking. In *ATVA*, LNCS, pages 317–332. Springer, 2012.
14. M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, 2nd edition, 1993.
15. D. Harel. Effective transformations on infinite trees, with applications to high undecidability, dominoes, and fairness. *J. ACM*, 33(1):224–248, 1986.
16. M. Kattenbelt, M. Z. Kwiatkowska, G. Norman, and D. Parker. A game-based abstraction-refinement framework for markov decision processes. *FMSD*, 2010.
17. M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Assume-guarantee verification for probabilistic systems. In *TACAS*, pages 23–37. Springer, 2010.
18. R. T. Marler and J. S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, 2004.
19. D. Martin. The determinacy of Blackwell games. *JSL*, 63(4):1565–1581, 1998.
20. A. Pnueli. Logics and models of concurrent systems. Springer, 1985.
21. L. S. Shapley. Stochastic games. *PNAS*, 39(10):1095, 1953.
22. B. Suman and P. Kumar. A survey of simulated annealing as a tool for single and multiobjective optimization. *J. Oper. Res. Soc.*, 57(10):1143–1160, 2005.
23. Y. Velner, K. Chatterjee, L. Doyen, T. A. Henzinger, A. Rabinovich, and J.-F. Raskin. The complexity of multi-mean-payoff and multi-energy games. *CoRR'12*.

A Reachability to expected reward

Here we present a reduction of the reachability problem to the one of expected total reward (a similar reduction was used by Etessami et al. [9]). Note that the presented reduction yields a game, which is exponential only in the number of non-terminal target states. Hence, when all states in the target sets are terminal, the reduction provides a game with polynomial number of reachable states.

Proposition 1. *Given a game $\mathcal{G} = \langle S, (S_\Box, S_\Diamond, S_\bigcirc), \Delta\rangle$ with a boolean combination ϕ of n reachability predicates $\Diamond T_i \bowtie_i v_i$ (for $1 \leq i \leq n$), there is a game $\mathcal{G}' = \langle S', (S'_\Box, S'_\Diamond, S'_\bigcirc), \Delta'\rangle$ of size $\mathcal{O}(|\mathcal{G}| \cdot 2^n)$ and an MQ φ combining the individual objectives $r_i \bowtie_i v_i$ in the same way as in ϕ, such that the query φ is achievable in $\mathcal{G}$ if and only if the query ϕ is achievable in $\mathcal{G}'$.*

Proof. We define $\mathcal{G}'$ so that a reward r_i of 1 is gained when a target set T_i is visited, and so that is not possible to visit any target set twice. Formally, we let $S' = \{(s, I, J) \mid s \in S, I, J \subseteq \{1, \ldots, n\}\}$, where in the partition of S', (s, I, J) is assigned to $S'_\Box$, $S'_\Diamond$ or $S'_\bigcirc$ if s is in $S_\Box$, $S_\Diamond$ or $S_\bigcirc$, respectively. The transition function Δ' is defined by $\Delta'(\langle (s, I, J), (s', I', J')\rangle) = x$ whenever

- $\Delta(s, s') = x$;
- $I' = I \cup J$; and
- $J' = \{i \mid s \in T_i\} \setminus I'$.

For all other values, Δ' returns 0. The reward functions r_i are defined by $r_i(s, I, J) = 1$ if $i \in J$ and $r_i(s, I, J) = 0$ otherwise for all $1 \leq i \leq n$.

To every run $s_0 s_1 \ldots$ in $\mathcal{G}$ corresponds a unique run $(s_0, I_0, J_0)(s_1, I_1, J_1)\ldots$ in $\mathcal{G}'$ initiated in $(s_0, \emptyset, \emptyset)$. These runs satisfy that for every $1 \leq i \leq n$ and $j \geq 0$, $r_i((s_j, I_j, J_j)) = 1$ if and only if $j \geq 1$, $s_{j-1} \in T_i$ and $s_\ell \notin T_i$ for any $\ell < j - 1$. The result easily follows by considering the correspondence between Player 1 strategies of $\mathcal{G}$ and $\mathcal{G}'$. $\Box$

B Proofs for Section 3

B.1 Proof of Theorem 2

We show that a strategy with infinite memory is needed for Player 1 to win the game from Figure 2 for $x = \frac{1}{4}$, when the objective is to ensure that for all σ we have $\Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_i) = \frac{1}{3}$ for $1 \leq i \leq 3$.

Let $h(k) = s_0(s_1s_2s_3s_0)^k$. Every strategy π for Player 1 determines (and is uniquely given by) an infinite sequence of vectors

$$
\begin{aligned}
\boldsymbol{p}^k &= (\pi(h(k)s_4)(s_4,1), \pi(h(k)s_4)(s_4,s_5) \cdot \frac{1}{4}, \pi(h(k)s_4)(s_4,s_5) \cdot \frac{3}{4}),\\
\boldsymbol{q}^k &= (\pi(h(k)s_1s_2s_8)(s_8,s_9) \cdot \frac{1}{2}, \pi(h(k)s_1s_2s_8)(s_8,2), \pi(h(k)s_1s_2s_8)(s_8,s_9) \cdot \frac{1}{2}),\\
\boldsymbol{w}^k &= (\frac{1}{3}, \frac{2}{3} \cdot \pi(h(k)s_1s_6s_7)(s_7,2), \frac{2}{3} \cdot \pi(h(k)s_1s_6s_7)(s_7,3)),\\
\boldsymbol{z}^k &= (\frac{2}{3} \cdot \pi(h(k)s_1s_2s_3s_{10}s_{11})(s_{11},1), \frac{1}{3}, \frac{2}{3} \cdot \pi(h(k)s_1s_2s_3s_{10}s_{11})(s_{11},1)).
\end{aligned}
$$

The intuitive interpretation of vector $\boldsymbol{p}^k$ (resp. $\boldsymbol{q}^k$, $\boldsymbol{w}^k$, $\boldsymbol{z}^k$) is that it represents the probability of reaching (T_1, T_2, T_3) from s_4 (resp. s_8, s_6, s_{10}) in the k-th step A (resp. B).

We define a winning strategy π by means of the vectors

$$
\begin{aligned}
\boldsymbol{p}^k &= (1 - \tfrac{1}{3\cdot 2^{k-1}}, \tfrac{1}{3\cdot 2^{k+1}}, \tfrac{1}{2^{k+1}}), & \boldsymbol{q}^k &= (\tfrac{1}{3\cdot 2^{k+1}}, 1 - \tfrac{1}{3\cdot 2^{k}}, \tfrac{1}{3\cdot 2^{k+1}}),\\
\boldsymbol{w}^k &= (\tfrac{1}{3}, +\tfrac{1}{6} + \tfrac{1}{2} - \tfrac{1}{3\cdot 2^{n+2}}), 1 - w_1^k - w_2^k), & \boldsymbol{z}^k &= (\tfrac{2}{3} - \tfrac{1}{3\cdot 2^{n+1}}, \tfrac{1}{3}, 1 - z_1^k - z_2^k).
\end{aligned}
$$

as follows. First, suppose Player 2 picks a strategy σ which does not take the *check* transition before the $n+1$-th visit to s_0. Then the probability of the runs that reach T_1 while visiting s_0 at most $n+1$ times is independent of σ and equal to $V(1,n) = \frac{1}{3} - \frac{1}{3\cdot 2^{2n+1}}$, as can be shown by the straightforward induction on n.

$$
\begin{aligned}
V(1,n) &= V(1,n-1) + \frac{1}{2^{2n}} \cdot (q_1^n + \frac{1}{2} \cdot p_1^{n+1})\\
&= V(1,n-1) + \frac{1}{2^{2n}} \cdot (\frac{1}{3\cdot 2^{n+1}} + \frac{1}{2}(1 - \frac{1}{3\cdot 2^{n}}))\\
&= V(1,n-1) + \frac{1}{2^{2n}} \cdot (\frac{1}{3\cdot 2^{n+1}} + \frac{1}{2} - \frac{1}{3\cdot 2^{n+1}})\\
&= V(1,n-1) + \frac{1}{2^{2n+1}}\\
&= \frac{1}{3} - \frac{1}{3} \cdot \frac{1}{2^{2n-1}} + \frac{1}{2^{2n+1}}\\
&= \frac{1}{3} + \frac{-4+3}{3\cdot 2^{2n+1}}\\
&= \frac{1}{3} - \frac{1}{3\cdot 2^{2n+1}}.
\end{aligned}
$$

Further, supposing Player 2 picks a strategy σ which does not take the *check* transition before the $n+1$-th visit to s_2, the probability of the runs that reach T_1 while visiting s_2 at most $n+1$ times is also independent of σ and equal to

$V(2,n) = \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+2}}$. This is again shown by an induction on n.

$$\begin{aligned}
V(2,n) &= V(2,n-1) + \frac{1}{2^{2n+1}} \cdot (p_2^{n+1} + \frac{1}{2} q_2^{n+1}) \\
&= V(2,n-1) + \frac{1}{2^{2n+1}} \cdot (\frac{1}{3 \cdot 2^{n+2}} + \frac{1}{2}(1 - \frac{1}{3 \cdot 2^{n+1}})) \\
&= V(2,n-1) + \frac{1}{2^{2n+1}} \cdot (\frac{1}{3 \cdot 2^{n+2}} + \frac{1}{2} - \frac{1}{3 \cdot 2^{n+2}})) \\
&= V(2,n-1) + \frac{1}{2^{2n+2}} \\
&= \frac{1}{3} - \frac{1}{3 \cdot 2^{2n}} + \frac{1}{2^{2n+2}} \\
&= \frac{1}{3} + \frac{-4+3}{3 \cdot 2^{2n+2}} \\
&= \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+2}}.
\end{aligned}$$

Now we are ready to show that π is winning by showing that probabilities of reaching T_1 and T_2 are both $\frac{1}{3}$ under any σ. By [19] it suffices to consider deterministic strategies σ. First, consider a strategy σ which never takes any transition labelled *check*. We have

$$\begin{aligned}
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_1) &= \lim_{n \to \infty} V(1,n) = \frac{1}{3}, \\
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_2) &= \lim_{n \to \infty} V(2,n) = \frac{1}{3}.
\end{aligned}$$

