Decidability Results for Multi-objective Stochastic Games

Romain Brenguier and Vojtěch Forejt

Department of Computer Science, University of Oxford, UK

Abstract. We study stochastic two-player turn-based games in which the objective of one player is to ensure several infinite-horizon total reward objectives, while the other player attempts to spoil at least one of the objectives. The games have previously been shown not to be determined, and an approximation algorithm for computing a Pareto curve has been given. The major drawback of the existing algorithm is that it needs to compute Pareto curves for finite horizon objectives (for increasing length of the horizon), and the size of these Pareto curves can grow unboundedly, even when the infinite-horizon Pareto curve is small.
By adapting existing results, we first give an algorithm that computes the Pareto curve for determined games. Then, as the main result of the paper, we show that for the natural class of stopping games and when there are two reward objectives, the problem of deciding whether a player can ensure satisfaction of the objectives with given thresholds is decidable. The result relies on an intricate and novel proof which shows that the Pareto curves contain only finitely many points.
As a consequence, we get that the two-objective discounted-reward problem for unrestricted class of stochastic games is decidable.

1 Introduction

Formal verification is an area of computer science which deals with establishing properties of systems by mathematical means. Many of the systems that need to be modelled and verified contain controllable decisions, which can be influenced by a user, and behaviour which is out of the user's control. The latter can be further split into events whose presence can be quantified, such as failure rate of components, and events which are considered to be completely adversarial, such as acts of an attacker who wants to break into the system.

Stochastic turn-based games are used as a modelling formalism for such systems [6]. Formally, a stochastic game comprises three kinds of states, owned by one of three players: Player 1, Player 2, and the stochastic player. In each state, one or more transitions to successor states are available. At the beginning of a play, a token is placed on a distinguished initial state, and the player who controls it picks a transition and the token is moved to the corresponding successor state. This is repeated ad infinitum and a path, comprising an infinite sequence of states, is obtained. Player 1 and Player 2 have a free choice of transitions, and

the recipe for picking them is called a strategy. The stochastic player is bound to pick each transition with a fixed probability that is associated with it.

The properties of systems are commonly expressed using rewards, where numbers corresponding to gains or losses are assigned to states of the system. The numbers along the infinite paths are then summed, giving the total reward of an infinite path, intuitively expressing the energy consumed or the profit made along a system's execution. Alternatively, the numbers can be summed with a discounting $\delta < 1$, giving discounted reward. It formalises the fact that immediate gains matter more than future gains, and it is particularly important in economics where money received early can be invested and yield interest.

Traditionally, the aim of one player is to make sure the expected (discounted) total reward exceeds a given bound, while the other player tries to ensure the opposite. We study the *multi-objective problem* in which each state is given a tuple of numbers, for example corresponding to both the profit made on visiting the state, and the energy spent. Subsequently, we give a bound on both profit and energy, and Player 1 attempts to ensure that the expected total profit and expected total energy exceed (or do not exceed) the given bound, while Player 2 tries to spoil this by making sure that at least one of the goals is not met.

The problem has been studied in [9], where it has been shown that Pareto optimal strategies might not exist, and the game might not be determined (for some bounds neither of the players have ε-optimal strategies). A value iteration algorithm has been given for approximating the Pareto curve of the game, i.e. the bounds Player 1 can ensure. The algorithm successively computes, for increasing n, the sets of bounds Player 1 can ensure if the length of the game is restricted to n steps. The approach has two major drawbacks. Firstly, the algorithm cannot decide, for given bounds, if Player 1 can achieve them. Secondly, it does not scale well since the representation of the sets can grow with increasing n, even if the ultimate Pareto curve is small.

The above limitations show that it is necessary to design alternative solution approaches. One of the promising directions is to characterise the shape of the set of achievable bounds, for computing it efficiently. The value iteration of [9] allows us to show that the sets are convex, but no further observations can be made, in particular it is not clear whether the sets are convex polyhedra, or if they can have infinitely many extremal points. The main result of our paper shows that for two-objective case and stopping games, the sets are indeed convex polyhedra, which directly leads to a decision algorithm. We believe that our proof technique is of interest on its own. It proceeds by assuming that there is an accumulation point on the Pareto curve, and then establishes that there must be an accumulation point in one of the successor states such that the *slope* of the Pareto curves in the accumulation points are equal. This allows us to obtain a cycle in the graph of the game in which we can "follow" the accumulation points and eventually revisit some of them infinitely many times. By further analysing slopes of points on the Pareto curves that are close to the accumulation point, we show that there are two points on the curve that are sufficiently far from each other yet have the same slope, which contradicts the assumption that they are near an accumulation point.

Our results also yield novel important contributions for non-stochastic games. Although there have recently been several works on non-stochastic games with multiple objectives, they a priori restrict to deterministic strategies, by which the associated problems become fundamentally different. It is easy to show that enabling randomisation of strategies extends the bounds Player 1 can achieve, and indeed, even in other areas of game-theory randomised strategies have been studied for decades: the fundamental theorem of game theory is that every finite game admits a *randomised* Nash equilibrium [15].

Related work. Single-objective problems are well studied for stochastic games. For reachability objectives the games are determined and the problem of existence of an optimal strategy achieving a given value is in NP∩co-NP [10]; same holds for total reward objectives. In the multi-objective setting, [9] gives a value iteration algorithm for the multi-objective total reward problem. Although value iteration converges to the correct result, it does so only in infinite number of steps. It is further shown in [9] that when Player 1 is restricted to only use deterministic strategies, the problem becomes undecidable; the proof relies fundamentally on the strategies being deterministic and it is not clear how it can be extended to randomised strategies. The works of [1, 2] extend the equations of [9] to expected energy objectives, and mainly concern a variant of multi-objective mean-payoff reward, where the objective is a "satisfaction objective" requiring that there is a set of runs of a given probability on which all mean payoff rewards exceed a given bound. [1] only studies existence of finite-memory strategies and the probability bound 1, and [2] in addition studies expectation objectives for multichain games, which is a very restricted class of games in which the expectation and probability-1 satisfaction objectives coincide. Very recently, [5] showed that quantitative satisfaction objective problem is coNP-complete.

In non-stochastic games, multi-objective optimisation has been studied for multiple mean-payoff objectives and energy games [18]. A comprehensive analysis of the complexity of synthesis of optimal strategies has been given in [7], and it has been shown that a variant of the problem is undecidable [17]. The work of [4] studies the complexity of problems related to exact computation of Pareto curves for multiple mean-payoff objectives. In [13], interval objectives are studied for total, mean-payoff and discounted reward payoff functions. The problems for interval objectives are a special kind of multi-objective problems that require the payoff to be within a given interval, as opposed to the standard single-objective setting where the goal is to exceed a given bound. As mentioned earlier, all the above works for non-stochastic games a priori restrict the players to use deterministic strategies, and hence the problems exhibit completely different properties than the problem we study.

Our contribution. We give the following novel decidability results. Firstly, we show that the problem for *determined* stochastic games is decidable. Then, as the main result of the paper, we show that for non-determined games which also satisfy the stopping assumption and for two objectives, the set of achievable bounds forms a convex polyhedron. This immediately leads to an algorithm for computing Pareto curves, and we obtain the following novel results as corollaries.

– Two-objective discounted-reward problem for stochastic games is decidable.
– Two-objective total-reward problem for stochastic stopping games is decidable.

Although we phrase our results in terms of stochastic games, to our best knowledge, the above results also yield novel decidability results for multi-objective *non-stochastic games* when randomisation of strategies is allowed.

