Playing Stochastic Games Precisely

Taolue Chen[1], Vojtěch Forejt[1], Marta Kwiatkowska[1], Aistis Simaitis[1], Ashutosh Trivedi[2], and Michael Ummels[3]

[1] Department of Computer Science, University of Oxford, Oxford, UK
[2] University of Pennsylvania, Philadelphia, USA
[3] Technische Universität Dresden, Germany

Abstract. We study stochastic two-player games where the goal of one player is to achieve *precisely* a given expected value of the objective function, while the goal of the opponent is the opposite. Potential applications for such games include controller synthesis problems where the optimisation objective is to maximise or minimise a given payoff function while respecting a strict upper or lower bound, respectively. We consider a number of objective functions including reachability, ω-regular, discounted reward, and total reward. We show that precise value games are not determined, and compare the memory requirements for winning strategies. For stopping games we establish necessary and sufficient conditions for the existence of a winning strategy of the controller for a large class of functions, as well as provide the constructions of compact strategies for the studied objectives.

1 Introduction

Two-player zero-sum stochastic games [13] naturally model controller synthesis problems [12] for systems exhibiting both the controllable and the uncontrollable nondeterminism coupled with stochastic behaviour. In such games two players—Min (the *controller*) and Max (the *environment*)—move a token along the edges of a graph, called a *stochastic game arena*, whose vertices are partitioned into those controlled by either of the players and the *stochastic* vertices. Player chooses an outgoing edge when the token is in a state controlled by her, while in a stochastic state the outgoing edge is chosen according to a state-dependent probability distribution. Starting from an initial state, choices made by players and at the stochastic vertices characterise a run in the game. Edge-selection choices of players are often specified by means of a *strategy*, which is a partial function from the set of finite runs to probability distributions over enabled edges. Fixing an initial state and strategies for both players determines a *probability space* on the runs of the stochastic game. In classical stochastic games players Min and Max are viewed as *optimisers* as their goals are to minimise and maximise, respectively, the expectation of a given real-valued function of the run called the *payoff function*. Payoff functions are often specified by annotating the vertices with rewards, and include total reward, discounted reward, average reward [8], and more recently ω-regular objectives [3].

In this paper we take a different stand from the well-established notion of viewing players as optimisers which, even though useful in many applications, is inadequate for the problems requiring precision. Among others, such precision requirements may stem from: a) controller design under strict regulatory or safety conditions, or b) optimal controller design minimising or maximising some payoff function while requiring that a given lower or upper bound is respected. For instance, consider the task of designing a gambling machine to maximise profit to the "house" while ensuring the minimum expected *payback* to the customers established by a law or a regulatory body [14,2]. Given that such a task can be cast as a controller synthesis problem using stochastic games, the objective of the controller is to ensure that the machine achieves the expected payback *exactly* equal to the limit set by the regulatory body—higher paybacks will result in a substantial decrease in profits, while lower paybacks will make the design illegal. There are examples from other domains, e.g., ensuring precise 'coin flipping' in a *security protocol* (e.g., Crowds), keeping the expected voltage constant in *energy grid*, etc.

In order to assist in designing the above-mentioned controllers, we consider the problem of achieving a *precise* payoff value in a stochastic game. More specifically, we study games played over a stochastic game arena between two players, Preciser and Spoiler, where the goal (the winning objective) of the Preciser is to ensure that the expected payoff is *precisely* a given payoff value, while the objective of the Spoiler is the contrary, i.e., to ensure that the expected value is anything but the given value. We say that the Preciser wins from a given state if he has a winning strategy, i.e., if he has a strategy such that, for all strategies of Spoiler, the expected payoff for the given objective function is *precisely* a given value x. Similarly, the Spoiler wins from a given state if she has a strategy such that, for all strategies of Preciser, the payoff for the given objective function is not equal to x. The winning region of a player is the set of vertices from which that player wins. Observe that the winning regions of Preciser and Spoiler are disjoint. We say that a game is *determined* if winning regions of the players form a partition of the states set of the arena. Our first result (Section 3.1) is that stochastic games with precise winning objectives are *not* determined even for reachability problems. Given the non-determinacy of the stochastic precise value games, we study the following two dual problems. For a fixed stochastic game arena $\mathcal{G}$, an objective function f, and a target value x,

- the *synthesis problem* is to decide whether there exists a strategy π of Preciser such that, for all strategies σ of Spoiler, the expected value of the payoff is equal to x, and to construct such a strategy if it exists;
- the *counter-strategy problem* is to decide whether, for a given strategy σ of Spoiler, there exists a counter-strategy π of Preciser such that the expected value of the payoff is equal to x[4].

[4] We do not consider the construction of π here. Note that the problem of constructing a counter-strategy is not well defined, because the strategy σ can be an arbitrary (even non-recursive) function.

Consider the case when Spoiler does not control any states, i.e., when the stochastic game arena is a *Markov decision process* [11]. In this case, both the synthesis and the counter-strategy problems overlap and they can be solved using optimisation problems for the corresponding objective function. Assuming that, for some objective function, Preciser achieves the value h while maximising, and value l while minimising, then any value $x \in [l, h]$ is precisely achievable by picking minimising and maximising strategies with probability θ and $(1 - \theta)$ respectively, where $\theta = \frac{h-x}{h-l}$ if $l \neq h$ and $\theta = 1$ if $l = h$. Notice that such a strategy will require just one bit of memory for all the objectives for which there exist memoryless strategies for the corresponding optimisation problems in a Markov decision process, including a large class of objective functions [11], such as expected reachability reward, discounted reward, and total reward objectives.

It seems natural to conjecture that a similar approach can be used for the game setting, i.e., Preciser can achieve any value between his minimising and maximising strategies by picking one of the strategies with an appropriate probability. Unfortunately, the same intuition does *not* carry over to stochastic games because, once Preciser fixes his strategy, Spoiler can choose any of her sub-optimal (i.e., not optimising) counter-strategies to ensure a payoff different from the target value. Intuitively, the strategy of Preciser may need to be responsive to Spoiler actions and, therefore, it should require memory.

Strategies are expressed as *strategy automata* [6,1] that consist of—i) a set of *memory elements*, ii) a *memory update function* that specifies how memory is updated as the transitions occur in the game arena, and iii) a *next move function* that specifies a distribution over the successors of game state, depending on the memory element. Memory update functions in strategy automata can be either deterministic or stochastic [1]. We show that the choice of how the memory is updated drastically influences the size of memory required. In Section 3.2 we show that deterministic update winning strategies require at least exponential memory size in precise value games. Although we are not aware of the exact memory requirement for deterministic memory update strategies, we show in Section 4 that, if *stochastic update* strategies are used, then memory need is linear in the size of the arena for the reachability, ω-regular properties and discounted and total reward objectives. We study precise value problems for these objectives and show necessary and sufficient conditions for the existence of winning strategies for controller synthesis problem in stopping games (Section 4) and counter-strategy problem in general (Section 5).