For a strategy σ which picks *check* on the $(n+1)$-th visit to s_1 we have

$$\begin{aligned}
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_1) &= V(1,n) + w_1^n = \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+1}} + \frac{1}{2^{2n+1}} \cdot \frac{1}{3} = \frac{1}{3}, \\
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_2) &= V(2,n) - \frac{1}{2^{2n+2}} \cdot (1 - \frac{1}{3 \cdot 2^{n+1}}) + \frac{1}{2^{2n+1}} \cdot w_2^n \\
&= \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+2}} - \frac{1}{2^{2n+2}} \cdot (1 - \frac{1}{3 \cdot 2^{n+1}}) + \frac{1}{2^{2n+1}} \cdot w_2^n \\
&= \frac{1}{3}.
\end{aligned}$$

Finally, for a strategy σ which picks *check* on the $(n+1)$-th visit to s_3 we have

$$\begin{aligned}
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_1) &= V(1,n) + \frac{1}{2^{2n+2}} \cdot \frac{1}{3 \cdot 2^{n+1}} + \frac{1}{2^{2n+2}} \cdot z_1^n \\
&= \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+1}} + \frac{1}{2^{2n+2}} \cdot \frac{1}{3 \cdot 2^{n+1}} + \frac{1}{2^{2n+2}} \cdot z_1^n \\
&= \frac{1}{3}, \\
\mathrm{Pr}_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_2) &= V(2,n) + z_2^n = \frac{1}{3} - \frac{1}{3 \cdot 2^{2n+2}} + \frac{1}{2^{2n+2}} \cdot \frac{1}{3} = \frac{1}{3}.
\end{aligned}$$

We have shown that σ ensures that each of the target sets T_1, T_2 and T_3 is reached with probability exactly $\frac{1}{3}$.

Now we show that there is no finite-memory strategy ensuring this. Let $\bar{\pi}$ be a finite memory strategy, determined by vectors $\bar{\boldsymbol{p}}^k$, $\bar{\boldsymbol{q}}^k$, $\bar{\boldsymbol{w}}^i$ and $\bar{\boldsymbol{z}}^i$. Since $\bar{\pi}$ is finite memory, there must be a k such that $\bar{\boldsymbol{p}}^k \neq \boldsymbol{p}$ or $\bar{\boldsymbol{q}}^k \neq \boldsymbol{q}$. Let k be the lowest such number. There are two possibilities:

– $\bar{\boldsymbol{p}}^k \neq \boldsymbol{p}^k$. Then necessarily $\bar{p}_1^k \neq p_1^k$. Also note that $w_1^k = \bar{w}_1^k = \frac{1}{3}$. We define the counter-strategy σ to take *check* on the (k+1)-th visit to s_1 and get (by minimality of k) that $\Pr_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_1) = \frac{1}{3} + \frac{1}{2^{2k\cdot+1}}(\bar{p}_1^k - p_1^k) \neq \frac{1}{3}$.
– $\bar{\boldsymbol{q}}^k \neq \boldsymbol{q}^k$. Then necessarily $\bar{q}_2^k \neq q_2^k$ and $z_1^k = \bar{z}_1^k = \frac{1}{3}$ and so we can define the counter-strategy σ to take *check* on the (k+1)-th visit to s_3. We get $\Pr_{\mathcal{G},s}^{\pi,\sigma}(\Diamond T_2) = \frac{1}{3} + \frac{1}{2^{2k\cdot+1}}(\bar{q}_2^k - q_2^k) \neq \frac{1}{3}$.

This completes the proof.

B.2 Proof of Theorem 3

We show the undecidability of the problem via a reduction to the termination problem of two-counter machines. The proof proceeds to establish that a two-counter machine $\mathcal{M}$ does not terminate if and only if there exists a winning strategy for the game $\mathcal{G}(\mathcal{M})$ constructed by the reduction from $\mathcal{M}$.

1. Formally a two-counter machine $\mathcal{M}$ consists of a sequence of instructions $l_1 : ins_1, \cdots, l_n : ins_n$, where each ins_i has one of the following forms:
   (a) $c_1 := c_2 := 0$ and goto l_j;
   (b) $c_1 = c_1 + 1$ and goto l_j;
   (c) $c_2 = c_2 + 1$ and goto l_j;
   (d) if $c_1 = 0$ then goto l_j else $c_1 = c_1 - 1$ and goto l_k;
   (e) if $c_2 = 0$ then goto l_j else $c_2 = c_2 - 1$ and goto l_k;
   (f) Terminate.

   The *state* of the two-counter machine is encoded by a location l and two counter values $c_1, c_2 \in \mathbb{N}$, i.e., $\langle l, c_1, c_2 \rangle$. Given an initial location l_0 with both counter values 0, the *termination problem* asks to determine whether a terminal location l_t is reached. The problem is known to be undecidable [15].
2. Let $\mathcal{M}$ be a Minsky machine. We construct a game $\mathcal{G}(\mathcal{M})$ and a CQ

$$\varphi = \Diamond T_{a_1} \geq \frac{1}{6} \wedge \Diamond T_{b_1} \geq \frac{1}{6} \wedge \Diamond T_{a_2} \geq \frac{1}{6} \wedge \Diamond T_{b_2} \geq \frac{1}{6} \wedge \Diamond T_c \geq \frac{1}{3} \geq \Diamond T_t \geq 1,$$

   where $T_{a_1} = \{a_1^t, a_1\}$, $T_{b_1} = \{b_1^t, b_1\}$, $T_{a_2} = \{a_2^t, a_2\}$, $T_{b_2} = \{b_2^t, b_2\}$, $T_c = \{c^t, c\}$, and $T_t = \{a_1^t, a_2^t, b_1^t, b_2^t, c^t\}$.
   We define the game $\mathcal{G}(\mathcal{M})$ incrementally. For each type of instructions, we have a corresponding gadget, i.e., Init, Terminate, Increment, and Decrement, which are shown in Figure 4. In this figure, Player 1 states with double

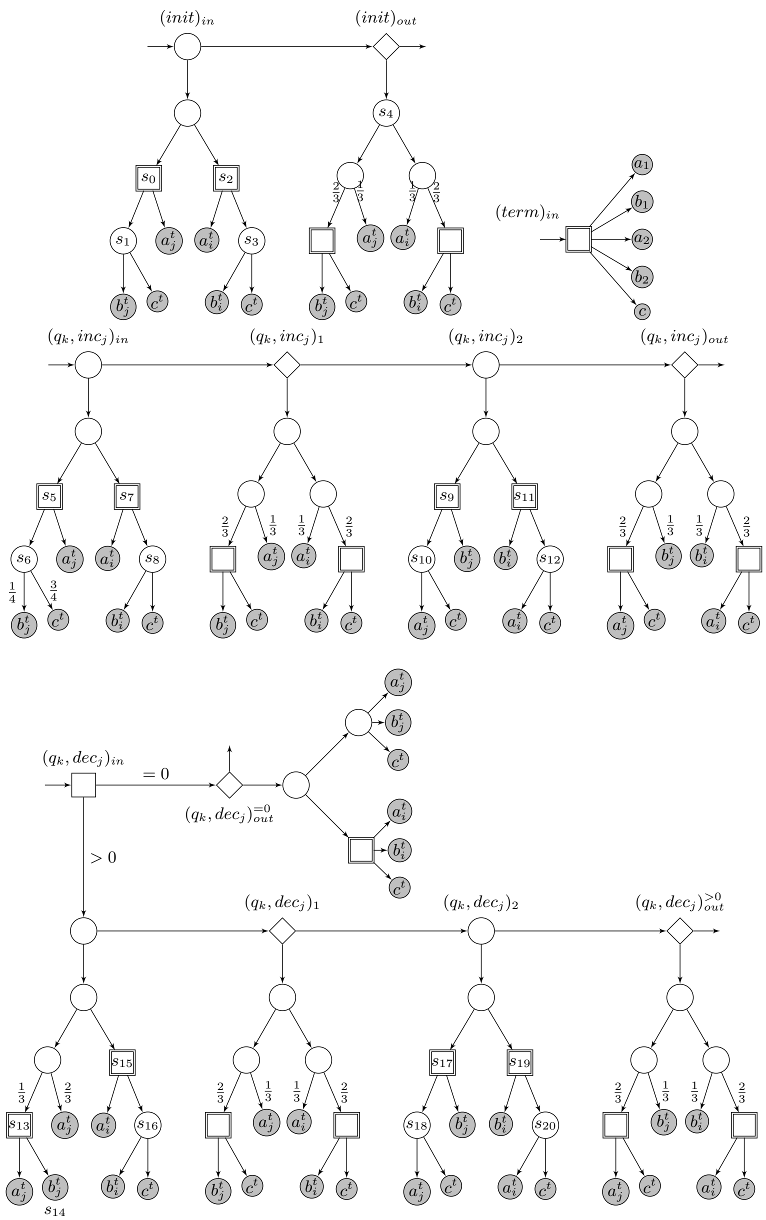
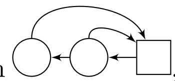


Fig. 4: Operations for counter j. Transition probabilities in stochastic states are uniform unless specified otherwise. Player 1 states with doubled border contain a gadget allowing to select arbitrary probability distributions even with deterministic strategies. Self-loops in target states omitted.

border are of the form , which allows Player 1 states to simulate any probability distribution even with deterministic strategies.