Outline of the paper. In Sec. 3, we show a simple algorithm that works for determined games and show how to decide whether a stopping game is determined. In Sec. 4, we give decidability results for two-objective stopping games.

2 Preliminaries on stochastic games

We begin this section by introducing the notation used throughout the paper. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we use $\boldsymbol{v}_i$ to refer to its i-th component, where $1 \leq i \leq n$. The comparison operator $\leq$ on vectors is defined to be the componentwise ordering: $\boldsymbol{u} \leq \boldsymbol{v} \Leftrightarrow \forall i \in [1,n].\ \boldsymbol{u}_i \leq \boldsymbol{v}_i$. We write $\boldsymbol{u} < \boldsymbol{v}$ when $\boldsymbol{u} \leq \boldsymbol{v}$ and $\boldsymbol{u} \neq \boldsymbol{v}$. Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, the *dot product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined by $\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} \boldsymbol{u}_i \cdot \boldsymbol{v}_i$.

The sum of two sets of vectors $U, V \subseteq \mathbb{R}^n$ is defined by $U + V = \{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \in U, \boldsymbol{v} \in V\}$. Given a set $V \in \mathbb{R}^n$, we define the *downward closure* of V as $\mathsf{dwc}(V) \stackrel{\text{def}}{=} \{\boldsymbol{u} \mid \exists \boldsymbol{v} \in V\,.\,\boldsymbol{u} \leq \boldsymbol{v}\}$, and we use $\mathsf{conv}(V)$ for the *convex closure* of V, i.e. the set of all $\boldsymbol{v}$ for which there are $\boldsymbol{v}^1, \ldots \boldsymbol{v}^n \in V$ and $w_1 \ldots w_n \in [0,1]$ such that $\sum_{i=1}^{n} w_i = 1$ and $\boldsymbol{v} = \sum_{i=1}^{n} w_i \cdot \boldsymbol{v}^i$. An *extremal point* of a set $X \subseteq \mathbb{R}^n$ is a vector $\boldsymbol{v} \in X$ that is not a convex combination of other points in X, i.e. $\boldsymbol{v} \notin \mathsf{conv}(X \setminus \{\boldsymbol{v}\})$.

A function $f\colon \mathbb{R} \to \mathbb{R}$ is concave whenever for all $x, y \in \mathbb{R}$ and $t \in [0,1]$ we have $f(t \cdot x + (1-t) \cdot y) \geq t \cdot f(x) + (1-t) \cdot f(y)$. Given $x \in \mathbb{R}$, the *left slope* of f in x is defined by $\mathsf{lslope}(f,x) \stackrel{\text{def}}{=} \lim_{x' \to x^-} \frac{f(x)-f(x')}{x-x'}$. Similarly the *right slope* is defined by $\lim_{x' \to x^+} \frac{f(x)-f(x')}{x-x'}$. Note that if f is concave then both limits are well-defined, because by concavity $\frac{f(x)-f(x')}{x-x'}$ is monotonic in x'; nevertheless, the left and right slope might still not be equal.

A point $\boldsymbol{p} \in \mathbb{R}^2$ is an *accumulation point* of f if $f(\boldsymbol{p}_1) = \boldsymbol{p}_2$ and for all $\varepsilon > 0$, there exists $x \neq \boldsymbol{p}_1$ such that $(x, f(x))$ is an extremal point of f and $|\boldsymbol{p}_1 - x| < \varepsilon$. Moreover, $\boldsymbol{p}$ is a *left (right) accumulation point* if in the above we in addition have $x < \boldsymbol{p}_1$ (resp. $x > \boldsymbol{p}_1$). We sometimes slightly abuse notation by saying that x is an extremal point when $(x, f(x))$ is an extremal point, and similarly for accumulation points.

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, and use $\mathsf{supp}(\mu) = \{s \in S \mid \mu(s) > 0\}$ for the *support set* of $\mu \in \mathcal{D}(S)$.

We now define turn-based stochastic two-player games together with the concepts of strategies and paths of the game. We then present the objectives that are studied in this paper and the associated decision problems.

Stochastic 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 $\Delta(s,t) \in \{0,1\}$ if $s \in S_\Box \cup S_\Diamond$ and $\sum_{t \in S} \Delta(s,t) = 1$ if $s \in S_\bigcirc$.

$S_\Box$ and $S_\Diamond$ represent the sets of states controlled by Player 1 and 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 $\Delta(s) \stackrel{\text{def}}{=} \{t \in S \mid \Delta(s,t){>}0\}$. 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 $\mathsf{Term} \stackrel{\text{def}}{=} \{s \in S \mid \Delta(s){=}\{s\}\}$.

Paths. An *infinite path* λ of a stochastic game $\mathcal{G}$ is a sequence $(s_i)_{i \in \mathbb{N}}$ of states such that $s_{i+1} \in \Delta(s_i)$ for all $i \geq 0$. A *finite path* is a prefix of such a 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-1} of $\lambda = s_0 s_1 \ldots$ and $\lambda_{\leq i}$ for the prefix of λ of length $i+1$. 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 $\Diamond T \stackrel{\text{def}}{=} \{\lambda \in \Omega_\mathcal{G} \mid \exists i \,.\, \lambda_i \in T\}$.

Strategies. We write $\Omega^\Box_\mathcal{G}$ and $\Omega^\Diamond_\mathcal{G}$ for the finite paths that end with a state of $S_\Box$ and $S_\Diamond$, respectively. A *strategy* of Player 1 is a function $\pi : \Omega^\Box_\mathcal{G} \to \mathcal{D}(S)$ such that $s \in \mathsf{supp}(\pi(\lambda))$ only if $\Delta(\mathsf{last}(\lambda), s) = 1$. 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 $\lambda \in \Omega^+_\mathcal{G}$, i.e. $\pi(\lambda)(s) = 1$ for some $s \in S$. A strategy σ for Player 2 is defined similarly replacing $\Omega^\Box_\mathcal{G}$ with $\Omega^\Diamond_\mathcal{G}$. We denote by Π and Σ the sets of all strategies for Player 1 and Player 2, respectively.

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

Total reward. A reward function $\varrho \colon S \to \mathbb{Q}$ assigns a reward to each state of the game. We assume the rewards are 0 in all terminal states. The *total reward* of a path λ is $\varrho(\lambda) \stackrel{\text{def}}{=} \sum_{j \geq 0} \varrho(\lambda_j)$. Given a game $\mathcal{G}$, an initial state s, a vector of n rewards $\boldsymbol{\varrho}$ and a vector of n bounds $\boldsymbol{z} \in \mathbb{R}^n$, we say that a pair of strategies (π, σ) *yields* an objective $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$ if $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{\varrho}_i] \geq \boldsymbol{z}_i$ for all $1 \leq i \leq n$. A strategy $\pi \in \Pi$ *achieves* $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$ if for all σ we have that (π, σ) yields $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$; the vector $\boldsymbol{z}$ is then called *achievable*, and we use $\mathcal{A}_s$ for the set of all achievable vectors. A strategy $\sigma \in \Sigma$ *spoils* $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$ if for no $\pi \in \Pi$, the tuple (π, σ) yields $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$. Note that lower bounds (objectives $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[\boldsymbol{\varrho}_i] \leq \boldsymbol{z}_i$) can be modelled by upper bounds after multiplying all rewards and bounds by -1.

A (lower) Pareto curve in s is the set of all maximal $\boldsymbol{z}$ such that for all $\varepsilon > 0$ there is $\pi \in \Pi$ that achieves the objective $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z} - \varepsilon)$. We use f_s for the Pareto curve, and for the two-objective case we treat it as a function, writing $f_s(x) = y$ when $(x, y) \in f_s$. We say that a game is *determined* if for all states, every bound can be spoiled or lies in the downward closure of the Pareto curve[1]. Note that the downward closure of the Pareto curve equals the closure of $\mathcal{A}_s$.