Contributions. The contributions of the paper can be summarised as follows.

- We show that stochastic games with precise value objectives are not determined even for reachability objectives, and we compare the memory requirements for different types of strategies.
- We solve the *controller synthesis* problem for precise value in stopping games for a large class of functions and provide a construction for compact winning strategies. We illustrate that for non-stopping games the problem is significantly harder to tackle.

– We solve the *counter strategy* as well as discounted reward controller synthesis problem for general games.

The proofs that have been omitted from this paper can be found in [5].

Related work. We are not aware of any other work studying precise value problem for any objective function. There is a wealth of results [8,10,3] studying two-player stochastic games with various objective functions where players optimise their objectives. The precise value problem studied here is a special case of *multi-objective optimisation*, where a player strives to fulfill several (in our case two) objectives at once, each with a certain minimum probability. Multi-objective optimisation has been studied for Markov decision processes with discounted rewards [4], long-run average rewards [1], as well as reachability and ω-regular objectives [7]; however, none of these works consider multi-player optimisation.

2 Preliminaries

We begin with some background on stochastic two-player games.

Stochastic Game Arena. Before we present the definition, we introduce the concept of discrete probability distributions. A *discrete probability 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 discrete 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 a distribution $\mu \in \mathcal{D}(S)$ is a *Dirac distribution* if $\mu(s) = 1$ for some $s \in S$. Sometimes we abuse the notation to identify a Dirac distribution μ with its unique element in $\mathsf{supp}(\mu)$.

We represent a discrete probability distribution $\mu \in \mathcal{D}(S)$ on a set $S = \{s_1, \ldots, s_n\}$ as a map $[s_1 \mapsto \mu(s_1), \ldots, s_n \mapsto \mu(s_n)] \in \mathcal{D}(S)$ and we omit the states outside $\mathsf{supp}(\mu)$ to improve presentation.

Definition 1 (Stochastic Game Arena). *A* stochastic game arena *is a tuple* $\mathcal{G} = \langle S, (S_\Box, S_\Diamond, S_\bigcirc), \Delta \rangle$ *where:*

– *S is a countable set of states partitioned into sets of states $S_\Box$, $S_\Diamond$, 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 $s \in S_\bigcirc$.*

A stochastic game arena is *finite* if S is a finite set. In this paper we omit the keyword "finite" as we mostly work with finite stochastic game arenas and explicitly use "countable" for the arenas for emphasise when they are not finite.

The sets $S_\Box$ and $S_\Diamond$ represent the sets of states controlled by players Preciser and Spoiler, respectively, while the set $S_\bigcirc$ is the set of stochastic states. A game arena is a *Markov decision process* if the set of states controlled by one of the players in an empty set, while it is a *Markov chain* if the sets of states controlled by both players are empty. 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 $s \in S$.

Paths. An infinite *path* λ of a stochastic game arena $\mathcal{G}$ is an infinite sequence $s_0 s_1 \ldots$ 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. Similarly, for $k \leq \mathsf{len}(\lambda)$ we denote the prefix of length k of the path λ by $\mathrm{Pref}(\lambda, k) \stackrel{\text{def}}{=} s_0 s_1 \dots s_{k-1}$. For a finite path $\lambda = s_0 s_1 \dots s_n$ we write $\mathsf{last}(\lambda)$ for the last state of the path, here $\mathsf{last}(\lambda) = s_n$. For a stochastic game arena $\mathcal{G}$ we write ${\Omega_{\mathcal{G}}}^+$ for the set of all finite paths, $\Omega_{\mathcal{G}}$ for the set of all infinite paths, $\Omega_{\mathcal{G},s}$ for the set of infinite paths starting in state s. If the starting state is given as a distribution $\alpha : S \to [0,1]$ then we write $\Omega_{\mathcal{G},\alpha}$ for the set of infinite paths starting from some state in $\mathsf{supp}(\alpha)$.

Strategy. Classically, a *strategy* of Preciser is a partial function $\pi\colon {\Omega_{\mathcal{G}}}^+ \to \mathcal{D}(S)$, 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 $\Delta(\langle \mathsf{last}(\lambda), s\rangle) = 1$. Such a strategy π is *memoryless* if $\mathsf{last}(\lambda) = \mathsf{last}(\lambda')$ implies $\pi(\lambda) = \pi(\lambda')$ for all $\lambda, \lambda' \in {\Omega_{\mathcal{G}}}^+$. If π is a memoryless strategy for Preciser then we identify it with a mapping $\pi\colon S_\Box \to \mathcal{D}(S)$. Similar concepts for a strategy σ of the Spoiler are defined analogously. In this paper we use an alternative formulation of strategy [1] that generalises the concept of strategy automata [6].

Definition 2. *A* strategy *of Preciser in a game arena* $\mathcal{G} = \langle S, (S_\Box, S_\Diamond, S_\bigcirc), \Delta\rangle$ *is a tuple* $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$*, where:*

- $\mathcal{M}$ *is a countable set of* memory elements*.*
- $\pi_u\colon \mathcal{M} \times S \to \mathcal{D}(\mathcal{M})$ *is a* memory update function*,*
- $\pi_n\colon S_\Box \times \mathcal{M} \to \mathcal{D}(S)$ *is a* next move function *such that* $\pi_n(s,m)[s'] = 0$ *for all* $s' \in S \setminus \Delta(s)$*,*
- $\alpha\colon S \to \mathcal{D}(\mathcal{M})$ *defines an* initial distribution *on the memory elements for a given initial state of* $\mathcal{G}$*.*

A strategy σ *for Spoiler is defined in an analogous manner. We denote the set of all strategies for Preciser and Spoiler by* Π *and* Σ*, respectively.*

A strategy is *memoryless* if $|\mathcal{M}| = 1$. We say that a strategy requires finite memory if $|\mathcal{M}| < \infty$ and infinite memory if $|\mathcal{M}| = \infty$. We also classify the strategies based on the use of randomisation. A strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ is *pure* if π_u, π_n, and α map to Dirac distributions; *deterministic update* if π_u and α map to Dirac distributions, while π_n maps to an arbitrary distributions; and *stochastic update* where π_u, π_n, and α can map to arbitrary distributions. Stochastic update strategies are convenient because, for example, they allow to randomly choose between several other strategies in α, thus making the implementation of exact value problem for MDPs (as discussed in the introduction) straightforward. Note that from an implementation point of view, the controller using a *stochastic update* or a *deterministic update* strategy where π_n uses randomisation has to be equipped with a random number generator to provide a correct realisation of the strategy.