Note that for Increment and Decrement gadgets, $j \in \{1, 2\}$ refers to the counter (c_1 or c_2) that under the operation in the instruction. Moreover we set $i = 3-j$. The game $\mathcal{G}(\mathcal{M})$ is then constructed by "gluing" the instructions together. Namely,

- for the init instruction $l_i : c_1 := c_2 := 0$ and goto l_k, use the Init gadget and link $(int)_{out}$ to $(q_k, op)_{in}$, where op is the operation type of l_k;
- for the increment instruction $l_i : c_j := c_j + 1$ and goto l_k, use the Increment gadget and link $(q_i, inc_j)_{out}$ to $(q_k, op)_{in}$, where op is the operation type of l_k;
- for the decrement instruction if $c_j = 0$ then goto l_k else $c_j = c_j - 1$ and goto $l_{k'}$, use the Decrement gadget and link $(q_i, dec_j)^{=0}_{out}$ to $(q_k, op_k)_{in}$, and link $(q_i, dec_j)^{>0}_{out}$ to $(q_{k'}, op_{k'})_{in}$.

Note that the $(init)_{in}$ is also the initial state s_0 of the whole game. In the sequel we denote the winning strategy of the game $\mathcal{G}(\mathcal{M})$ by π^*.

3. First observe that, since T_{a_1}, T_{b_1}, T_{a_2}, T_{b_2}, and T_c form a partition of the terminal states of $\mathcal{G}(\mathcal{M})$, for any pair of strategies π and σ, $\Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{a_1}) + \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{a_2}) + \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{b_1}) + \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{b_2}) + \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_c) = 1$. It follows that for any winning strategy π, it must be the case that for any Player 2 strategy σ, $\Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{a_1}) = \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{a_2}) = \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{b_1}) = \Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_{b_2}) = \frac{1}{6}$ and $\Pr^{\pi,\sigma}_{\mathcal{G},s_0}(\Diamond T_c) = \frac{1}{3}$.

   We then show that, in $\mathcal{G}(\mathcal{M})$, π^* must guarantee that, under any Player 2 strategy σ, the following properties hold:

   (a) For each state $(q_k, inc_j)_{in}$, $(q_k, dec_j)_{in}$, the reachability probability to T_{b_1} and the reachability probability to T_{b_2} both must be exactly $\frac{1}{6}$.

   (b) For each state $(q_k, inc_j)_2$ and $(q_k, dec_j)^{>0}_2$, the reachability probability to T_{a_1} and the reachability probability to T_{a_2} both must be exactly $\frac{1}{6}$.

   To see (a), we examine each gadget, in particular, the "out" states $(q_k, \star)_{out}$, where $\star \in \{dec_j, inc_j\}$. Consider any two different Player 2 strategies σ_1 and σ_2 which select, at $(q_k, \star)_{out}$, the horizontal and the vertical edge respectively. As π^* has to guarantee that for any Player 2 strategy the probability to reach T_{b_1} is the same for σ_1 and σ_2 (namely $\frac{1}{6}$), at $(q_k, \star)_{out}$, the strategy pairs π^*, σ_1 and π^*, σ_2 must give the same probability to reach T_{b_1} as well. From the gadget, the probability to reach T_{b_1} following σ_2 is $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$, hence the claim. The same holds for T_{b_2}.

   To see (b), we examine each gadget, in particular, the state labelled by $(q_k, inc_j)_1$ or $(q_k, dec_j)^{>0}_1$. By the same argument as (a), the probability to T_{a_1} must be $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$. The same holds for T_{a_2}.

4. Below we show some properties for each gadget separately.

   Init. A basic observation is that at Player 1 state s_0, π must select the edge $\langle s_0, s_1 \rangle$ with probability $x = \frac{2}{3} = \frac{2}{3 \cdot 2^0}$. To see this, consider the strategy σ for Player 2 which selects s_4 at state $(init)_{out}$. As the probability of reaching

T_{a_1} under π^* and σ is $\frac{1}{6}$, we have that

$$\frac{1}{2} \cdot \frac{1}{2} \cdot (1-x) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6},$$

yielding that $x = \frac{2}{3}$, as desired.
By a similar argument, at state s_2, for π the probability of selecting the edge $\langle s_2, s_3 \rangle$ must be $\frac{2}{3} = \frac{2}{3 \cdot 2^0}$.
Increment. A basic observation is that when the probability of selecting edge $\langle s_5, s_6 \rangle$ for π^* is $\frac{2}{3 \cdot 2^{c_j}}$, then the probability of the edge $\langle s_9, s_{10} \rangle$ must be $\frac{2}{3 \cdot 2^{c_j+1}}$. To see this, suppose the probability of the edge $\langle s_9, s_{10} \rangle$ is x, and consider a Player 2 strategy σ which selects the vertical edge at $(q_k, inc_j)_{out}$. By (a), the reachability probability to T_{b_j} must be $\frac{1}{6}$. This entails that

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3 \cdot 2^{c_j}} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot (1-x) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6},$$

which implies that $x = \frac{2}{3 \cdot 2^{c_j+1}}$, as desired.
Similarly, if the probability of selecting edge $\langle s_7, s_8 \rangle$ for π^* is $\frac{2}{3 \cdot 2^{c_i}}$, then the probability of selecting edge $\langle s_{11}, s_{12} \rangle$ must be $\frac{2}{3 \cdot 2^{c_i}}$ as well. To see this, we repeat the same argument as the previous case and consider the the reachability probability to T_{b_i} which yields, by (a), that

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3 \cdot 2^{c_i}} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot (1-x) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6},$$

where x is the probability of selecting edge $\langle s_{11}, s_{12} \rangle$ for π^*. This implies that $x = \frac{2}{3 \cdot 2^{c_i}}$, as desired.
Decrement. A basic observation is that when entering the state $(q_k, dec_j)_{in}$, suppose that π^* selects the edge labelled by "> 0," and that the probability of selecting the edge $\langle s_{13}, s_{14} \rangle$ is $\frac{2}{2^{c_j}}$, then the probability of selecting edge $\langle s_{17}, s_{18} \rangle$ must be $\frac{2}{3 \cdot 2^{c_j-1}}$. To see this, suppose the probability of the edge $\langle s_{17}, s_{18} \rangle$ is x, and consider a Player 2 strategy σ which selects the vertical edge at $(q_k, dec_j)_{out}$. It follows that the reachability probability to T_{b_j} must satisfy

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{2^{c_j}} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot (1-x) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot = \frac{1}{6},$$

which implies that $x = \frac{2}{3 \cdot 2^{c_j-1}}$, as desired.
Similarly, suppose that π^* selects the edge labelled by "> 0," and that the probability of selecting edge $\langle s_{15}, s_{16} \rangle$ is $\frac{2}{3 \cdot 2^{c_i}}$, then the probability of selecting edge $\langle s_{19}, s_{20} \rangle$ must be $\frac{2}{3 \cdot 2^{c_i}}$ as well. To see this, we we repeat the same argument as the previous case and consider the reachability probability to T_{b_i} which yields

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3 \cdot 2^{c_i}} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot (1-x) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

This implies that $x = \frac{2}{3 \cdot 2^{c_i}}$, as desired.

5. As the next step, we shall verify that when two instructions are "glued" together, the counter values do not change.
Init+Increment. We show that the probabilities of selecting edges $\langle s_5, s_6\rangle$ and $\langle s_7, s_8\rangle$ for π^* must be $x = \frac{2}{3}$. To see this, consider the Player 2 strategy σ which selects the vertical edge at state $(q_k, inc_j)_1$. Since from $(init)_{in}$ the reachability to T_{a_j} must be $\frac{1}{6}$, we have that
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot(1-x)+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6},$$
which implies that $x = \frac{2}{3}$, as desired.
Init+Decrement. We show that at state $(q_k, dec_j)_{in}$, π^* must choose the edge labelled by "$= 0$." To see this, suppose the opposite, i.e., π^* chooses the edge labelled by "> 0." The reachability probability to T_{b_j} is
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot(\frac{2}{3}+\frac{1}{3}\cdot x)>\frac{1}{6},$$
which contradicts (b).
Increment+Increment. The first instruction is $l_h : c_{j'} := c_{j'} + 1$, goto l_k, and the second instruction is $l_k : c_j := c_j + 1$. We show that the probability of selecting edge $\langle s_5, s_6\rangle$ for π^* must be $\frac{2}{3\cdot 2^{c_j}}$, and the probability for edge $\langle s_7, s_8\rangle$ must be $x = \frac{2}{3\cdot 2^{c_i}}$. By (b), from $(q_h, inc_{j'})_2$ the reachability probability to T_{a_j} must be $\frac{1}{6}$, which stipulates
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3\cdot 2^{c_j}}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot(1-x)+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6},$$
yielding that $x = \frac{2}{3\cdot 2^{c_j}}$, as desired.
Increment+Decrement. The first instruction is $l_h : c_{j'} := c_{j'} + 1$, goto l_k, and the second instruction is $l_k : if\ c_j = 0 \cdots$. If $c_j > 0$ when executing instruction l_k, we show that π^* must choose the edge labelled by "> 0." Assume that this is not the case, and we immediately have that the probability to reach T_{a_j} is
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3\cdot 2^{c_j+1}}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{6}<\frac{1}{6},$$
which contradicts (b). Hence, the edge labelled by "> 0" is taken. Then by (b), from $(q_h, inc_{j'})_2$ the probability to reach T_{a_j} must be $\frac{1}{6}$, which gives
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3\cdot 2^{c_j}}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}\cdot(1-x)+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6},$$
yielding that the probability of π^* selecting the edge $\langle s_{13}, s_{14}$ is $x = \frac{2}{2^{c_j}}$.
For the counter i, again by (b), from $(q_h, inc_{j'})_2$ the probability to reach T_{a_j} must be $\frac{1}{6}$, which gives
$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3\cdot 2^{c_i}}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot(1-x)+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6},$$

yielding that $x = \frac{2}{3 \cdot 2^{c_i}}$, as desired.
If $c_j = 0$ when executing instruction l_k, we have that π^* must choose the edge labelled by $= 0$, by exactly the same argument as for the Init+Decrement case.
Decrement+Decrement.
Here we verify two cases: the first case is that $(q_h, dec_{j'})^{=0}_{out}$ is linked with $(q_k, dec_j)_{in}$, which is the same as for the Init+Decrement case; the second case is that $(q_h, dec_{j'})^{>0}_{out}$ is linked with $(q, dec_j)_{in}$, which is the same as for the Increment+Decrement case.
Decrement+Increment.
Here we verify two cases: the first case is that $(q_h, dec_{j'})^{=0}_{out}$ is linked with $(q_k, inc_j)_{in}$, which is the same as for the Init+Increment case; the second case is that $(q_h, dec_{j'})^{>0}_{out}$ is linked with $(q_h, inc_{j'})_{in}$, which is the same as for the Increment+Increment case.

