Department of Computer Science

Strategy Synthesis for Stochastic Games with Multiple Long-Run Objectives

Nicolas Basset, Marta Kwiatkowska, Ufuk Topcu, and Clemens Wiltsche

CS-RR-14-10

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

Strategy Synthesis for Stochastic Games with Multiple Long-Run Objectives

Nicolas Basset*, Marta Kwiatkowska*, Ufuk Topcu†, and Clemens Wiltsche*

* Department of Computer Science, University of Oxford, United Kingdom
† Department of Electrical and Systems Engineering, University of Pennsylvania, USA

Abstract. We consider turn-based stochastic games whose winning conditions are conjunctions of satisfaction objectives for long-run average rewards, and address the problem of finding a strategy that almost surely maintains the averages above a given multi-dimensional threshold vector. We show that strategies constructed from Pareto set approximations of expected energy objectives are ε-optimal for the corresponding average rewards. We further apply our methods to compositional strategy synthesis for multi-component stochastic games that leverages composition rules for probabilistic automata, which we extend for long-run ratio rewards with fairness. We implement the techniques and illustrate our methods on a case study of automated compositional synthesis of controllers for aircraft primary electric power distribution networks that ensure a given level of reliability.

1 Introduction

Reactive systems must continually interact with the changing environment. Since it is assumed that they should never terminate, their desirable behaviours are typically specified over infinite executions. Reactive systems are naturally modelled using games, which distinguish between the controllable and uncontrollable events. Stochastic games [18], in particular, allow one to specify uncertainty of outcomes by means of probability distributions. When such models are additionally annotated by rewards that represent, e.g., energy usage and time passage, quantitative objectives and analysis techniques are needed to ensure their correctness. Often, not just a single objective is under consideration, but several, potentially conflicting, objectives must be satisfied, for example maximising both throughput and latency of a network.

In our previous work [7,8], we formulated multi-objective expected total reward properties for stochastic games with certain terminating conditions and showed how ε-optimal strategies can be approximated. Expected total rewards, however, are unable to express long-run average (also called mean-payoff) properties of reactive systems. Another important class of properties are ratio rewards, with which one can state, e.g., speed (distance per time unit) or fuel efficiency (distance per unit of fuel). In this paper we consider controller synthesis for the general class of turn-based stochastic games whose winning conditions are conjunctions of satisfaction objectives for long-run average rewards. We represent the controllable and uncontrollable actions by Player $\Diamond$ and Player $\Box$, respectively, and address the problem of finding a strategy to satisfy such long-run objectives almost surely for Player $\Diamond$ against all choices of Player $\Box$. These objectives can be used to specify behaviours that guarantee that the probability density is above a threshold, in several dimensions, and the executions actually satisfy the objective we are interested in, which is important for, e.g., reliability and availability analysis. In contrast, expected rewards average the reward over different probabilistic outcomes, possibly with arbitrarily high variance, and thus it may be the case that none of the paths actually satisfy the objective.

Satisfaction Objectives. The specifications we consider are quantitative, in the sense that they are required to maintain the rewards above a certain threshold, and we are interested in almost sure satisfaction, that is, this condition on the rewards is satisfied with probability one. The problem we study generalises the setting of stopping games with multiple satisfaction objectives, which for LTL specifications can be solved via reduction to expected total rewards [8], while our methods are applicable to general turn-based stochastic games. In stopping games, objectives defined using total rewards are appropriate, since existence of the limits is ensured by termination; however, total rewards may diverge for reactive systems, and hence we cannot reduce our problem to total rewards.

Strategy Synthesis. Stochastic games with multiple objectives have been studied in [12], where determinacy under long-run objectives (including ours) is shown (but without strategy construction). However, in general, the winning strategies are history-dependent, requiring infinite memory, which is already the case for Markov decision processes [5]. We restrict to finite memory strategies and utilise the stochastic memory update representation of [7]. For approximating expected total rewards in games, one can construct strategies (in particular, their memory update representation) after finitely many iterations from the difference between achievable values of successive states [8], but long-run properties erase all transient behaviours, and so, in general, we cannot use the achievable values for strategy construction. Inspired by [6], we use expected energy objectives to compute the strategies. These objectives are meaningful in their own right to express that, at every step, the average over some resource requirement does not exceed a certain budget, i.e. some sequences of operations are allowed to violate the budget constraint, as long as they are balanced by other sequences of operations. Consider, for example, sequences of stock market transactions: it is desirable that the expected capital never drops below zero (or some higher value), which can be balanced by credit for individual transactions below the threshold. Synthesis via expected energy objectives yields strategies that not only achieve the required target, but we also obtain a bound on the maximum expected deviation at any step by virtue of the bounded energy. Then, given an achievable target $\boldsymbol{v}$ for mean-payoff, the target $\mathbf{0}$ is ε-achievable by an energy objective with rewards shifted by $-\boldsymbol{v}$, and the same strategy achieves $\boldsymbol{v} - \varepsilon$ for the mean-payoff objective under discussion.