Markov chain induced by strategy pairs. Given a stochastic game arena $\mathcal{G}$ and an initial state distribution α, a strategy $\pi = \langle \mathcal{M}_1, \pi_u, \pi_n, \alpha_1\rangle$ of Preciser and a strategy $\sigma = \langle \mathcal{M}_2, \sigma_u, \sigma_n, \alpha_2\rangle$ of Spoiler induce a countable Markov chain $\mathcal{G}(\alpha, \pi, \sigma) = \langle S', (\emptyset, \emptyset, S'), \Delta'\rangle$ with starting state distribution $\alpha(\pi, \sigma)$ where

– $S' = S \times \mathcal{M}_1 \times \mathcal{M}_2$,
– $\Delta' \colon S' \times S' \to [0,1]$ is such that for all $(s, m_1, m_2), (s', m_1', m_2') \in S'$ we have

$$\Delta'(\langle (s, m_1, m_2), (s', m_1', m_2') \rangle) = \begin{cases} \pi_n(s, m_1)[s'] \cdot \pi_u(m_1, s')[m_1'] \cdot \sigma_u(m_2, s')[m_2'] & \text{if } s \in S_\Box, \\ \sigma_n(s, m_2)[s'] \cdot \pi_u(m_1, s')[m_1'] \cdot \sigma_u(m_2, s')[m_2'] & \text{if } s \in S_\Diamond, \\ \Delta(\langle s, s' \rangle) \cdot \pi_u(m_1, s')[m_1'] \cdot \sigma_u(m_2, s')[m_2'] & \text{if } s \in S_\bigcirc. \end{cases}$$

– $\alpha(\pi, \sigma) : S' \to [0,1]$ is defined such that for all $(s, m_1, m_2) \in S'$ we have that $\alpha(\pi, \sigma)[s, m_1, m_2] = \alpha[s] \cdot \alpha_1(s)[m_1] \cdot \alpha_2(s)[m_2]$.

To analyze a stochastic game $\mathcal{G}$ under a strategy pair $(\pi, \sigma) \in \Pi \times \Sigma$ and a starting state distribution α we define the probability measure over the set of paths $\Omega_{\mathcal{G},\alpha}^{\pi,\sigma}$ of $\mathcal{G}(\alpha, \pi, \sigma)$ with starting state distribution $\alpha(\pi, \sigma)$ in the following manner. The basic open sets of $\Omega_{\mathcal{G},\alpha}^{\pi,\sigma}$ are the *cylinder sets* $\mathrm{Cyl}(P) \stackrel{\text{def}}{=} P \cdot S'^{\omega}$ for every finite path $P = s_0' s_1' \dots s_k'$ of $\mathcal{G}(s, \pi, \sigma)$, and the probability assigned to $\mathrm{Cyl}(P)$ equals $\alpha(\pi, \sigma)[s_1'] \cdot \prod_{i=0}^{k} \Delta'(\langle s_i', s_{i+1}' \rangle)$. This definition induces a probability measure on the algebra of cylinder sets which, by Carathéodory's extension theorem, can be extended to a unique probability measure on the σ-algebra $\mathfrak{B}'$ generated by these sets. We denote the resulting probability measure by $\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}$. Often, we are only interested in the states visited on a path through $\mathcal{G}(s, \pi, \sigma)$ and not the memory contents. Let $\mathfrak{B}$ be the σ-algebra generated by the cylinder subsets of S^{ω}. We obtain a probability measure P on $\mathfrak{B}$ by setting $P(A) = \Pr_{\mathcal{G},\alpha}^{\pi,\sigma}(\rho^{-1}(A))$, where ρ is the natural projection from S'^{ω} to S^{ω}. We abuse notation slightly and denote this probability measure also by $\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}$. Our intended measurable space will always be clear form the context.

The xpected value of a measurable function $f \colon S'^{\omega} \to \mathbb{R} \cup \{\infty\}$ or $f \colon S^{\omega} \to \mathbb{R} \cup \{\infty\}$ under a strategy pair $(\pi, \sigma) \in \Pi \times \Sigma$ and a starting state distribution α is defined as $\mathbb{E}_{\mathcal{G},\alpha}^{\pi,\sigma}[f] \stackrel{\text{def}}{=} \int f \, d\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}$. The *conditional expectation* of a measurable function f *given* an event $A \in \mathfrak{B}$ ($A \in \mathfrak{B}'$) such that $\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}(A) > 0$ is defined analogously, i.e. $\mathbb{E}_{\mathcal{G},\alpha}^{\pi,\sigma}[f \mid A] = \int f \, d\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}(\cdot \mid A)$, where $\Pr_{\mathcal{G},\alpha}^{\pi,\sigma}(\cdot \mid A)$ denotes the usual conditional probability measure (conditioned on A).

3 Stochastic Games with Precise Objectives

We start this section by providing generic definitions of the two types of problems that we consider – *controller synthesis* and *counter strategy*. Then we show that the games are not determined even for reachability objectives and discuss the memory requirements for deterministic update strategies.

In a stochastic game with precise objective on arena $\mathcal{G}$, with starting state s, *objective function* $f \colon \Omega_{\mathcal{G},s} \to \mathbb{R}$, and target value $x \in \mathbb{Q}$, we say that a strategy π of player Preciser is *winning* if $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[f] = x$ for all $\sigma \in \Sigma$. Analogously, a strategy σ of player Spoiler is winning if $\mathbb{E}_{\mathcal{G},s}^{\pi,\sigma}[f] \neq x$ for all $\pi \in \Pi$. It is straightforward to see that for every starting state at most one player has a winning strategy. In Section 3.1 we show via an example that there are games where no

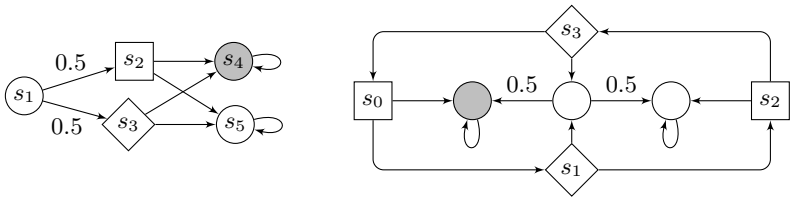

Fig. 1. Two stochastic game arenas where we depict stochastic vertices as circles and vertices of players Preciser and Spoiler as boxes and diamonds, respectively.

player has a winning strategy from some given state, i.e. stochastic games with precise objective are in general not determined. Hence, we study the following two problems with applications in controller synthesis of systems.