Discounted reward. Discounted games play an important role in game theory. In these games, the rewards have a discount factor $\delta \in (0, 1)$ meaning that the reward received after j steps is multiplied by δ^j, and so a discounted reward of a path λ is then $\varrho(\lambda, \delta) = \sum_{j \geq 0} \varrho(\lambda_j) \cdot \delta^j$. We define the notions of achieving, spoiling and Pareto curves for discounted reward $\mathsf{disrew}(\boldsymbol{\varrho}, \delta, \boldsymbol{z})$ in the same way as for total reward. Since the problems for discounted reward can easily be encoded using the total reward framework (by adding before each state a stochastic state from which with probability $(1 - \delta)$ we transition to a terminal state), from now on we will concentrate on total reward, unless specified otherwise.

The problems. In this paper we study the following decision problems.

Definition 1 (Total-reward problem). *Given a stochastic game $\mathcal{G}$, an initial state s_0, and vectors of n reward functions $\boldsymbol{\varrho}$ and thresholds $\boldsymbol{z}$, is $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$ achievable from s_0?*

Definition 2 (Discounted-reward problem). *Given a stochastic game $\mathcal{G}$, an initial state s_0, vectors of n reward functions $\boldsymbol{\varrho}$ and thresholds $\boldsymbol{z}$, and a discount factor $\delta \in (0, 1)$, is $\mathsf{disrew}(\boldsymbol{\varrho}, \delta, \boldsymbol{z})$ achievable from s_0?*

In the particular case when n above is 2, we speak about *two-objective* problems.

Simplifying assumption. In order to keep the presentation of the proofs simple, we will assume that each non-terminal state has exactly two successors and that only the states controlled by Player 2 have weights different from 0. Note that any stochastic game can be transformed into an equivalent game with this property in polynomial time, so we do not lose generality by this assumption.

Example 3 (Floor heating problem). As an example illustrating the definitions, as well as possible applications of our results, we consider a simplified version of the smart-house case study presented in [14] with a difference that we model both user comfort and energy consumption. Player 1, representing a controller, decides which rooms are heated, while the Player 2 represents the configuration of the house, for instance which door and windows are open, which cannot be influenced by the controller. The temperature in another room changes based on additional probabilistic factors. We illustrate this example in Fig. 1 and a simple model as a stochastic game is given in Fig. 2 (left). We have to control

[1] The reader might notice that in some works, games are said to be determined when each vector can be either achieved by one player, or spoiled by the other. This is not the case of our definition, where the notion of determinacy is *weaker* and only requires ability to spoil or achieve up to arbitrarily small ε.

the floor heating of two rooms in a house, by opening at most one of the valves V_1 and V_2 at a time.

The state of each room is either cold or hot, for instance in state H, C, the first room is warm while the second one is cold, and the third room has unknown temperature. Weights on the first dimension represent the energy consumption of the system while the second represent the comfort inside the house. Player 2 controls whether the door D between the second room and a third one is open or not. The temperature T in the other room of the house is controlled by stochastic transitions. For instance in the initial state (C, C), the controller can choose either to switch on the heating in room 1 or room 2. Then the second player chooses whether the door is opened or not and stochastic states determine the contribution of the other rooms: for instance from (H, C) if the second player chooses that the door is opened then depending on whether the temperature of the other room is low or high, room 2 can either stay cold or get heated through the door, and the next state in that case is (H, H) which is the terminal state. The objective is to optimise energy consumption and comfort until both rooms are warm. The Pareto curve for a few states of the game is depicted in Fig. 2 (right).

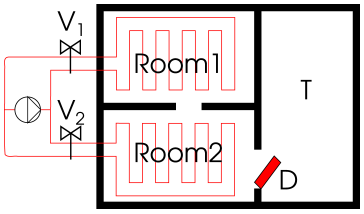

Fig. 1: A house with controllable floor heating in two rooms.

2.1 Equations for lower value

We recall the results of [9, 1] showing that for stopping games the sets of achievable points $\mathcal{A}_s$ are the unique solution to the sets of equations defined as follows:

$$X_s = \begin{cases} \mathsf{dwc}(\{(0,\dots,0)\}) & \text{if } s \in \mathsf{Term} \\ \mathsf{dwc}(\mathsf{conv}(\bigcup_{t\in\Delta(s)} X_t)) & \text{if } s \in S_\Box \\ \boldsymbol{\varrho}(s) + \mathsf{dwc}(\bigcap_{t\in\Delta(s)} X_t) & \text{if } s \in S_\Diamond \\ \mathsf{dwc}(\sum_{t\in\Delta(s)} \Delta(s,t)\cdot X_t) & \text{if } s \in S_\bigcirc \end{cases}$$

The equations can be used to design a value-iteration algorithm that iteratively computes sets X_s^i for increasing i: As a base step we have $X_s^0 = \mathsf{dwc}(\mathbf{0})$ (where $\mathbf{0} = (0,\dots,0)$); we then substitute X_s^i for X_s on the right-hand side of the equations, and obtain X_s^{i+1} as X_s on the left-hand side. The sets X_s^i so obtained converge to the least fixpoint of the equations above [9, 1]. As we will show later, the sets X_s^i might be getting increasingly complex even though the actual solution X_s only comprises two extremal points.

3 Determined games

In this section we present a simple algorithm which works under the assumption that the game is determined. For stopping games, we then give a procedure to decide whether a game is determined.

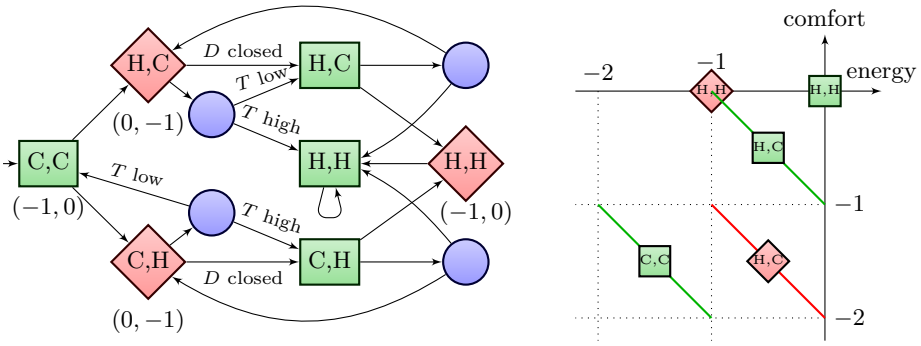

Fig. 2: A stochastic two-player game modelling the floor heating problem. Vectors under states denote a reward function when it is not $(0, 0)$. All probabilistic transitions have probability $\frac{1}{2}$. Pareto curves of a few states of the game are depicted on the right.

Theorem 3. *There is an algorithm working in exponential time, which given a determined stochastic two-player game, computes its Pareto-curve.*

For the proof of the theorem we will make use of the following:

Theorem 4 ([9, Thm. 7]). *Suppose Player 2 has a strategy σ such that for all π of Player 1 there is at least one $1 \leq i \leq n$ with $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}(\boldsymbol{\varrho}_i) < \boldsymbol{z}_i$. Then Player 2 has a memoryless deterministic strategy with the same properties.*

From the above theorem we obtain the following lemma.