6. We are now in a position to show the main claim which establishes the correctness of the construction, namely, that Player 1 has a winning strategy in $\mathcal{G}(\mathcal{M})$ if and only if $\mathcal{M}$ does *not* terminate. We show two directions:
"$\Leftarrow$". Suppose that $\mathcal{M}$ does not terminate, then consider a Player 1 strategy π^* for $\mathcal{G}(\mathcal{M})$. We can pick π^* such that it follows the counter update, i.e., π^* must perform the following:
   - For the Init gadget, at state s_0, the probability of selecting edge $\langle s_0, s_1\rangle$ is $\frac{2}{3}$, and at state s_2, the probability of selecting edge $\langle s_2, s_3\rangle$ is $\frac{2}{3}$.
   - For each Increment gadget with index k, if the counter values are c_1 and c_2 respectively, then
     - $\langle s_5, s_6\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_j}}$;
     - $\langle s_7, s_8\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_i}}$;
     - $\langle s_9, s_{10}\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_j+1}}$; and
     - $\langle s_{11}, s_{12}\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_i}}$.
   - For each Decrement gadget with index k, suppose the counter values are c_1 and c_2 respectively. Then, if $c_j = 0$, then at state $(q_k, dec_j)_{in}$, π^* selects the edge labelled with "$= 0$," and if $c_j > 0$, then at state $(q_k, dec_j)_{in}$, π^* selects the edge labelled with "> 0," and
     - $\langle s_{13}, s_{14}\rangle$ is chosen with probability $\frac{2}{2^{c_j}}$;
     - $\langle s_{15}, s_{16}\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_i}}$;
     - $\langle s_{17}, s_{18}\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_j-1}}$; and
     - $\langle s_{19}, s_{20}\rangle$ is chosen with probability $\frac{2}{3 \cdot 2^{c_i}}$.

   It is not difficult to verify that π achieves the first five objectives. Furthermore, as $\mathcal{M}$ does not terminate, under any σ, T_t is reached with probability 1. This is because the only way to reach terminal states a_1, b_1, a_1, a_2 or c is by reaching the Termination gadget.
"$\Rightarrow$". For the other direction, suppose that there is a winning Player 1 strategy π^*. Then in order to satisfy the first five objectives, π^* must follow the counter update, as described above. However, in order to satisfy the last objective, i.e. reaching T_t with probability one, π^* must ensure that the probability to reach terminals a_1, b_1, a_1, a_2 and c is zero. This is only possible if the Terminal gadget is never reached, implying that $\mathcal{M}$ does not terminate.

B.3 Proof of Theorem 4

We define the set of vectors than can be achieved by Player 1 strategy π in k steps as

$$R^{\pi}_{s,k} \stackrel{\text{def}}{=} \{\boldsymbol{y} \in \mathbb{R}^n \mid \forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew^{\leq k}(\boldsymbol{r})] \geq \boldsymbol{y}\},$$

where $rew^{\leq k}(\boldsymbol{r})(\lambda) \stackrel{\text{def}}{=} \sum_{j=0}^{k} \boldsymbol{r}(\lambda_j)$, and we let $R_{s,k} \stackrel{\text{def}}{=} \bigcup_{\pi \in \Pi} R^{\pi}_{s,k}$. For all $s \in S$, let X^k_s be the k-th iteration of the functional given by the equations from Theorem 4, starting with $X^0_s = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \leq \boldsymbol{r}(s)\}$.

Proposition 2. *For all $k \geq 0$, it is the case that $R_{s,k} = X^k_s$.*

Proof. We prove the claim by induction on k. The induction hypothesis is

$$\forall s \in S \,.\, \bigcup_{\pi \in \Pi} R^{\pi}_{s,k-1} = X^{k-1}_s,$$

and we want to show that

$$\forall s \in S \,.\, \bigcup_{\pi \in \Pi} R^{\pi}_{s,k} = X^{k}_s.$$

– **Base case.** Let $k = 0$. We have that for all $s \in S$ and all strategies $\pi \in \Pi$,

$$\begin{aligned} R^{\pi}_{s,0} &= \{\boldsymbol{x} \in \mathbb{R}^n \mid \forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew^{\leq 0}(\boldsymbol{r})] \geq \boldsymbol{x}\} \\ &= \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \leq \boldsymbol{r}(s)\} \\ &= X^0_s. \end{aligned}$$

Hence, $\forall s \in S \,.\, \bigcup_{\pi \in \Pi} R^{\pi}_{s,0} = X^0_s$.

– **Induction step.** Suppose the claim holds for $k-1$, i.e. for all $s \in S$ we have that $\bigcup_{\pi \in \Pi} R^{\pi}_{s,k-1} = X^{k-1}_s$. We suppose w.l.o.g. that s has exactly two successors s_1 and s_2. Furthermore, for $\ell \in \{1,2\}$ we define π^{ℓ} to be the strategy π conditioned on picking the edge $\langle s, s_\ell \rangle$, i.e. $\pi^{\ell}(s_\ell \cdot \lambda) \stackrel{\text{def}}{=} \pi(s \cdot s_\ell \cdot \lambda)$. We now distinguish several cases for $s \in S$.

  - $s \in S_\Box$. For any $\pi \in \Pi$ we have that Player 1 picks s_1 with some probability $p \in [0,1]$ and s_2 with probability $1-p$. Hence in s, Player 1 can achieve all points that can be achieved by some convex combination of some points in the successors of s. This can be stated formally as

$$R^{\pi}_{s,k} = \mathsf{dwc}(\bigcup_{p \in [0,1]} (p \times R^{\pi^1}_{s_1,k-1} + (1-p) \times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s)). \quad (1)$$

  Further, for any convex sets $X_\ell \subseteq \mathbb{R}^n$ for $\ell \in \{1,2\}$, by the definition of the convex hull,

$$\bigcup_{p \in [0,1]} (p \times X_1 + (1-p) \times X_2) = \mathsf{conv}(\bigcup_{\ell} X_\ell). \quad (2)$$

Now, from the induction hypothesis and the definition of X_s^k, we get that

$$\begin{aligned}
&\bigcup_{\pi\in\Pi} R^{\pi}_{s,k}\\
&\quad\stackrel{(1)}{=} \bigcup_{\pi\in\Pi} \mathsf{dwc}(\bigcup_{p\in[0,1]}(p\times R^{\pi^1}_{s_1,k-1} + (1-p)\times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s))\\
&\quad= \mathsf{dwc}(\bigcup_{p\in[0,1]}\bigcup_{\pi\in\Pi}(p\times R^{\pi^1}_{s_1,k-1} + (1-p)\times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s))\\
&\quad= \mathsf{dwc}(\bigcup_{p\in[0,1]}(p\times(\bigcup_{\pi\in\Pi} R^{\pi^1}_{s_1,k-1}) + (1-p)\times(\bigcup_{\pi\in\Pi} R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s)))\\
&\quad\stackrel{(2)}{=} \mathsf{dwc}(\mathsf{conv}(\bigcup_{\ell\in\{1,2\}}\bigcup_{\pi\in\Pi} R^{\pi^\ell}_{s_\ell,k-1}) + \boldsymbol{r}(s))\\
&\quad\stackrel{IH}{=} \mathsf{dwc}(\mathsf{conv}(\bigcup_{\ell\in\{1,2\}} X_s^{k-1}) + \boldsymbol{r}(s))\\
&\quad\stackrel{def}{=} X_s^k.
\end{aligned}$$

- $s \in S_\Diamond$. For any $\pi \in \Pi$ we have that Player 2 picks s_1 with some probability $p \in [0,1]$ and s_2 with probability $1-p$. Hence in s Player 1 can only achieve points that can be achieved by any convex combination of some points in the successors of s. This can be stated formally as

$$R^{\pi}_{s,k} = \mathsf{dwc}(\bigcap_{p\in[0,1]}(p\times R^{\pi^1}_{s_1,k-1} + (1-p)\times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s)). \tag{3}$$

Further, for any sets $X_\ell \subseteq \mathbb{R}^n$ for $\ell \in \{1,2\}$,

$$\bigcap_{p\in[0,1]}(p\times X_1 + (1-p)\times X_2) = \bigcap_{\ell} X_\ell, \tag{4}$$

which can be justified as follows:
  - $*$ For any $\boldsymbol{x} \in \bigcap_{p\in[0,1]}(p\times X_1 + (1-p)\times X_2)$, let p be either 1 or 0, we obtain that $\boldsymbol{x} \in X_1$ and $\boldsymbol{x} \in X_2$ respectively.
  - $*$ For $\boldsymbol{x} \in X_1 \cap X_2$, we have that for all $p \in [0,1]$, $p\boldsymbol{x} \in p\times X_1$ and $(1-p)\boldsymbol{x} \in (1-p)\times X_2$, so $\boldsymbol{x} = p\boldsymbol{x} + (1-p)\boldsymbol{x} \in (p\times X_1 + (1-p)\times X_2)$.