Definition 3 (Controller synthesis problem). *Given a game $\mathcal{G}$, a state s, an* objective function $f\colon \Omega_{\mathcal{G},s} \to \mathbb{R}$*, and a target value $x \in \mathbb{Q}$, the* controller synthesis problem *is to decide whether player Preciser has a winning strategy.*

Definition 4 (Counter-strategy problem). *Given a game $\mathcal{G}$, a state s, an objective function $f\colon \Omega_{\mathcal{G},s} \to \mathbb{R}$, and a target value $x \in \mathbb{Q}$, the* counter-strategy problem *asks whether Spoiler has no winning strategy, i.e., whether for every strategy σ of Spoiler there exists a strategy π of Preciser such that $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] = x$.*

In this paper we study the study controller synthesis and counter-strategy problems for the following objective functions:

- *Reachability* (with respect to a target set $T \subseteq S$) defined as $f^T_{\text{reach}}(\lambda) \stackrel{\text{def}}{=} 1$ if $\exists i \in \mathbb{N} : \lambda_i \in T$, and $f^T_{\text{reach}}(\lambda) \stackrel{\text{def}}{=} 0$ otherwise.
- *ω-regular* (with respect to an ω-regular property given as a deterministic parity automaton $\mathcal{A}$ [9]; we write $\mathcal{L}(\mathcal{A})$ for the language accepted by $\mathcal{A}$) defined as $f^{\mathcal{A}}_{\text{omega}}(\lambda) \stackrel{\text{def}}{=} 1$ if $\lambda \in \mathcal{L}(\mathcal{A})$, and $f^{\mathcal{A}}_{\text{omega}}(\lambda) \stackrel{\text{def}}{=} 0$ otherwise.
- *Total reward* (with respect to a *reward structure* $r\colon S \to \mathbb{R}^{\geq 0}$) defined as $f^r_{\text{total}}(\lambda) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} r(\lambda_i)$.
- *Discounted reward* (with respect to a discount factor $\delta \in [0,1)$ and a reward structure $r\colon S \to \mathbb{R}^{\geq 0}$) defined as $f^{\delta,r}_{\text{disct}}(\lambda) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} r(\lambda_i) \cdot \delta^i$.

3.1 Determinacy

In this section, we show that our games are, in general, *not determined*, i.e., a positive answer to the counter-strategy problem does not imply a positive answer to the controller synthesis problem. To see this, consider the game arena $\mathcal{G}$ given in Figure 1 (left) w.r.t the reachability function f^T_{reach} with target set $T = \{s_4\}$.

Proposition 1. *Preciser has no winning strategy on $\mathcal{G}$ from state s_1 for objective function f^T_{reach} and target value $x = 0.5$.*

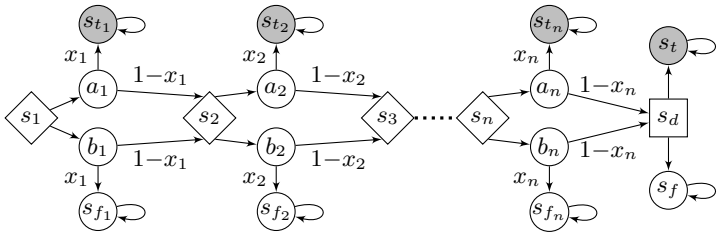

Fig. 2. Exponential deterministic update memory for Preciser

Proof. Assume that $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha \rangle$ is a solution the controller synthesis problem. We define two memoryless Spoiler strategies $\sigma = \langle \mathcal{M}_2, \sigma_u, \sigma_n, \alpha_2 \rangle$ and $\sigma' = \langle \mathcal{M}_2, \sigma_u, \sigma'_n, \alpha_2 \rangle$, where $\mathcal{M}_2 = \{init\}$, $\sigma_u(init, s_1) = \alpha_2(s_1) = init$, $\sigma_n(s_3, init) = s_4$, and $\sigma'_n(s_3, init) = s_5$. From the strategy construction and the fact that 0.5 of the probability mass is under control of Spoiler in s_3, we get that

$$\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_1}[f^T_{\text{reach}}] - \mathbb{E}^{\pi,\sigma'}_{\mathcal{G},s_1}[f^T_{\text{reach}}] = 0.5 \implies \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_1}[f^T_{\text{reach}}] \neq 0.5 \text{ or } \mathbb{E}^{\pi,\sigma'}_{\mathcal{G},s_1}[f^T_{\text{reach}}] \neq 0.5,$$

and thus π cannot be a solution to the controller synthesis problem. □

Proposition 2. *Spoiler has no winning strategy on $\mathcal{G}$ from state s_1 for objective function f^T_{reach} and target value $x = 0.5$.*

Proof. Let $\sigma = \langle \mathcal{M}, \sigma_u, \sigma_n, \alpha \rangle$ be any strategy for Spoiler. Then any strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha \rangle$ for Preciser with $\pi_u(m, s_2) = \sigma_u(m, s_3)$ and $\pi_n(s_2, m)[s_4] = \sigma_n(s_3, m)[s_5]$ for all $m \in \mathcal{M}$ satisfies $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_1}[f^T_{\text{reach}}] = 0.5$. □

3.2 Memory requirements

In this section we show that if *deterministic update* strategies are used, then the required size of the memory may be exponential in the size of the game. On the other hand, we later prove that *stochastic update* strategies require memory linear in the size of the game arena.

Proposition 3. *In the* controller synthesis *problem, Preciser may need memory exponential in the size of the game while using* deterministic update *strategy.*

Proof. Consider the game $\mathcal{G}$ in Figure 2 with the target set T shaded in gray, and constants x_i set to $2^{-(i+1)}$. Observe that under any strategy of Spoiler, the probability of runs that end in state s_{t_i} or s_{f_i} is exactly $\sum_{i=1}^{n} x_i \cdot \beta(i-1)$, where $\beta(k) = \prod_{j=1}^{k}(1 - x_j)$.