Lemma 5. *The following two statements are equivalent for determined games:*

- *A given point $\boldsymbol{z}$ lies in the downward closure of the Pareto curve for s.*
- *For all memoryless deterministic strategies σ of Player 2, there is a strategy π of Player 1 such that (π, σ) yield $\mathsf{totrew}(\boldsymbol{\varrho}, \boldsymbol{z})$.*

Thus, to compute the Pareto curve for a determined game $\mathcal{G}$, it is sufficient to consider all memoryless deterministic strategies $\sigma_1, \sigma_2, \ldots, \sigma_m$ of Player 2 and use [11] to compute the Pareto curves $f^{\sigma_i}_s$ for the games $\mathcal{G}^{\sigma_i}$ induced by $\mathcal{G}$ and σ_i (i.e. $\mathcal{G}^{\sigma_i}$ is obtained from $\mathcal{G}$ by turning all $s \in S_\Diamond$ to stochastic vertices and stipulating $\Delta(s, t) = \sigma_i(s)$ for all successors t of s; in turn, $\mathcal{G}^{\sigma_i}$ is a Markov decision process), and obtain the Pareto curve for $\mathcal{G}$ as the pointwise minimum $V_s := \min_{1 \leq i \leq m} f^{\sigma_i}_s$.

To decide if a stopping game is determined, it is sufficient to take the downward closures of solutions V_s and check if they satisfy the equations from Sec. 2.1. Since in stopping games the solution of the equations is unique, if the sets are a solution they are also the Pareto curves and the game is determined. If any of the equations are not satisfied, then V_s are not the Pareto curves and the game is not determined. Note that for non-stopping games the above approach does not work: even if the sets do not change by applying one step of value iteration, it is still possible that the solution is not the least fixpoint, and so we cannot infer any conclusion.

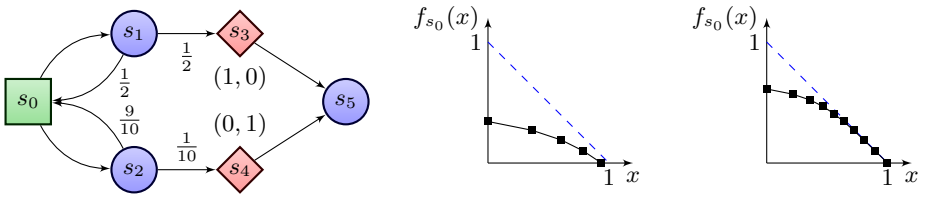

Fig. 3: An example showing that value iteration might produce Pareto curves with unboundedly many extremal points.

4 Games with two objectives

We start this section by showing that the existing value iteration algorithm presented in Sec. 2.1 might iteratively compute sets X_s^i with increasing number of extremal points, although the actual resulting set X_s (and the associated Pareto curve f_s) is very simple. Consider the game from Fig. 3 (left). Applying the value-iteration algorithm given by the equations from Sec. 2.1 for n steps gives a Pareto curve in s_0 with $n-1$ extremal points. Each extremal point corresponds to a strategy π_i that in s_0 chooses to go to s_2 when the number of visits of s_0 is less than i, and after that chooses to go to s_1. The upper bounds of the sets X_s^n for $n = 5$ and $n = 10$ are drawn in Fig. 3 (centre and right, respectively) using solid line, and their extremal points are marked with dots. The Pareto curve f_s is drawn with dashed blue line, and it consists of two extremal points, $(0, 1)$ and $(1, 0)$.

We now proceed with the main result of this section, the decidability of the two-objective strategy synthesis problem for stopping games. The result can be obtained from the following theorem.

Theorem 6. *If $\mathcal{G}$ is a stopping stochastic two-player game with two objectives, and s a state of $\mathcal{G}$ then the Pareto curve f_s has only finitely many extremal points.*

The above theorem can be used to design the following algorithm. For a fixed number k, we create a formula φ_k over $(\mathbb{R}, +, \cdot, \leq)$ which is true if and only if for each $s \in S$ there are points $\boldsymbol{p}^{s,1}, \ldots, \boldsymbol{p}^{s,k}$ such that the sets $V_s \stackrel{\text{def}}{=} \mathsf{dwc}(\mathsf{conv}(\{\boldsymbol{p}^{s,1}, \ldots \boldsymbol{p}^{s,k}\}))$ satisfy the equations from Sec. 2.1. Using [16] we can then successively check validity of φ_k for increasing k, and Thm. 6 guarantees that we will eventually get a formula which is valid, and it immediately gives us the Pareto curve. We get the following result as a corollary.

Corollary 7. *Two-objective total reward problem is decidable for stopping stochastic games, and two-objective discounted-reward problem is decidable for stochastic games.*

Outline of the proof of Thm. 6. The proof of Thm. 6 proceeds by assuming that there are infinitely many extremal points on the Pareto curve, and

then deriving a contradiction. Firstly, because the game is stopping, an upper bound on the expected total reward that can be obtained with respect to a single total reward objective is $M := \sum_{i=0}^{\infty}(1 - p_{min}^{|S|})^i \cdot \varrho_{max}^{|S|}$ where $p_{min} = \min\{\Delta(s,s') \mid \Delta(s,s') > 0\}$ is the smallest transition probability, and $\varrho_{max} = \max_{i\in\{1,2\}} \max_{s\in S} \boldsymbol{\varrho}_i(s)$ is the maximal reward assigned to a state. Thus, the Pareto curve is contained in a compact set, and this implies that there is an accumulation point on it. In Sec. 4.1, we show that we can follow one accumulation point $\boldsymbol{p}$ from one state to one of its successors, while preserving the same left slope. Moreover, in the neighbourhood of the accumulation point the rate at which the right slope decreases is quite similar to the decrease in the successors, in a way that is made precise in Lem. 9, 10, and 11. This is with the exception of some stochastic states for which the decrease strictly slows down when going to the successors: we will exploit this fact to get a contradiction. We construct a transition system $T_{s_0,\boldsymbol{p}}$, which keeps all the paths obtained by following the accumulation point $\boldsymbol{p}$ from s_0. We show that if $\mathcal{G}$ is a stopping game, then we can obtain a path in $T_{s_0,\boldsymbol{p}}$ which visits stochastic states for which the decrease of the right slope strictly slows down. This relies on results for *inverse betting games*, which are presented in Sec. 4.2. Since this decrease can be repeated and there are only finitely many reachable states in $T_{s_0,\boldsymbol{p}}$, we show in Sec. 4.3 that the decrease of the right slope must be zero somewhere, meaning that the curve is constant in the neighbourhood of an accumulation point, which is a contradiction.

We will rely on the properties of the equations from Sec. 2.1 and the left and right slopes of the Pareto curve. Note that we introduced the notion of slope only for two-dimensional sets, and so our proofs only work for two dimensions. Generalisations of the concept of slopes exist for higher dimensions, but simple generalisation of our lemmas would not be valid, as we will show later. Hence, in the remainder of this section, we focus on the two-objective case. For the simplicity of presentation, we will present all claims and proofs for *left* accumulation points. The case of right accumulation points is analogous.

4.1 Mapping accumulation points to successor states

We start by enumerating some basic but useful properties of the Pareto curve and its slopes. First notice that it is a continuous concave function and we can prove the following:

Lemma 8. *Let f be a continuous concave function defined on $[a,b]$.*

1. *If $a < x < x' \leq b$ are two reals for which $\mathsf{lslope}(f)$ is defined, then $\mathsf{lslope}(f,x) \geq \mathsf{rslope}(f,x) \geq \mathsf{lslope}(f,x')$.*
2. *If (x,x') contains an extremal point of f then $\mathsf{lslope}(f_s,x) \neq \mathsf{lslope}(f_s,x')$.*
3. *If $x \in (a,b]$, then $\lim_{x'\to x^-} \mathsf{lslope}(f,x') = \lim_{x'\to x^-} \mathsf{rslope}(f,x') = \mathsf{lslope}(f,x)$.*

To prove Thm. 6, we will use the equations from Sec. 2.1 to describe how accumulation points on a Pareto curve for s "map" to accumulation points on successors.