We now show that

$$\bigcup_{\pi\in\Pi}\bigcap_{\ell} R^{\pi^\ell}_{s_\ell,k-1} = \bigcap_{\ell}\bigcup_{\pi\in\Pi} R^{\pi}_{s_\ell,k-1}. \tag{5}$$

  - $*$ $\subseteq$. Take $\boldsymbol{x} \in \bigcap_{\ell\in\{1,2\}} R^{\pi^\ell}_{s_\ell,k-1}$ for some $\pi \in \Pi$. Then for any $\ell \in \{1,2\}$, $\boldsymbol{x} \in R^{\pi^\ell}_{s_\ell,k-1} \subseteq \bigcup_{\pi\in\Pi} R^{\pi}_{s_\ell,k-1}$. Hence $\boldsymbol{x} \in \bigcap_\ell \bigcup_{\pi\in\Pi} R^{\pi}_{s_\ell,k-1}$.
  - $*$ $\supseteq$. Take $\boldsymbol{x} \in \bigcap_\ell \bigcup_{\pi\in\Pi} R^{\pi}_{s_\ell,k-1}$. Therefore, for each $\ell \in \{1,2\}$ have a strategy π_ℓ such that $\boldsymbol{x} \in R^{\pi_\ell}_{s_\ell,k-1}$. We construct a strategy π from

π_1 and π_2 as follows: $\pi(s \cdot s_\ell \cdot \lambda) \stackrel{\text{def}}{=} \pi_\ell(s_\ell \cdot \lambda)$ for all ℓ. Then $\pi^\ell = \pi_\ell$, and hence $R^{\pi^\ell}_{s_\ell,k-1} = R^{\pi_\ell}_{s_\ell,k-1}$. Therefore, we have that π satisfies $\boldsymbol{x} \in \bigcap_{\ell\in\{1,2\}} R^{\pi^\ell}_{s_\ell,k-1}$ and hence $\boldsymbol{x} \in \bigcup_{\pi\in\Pi} \bigcap_\ell R^{\pi^\ell}_{s_\ell,k-1}$.

Now, from the induction hypothesis and the definition of X^k_s, we get that

$$\begin{aligned}\bigcup_{\pi\in\Pi} R^\pi_{s,k} &\stackrel{(3),(4)}{=} \bigcup_{\pi\in\Pi} \mathsf{dwc}(\bigcap_\ell R^{\pi^\ell}_{s_\ell,k-1} + \boldsymbol{r}(s)) \\ &= \mathsf{dwc}(\bigcup_{\pi\in\Pi} \bigcap_\ell R^\pi_{s_\ell,k-1} + \boldsymbol{r}(s)) \\ &\stackrel{(5)}{=} \mathsf{dwc}(\bigcap_\ell \bigcup_{\pi\in\Pi} R^\pi_{s_\ell,k-1} + \boldsymbol{r}(s)) \\ &\stackrel{IH}{=} \mathsf{dwc}(\bigcap_\ell (X^{k-1}_s + \boldsymbol{r}(s))) \\ &\stackrel{def}{=} X^k_s.\end{aligned}$$

- $s \in S_\bigcirc$. We have that s_ℓ is picked with probability $\Delta(\langle s, s_\ell\rangle)$. Hence in s Player 1 can achieve all points that can be achieved by the convex combination with coefficients $\Delta(\langle s, s_\ell\rangle)$ of some points in the successors of s. This can be stated formally as

$$R^\pi_{s,k} = \mathsf{dwc}(\Delta(\langle s, s_1\rangle) \times R^{\pi^1}_{s_1,k-1} + \Delta(\langle s, s_2\rangle) \times R^{\pi^2}_{s_2,k-1} + \boldsymbol{r}(s)). \quad (6)$$

Now, from the induction hypothesis and the definition of X^k_s, we get that

$$\begin{aligned}&\bigcup_{\pi\in\Pi} R^\pi_{s,k} \\ &\stackrel{(6)}{=} \bigcup_{\pi\in\Pi} \mathsf{dwc}((\Delta(\langle s, s_1\rangle) \times R^{\pi^1}_{s_1,k-1} + \Delta(\langle s, s_2\rangle) \times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s)) \\ &= \mathsf{dwc}(\bigcup_{\pi\in\Pi} (\Delta(\langle s, s_1\rangle) \times R^{\pi^1}_{s_1,k-1} + \Delta(\langle s, s_2\rangle) \times R^{\pi^2}_{s_2,k-1}) + \boldsymbol{r}(s)) \\ &= \mathsf{dwc}(\Delta(\langle s, s_1\rangle) \times \bigcup_{\pi\in\Pi} R^{\pi^1}_{s_1,k-1} + \Delta(\langle s, s_2\rangle) \times \bigcup_{\pi\in\Pi} R^{\pi^2}_{s_2,k-1} + \boldsymbol{r}(s)) \\ &\stackrel{IH}{=} \mathsf{dwc}(\Delta(\langle s, s_1\rangle) \times X^{k-1}_{s_1} + \Delta(\langle s, s_2\rangle) \times X^{k-1}_{s_2} + \boldsymbol{r}(s)) \\ &\stackrel{def}{=} X^k_s.\end{aligned}$$

□

Proposition 3. *Given a game $\mathcal{G}$, an n-dimensional reward function $\boldsymbol{r}$, and $\varepsilon > 0$, after $k = |S| + \lceil |S| \cdot \frac{\ln(\varepsilon\cdot(n\cdot M)^{-1})}{\ln(1-\delta)} \rceil$ iterations of the functional F from Theorem 4, for any state $s \in S$, the set X^k_s is an ε-approximation of the Pareto set for $\boldsymbol{r}$ of achievable points at state s, where $M = |S| \cdot \frac{\max_{s\in S,i} |\boldsymbol{r}(s)_i|}{\delta}$ for $\delta = p^{|S|}_{\min}$ and $p_{\min}$ being the smallest positive probability in $\mathcal{G}$.*

Proof. From Proposition 2 we know that $X_s^k = R_{s,k}$ for all k, i.e. the Pareto set of points achievable by Player 1 in k steps is computed by k iterations of F.

From the stopping game assumption we know that after $|S|$ steps, the game has terminated with probability at least $\delta = p_{\min}^{|S|}$, where $p_{\min}$ is the minimum positive probability in $\mathcal{G}$. Hence, the maximum change to any dimension to any vector in X_s^k after k steps of the iteration is less than $M \cdot (1-\delta)^{\lfloor \frac{k}{|S|} \rfloor}$, which is also the maximum change that any strategy can make over a strategy that is optimal for k steps.

Hence, for ε-optimality after k steps, we need to pick a k such that $\varepsilon > n{\cdot}M \cdot (1-\delta)^{\lfloor \frac{k}{|S|} \rfloor}$. The factor n is because ε-optimality requires that the strategy achieves a point that is ε-close in each of the n dimensions individually. We get that

$$\begin{aligned}
\varepsilon > n{\cdot}M \cdot (1-\delta)^{\lfloor \frac{k}{|S|} \rfloor} &\Leftrightarrow \ln(\varepsilon) > \ln(n{\cdot}M) + \left\lfloor \frac{k}{|S|} \right\rfloor \cdot \ln(1-\delta) \\
&\Leftrightarrow \frac{\ln(\varepsilon \cdot (n{\cdot}M)^{-1})}{\ln(1-\delta)} < \left\lfloor \frac{k}{|S|} \right\rfloor \\
&\Leftarrow \frac{\ln(\varepsilon \cdot (n{\cdot}M)^{-1})}{\ln(1-\delta)} < \frac{k}{|S|} - 1 \\
&\Leftarrow |S| + |S| \cdot \frac{\ln(\varepsilon \cdot (n{\cdot}M)^{-1})}{\ln(1-\delta)} < k
\end{aligned}$$

Set $k = |S| + \lceil |S| \cdot \frac{\ln(\varepsilon \cdot (n \cdot M)^{-1})}{\ln(1-\delta)} \rceil$. Note that the Pareto set for φ at state s is defined by

$$R_s = \{\boldsymbol{y} \in \mathbb{R}^n \mid \exists \pi \in \varPi \,.\, \forall \sigma \in \varSigma \,.\, \mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[rew(\boldsymbol{r})] \geq \boldsymbol{y}\},$$

which is the set whose approximation we aim to compute. We have the following:

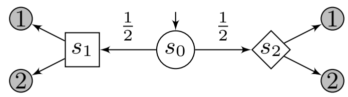
- For any point $\boldsymbol{x} \in R_s$ there is (by definition) a strategy π which achieves $\boldsymbol{x}$, i.e. for all σ we have $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[rew(\boldsymbol{r})] \geq \boldsymbol{x}$. Above we argued that we can find a Player 1 strategy π that after k steps achieves a point that differs from $\boldsymbol{x}$ by at most ε in each dimension. Hence, there is a Player 1 strategy π such that for all Player 2 strategies σ we have that $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[rew^{\leq k}(\boldsymbol{r})] \geq \boldsymbol{x} - \varepsilon$, which means that $\boldsymbol{x} - \varepsilon \in R_{s,k} = X_s^k$.
- For any point $\boldsymbol{x} \in X_s^k = R_{s,k}$, let π be the strategy that ensures $\boldsymbol{x}$ is achieved in k steps, i.e. for all σ we have $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[rew^{\leq k}(\boldsymbol{r})] \geq \boldsymbol{x}$. Again, by the above argument the point $\boldsymbol{x}$ achieved by π in k steps may only change by at most ε in each dimension by any other strategy. Hence we have $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[rew(\boldsymbol{r})] \geq \boldsymbol{x} - \varepsilon$ for all σ, and so $\boldsymbol{x} - \varepsilon \in R_s$. $\square$

C Proofs for Section 4

C.1 Nondeterminacy for disjunctive objectives

Theorem 9 (Nondeterminacy). *Stochastic games with disjunctive objectives are in general not determined already for two objectives.*

Proof. Consider the game $\mathcal{G}$ with initial state s_0 shown below, where the (reachability) objective for Player 1 is to reach state 1 or 2 with probability at least $\frac{3}{4}$.