We now construct a deterministic update strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha \rangle$, which ensures that the probability to reach T is exactly 0.5. Intuitively, the strategy remembers the exact history, and upon arriving to s_d it looks at which states a_i for $1 \leq i \leq n$ were visited on a prefix of a history (and hence how much of the probability mass was directed to s_{t_i}), and sets the probability of going to s_f so that it "compensates" for these paths to target states to get the overall probability to reach target equal to 0.5. Formally,

– $\mathcal{M} = \{\mathrm{Pref}(\lambda, k) \colon \lambda \in \Omega_{\mathcal{G},s_1}, k \in \{1, \ldots, 2n\}\}$
– $\pi_u(m, s)$ equals $[m{\cdot}s \mapsto 1]$ if $m{\cdot}s \in \mathcal{M}$, and $[m \mapsto 1]$ otherwise.
– $\pi_n(s_d, m) = [s_t \mapsto p, s_f \mapsto 1-p]$, s. t. $p \cdot \beta(n) + \sum_{a_i \in m} x_i \cdot \beta(i-1) = 0.5$
– $\alpha(s) = [s \mapsto 1]$

Note that p above surely exists, because $\beta(n) \geq \beta(\infty) > \frac{1}{2}$. We argue that any strategy needs at least 2^n memory elements to achieve 0.5. Otherwise, there are two different histories $s_1 t_1 s_2 t_2 \ldots s_n t_n s_d$ and $s_1 t'_1 s_2 t'_2 \ldots s_n t'_n s_d$ where $t_i, t'_i \in \{a_i, b_i\}$ after which π assigns the same distribution $[s_t \mapsto y, s_f \mapsto 1-y]$. Let k be the smallest number such that $t_k \neq t'_k$, and w.l.o.g. suppose $t_k = a_k$. Let $\sigma \in \Sigma$ be a deterministic strategy that chooses to go to t_i in s_i, and let $\sigma' \in \Sigma$ be a deterministic strategy that chooses to go to t'_i in s_i. Then the probability to reach a target state under π and σ is at least $\sum_{i<k, t_i=a_i} x_i{\cdot}\beta(i-1) + x_k{\cdot}\beta(k-1) + y{\cdot}\beta(n)$, and under π and σ' — at most $\sum_{i<k, t_i=a_i} x_i{\cdot}\beta(i-1) + \sum_{k<i\leq n} x_i{\cdot}\beta(i-1) + y{\cdot}\beta(n)$. Because $x_k{\cdot}\beta(k-1) > (\sum_{k<i\leq n} x_i){\cdot}\beta(k-1) > \sum_{k<i\leq n} x_i{\cdot}\beta(i-1)$, we obtain a contradiction.

Note that by replacing the states a_i and b_i with gadgets of $i+1$ stochastic states the example can be altered so that the only probabilities assigned by the probabilistic transition function are 0, 1 and $\frac{1}{2}$.

4 Controller Synthesis Problem

In this section we present our results on controller synthesis problem. We say that a state is *terminal* if no other state is reachable from it under any strategy pair. We call a stochastic game *stopping* if a terminal state is reached with probability 1 under any pair of strategies. We define conditions under which the controller synthesis problem has a solution for a general class of functions, the so-called linearly bounded functions—under stopping games assumption. We say that an objective function is *linearly bounded* if there are x_1 and x_2 such that for any ω that contains k nonterminal states we have $|f(\omega)| \leq x_1 \cdot k + x_2$. We observe that objective functions define in previous section are linearly-bounded and present compact winning strategies for those objective.

4.1 Conditions for the existence of winning strategies

We define $\mathit{Exact}_{\mathcal{G}}(s, f) \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid \exists \pi \in \Pi \,.\, \forall \sigma \in \Sigma : \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] = x\}$ to be the set of values for which Preciser has a winning strategy on $\mathcal{G}$ from s with objective function f. Given a function $f : \Omega_{\mathcal{G}} \to \mathbb{R}$, a finite path $u{\cdot}s \in {\Omega_{\mathcal{G}}}^+$ and an infinite path $v \in \Omega_{\mathcal{G}}$, we define a curried function $f_{u\cdot s}(s \cdot v) = f(u{\cdot}s{\cdot}v)$, where $s \in S$. Given a finite path as the history of the game, the following lemma presents conditions under which player Preciser cannot win the game for any value.

Lemma 1. *Given a game $\mathcal{G}$, a finite path $w{\cdot}s \in {\Omega_{\mathcal{G}}}^+$, where $s \in S$ and a function f, if* $\inf\limits_{\pi\in\Pi} \sup\limits_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}] > \sup\limits_{\pi\in\Pi} \inf\limits_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}]$, *then Preciser cannot achieve any exact value after that path, i.e., $\mathit{Exact}_{\mathcal{G}}(s, f_{w\cdot s}) = \emptyset$.*

Proof. For every Preciser strategy $\pi \in \Pi$, we have that

$$\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}] \leq \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}] < \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}] \leq \sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}].$$

Hence, for any of the strategy π of Preciser, Spoiler can ensure one of the two *distinct* values $\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}]$ or $\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}]$, and therefore Preciser cannot guarantee any exact value after history $w\cdot s$, so $Exact_{\mathcal{G}}(s, f_{w\cdot s}) = \emptyset$. □

Let $\Omega^{\text{no}}_{\mathcal{G},f} \subseteq {\Omega_{\mathcal{G}}}^+$ be a set paths in $\mathcal{G}$ such that a path w is in $\Omega^{\text{no}}_{\mathcal{G},f}$ if and only if w satisfies the condition in Lemma 1, i.e., after w Preciser cannot guarantee any exact value for a function f. The above proposition characterises the states from which Preciser cannot achieve any exact value.

Proposition 4. *In a game $\mathcal{G}$, and a state $s \in S$, if for any strategy of Preciser, Spoiler has a strategy to make sure that at least one path from $\Omega^{no}_{\mathcal{G},f}$ has positive probability, then $Exact_{\mathcal{G}}(s, f) = \emptyset$, i.e.,*

$$\forall\pi \in \Pi\,.\,\exists\sigma \in \Sigma : \Pr^{\pi,\sigma}_{\mathcal{G},s}(\bigcup_{w\in\Omega^{no}_{\mathcal{G},f}} \text{Cyl}(w)) > 0 \Rightarrow Exact_{\mathcal{G}}(s, f) = \emptyset.$$

In the next theorem we complement the proposition by describing the states with nonempty sets $Exact_{\mathcal{G}}(s, f)$, for the class of linearly-bounded objective functions.

Theorem 1. *Given a stopping game $\mathcal{G}$, a linearly bounded objective function f satisfying $\Omega^{no}_{\mathcal{G},f} = \emptyset$, a state $s \in S$, and a value $x \in \mathbb{R}$,*

$$x \in Exact_{\mathcal{G}}(s, f) \iff \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f].$$

Proof (Sketch). The "⇒" direction of the theorem is straightforward. To show "⇐" direction, we construct a strategy to achieve any given probability x.