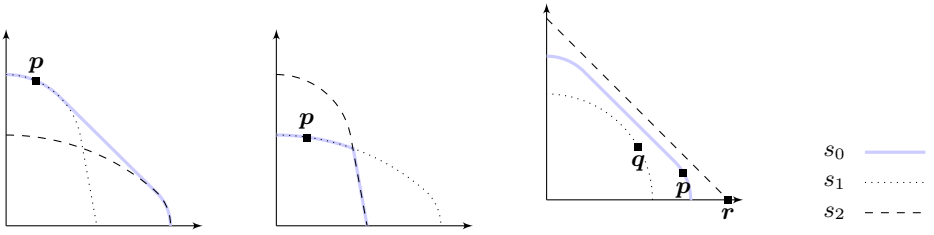

Fig. 4: An example of Pareto curve in a state s_0 with two successors s_1 and s_2, for the case of $s_0 \in S_\Box$ (left), $s_0 \in S_\Diamond$ (centre), and $s_0 \in S_\bigcirc$ with uniform probabilities on transitions (right). In each case, the curve has infinitely many accumulation points.

Lemma 9. *Let s_0 be a Player 1 state with two successors s_1 and s_2, and let $\boldsymbol{p}$ be a left accumulation point of f_{s_0}. Then there is $\eta(s_0, \boldsymbol{p}) > 0$ such that for all $\varepsilon \in (0,\ \eta(s_0, \boldsymbol{p})\ )$, there is $s' \in \{s_1, s_2\}$ such that: 1. $\boldsymbol{p}$ is a left accumulation point in $f_{s'}$; 2. $\mathsf{lslope}(f_{s_0}, \boldsymbol{p}_1) = \mathsf{lslope}(f_{s'}, \boldsymbol{p}_1)$; 3. $f_{s_0}(\boldsymbol{p}_1 - \varepsilon) \geq f_{s'}(\boldsymbol{p}_1 - \varepsilon)$ and $\mathsf{rslope}(f_{s_0}, \boldsymbol{p}_1 - \varepsilon) \geq \mathsf{rslope}(f_{s'}, \boldsymbol{p}_1 - \varepsilon)$.*

Proof (Sketch). The point 1. follows from the fact that every extremal point in the Pareto curve for s_0 must be an extremal point in one of the successors. This is illustrated in Fig. 4 (left): $\boldsymbol{p}$ which is an extremal point for s_0 is also an extremal point for s_1. The point 2. follows because from a sequence of extremal points $(\boldsymbol{p}^i)_{i \geq 0}$ on the Pareto curve of s_0 that converge to $\boldsymbol{p}$, we can select a subsequence that gives extremal points on s' that converge to the left accumulation point $\boldsymbol{p}$ on s'. Finally, to prove 3. we use the fact that the right slope of f_{s_0} is always between those of f_{s_1} and of f_{s_2}. □

Lemma 10. *Let s_0 be a Player 2 state with two successors s_1 and s_2, and let $\boldsymbol{p}$ be a left accumulation point of f_{s_0}. There is $\eta(s_0, \boldsymbol{p}) > 0$ such that for all $\varepsilon \in (0,\ \eta(s_0,\ \boldsymbol{p})\ )$, there is $s' \in \{s_1, s_2\}$, such that: 1. $\boldsymbol{p} - \boldsymbol{\varrho}(s_0)$ is a left accumulation point in $f_{s'}$; 2. $\mathsf{lslope}(s_0, \boldsymbol{p}_1) = \mathsf{lslope}(s', \boldsymbol{p}_1 - \boldsymbol{\varrho}_1(s_0))$; 3. $f_{s_0}(\boldsymbol{p}_1 - \varepsilon) = f_{s'}(\boldsymbol{p}_1 - \varepsilon - \boldsymbol{\varrho}_1(s_0))$ and $\mathsf{rslope}(f_{s_0}, \boldsymbol{p}_1 - \varepsilon) = \mathsf{rslope}(f_{s'}, \boldsymbol{p}_1 - \varepsilon - \boldsymbol{\varrho}_1(s_0))$.*

Proof (Sketch). A crucial observation here is that $f_{s_0}(\boldsymbol{p}_1^i)$ is either $\boldsymbol{\varrho}_2(s_0) + f_{s_1}(\boldsymbol{p}_1^i - \boldsymbol{\varrho}_1(s_0))$ or $\boldsymbol{\varrho}_2(s_0) + f_{s_2}(\boldsymbol{p}_1^i - \boldsymbol{\varrho}_1(s_0))$. This is illustrated in Fig. 4 (center): $f_{s_0}(\boldsymbol{p}_1) = \boldsymbol{\varrho}_2(s_0) + f_{s_1}(\boldsymbol{p}_1 - \boldsymbol{\varrho}_1(s_0))$ (there $\boldsymbol{\varrho}(s_0) = (0,0)$). Hence when we take a sequence $(\boldsymbol{p}_1^i)_{i \in \mathbb{N}}$, for some $\ell \in \{1,2\}$ the value $f_{s_0}(\boldsymbol{p}_1^i)$ equals $\boldsymbol{\varrho}_2(s_0) + f_{s_\ell}(\boldsymbol{p}_1^i - \boldsymbol{\varrho}_1(s_0))$ infinitely many times. From this we get a converging sequence of points in s_ℓ, and obtain that the left slopes are equal in s_0 and s_ℓ. By further arguing that in any left neighbourhood of $\boldsymbol{p}_1^i - \boldsymbol{\varrho}_1(s_0)$ we can find infinitely many points with different left slopes, we obtain that there are also infinitely many extremal points in the neighbourhood and hence $\boldsymbol{p}_1^i - \boldsymbol{\varrho}_1(s_0)$ is a left accumulation point.

As for the last item, the important observation here is that if at some point $\boldsymbol{p}'$, f_{s_1} is strictly below f_{s_2} then the right slope of f_{s_0} corresponds to that of f_{s_1}, and if f_{s_1} equals f_{s_2} then the right slope of f_{s_0} corresponds to the minimum of the right slopes of f_{s_1} and f_{s_2} (it is also interesting to note that the left slope corresponds to the maximum of the two). □

Lemma 11. *Let s_0 be a stochastic state with two successors s_1 and s_2, and $\boldsymbol{p}$ a left accumulation point of f_{s_0}. There are points $\boldsymbol{q}$ and $\boldsymbol{r}$ on f_{s_1} and f_{s_2} respectively such that $\boldsymbol{p} = \Delta(s_0, s_1) \cdot \boldsymbol{q} + \Delta(s_0, s_2) \cdot \boldsymbol{r}$. Moreover:*