There does not exist a Player 1 strategy π, which, for any Player 2 strategy σ achieves this objective. To see this, consider the strategy σ for Player 2 given by $\sigma(s_0 s_2) = [\langle s_2, 1\rangle \mapsto 1 - \eta, \langle s_2, 2\rangle \mapsto \eta]$, where $\eta = \pi(s_0 s_1)[\langle s_1, 1\rangle]$, i.e., the probability that π selects the edge $\langle s_1, 1\rangle$ at s_1. Note that σ may depend on π. With this Player 2 strategy, Player 1 may reach both objectives with at most probability $\frac{1}{2}$, and hence neither of the objectives is satisfied. However, for every Player 2 strategy σ, there exists a Player 1 strategy π (depending on σ) to win this game, for example $\pi(s_0 s_1)[\langle s_1, 1\rangle] = 0$ if $\sigma(s_0 s_2)[\langle s_2, 1\rangle] < \frac{1}{2}$, and $\pi(s_0 s_1)[\langle s_1, 1\rangle] = 1$ otherwise. This ensures that the probability to reach one of the targets is at least $\frac{3}{4}$. □

C.2 PSPACE-hardness of boolean combinations of objectives

We prove the PSPACE-hardness of the problem of deciding the existence of a winning strategy for Player 1 to achieve a boolean combination of objectives by reduction from satisfiability of quantified boolean formula (QBF), which is known to be PSPACE-complete. Consider QBF with n variables and m clauses

$$\psi = \exists x_1 \forall x_2 \exists x_3 \ldots \forall x_n \,.\, c_1 \wedge c_2 \wedge \cdots \wedge c_m,$$

where each $c_i = (l_1^i \vee l_2^i \vee l_3^i)$ and $l_j^i \in \{x_1, \neg x_1, \ldots, x_n, \neg x_n\}$. We assume that every clause contains at most one literal for any given variable. The stochastic game that we use in the reduction is shown in Figure 5. Consider the following MQ

$$\varphi = \bigwedge_{i=1}^{m} \Diamond C_i \geq \frac{1}{2^{2\cdot n}} \wedge \tag{7}$$

$$\bigwedge_{i=\{1,3,\ldots,n-1\}} (\Diamond\{p_i\} \leq 0 \vee \Diamond\{n_i\} \leq 0) \tag{8}$$

where set C_i contains state x_j^+ if clause c_i of ψ contains literal x_j, and state x_j^- if c_i contains literal $\neg x_j$, for all j.

First observe that in order to win the game, Player 1 has to use a deterministic strategy. This is ensured by the conjunction in (8), which makes sure that if Player 1 has a winning strategy, then this strategy has to pick either p_i or n_i in state s_i for all i.

We show that ψ is true if and only if Player 1 has a winning strategy for φ.

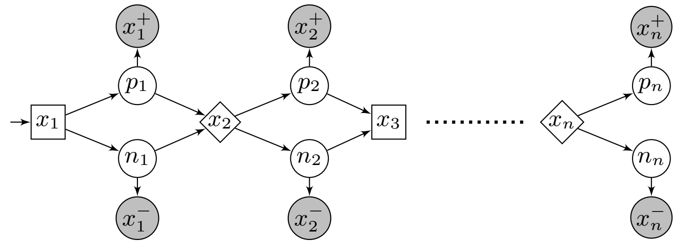


Fig. 5: Game illustrating PSPACE-hardness. Probabilities in stochastic states are uniform.

For the "⇒" direction, let us assume there are functions $q_i : \mathbb{B}^{i-1} \to \mathbb{B}$ for $i \in \{1, 3, \ldots n-1\}$ such that for any $v_2, v_4 \ldots, v_n \in \mathbb{B}$ the formula $c_1 \wedge \ldots \wedge c_m$ is satisfied under the assignment ν defined inductively by

$$\nu(x_i) = \begin{cases} q_i(\nu(x_1), \ldots, \nu(x_{i-1})) & \text{if } i \in \{1, 3, \ldots, n-1\} \\ v_i & \text{if } i \in \{2, 4, \ldots, n\} \end{cases}$$

This can be directly transformed into a Player 1 strategy in the game from Figure 5 where $\star_i \in \{p_i, n_i\}$, $b(p_i) = 1$ and $b(n_i) = 0$:

$$\pi(x_1 \star_1 \ldots x_{i-1} \star_{i-1}) = \begin{cases} p_i & \text{if } q_i(b(\star_1), \ldots, b(\star_{i-1})) = 1 \\ n_i & \text{otherwise.} \end{cases}$$

Let σ be an arbitrary strategy for Player 2. Let us consider a path $x_1 \star_1 \ldots x_n \star_n$ such that, for every $i \in \{1, 3, \ldots n-1\}$ we have $\pi(x_1 \star_1 \ldots x_i) = \star_i$, and for every $i \in \{2, 4, \ldots n\}$ we have $\sigma(x_1 \star_1 \ldots x_i)(\star_i) \geq 0.5$. Note that such a path always exsits since Player 2 has exactly two choices in every state it controls. By the properties of the functions q_i and by the construction of π we have that the valuation μ which to x_i assigns $b(\star_i)$ satisfies every $c_1 \wedge \ldots \wedge c_m$. Fix c_j for $1 \leq j \leq m$, there must be a literal which makes c_j satisfied under μ, let x_k be such a variable. By the definition of the game we have that the state x_k^+ (resp. x_k^-) is in the set C_j if c_j contains x_k (resp. $\neg x_k$). Thus, C_j is reached at least with probability

$$\Big(\prod_{i=1}^{k} \frac{1}{2}\Big) \cdot \Big(\prod_{i \in \{2,4,\ldots e(k)\}} \sigma(x_1 \star_1 \ldots x_i)(\star_i) \Big) \quad \geq \quad \frac{1}{2^{2 \cdot k}} \geq \quad \frac{1}{2^{2 \cdot n}}$$

which is the probability of the path $x_1 \star_1 \ldots x_k p_k x_k^+$ (resp. $x_1 \star_1 \ldots x_k n_k x_k^-$), where $e(k) = k$ if k is even and $e(k) = k-1$ if k is odd.

The other direction "⇐" can be proved by directly constructing assignment functions q_i and q_i' from the winning strategy π for Player 1. This can be achieved because the winning strategy must be deterministic, as discussed above. $\square$

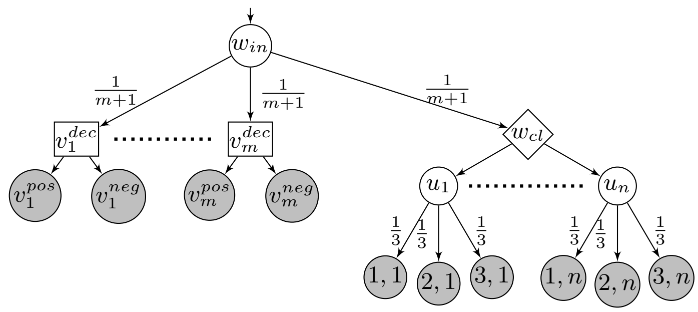


Fig. 6: Game illustrating NP-hardness. State label "i, j" corresponds to v_i^{pos} (resp. v_i^{neg}) if the ith variable in clause c_j is positive (resp. negative).

C.3 NP-hardness of disjunctive objectives

Lemma 2. *The problem of deciding the existence of the winning* Player 1 *strategy in a stochastic game with a disjunctive objective is NP-hard.*

Proof. We reduce 3SAT to the problem. Let Ψ be a 3CNF formula with clauses $c_1, \ldots, c_n$ and variables $x_1, \ldots, x_m$. We construct the game shown in Figure 6, where the terminal states are v_i^{pos} and v_i^{neg} for all $1 \leq i \leq m$, corresponding to the valuations of the variables. We further construct $2m$ target sets, each a singleton containing either v_i^{pos} or v_i^{neg}. We claim that there is a satisfying assignment to Ψ if and only if there is a strategy π which reaches at least one of the target sets with probability at least $q = \frac{1}{m+1} + \frac{1}{m+1} \cdot \frac{1}{n} \cdot \frac{1}{3}$.

For the "$\Rightarrow$" direction, given a satisfying assignment μ, we define a strategy π that goes go to v_i^{pos} from v_i^{dec} if and only if $\mu(x_i) = 1$ and to v_i^{neg} otherwise, for all i. Consider any strategy σ for Player 2, and let j be such that σ picks u_j with probability at least $\frac{1}{n}$ in w_{cl} (such j surely exists). There must be a literal in c_j which is satisfied under μ. Let x_i be a variable in this literal. If the literal is of the form x_i, then we get that the state v_i^{pos} is reached on a path $w_{in} v_i^{dec} v_i^{pos}$ with probability $\frac{1}{m+1}$ and on a path $w_{in} w_{cl} u_j v_i^{pos}$ with probability at least $\frac{1}{m+1} \cdot \frac{1}{n} \cdot \frac{1}{3}$, and so the objective is satisfied. Similarly, if the literal is of the form $\neg x_i$, we get the same line of argument, replacing v_i^{pos} with v_i^{neg}.

For the "$\Leftarrow$" direction, we assume that π is memoryless deterministic (see Theorem 7). Define a valuation μ by $\mu(x_i) = 1$ if and only if v_i^{pos} is reached from v_i^{dec}. Let c_j be an arbitrary clause in Ψ, and consider a strategy σ which goes deterministically to u_j in w_{cl}. There must be a target set T satisfying $\mathrm{Pr}_{w_{in}}^{\pi,\sigma}(\Diamond T) \geq q$. Fix one such set T, and suppose that $T = \{v_i^{pos}\}$. This set can be reached by the path $w_{in} v_i^{dec} v_i^{pos}$ and the paths starting with $w_{in} w_{cl}$. Since the first path has probability only $\frac{1}{m+1}$, the other paths must have a non-zero probability. But since σ is deterministic and selects u_j, there must be a path $w_{in} w_{cl} u_j v_i^{pos}$, which means that the literal x_i is in c_j under μ. Since this literal is true under μ, c_j is satisfied. For $T = \{v_i^{neg}\}$ we proceed similarly. $\square$

C.4 Multiobjective queries in CNF

We present the proof of Theorem 6.