Let π^- and π^+ be minimising and maximising pure deterministic update strategies [5]. Let $w\cdot s \in {\Omega_{\mathcal{G}}}^+$. We define minimum and maximum expected values achievable by Preciser after a finite path $w\cdot s$ as:

$$\text{val}^-(w\cdot s) = \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}] \quad \text{and} \quad \text{val}^+(w\cdot s) = \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w\cdot s}].$$

We will now construct a stochastic update strategy for Preciser, which is winning from all $s \in S$. Given any $l \leq y \leq h$, we define $c(y, l, h)$ as $\frac{h-y}{h-l}$ if $l \neq h$ and 1 otherwise. For a finite path $w \in {\Omega_{\mathcal{G}}}^+$ such that $\text{val}^-(w) \leq y \leq \text{val}^+(w)$, we define $\beta(y, w) = c(y, \text{val}^-(w), \text{val}^+(w))$. The strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ is defined by

– $\mathcal{M} = \{\langle w, \text{val}^-(w)\rangle, \langle w, \text{val}^+(w)\rangle \mid w \in {\Omega_{\mathcal{G}}}^+\}$,

[5] Note that thanks to our restrictions on f and $\mathcal{G}$ these always exist.

– $\pi_u(\langle w{\cdot}s, y\rangle, t) = \begin{cases} \langle w{\cdot}s{\cdot}t, y\rangle, & \text{if } s \in S_{\Box}, \\ [\langle w{\cdot}s{\cdot}t, \mathrm{val}^-(w{\cdot}s{\cdot}t)\rangle \mapsto \beta(y, w{\cdot}s{\cdot}t), & \\ \ \langle w{\cdot}s{\cdot}t, \mathrm{val}^+(w{\cdot}s{\cdot}t)\rangle \mapsto 1-\beta(y, w{\cdot}s{\cdot}t)], & \text{if } s \in S_{\Diamond}, \\ \langle w{\cdot}s{\cdot}t, \mathrm{val}^-(w{\cdot}s{\cdot}t)\rangle, & \text{if } s \in S_{\bigcirc} \text{ and } y = \mathrm{val}^-(w{\cdot}s), \\ \langle w{\cdot}s{\cdot}t, \mathrm{val}^+(w{\cdot}s{\cdot}t)\rangle, & \text{if } s \in S_{\bigcirc} \text{ and } y = \mathrm{val}^+(w{\cdot}s), \end{cases}$

– $\pi_n(s, \langle w, y\rangle) = \begin{cases} \pi^-(w) & \text{if } y = \mathrm{val}^-(w), \\ \pi^+(w) & \text{otherwise} \end{cases}$

– $\alpha(s) = [\langle s, \mathrm{val}^-(s)\rangle \mapsto \beta(x, s), \langle s, \mathrm{val}^+(s)\rangle \mapsto 1-\beta(x, s)]$,

for all $w \in {\Omega_{\mathcal{G}}}^+$, $s, t \in S$, and $\langle w, y\rangle, \langle w{\cdot}s, y\rangle \in \mathcal{M}$. The correctness of the strategy follows from the proof in [5]. □

4.2 Compact Strategies for Objective Functions

In this section, using the results from Theorem 1, we construct stochastic update strategies for the functions defining reachability, total expected reward, discounted reward and ω-regular objectives, all of which are linearly bounded. For all games, and objective functions in this section we assume that $\Omega^{\mathrm{no}}_{\mathcal{G},f} = \emptyset$.

Proposition 5. *Reachability, ω-regular, total reward and discounted reward objectives are linearly-bounded.*

From Theorem 1 and Proposition 5 it follows that for in a game $\mathcal{G}$, a state s and value x, if f is reachability, ω-regular, total reward or discounted reward objectives satisfying the assumptions of Theorem 1, then player Preciser has a winning strategy if and only if $\inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f]$. The construction from Theorem 1 only provides strategy having countable memory. In this section we show that these objectives allow for a compact strategy.

Proposition 6 (Reachability). *If there exists a winning strategy for Preciser in stopping game $\mathcal{G}$ for reachability function f^T_{reach}, then there exists a stochastic update winning strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ such that $|\mathcal{M}| \leq 2{\cdot}|S|$.*

Proof (Sketch). Let $f = f^T_{\mathrm{reach}}$ and π^- and π^+ be the pure memoryless deterministic update strategies achieving, for every $w{\cdot}s \in {\Omega_{\mathcal{G}}}^+$, the minimum and maximum expected value for f. By Theorem 1 there exists a stochastic update strategy $\pi^\star$, which achieves the precise reachability probability. However, the construction only provides a strategy having countable memory. We will construct a stochastic update strategy which is equivalent to $\pi^\star$, but has memory size at most $2{\cdot}|S|$. The strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ is defined as follows:

– $\mathcal{M} = \{\langle s, \mathrm{val}^-(s)\rangle, \langle s, \mathrm{val}^+(s)\rangle \mid s \in S\}$,

– $\pi_u(\langle s, y\rangle, t) = \begin{cases} \langle t, y\rangle & \text{if } s \in S_{\Box}, \\ [\langle t, \mathrm{val}^-(t)\rangle \mapsto \beta(y, t), & \\ \ \langle t, \mathrm{val}^+(t)\rangle \mapsto 1-\beta(y, t)] & \text{if } s \in S_{\Diamond}, \\ \langle t, \mathrm{val}^-(t)\rangle & \text{if } s \in S_{\bigcirc} \text{ and } y = \mathrm{val}^-(s), \\ \langle t, \mathrm{val}^+(t)\rangle & \text{if } s \in S_{\bigcirc} \text{ and } y = \mathrm{val}^+(s), \end{cases}$

– $\pi_n(s, \langle s, y\rangle) = \begin{cases} \pi^-(s) & \text{if } y = \mathrm{val}^-(s), \\ \pi^+(s) & \text{otherwise} \end{cases}$

– $\alpha(s) = [\langle s, \mathrm{val}^-(s)\rangle \mapsto \beta(x, s), \langle s, \mathrm{val}^+(s)\rangle \mapsto 1 - \beta(x, s)]$,

for all $s, t \in S$, and $\langle s, y\rangle \in \mathcal{M}$.

Let us look at the functions of the strategy individually. The initial distribution functions of $\pi^\star$ and π are the same. For the next move functions, since π^- and π^+ are memoryless, we have that for any path $w{\cdot}s \in {\Omega_\mathcal{G}}^+$, $\pi^-(w{\cdot}s) = \pi^-(s)$ and $\pi^+(w{\cdot}s) = \pi^+(s)$. It follows that $\pi_n(s, \langle w{\cdot}s, y\rangle) = \pi_n(s, \langle s, y\rangle)$. For the memory update function π_u, it is equivalent to the memory update function of $\pi^\star$ (i.e., produces the same distributions for all paths) if the target states are treated as terminal, i.e., for reachability function it does not matter what actions are played after the target has been reached. □

The proofs for the following two propositions are similar (see [5] for details).