1. *there is $(s', \boldsymbol{t}) \in \{(s_1, \boldsymbol{q}), (s_2, \boldsymbol{r})\}$ such that $\boldsymbol{t}$ is a left accumulation point of $f_{s'}$ and $\mathsf{lslope}(f_{s_0}, \boldsymbol{p}_1) = \mathsf{lslope}(f_{s'}, \boldsymbol{t}_1)$;*
2. *there is $\eta(s_0, \boldsymbol{p}) > 0$ such that for all $\varepsilon \in (0,\ \eta(s_0, \boldsymbol{p})\ )$:*
   - *there are $\varepsilon_1 \geq 0, \varepsilon_2 \geq 0$ such that $\mathsf{rslope}(f_{s_0}, \boldsymbol{p}_1 - \varepsilon) \geq \mathsf{rslope}(f_{s_1}, \boldsymbol{q}_1 - \varepsilon_1)$, $\mathsf{rslope}(f_{s_0}, \boldsymbol{p}_1 - \varepsilon) \geq \mathsf{rslope}(f_{s_2}, \boldsymbol{r}_1 - \varepsilon_2)$, and $\varepsilon = \Delta(s_0, s_1) \cdot \varepsilon_1 + \Delta(s_0, s_2) \cdot \varepsilon_2$;*
   - *if $\boldsymbol{r}$ is not a left accumulation point in f_{s_2}, or $\mathsf{lslope}(f_{s_0}, \boldsymbol{p}_1) \neq \mathsf{lslope}(f_{s_2}, \boldsymbol{r}_1)$, then $f_{s_0}(\boldsymbol{p}_1 - \varepsilon) = \Delta(s_0, s_1) \cdot f_{s_1}\left(\frac{\boldsymbol{p}_1 - \varepsilon - \Delta(s_0, s_2) \cdot \boldsymbol{r}_1}{\Delta(s_0, s_1)}\right) + \Delta(s_0, s_2) \cdot \boldsymbol{r}_2$;*
   - *if $\boldsymbol{q}$ is not a left accumulation point in f_{s_1}, or $\mathsf{lslope}(f_{s_0}, \boldsymbol{p}_1) \neq \mathsf{lslope}(f_{s_1}, \boldsymbol{q}_1)$, then $f_{s_0}(\boldsymbol{p}_1 - \varepsilon) = \Delta(s_0, s_1) \cdot \boldsymbol{q}_2 + \Delta(s_0, s_2) \cdot f_{s_1}\left(\frac{\boldsymbol{p}_1 - \varepsilon - \Delta(s_0, s_1) \cdot \boldsymbol{q}_1}{\Delta(s_0, s_2)}\right)$.*

Proof (Sketch). We use the fact that for every extremal point $\boldsymbol{p}'$ there are unique extremal points $\boldsymbol{q}'$ and $\boldsymbol{r}'$ on f_{s_1} and f_{s_2}, respectively, such that $\boldsymbol{p}' = \Delta(s_0, s_1) \cdot \boldsymbol{q}' + \Delta(s_0, s_2) \cdot \boldsymbol{r}'$.

To prove item 1, we show that for all extremal points $\boldsymbol{p}'$, $\mathsf{lslope}(s_0, \boldsymbol{p}') = \min(\mathsf{lslope}(s_1, \boldsymbol{q}'), \mathsf{lslope}(s_2, \boldsymbol{r}'))$, which can be surprising at first glance since one could have expected a weighted sum of the left slopes. This fact is illustrated in Fig. 4 (right): $\mathsf{lslope}(s_0, \boldsymbol{p}') = \mathsf{lslope}(s_1, \boldsymbol{q}') \leq \mathsf{lslope}(s_2, \boldsymbol{r}')$. The inequality $\mathsf{lslope}(s_0, \boldsymbol{p}) \leq \mathsf{lslope}(s_1, \boldsymbol{q})$ (and similarly $\mathsf{lslope}(s_0, \boldsymbol{p}) \leq \mathsf{lslope}(s_2, \boldsymbol{r})$), follows from concavity of f_{s_0}: because for all $\varepsilon > 0$ the inequality $f_{s_0}(\boldsymbol{p}_1 - \varepsilon) \geq \Delta(s_0, s_1) \cdot f_{s_1}(\boldsymbol{q}_1 - \frac{\varepsilon}{\Delta(s_0, s_1)}) + \Delta(s_0, s_2) \cdot f_{s_2}(\boldsymbol{r}_1)$ holds true, from which we obtain $\lim_{\varepsilon \to 0^+} \frac{f_{s_0}(\boldsymbol{p}_1) - f_{s_0}(\boldsymbol{p}_1 - \varepsilon)}{\varepsilon} \leq \lim_{\varepsilon \to 0^+} \frac{f_{s_1}(\boldsymbol{q}_1) - f_{s_1}(\boldsymbol{q}_1 - \varepsilon)}{\varepsilon}$. Showing that the left slope is at least the minimum of the successors' slopes is significantly more demanding and technical, and we give the proof in the long version of this paper [3].

Proving the second point, is based on the observation that a point on the Pareto curve f_{s_0} is a combination of points of f_{s_1} and f_{s_2} that share a common tangent: in other words they maximize the dot product with a specific vector on their respective curves. From this observation it is possible to link the right slopes of these curves. The last two points hold because with the assumption, extremal points that converge to $\boldsymbol{p}$ from the left can be obtained as a combination from a fixed $\boldsymbol{r}$ and points on f_{s_2}. □

Now we will prove that there are no left accumulation points on the Pareto curve. To do that, we will try to follow one in the game: if there is a left accumulation point in one state then at least one of its successors also has one, as the above lemmas show. By using the fact that the left slopes of left accumulation points are preserved we show that the number of reachable combinations $(s, \boldsymbol{p})$, where $s \in S$ and $\boldsymbol{p}$ is a left accumulation point, is finite. We then look at points slightly to the left of the accumulation points, their distance to the accumulation point and right slopes are also mostly preserved except in stochastic states, where if only one successor has a left accumulation point, the decrease of the right slope accelerates (by Lem. 11.2). By using the fact that in stopping games

we can ensure visiting such stochastic states, we will show that for some states the right slope is constant on the left neighbourhood of the left accumulation point, which is a contradiction.

Assume we are given a state s_0 and a left accumulation point $\boldsymbol{p}^0$ of f_{s_0}. We construct a transition system $T_{s_0,\boldsymbol{p}^0}$ where the initial state is $(s_0, \boldsymbol{p}^0)$, and the successors of a given configuration $(s, \boldsymbol{p})$ are the states $(s', \boldsymbol{p}')$ such that s' is a successor of s, and $\boldsymbol{p}'$ is a left accumulation point of s with the same left slope on $f_{s'}$ as $\boldsymbol{p}$ on f_s. Lem. 9, 10, and 11, ensure that all the reachable states have at least one successor.

Lemma 12. *For all reachable states $(s, \boldsymbol{p})$ and $(s', \boldsymbol{p}')$ in the transition system $T_{s_0,\boldsymbol{p}^0}$, if $s = s'$, then $\boldsymbol{p} = \boldsymbol{p}'$.*

Proof. Assume $s = s'$. By construction of $T_{s_0,\boldsymbol{p}^0}$, the left slope in s of $\boldsymbol{p}$ and $\boldsymbol{p}'$ is the same: $\mathsf{lslope}(s, \boldsymbol{p}_1) = \mathsf{lslope}(s_0, \boldsymbol{p}^0_1) = \mathsf{lslope}(s_2, \boldsymbol{p}'_1)$. Assume towards a contradiction that $\boldsymbol{p} < \boldsymbol{p}'$; the proof would work the same for $\boldsymbol{p}' < \boldsymbol{p}$. Since $\boldsymbol{p}'$ is a left accumulation point, there is an extremal point in $(\boldsymbol{p}_1, \boldsymbol{p}'_1)$. Lem. 8.2 tells us that $\mathsf{lslope}(s_1, \boldsymbol{p}_1) \neq \mathsf{lslope}(s_2, \boldsymbol{p}'_1)$ which is a contradiction. Hence $\boldsymbol{p} = \boldsymbol{p}'$. □

As a corollary of this lemma, the number of states that are reachable in $T_{s_0,\boldsymbol{p}^0}$ is finite and bounded by $|S|$.