By $\boldsymbol{q}_i$ and $\boldsymbol{u}_i$ we denote the vectors $(q_{i,1},\dots q_{i,m})$ and $(u_{i,1},\dots u_{i,m})$. Fix $\pi \in \Pi$. For the "⇒" direction, for all $1 \le i \le n$ define $R^\pi_s[i] \stackrel{\text{def}}{=} \{\boldsymbol{y} \in \mathbb{R}^m_{\pm\infty} \mid \exists \sigma \in \Sigma . \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(\boldsymbol{q}_i)] = \boldsymbol{y}\}$. Fix $\varepsilon > 0$ and $1 \le i \le n$. Because π achieves ψ it must also achieve $\bigvee_{j=1}^m q_{i,j} \ge u_{i,j}$. Hence, for every $\boldsymbol{y} \in \mathsf{up}(R^\pi_s[i])$ there is a j satisfying $y_j > u_{i,j} - \frac{\varepsilon}{2}$, and so $\boldsymbol{u}_i - \frac{\varepsilon}{2} \notin R^\pi_s[i]$. By Lemma 1, since $\boldsymbol{u}_i - \varepsilon$ is not in the closure of $\mathsf{up}(R^\pi_s[i])$, and since $R^\pi_s[i]$ satisfies the conditions of the lemma, we can obtain a vector $\boldsymbol{x}_i$ for $\mathsf{up}(R^\pi_s[i])$ and $\boldsymbol{u}_i - \varepsilon$. Fix any strategy σ. We have $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(\boldsymbol{q}_i)] \in \mathsf{up}(R^\pi_s)$, and it follows that

$$\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(r_i)] = \int\limits_{\lambda \in \Omega_{\mathcal{G},s}} \sum_{k=0}^{\infty}\sum_{j=1}^{m} x_{i,j} \cdot q_{i,j}(\lambda_k)\, d\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s} \stackrel{(*)}{=} \int\limits_{\lambda \in \Omega_{\mathcal{G},s}} \sum_{j=1}^{m} x_{i,j} \cdot \sum_{k=0}^{\infty} q_{i,j}(\lambda_k)\, d\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s}$$
$$= \sum_{j=1}^{m} x_{i,j} \cdot \int\limits_{\Omega_{\mathcal{G},s}} \sum_{k=0}^{\infty} q_{i,j}\, d\mathrm{Pr}^{\pi,\sigma}_{\mathcal{G},s} = \boldsymbol{x}_i \cdot \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(\boldsymbol{q}_i)] \ge \boldsymbol{x}_i \cdot (\boldsymbol{u}_i - \varepsilon) = v_i.$$

The equality marked with $(*)$ holds because $\sum_{k=0}^\infty \sum_{j=1}^m x_{i,j} \cdot q_{i,j}(\lambda_k) = \sum_{j=1}^m x_{i,j} \cdot \sum_{k=0}^\infty q_{i,j}(\lambda_k)$ for almost every λ; this is true because for every j we either have $x_{i,j} = 0$, or the sum $\sum_{k=0}^\infty q_{i,j}(\lambda_k)$ is strictly greater than $-\infty$ for almost all λ.

For the "⇐" direction, for each $\varepsilon > 0$ we have non-zero vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n \in \mathbb{R}^m_{\ge 0}$ such that π achieves φ. Assume for the sake of contradiction that this π does not achieve ψ. Then there exists a Player 2 strategy σ and an index i such that for all j we have that $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(q_{i,j})] = u_{i,j} - \tau_j < u_{i,j}$ for some $\tau_j > 0$ (and possibly $\tau_j = \infty$). Now fix such a strategy σ, a corresponding index i, and let $\varepsilon = \frac{\min_j \tau_j}{2} < \infty$. We can pick $\boldsymbol{x}_i$ such that π achieves φ, and hence $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(r_i)] \ge \boldsymbol{x}_i \cdot (\boldsymbol{u}_i - \varepsilon)$. Consequently $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(r_i)] = \boldsymbol{x}_i \cdot \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(\boldsymbol{q}_i)]$ by the same argument as above. Thus $\boldsymbol{x}_i \cdot \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(\boldsymbol{q}_i)] \ge \boldsymbol{x}_i \cdot (\boldsymbol{u}_i - \varepsilon)$, and because $\boldsymbol{x}_i$ is non-zero and has no negative components, there must be a j such that $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(q_{i,j})] \ge u_{i,j} - \varepsilon > u_{u,j} - \tau_j = \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[rew(q_{i,j})]$, a contradiction. □

C.5 Memoryless deterministic strategies for DQs

We present the proof of Theorem 7.

Assume that there exists a strategy achieving the DQ $\varphi = \bigvee_{j=1}^m r_j \ge v_j$. Then by Theorem 6 we know that for all $\varepsilon > 0$ there exists a winning strategy π_ε which achieves the single objective $\phi_\varepsilon = \forall \sigma \in \Sigma . \mathbb{E}^{\pi_\varepsilon,\sigma}_{\mathcal{G},s}[\boldsymbol{x}_\varepsilon \cdot rew(\boldsymbol{r})] \ge \boldsymbol{x}_\varepsilon \cdot (\boldsymbol{v} - \varepsilon)$ for some $\boldsymbol{x}_\varepsilon \in \mathbb{R}^m$. We can assume π_ε is memoryless deterministic (MD), because such strategies suffice to achieve a single-objective expected total reward in stochastic games [11]. Define a (countable) set $\Gamma = \{k^{-1} \mid k \in \mathbb{N}\}$. We know that for every $\varepsilon \in \Gamma$ there exists an MD strategy π_ε achieving ϕ_ε. Because the number of MD strategies if finite, there must exists some π^*, which is MD and winning for infinitely many $\varepsilon \in \Gamma$. We prove that this π^* actually achieves ϕ_ε for all $\varepsilon > 0$. Assume for a contradiction that there is some $\delta > 0$ such that

$$\forall \boldsymbol{x}_\delta \in \mathbb{R}^m_{\ge 0} . \exists \sigma \in \Sigma . \mathbb{E}^{\pi^*,\sigma}_{\mathcal{G},s}[\boldsymbol{x}_\delta \cdot rew(\boldsymbol{r})] < \boldsymbol{x}_\delta \cdot (\boldsymbol{v} - \delta). \tag{9}$$

Pick $\varepsilon \in \Gamma$ such that $\varepsilon < \delta$ and

$$\exists \boldsymbol{x}_\varepsilon \in \mathbb{R}^m_{\geq 0} \,.\, \forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi^*,\sigma}_{\mathcal{G},s}[\boldsymbol{x}_\varepsilon \cdot \mathit{rew}(\boldsymbol{r})] \geq \boldsymbol{x}_\varepsilon \cdot (\boldsymbol{v} - \varepsilon).$$

But $(\boldsymbol{v} - \varepsilon) > (\boldsymbol{v} - \delta)$, and hence

$$\forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi^*,\sigma}_{\mathcal{G},s}[\boldsymbol{x}_\varepsilon \cdot \mathit{rew}(\boldsymbol{r})] > \boldsymbol{x}_\varepsilon \cdot (\boldsymbol{v} - \delta),$$

which contradicts (9). □

C.6 Extensions of separating hyperplane theorem

Proof of Lemma 1 Let $I \subseteq \{1, \ldots, m\}$ be the set of indices such that all $\boldsymbol{w} \in W$ satisfy $\mathrm{sgn}(w_i) \leq 0$. Let U be the closure of $\mathsf{up}(W) \cap \mathbb{R}^m$.

If $U = \emptyset$, then we define $\boldsymbol{x}$ by $x_i = 0$ if $i \in I$ and $x_i = 1$ otherwise. For any $\boldsymbol{w} \in W$, we have

$$\boldsymbol{w} \cdot \boldsymbol{x} = \sum_{i \in I} w_i \cdot x_i + \sum_{i \notin I} w_i \cdot x_i$$

where the left summand is 0. We argue that the right summand must be positive. Suppose otherwise, then it must be the case that all w_i for $i \notin I$ are real numbers. But then we can replace any $-\infty$ in $\boldsymbol{w}$ by any real number, and get a vector which by the definition of U lies in U, contradicting the property that $U = \emptyset$.

Suppose $U \neq \emptyset$. First we argue that U is convex. Let $\boldsymbol{a}, \boldsymbol{b} \in U$, and let $t \in (0, 1)$, we show that $\boldsymbol{c} := t\boldsymbol{a} + (1-t)\boldsymbol{b} \in U$. Let $\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}$ be vectors with all components non-negative such that $\boldsymbol{a} - \bar{\boldsymbol{a}}$ and $\boldsymbol{b} - \bar{\boldsymbol{b}}$ are in W. Then by convexity of W the vector

$$t \cdot (\boldsymbol{a} - \bar{\boldsymbol{a}}) + (1-t) \cdot (\boldsymbol{b} - \bar{\boldsymbol{b}}) = \boldsymbol{c} - (t \cdot \bar{\boldsymbol{a}} + (1-t)\bar{\boldsymbol{b}})$$

is in W, and since we have $\boldsymbol{c} - (t \cdot \bar{\boldsymbol{a}} + (1-t)\bar{\boldsymbol{b}}) \leq \boldsymbol{c}$, we get that $\boldsymbol{c} \in U$.

Let $\tau > 0$ be the smallest number such that $\boldsymbol{z} + \tau$ lies in the closure of U. Denote $\bar{\boldsymbol{z}} := \boldsymbol{z} + \tau$. By the separating hyperplane theorem [14], there is some non-zero vector $\boldsymbol{y} \in \mathbb{R}^m$, s.t. for all $\boldsymbol{w} \in U$, $\boldsymbol{w} \cdot \boldsymbol{y} \geq \bar{\boldsymbol{z}} \cdot \boldsymbol{y}$.