Proposition 7 (ω-regular). *If there exists a winning strategy for Preciser in stopping game $\mathcal{G}$ for ω-regular objective function $f^{\mathcal{A}}_{\mathrm{omega}}$ and objective given as a deterministic parity automaton $\mathcal{A}$, then there exists a stochastic update winning strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ such that $|\mathcal{M}| \leq 2{\cdot}|S|{\cdot}|\mathcal{A}|$.*

Proposition 8 (Total reward). *If there exists a winning strategy for Preciser in a stopping game $\mathcal{G}$ for total reward function f^r_{total}, then there exists a stochastic update winning strategy $\pi = \langle \mathcal{M}, \pi_u, \pi_n, \alpha\rangle$ such that $|\mathcal{M}| \leq 2{\cdot}|S|$.*

Since discounted objective implicitly mimics stopping mechanism, using Proposition 8 and Theorem 1 we show that for the discounted objectives we can construct compact strategies for arbitrary finite games without the stopping assumption.

Theorem 2. *Given a game arena $\mathcal{G}$, a discounted reward function $f = f^{\delta,r}_{\mathrm{disct}}$, satisfying $\Omega^{no}_{\mathcal{G},f} = \emptyset$, a state $s \in S$, and a value $x \in \mathbb{R}$.*

$$x \in \mathit{Exact}_\mathcal{G}(s, f) \iff \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f].$$

Proof. The proof employs a standard construction [11] that reduces the expected discounted reward problem to expected total reward problem. Let $\mathcal{G} = \langle S, (S_\Box, S_\Diamond, S_\bigcirc), \Delta\rangle$, and let $f^{\delta,r}_{\mathrm{disct}}$ be given by a reward structure r and a discount factor $0 < \delta < 1$, we define a stopping game $\mathcal{G}' = \langle S \cup S', (S_\Box, S_\Diamond, S_\bigcirc \cup S'), \Delta'\rangle$ and a *total reward* objective function f^r_{total} as follows. The set S' contains states $\bar{s}$ for all $s \in S$ and a distinguished state $\star$. The set Δ' is defined as follows: for all s, t we define $\Delta'(s, \bar{t}) = \Delta(s, t)$, $\Delta(\bar{t}, t) = 1 - \delta$ and $\Delta(\bar{t}, \star) = \delta$. We make the state $\star$ terminal by putting $\Delta'(\star, \star) = 1$. The reward structure r' for f^r_{total} in $\mathcal{G}'$ is defined by $r'(s) = r(s)$ for all $s \in S$ and $r(s') = 0$ otherwise. There is a straightforward bijection between the strategies of $\mathcal{G}$ and $\mathcal{G}'$ that for any π and σ returns π' and σ' such that $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f^{\delta,r}_{\mathrm{disct}}] = \mathbb{E}^{\pi',\sigma'}_{\mathcal{G}',s}[f^r_{\mathrm{total}}]$. The theorem is then obtained by using Theorem 1 and Proposition 8. □

4.3 Complexity

We discuss the complexity of the controller synthesis problem for the objectives considered in Section 4.2 where compact strategies do exist. As we discussed in previous section, controller synthesis essentially boils down to computing the extreme values of the corresponding game. Assume that we have an oracle to decide the following: (1) given any state of the game s, whether $\sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \geq \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f]$ and (2) given any state of the game s, whether $\inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f]$. By Proposition 4 and Theorem 1, together with Proposition 6 – 8 the controller synthesis problem is decidable in polynomial time if we have oracles for (1) and (2).

For the considered objectives, (1) and (2) are decidable in NP $\cap$ co-NP since games with these objectives admit pure memoryless strategies for both players (in the product of the game with the deterministic parity automaton at least in the case of ω-regular objectives; cf. [3]). It is easy to see that $\mathrm{P}^{\mathrm{NP}\cap\text{co-NP}} = \mathrm{NP}\cap\text{co-NP}$. Hence, we can conclude that the controller-synthesis problem is in NP $\cap$ co-NP for the objectives studied in Section 4.2.

4.4 Non-stopping games

It is natural to ask whether the result of Theorem 1 can be transferred to non-stopping games. The following proposition provides a negative answer.

Proposition 9. *There is a game $\mathcal{G}$ and a reachability objective f, a state $s \in S$ and a number $\inf\limits_{\pi\in\Pi}\sup\limits_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \sup\limits_{\pi\in\Pi}\inf\limits_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f]$ such that $\Omega^{no}_{\mathcal{G},f} = \emptyset$ and $x \notin \mathit{Exact}_{\mathcal{G}}(s,f)$.*

To prove Proposition 9, consider the game $\mathcal{G}$ from Figure 1 (right), where the target state is marked with gray colour. For each state $s \in \{s_0, s_1, s_3\}$ we have $\inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] = 0.5$, and for state s_2 we have $\inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_2}[f] = 0.0$. On the other hand, for each state $s \in \{s_1, s_2, s_3\}$ we have that $\sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] = 0.5$, while for state s_0 we have $\sup_{\pi\in\Pi}\inf_{\sigma\in\Sigma}\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_0}[f] = 1$. However, for example in state s_0 we get $\mathit{Exact}(s,f) = \{1\}$. For any value $0.5 \leq x < 1$, any strategy π that should achieve x must in s_0 pick the transition to the terminal state with probability $2 \cdot x - 1$, because otherwise Spoiler could propose a counter strategy σ which deterministically goes up from s_1, and thus $\mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \neq x$. Let us suppose that π has this property, then it must further ensure that from s_1 the target state is reached with probability 0.5, which means that it can *never* randomise in s_0 or s_1, except for the very first step: if it randomised, Spoiler could propose a winning counter-strategy that would go to central vertex immediately after the first randomisation took place. But this means that the strategy π must always keep going from s_2 to s_3 and from s_0 to s_1 deterministically, to which Spoiler can respond by a strategy σ that always goes from s_1 to s_2 and from s_3 to s_0 deterministically, hence avoiding to enter the target state at all.

An interesting point to make is that even though Preciser has not any strategy that would ensure reaching the target from s_0 in $\mathcal{G}$ with probability x for a

given $0.5 \leq x < 1$, he has got an "ε-optimal" strategy for any $\varepsilon > 0$, i.e. for any x there is a strategy π of Preciser such that for all σ of Spoiler we get $x - \varepsilon \leq \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s_0}[f] \leq x + \varepsilon$. For example, if $x = 0.5$, the strategy π can be defined so that in s_0 it picks the transition to s_1 with probability $1 - \varepsilon$, and the other available transition with probability ε, while in s_2 it takes the transition to s_3 with probability $1 - \frac{\varepsilon}{1-\varepsilon}$, and the other available transition with probability $\frac{\varepsilon}{1-\varepsilon}$.