4.2 Inverse betting game

To show a contradiction, we will follow a path with left accumulation points. We want this path to visit stochastic states which have only one successor in $T_{s_0,\boldsymbol{p}^0}$. For that, we will prove a property of an intermediary game that we call an inverse betting game.

An *inverse betting game* is a two player game, given by $\langle V_\exists, V_\forall, E, (v_0, c_0), w\rangle$ where $V_\exists$ and $V_\forall$ are the set of vertices controlled by Eve and Adam, respectively, $\langle V_\exists \cup V_\forall, E\rangle$ is a graph whose each vertex has two successors, $(v_0, c_0) \in V \times \mathbb{R}$ is the initial configuration, and $w\colon E \to \mathbb{R}$ is a weight function such that for all $v \in V$: $\sum_{v'|(v,v')\in E} w(v, v') = 1$.

A configuration of the game is a pair $(v, c) \in V \times \mathbb{R}$ where v is a vertex and c a credit. The game starts in configuration (v_0, c_0) and is played by two players Eve and Adam. At each step, from a configuration (v, c) controlled by Eve, Adam suggests a valuation $d\colon E \to \mathbb{R}$ for the outgoing edges of v such that $\sum_{v'|(v,v')\in E} w(v, v') \cdot d(v, v') = c$. Eve then chooses a successor v' such that $(v, v') \in E$ and the game continues from configuration $(v', d(v, v'))$. From a configuration (v, c) controlled by Adam, Adam chooses a successor v' of v and keeps the same credit, hence the game continues from (v', c).

Intuitively, Adam has some credit, and at each step he has to distribute it by betting over the possible successors. Then Eve chooses the successor and Adam gets a credit equal to its bet divided by the probability of this transition. The game is *inverse* because Eve is trying to maximize the credit of Adam.

Theorem 13. *Let $\langle V_\exists, V_\forall, E, (v_0, c_0), w\rangle$ be an inverse betting game. Let $T \subseteq V_\exists \cup V_\forall$ be a target set and $B \in \mathbb{R}$ a bound. If from every vertex $v \in V$, Eve has a strategy to ensure visiting T then she has one to ensure visiting it with a credit $c \geq 1$ or to exceed the bound, that is, she can force a configuration in $(T \times [c_0, +\infty)) \cup (V \times [B, +\infty))$.*

Our next step is transforming the transition system $T_{s_0,\boldsymbol{p}^0}$ into such a game. Consider the inverse betting game $\mathcal{B}$ on the structure given by $T_{s_0,\boldsymbol{p}^0}$ where $V_\exists = S_\bigcirc$ are the states controlled by Eve, $V_\forall = S_\Box \cup S_\Diamond$ are controlled by Adam, $w((s,\boldsymbol{p}),(s',\boldsymbol{p}')) = \Delta(s,s')$ is a weight on edges and the initial configuration is $((s_0,\boldsymbol{p}^0),\varepsilon_0)$. Let $U_{s_0,\boldsymbol{p}^0}$ the set of terminal states and of stochastic states that have only one successor in $T_{s_0,\boldsymbol{p}^0}$. We show that in the inverse betting game obtained from a stopping game $\mathcal{G}$, Eve can ensure visiting $U_{s_0,\boldsymbol{p}^0}$.

Lemma 14. *If $\mathcal{G}$ is stopping, there is a strategy for Eve in $\mathcal{B}$ such that from every vertex $v \in V$, all outcomes visit $U_{s_0,\boldsymbol{p}^0}$.*

Proof. Assume towards a contradiction that this is not the case, then by memoryless determinacy of turn-based reachability games (see e.g. [12]) there is a vertex v and a memoryless deterministic strategy σ_{Adam} of Adam, such that no outcomes of σ_{Adam} from v visit $U_{s_0,\boldsymbol{p}^0}$. Let π and σ be the strategies of Player 1 and Player 2 respectively corresponding to σ_{Adam}. Formally, if $h \in \Omega_\mathcal{G}^\Box$ then $\pi(h) = \sigma_{\mathsf{Adam}}(h)$ and if $h \in \Omega_\mathcal{G}^\Diamond$ then $\sigma(h) = \sigma_{\mathsf{Adam}}(h)$. We prove that all outcomes λ in $\mathcal{G}$ of π, σ from v are outcomes of σ_{Adam} in $\mathcal{B}$. This is by induction on the prefixes $\lambda_{\leq i}$ of the outcomes. It is clear when $\lambda_{\leq i}$ ends with states that are controlled by Player 1 and Player 2 by the way we defined π and σ, that $\lambda_{\leq i+1}$ is also compatible with σ_{Adam} in $\mathcal{B}$. For a finite path $\lambda_{\leq i}$ ending with a stochastic state s in $\mathcal{G}$, two successors are possible. With the induction hypothesis that $\lambda_{\leq i}$ is compatible with σ_{Adam}, and by the assumption on σ_{Adam}, s does not belong to $U_{s_0,\boldsymbol{p}^0}$. Therefore, both successors of s are also in $T_{s_0,\boldsymbol{p}^0}$, and $\lambda_{\leq i+1}$ is compatible with σ_{Adam} in $\mathcal{B}$. This shows that outcomes in $\mathcal{G}$ of (π,σ) are also outcomes of σ_{Adam} in $\mathcal{B}$. Therefore, π and σ ensure that from v, we visit no state of $U_{s_0,\boldsymbol{p}^0}$ and thus no terminal state. This contradicts that the game is stopping. □

Putting Thm. 13 and Lem. 14 together we can conclude the following:

Corollary 15. *If $\mathcal{G}$ is stopping then in $\mathcal{B}$, for any bound B, Eve has a strategy to ensure visiting $U_{s_0,\boldsymbol{p}^0}$ with a credit $c \geq 1$ or making c exceed B.*

4.3 Contradicting sequence

We define $\theta(s_0,\boldsymbol{p}^0) = \min\{\eta(s,\boldsymbol{p}) \mid (s,\boldsymbol{p}) \text{ reachable in } T_{s_0,\boldsymbol{p}^0}\}$, and consider a sequence of points that are $\theta(s_0,\boldsymbol{p}^0)$ close to $\boldsymbol{p}^0$ and with a right slope that is decreasing at least as fast as that of their predecessors.

Lemma 16. *For stopping games, given $s_0 \in S$, $\boldsymbol{p}^0 \in \mathbb{R}^2$, and $\varepsilon_0 > 0$, such that $\varepsilon_0 < \theta(s_0,\boldsymbol{p}^0)$, there is a finite sequence $\pi(s_0,\boldsymbol{p}^0,\varepsilon_0) = (s_i,\boldsymbol{p}^i,\varepsilon_i)_{i\leq j}$ such that:*

– $(s_i, \boldsymbol{p}^i)_{i \leq j}$ *is a path in* $T_{s_0,\boldsymbol{p}^0}$*;*
– *for all* $i \leq j$, $\mathit{rslope}(f_{s_i}, \boldsymbol{p}_1^i - \varepsilon_i) \geq \mathit{rslope}(f_{s_{i+1}}, \boldsymbol{p}_1^{i+1} - \varepsilon_{i+1})$.
– *either* $\varepsilon_j \geq \theta(s_0, \boldsymbol{p}^0)$ *or* $s_j \in U_{s_0,\boldsymbol{p}^0}$ *and* $\varepsilon_j \geq \varepsilon_0$.