We show that the vector $\boldsymbol{y}$ satisfies the condition 1, i.e. that all components of $\boldsymbol{y}$ are non-negative. Assume for the sake of contradiction that for some $1 \leq j \leq m$ we have $y_j < 0$. Let $\boldsymbol{w}$ be any point from U. Let $d = \boldsymbol{w} \cdot \boldsymbol{y} - \bar{\boldsymbol{z}} \cdot \boldsymbol{y}$, and let $\boldsymbol{w'}$ be the vector which is obtained from $\boldsymbol{w}$ by replacing jth coordinate with $w_j + \frac{d+1}{-y_j}$. Since $\frac{d+1}{-y_j}$ is positive and U is upwards closed in $\mathbb{R}^m_{\pm\infty}$, we have $\boldsymbol{w'} \in U$. So

$$\begin{aligned} \boldsymbol{w'} \cdot \boldsymbol{y} = \sum_h w'_h \cdot y_h &= \frac{d+1}{-y_j} \cdot y_j + \sum_h w_h \cdot y_h \\ &= -(d+1) + \boldsymbol{w} \cdot \boldsymbol{y} = \bar{\boldsymbol{z}} \cdot \boldsymbol{y} - 1, \end{aligned}$$

which is a contradiction, since $\bar{\boldsymbol{z}} \cdot \boldsymbol{y} \leq \boldsymbol{w'} \cdot \boldsymbol{y}$.

Let $\varepsilon := \bar{\boldsymbol{z}} \cdot \boldsymbol{y} - \boldsymbol{z} \cdot \boldsymbol{y}$, we have $\varepsilon > 0$. We define $\boldsymbol{x}$ by putting $x_i = y_i$ for $i \in I$, and $x_i = y_i + \frac{\varepsilon}{|\sum_{j=1}^m z_j|+1}$ for $i \notin I$. The vector $\boldsymbol{x}$ obviously satisfies the condition 1.

We show that $\boldsymbol{x}$ satisfies the condition 2. Let $L \subseteq \{1, \ldots, m\}$ be the set of indices such that $l \in L$ if and only if there is $\boldsymbol{u} \in W$ with $u_l = -\infty$. Note that $L \subseteq I$. Since $x_l = y_l$ for all $l \in L$, it suffices to show that $y_l = 0$ for all $l \in L$. If $L = \emptyset$, there is nothing we need to prove. Otherwise, because W is convex, there is a vector $\boldsymbol{u} \in W$ with $u_l = -\infty$ for all $l \in L$, and so for arbitrary $\alpha \in \mathbb{R}^m$ the set U contains the vector $\boldsymbol{u}^\alpha$ defined by $u^\alpha_l = u_l$ if $l \in L$ and $u^\alpha_l = \alpha$ otherwise. Then $\lim_{\alpha \to -\infty} \boldsymbol{x} \cdot \boldsymbol{u}^\alpha = -\infty$ if $y_l > 0$ for any $l \in L$, contradicting that $\boldsymbol{y} \cdot \boldsymbol{u}^\alpha \geq \boldsymbol{y} \cdot \boldsymbol{z}$ for all α.

Finally, we prove the condition 3. Let $\boldsymbol{w} \in W$. The product $\boldsymbol{w} \cdot \boldsymbol{x}$ is defined by the condition 2. Also,

$$
\begin{aligned}
\boldsymbol{w} \cdot \boldsymbol{x} &= \sum_{i \in I} w_i \cdot x_i + \sum_{i \notin I} w_i \cdot x_i \\
&= \sum_{i \in I} w_i \cdot y_i + \sum_{i \notin I} w_i \cdot (y_i + \frac{\varepsilon}{|\sum_{j=1}^m z_j| + 1}) \\
&= \boldsymbol{w} \cdot \boldsymbol{y} + \sum_{i \notin I} w_i \cdot \frac{\varepsilon}{|\sum_{j=1}^m z_j| + 1} \\
&\geq \boldsymbol{w} \cdot \boldsymbol{y} \geq \bar{\boldsymbol{z}} \cdot \boldsymbol{y} = \boldsymbol{z} \cdot \boldsymbol{y} + \varepsilon \\
&= \Big(\sum_{i \in I} z_i \cdot x_i + \sum_{i \notin I} z_i \cdot (x_i - \frac{\varepsilon}{|\sum_{j=1}^m z_j| + 1}) \Big) + \varepsilon \\
&= \boldsymbol{z} \cdot \boldsymbol{x} - \Big(\sum_{i \notin I} z_i \frac{\varepsilon}{|\sum_{j=1}^m z_j| + 1} \Big) + \varepsilon \\
&\geq \boldsymbol{z} \cdot \boldsymbol{x}.
\end{aligned}
$$

where the first inequality follows because all w_i are positive for $i \notin I$.

Extension of Lemma 1 to boundary points By a modification of the proof of Lemma 1 we can obtain the following lemma, which establishes an existence of separating hyperplanes for points on a boundary of some set in Euclidean space.

Lemma 3. *Let $W \subseteq \mathbb{R}^m$ be a convex set satisfying that for all j, whenever $\boldsymbol{x} \in W$ and $sgn(x_j) \geq 0$ (resp. $sgn(x_j) \leq 0$), then $sgn(y_j) \geq 0$ (resp. $sgn(y_j) \leq 0$) for all $\boldsymbol{y} \in W$. Let $\boldsymbol{z} \in \mathbb{R}^m$ be a point which does not lie in the* interior *of* $\mathsf{up}(W)$. *Then there is a non-zero vector $\boldsymbol{x} \in \mathbb{R}^m$ such that the following conditions hold:*

1′. for all $1 \leq j \leq m$ we have $x_j \geq 0$;
3′. for all $\boldsymbol{w} \in W$ we have $\boldsymbol{w} \cdot \boldsymbol{x} \geq \boldsymbol{z} \cdot \boldsymbol{x}$.

Proof. We can obtain the proof by the following modifications of the proof of Lemma 1: Since $\boldsymbol{z}$ possibly lies on the boundary of $\mathsf{up}(W)$, we might get $\tau = 0$

and so $\varepsilon = 0$. Nevertheless this does not cause any problems since the part of the proof proving condition 2 of Lemma 1 will be omitted, and the remaining parts, proving conditions 1 and 3 carry over without any change.

C.7 Pareto set approximation

In this section we provide the proof of Theorem 8.

First, observe that for stopping games, the maximum expected reward is a real number (i.e., $M \in \mathbb{R}$). Hence by Theorem 6 and Lemma 3 (see Appendix C.6), we have that a DQ $\varphi = \bigvee_{j=1}^{m} r_j \geq v_j$ is achievable if and only if there exists π and $\boldsymbol{x} \in \mathbb{R}^m_{\geq 0}$ such that $\forall \sigma \in \Sigma \,.\, \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{x} \cdot rew(\boldsymbol{r})] \geq \boldsymbol{x} \cdot \boldsymbol{v}$, which is a single-objective query decidable by a NP$\cap$coNP oracle.

Given a finite set $X \subseteq \mathbb{R}^m$, we can compute values $d_{\boldsymbol{x}} = \sup_\pi \inf_\sigma \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{x} \cdot rew(\boldsymbol{r})]$ for all $\boldsymbol{x} \in X$, and define $U_X = \bigcup_{\boldsymbol{x} \in X} \{\boldsymbol{p} \mid \boldsymbol{x} \cdot \boldsymbol{p} \leq d_{\boldsymbol{x}}\}$. It is not difficult to see that U_X yields an under-approximation of achievable points.

Let $\tau = \frac{\varepsilon}{2 \cdot m^2 \cdot (M+1)}$. We argue that when we let X be the set of all non-zero vectors $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 1$, and where all x_i are of the form $\tau \cdot k_i$ for some $k_i \in \mathbb{N}$, we obtain an ε-approximation of the Pareto set by taking all Pareto points on U_X. We show that, for any point in the Pareto set, there is an ε-close point in U_X. Consider any point $\boldsymbol{p}$ in the Pareto set and let π be a strategy which achieves this point. Note that for some $\boldsymbol{y} \in \mathbb{R}^m_{\geq 0}$ such that $\|\boldsymbol{y}\| = 1$ we have $\boldsymbol{p} \cdot \boldsymbol{y} = d_{\boldsymbol{y}}$, since otherwise $\boldsymbol{p}$ would not be a Pareto point. Let $\boldsymbol{x} = \mathrm{argmin}_{\boldsymbol{z} \in X} \|\boldsymbol{z} - \boldsymbol{y}\|$ be a vector in X, which is closest to $\boldsymbol{y}$. Note that $d_{\boldsymbol{y}} - d_{\boldsymbol{x}} \leq m \cdot M \cdot \tau$ and thus $d_{\boldsymbol{x}} \geq d_{\boldsymbol{y}} - m \cdot M \cdot \tau$. For the point $\boldsymbol{q} = \boldsymbol{p} - \frac{\varepsilon}{m}$, we have $\boldsymbol{q} \cdot \boldsymbol{x} \leq d_{\boldsymbol{x}}$ because

$$
\begin{aligned}
\boldsymbol{q} \cdot \boldsymbol{x} = \boldsymbol{p} \cdot \boldsymbol{x} - \frac{\boldsymbol{\varepsilon}}{m} \cdot \boldsymbol{x} \leq \boldsymbol{p} \cdot \boldsymbol{y} + \boldsymbol{p} \cdot \boldsymbol{\tau} - (\frac{\boldsymbol{\varepsilon}}{m} \cdot \boldsymbol{y} - \frac{\boldsymbol{\varepsilon}}{m} \cdot \boldsymbol{\tau}) \leq d_{\boldsymbol{y}} + \boldsymbol{M} \cdot \boldsymbol{\tau} - \frac{\varepsilon}{m} + m \cdot \tau \\
\leq d_{\boldsymbol{y}} + m \cdot (M+1) \cdot \tau - \frac{\varepsilon}{m} \leq d_{\boldsymbol{y}} - m \cdot M \cdot \tau \leq d_{\boldsymbol{x}},
\end{aligned}
$$

and so $\boldsymbol{q} \in U_X$. Since $\|\boldsymbol{p} - \boldsymbol{q}\| \leq \varepsilon$, this concludes the proof. The result follows from the fact that $|X| \leq (\frac{2 \cdot m^2 \cdot (M+1)}{\varepsilon})^{m-1}$. The other direction can be proved similarly. $\square$