Again, one might ask whether ε-optimal strategies always exist. Unfortunately, this is also not the case, as can be seen when the transition from s_0 to the target state is redirected to the non-target terminal state.

5 Counter-Strategy Problem

In this section we discuss the *counter-strategy* problem, which, given a game $\mathcal{G}$, a state s, and an objective function f, asks whether for any strategy of Spoiler there exists a counter-strategy for Preciser such that the expected value of f is exactly x. Let us characterise the set of all values for which counter-strategy exists by defining $CExact_{\mathcal{G}}(s, f) = \{x \in \mathbb{R} \mid \forall \sigma \in \Sigma \,.\, \exists \pi \in \Pi : \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] = x\}$.

Lemma 2. *Given a game $\mathcal{G}$, a finite path $w \cdot s \in {\Omega_{\mathcal{G}}}^+$, where $s \in S$ and a function f, if $\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w \cdot s}] > \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w \cdot s}]$, then Preciser cannot achieve any exact value after that path, i.e., $CExact_{\mathcal{G}}(s, f_{w \cdot s}) = \emptyset$.*

Proof. Let $x^* = \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w \cdot s}]$ and $x_* = \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f_{w \cdot s}]$ such that $x^* > x_*$; and let $\sigma^*, \sigma_* \in \Sigma$ be the corresponding strategies of Spoiler. Notice that for any arbitrary strategy π of Preciser we have that

$$\mathbb{E}^{\pi,\sigma_*}_{\mathcal{G},w_0}[f_{w \cdot s}] \leq x_* < x^* \leq \mathbb{E}^{\pi,\sigma^*}_{\mathcal{G},w_0}[f_{w \cdot s}].$$

Hence, if $x \leq x_*$ then there is no strategy of Preciser that yields expectation at most x against σ^*, while if $x > x_*$ then there is no strategy of Preciser that yields expectation at least x against σ_*. Hence, $CExact_{\mathcal{G}}(s, f_{w \cdot s}) = \emptyset$. □

Let $\Omega^{\mathrm{no_c}}_{\mathcal{G},f} \subseteq {\Omega_{\mathcal{G}}}^+$ be a set paths in $\mathcal{G}$ such that a path w is in $\Omega^{\mathrm{no_c}}_{\mathcal{G},f}$ if and only if w satisfies the condition in Lemma 2, i.e., after w Preciser cannot propose a counter strategy to achieve any exact value, for a function f.

Proposition 10. *In an game $\mathcal{G}$, and a state $s \in S$, if there exists a strategy of Spoiler, such that for all strategies of Preciser at least one path from $\Omega^{no_c}_{\mathcal{G},f}$ has a positive probability, then $CExact_{\mathcal{G}}(s, f) = \emptyset$, i.e.,*

$$\exists \sigma \in \Sigma \,.\, \forall \pi \in \Pi : \Pr^{\pi,\sigma}_{\mathcal{G},s}\Big(\bigcup_{w \in \Omega^{no_c}_{\mathcal{G},f}} \mathrm{Cyl}(w)\Big) > 0 \Rightarrow CExact_{\mathcal{G}}(s, f) = \emptyset.$$

Using the results above we are now ready to characterise the states from which Preciser has, for any Spoiler strategy, a winning counter strategy to achieve exactly the specified value x. The following theorem is proved in [5].

Theorem 3. *In a game $\mathcal{G}$ with $\Omega^{no_c}_{\mathcal{G},f} = \emptyset$, and a state $s \in S$, $x \in CExact_{\mathcal{G}}(s, f)$ if and only if $\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f] \leq x \leq \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi} \mathbb{E}^{\pi,\sigma}_{\mathcal{G},s}[f]$.*

6 Conclusion and future work

In this paper we studied a novel kind of objectives for two-player stochastic games, in which the role of one player is to achieve exactly a given expected value, while the role of the other player is to get any other value. We settled the controller synthesis problem for stopping games with linearly bounded objective functions and for arbitrary finite games with discounted reward objective. We solved the counter strategy problem for arbitrary finite games and arbitrary payoff functions. There are two main directions for future work: 1. relaxing the restrictions on the game arenas, i.e., studying the controller-synthesis problem for non-stopping games; 2. modifying the problem so that the role of preciser is to reach a value from certain interval, rather than one specific number.

Acknowledgments. The authors are part supported by ERC Advanced Grant VERIWARE and EPSRC grant EP/F001096/1. Vojtěch Forejt is supported by a Royal Society Newton Fellowship. Ashutosh Trivedi is supported by NSF awards CNS 0931239, CNS 1035715, CCF 0915777. Michael Ummels is supported by the DFG project SYANCO.

References

1. T. Brázdil, V. Brožek, K. Chatterjee, V. Forejt, and A. Kučera. Two views on multiple mean-payoff objectives in Markov decision processes. In *LICS*, pages 33–42, 2011.
2. A. N. Cabot and R. C. Hannum. Gaming regulation and mathematics: A marriage of necessity. *John Marshall Law Review*, 35(3):333–358, 2002.
3. K. Chatterjee and T. A. Henzinger. A survey of stochastic ω-regular games. *J. Comput. Syst. Sci.*, 78(2):394–413, 2012.
4. K. Chatterjee, R. Majumdar, and T. A. Henzinger. Markov decision processes with multiple objectives. In *STACS'06*, volume 3884 of *LNCS*, pages 325–336, 2006.
5. T. Chen, V. Forejt, M. Kwiatkowska, A. Simaitis, A. Trivedi, and M. Ummels. Playing stochastic games precisely. Technical Report No. CS-RR-12-03, Department of Computer Science, University of Oxford, June 2012.
6. S. Dziembowski, M. Jurdzinski, and I. Walukiewicz. How much memory is needed to win infinite games? In *LICS*, pages 99–110. IEEE Computer Society, 1997.
7. K. Etessami, M. Z. Kwiatkowska, M. Y. Vardi, and M. Yannakakis. Multi-objective model checking of Markov decision processes. *LMCS*, 4(4), 2008.
8. J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.
9. E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games. A Guide to Current Research*, volume 2500 of *LNCS*. Springer, 2002.
10. A. Neyman and S. Sorin, editors. *Stochastic Games and Applications*, volume 570 of *NATO Science Series C*. Kluwer Academic Publishers, 2004.
11. M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.
12. P. Ramadge and W. Wonham. The control of discrete event systems. In *Proc. IEEE*, volume 77(1), 1989.
13. L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. U.S.A.*, 39, 1953.
14. State of New Jersey, 214th legislature, as amended by the General Assembly on 01/10/2011. `http://www.njleg.state.nj.us/2010/Bills/S0500/12_R4.PDF`, November 2010.