The idea of the proof is that in $\mathcal{B}$, thanks to Lem. 9, 10, and 11, Adam can always choose a successor such that $\mathsf{rslope}(f_{s_i}, \boldsymbol{p}_1^i - \varepsilon_i) \geq \mathsf{rslope}(f_{s_{i+1}}, \boldsymbol{p}_1^{i+1} - \varepsilon_{i+1})$. Then thanks to Cor. 15, there is a strategy for Eve to reach $(U_{s_0,\boldsymbol{p}^0} \times [c_0, +\infty)) \cup (V \times [B, +\infty))$. By combining the two strategies, we obtain an outcome that satisfies the desired properties.

We use the path obtained from this lemma to show that no matter how small ε_0 we choose, ε_i can grow to reach $\theta(s_0, \boldsymbol{p}^0)$.

Lemma 17. *For all states* s *with a left accumulation point* $\boldsymbol{p}$ *and for all* $0 < \varepsilon < \theta(s, \boldsymbol{p})$*, there is some* $(s', \boldsymbol{p}')$ *reachable in* $T_{s,\boldsymbol{p}}$ *such that* $\mathit{rslope}(f_{s'}, \boldsymbol{p}'_1 - \theta(s, \boldsymbol{p})) \leq \mathit{rslope}(f_s, \boldsymbol{p}_1 - \varepsilon)$.

Thanks to this lemma, we can now prove Thm. 6. Assume towards a contradiction that there is a left accumulation point $\boldsymbol{p}$ in the state s. Let $m = \min\{\mathsf{lslope}(f_{s'}, \boldsymbol{p}'_1 - \theta(s, \boldsymbol{p})) \mid (s', \boldsymbol{p}')$ reachable in $T_{s,\boldsymbol{p}}\}$ and $(s', \boldsymbol{p}')$ the configuration of $T_{s,\boldsymbol{p}}$ for which this minimum is reached (it is reached because the number of reachable configurations is finite: this is a corollary of Lem. 12). Because of Lem. 17, $\mathsf{rslope}(f_s, \boldsymbol{p}_1 - \varepsilon)$ is greater than m. By Lem. 8.3, when ε goes towards 0, $\mathsf{rslope}(f_s, \boldsymbol{p}_1 - \varepsilon)$ converges to $\mathsf{lslope}(f_s, \boldsymbol{p}_1)$. This means that $\mathsf{lslope}(f_s, \boldsymbol{p}_1) \geq m$. Moreover, by construction of $T_{s,\boldsymbol{p}}$, we also have that $\mathsf{lslope}(f_{s'}, \boldsymbol{p}'_1) = \mathsf{lslope}(f_s, \boldsymbol{p}_1)$, so $\mathsf{lslope}(f_{s'}, \boldsymbol{p}'_1) \geq m$. Because the slopes are decreasing (Lem. 8.1), $m = \mathsf{rslope}(f_{s'}, \boldsymbol{p}'_1 - \theta(s, \boldsymbol{p})) \geq \mathsf{lslope}(f_{s'}, \boldsymbol{p}'_1) \geq m$. Hence, the left and right slopes of $f_{s'}$ are constant on $[\boldsymbol{p}'_1 - \theta(s, \boldsymbol{p}), \boldsymbol{p}'_1]$, and Lem. 8.2 implies that there are no extremal point in $(\boldsymbol{p}'_1 - \theta(s, \boldsymbol{p}), \boldsymbol{p}'_1)$. This contradicts the fact that $\boldsymbol{p}'$ is a left accumulation point: there should be an extremal point in any neighbourhood on the left of $\boldsymbol{p}'$. Hence, f_s contains no accumulation point.

Remark 18. One might attempt to extend the proof of Thm. 6 to three or more objectives, but this does not seem to be easily doable. Although it is possible to use directional derivative (or pick a subgradient) instead of using left and right slope in such setting, an analogue of Lem. 8.2 cannot be proved because in multiple dimensions, two accumulation points can share the same directional derivative, for a fixed direction. It is also not easily possible to avoid this problem by following several directional derivatives instead of just one. This is because the slope in one direction may be inherited from one successor while the slope in another direction comes from another successor. We give more details and example of convex sets that would contradict generalisations of Lem. 8.2 and Lem. 10 in the long version of this paper [3].

5 Conclusions

We have studied stochastic games under multiple objectives, and have provided decidability results for determined games and for stopping games with two objectives. Our results provide an important milestone towards obtaining decidability

for the general case, which is a major task which will require further novel insights into the problem. Another research direction concerns establishing an upper bound on the number of extremal points of a Pareto curve; such result would allow us to give upper complexity bounds for the problem.

Acknowledgements. The authors would like to thank Aistis Šimaitis and Clemens Wiltsche for useful discussions on the topic. The work was supported by EPSRC grant EP/M023656/1. Vojtěch Forejt is also affiliated with Masaryk University, Czech Republic.

References

1. N. Basset, M. Kwiatkowska, U. Topcu, and C. Wiltsche. Strategy synthesis for stochastic games with multiple long-run objectives. In *TACAS*. Springer, 2015.
2. N. Basset, M. Kwiatkowska, and C. Wiltsche. Compositional strategy synthesis for stochastic games with multiple objectives. Technical report, Department of Computer Science, Oxford, UK, 2016.
3. R. Brenguier and V. Forejt. Decidability results for multi-objective stochastic games. *arXiv preprint arXiv:1605.03811*, 2016.
4. R. Brenguier and J. Raskin. Pareto curves of multidimensional mean-payoff games. In *CAV*, 2015.
5. K. Chatterjee and L. Doyen. Perfect-information stochastic games with generalized mean-payoff objectives. In *LICS*, 2016. To appear.
6. K. Chatterjee and T. A. Henzinger. A survey of stochastic ω-regular games. *J. Comput. Syst. Sci.*, 78(2), 2012.
7. K. Chatterjee, M. Randour, and J. Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Inf.*, 51(3-4), 2014.
8. T. Chen, V. Forejt, M. Kwiatkowska, A. Simaitis, A. Trivedi, and M. Ummels. Playing stochastic games precisely. In *CONCUR*, 2012.
9. T. Chen, V. Forejt, M. Kwiatkowska, A. Simaitis, and C. Wiltsche. On stochastic games with multiple objectives. In *MFCS*. Springer, 2013.
10. A. Condon. The complexity of stochastic games. *Inf. Comput.*, 96(2), 1992.
11. K. Etessami, M. Kwiatkowska, M. Vardi, and M. Yannakakis. Multi-objective model checking of Markov decision processes. *LMCS*, 4(4), 2008.
12. E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics, and Infinite Games: A Guide to Current Research*, volume 2500. Springer, 2003.
13. P. Hunter and J. Raskin. Quantitative games with interval objectives. In *FSTTCS*, 2014.
14. K. G. Larsen, M. Mikucionis, M. Muñiz, J. Srba, and J. H. Taankvist. Online and compositional learning of controllers with application to floor heating. In *TACAS*, 2016.
15. J. F. Nash, Jr. Equilibrium points in n-person games. *Proc. National Academy of Sciences of the USA*, 36(1), Jan. 1950.
16. A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, Berkeley, 1951.
17. Y. Velner. Robust multidimensional mean-payoff games are undecidable. In *FoSSaCS*. Springer, 2015.
18. Y. Velner, K. Chatterjee, L. Doyen, T. A. Henzinger, A. M. Rabinovich, and J. Raskin. The complexity of multi-mean-payoff and multi-energy games. *Inf. Comput.*, 241, 2015.