Compositional Synthesis. In our previous work [4], we proposed a synchronising parallel composition for stochastic games that enables a compositional approach to controller synthesis that significantly outperforms the monolithic method. The strategy for the composition of games is derived from the strategies synthesised for the individual components. To apply these methods for a class of objectives (e.g. total rewards), one must (i) show that the objectives are defined on traces, i.e. synchronisation of actions is sufficient for information sharing; (ii) provide compositional verification rules for probabilistic automata (e.g. assume-guarantee rules); and (iii) provide synthesis methods for single component games. We address these points for long-run average objectives, extending [13] for (ii), enabling compositional synthesis for ratio rewards. A key characteristic of the rules is the use of fairness, which requires that no component is prevented from making progress. The methods of [4] were presented with total rewards, where (trivial) fairness was only guaranteed through synchronised termination.

Case Study. We implement the methods and demonstrate their scalability and usefulness via a case study that concerns the control of the electric power distribution on aircraft [15]. In avionics, the transition to more-electric aircraft has been brought about by advances in electronics technology, reducing take-off weight and power consumption. We extend the (non-quantitative) game-theoretic approach of [21] to the stochastic games setting with multiple long-run satisfaction objectives, where the behaviour of generators is described stochastically. We demonstrate how our approach yields controllers that ensure given reliability levels and higher uptimes than those reported in [21].

Contributions. Our main contributions are as follows.

- We show that expected energy objectives enable synthesis of ε-optimal finite-memory strategies for almost sure satisfaction of average rewards (Theorem 2).
- We propose a semi-algorithm to construct ε-optimal strategies using stochastically updated memory (Theorem 1).
- We extend compositional rules to specifications defined on traces, and hence show how to utilise ratio rewards in compositional synthesis (Theorem 3).
- We demonstrate compositional synthesis using long-run objectives via a case study of an aircraft electric power distribution network.

Related Work. For Markov decision processes (MDPs), multi-dimensional long-run objectives for satisfaction and expectation were studied in [5], and expected ratio rewards in [20]. Satisfaction for long-run properties in stochastic games is the subject of [12]; in particular, they present algorithms for combining a single mean-payoff with a Büchi objective, which rely on the non-quantitative nature of the Büchi objective, and hence cannot be straightforwardly extended to several mean-payoff objectives that we consider. Non-stochastic games with energy objectives have been considered, for example, in [6], where it is assumed that Player $\Box$ plays deterministically, in contrast to our approach that permits the use of stochasticity. Our almost sure satisfaction objectives are related to the concept of quantiles in [1], in that they correspond to 1-quantiles, but here we consider mean-payoff objectives for games. An extended version of this paper, including proofs, can be found in [3].

2 Preliminaries

Notation. A *discrete probability distribution* (or *distribution*) over a (countable) set Q is a function $\mu : Q \to [0,1]$ such that $\sum_{q\in Q} \mu(q) = 1$; its *support* $\mathsf{supp}(\mu)$ is $\{q \in Q \mid \mu(q) > 0\}$. We denote by $\mathcal{D}(Q)$ the set of all distributions over Q with finite support. A distribution $\mu \in \mathcal{D}(Q)$ is *Dirac* if $\mu(q) = 1$ for some $q \in Q$, and if the context is clear we just write q to denote such a distribution μ.

We work with the usual metric-space topology on $\mathbb{R}^n$. The *downward closure* of a set X is defined as $\mathsf{dwc}(X) \stackrel{\text{def}}{=} \{\boldsymbol{y} \mid \exists \boldsymbol{x} \in X \,.\, \boldsymbol{y} \leq \boldsymbol{x}\}$. A set $X \subseteq \mathbb{R}^n$ is *convex* if for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$, and all $\alpha \in [0,1]$, $\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in X$; its *convex hull* $\mathsf{conv}(X)$ is the smallest convex set containing X. Given a set X, $\alpha \times X$ denotes the set $\{\alpha \cdot \boldsymbol{x} \mid \boldsymbol{x} \in X\}$. The *Minkowski sum* of sets X and Y is $X + Y \stackrel{\text{def}}{=} \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in X, \boldsymbol{x} \in Y\}$. We refer to the sth component of a vector $\boldsymbol{v}$ by v_s and $[\boldsymbol{v}]_s$. We write $\boldsymbol{\varepsilon}$ to denote the vector $(\varepsilon, \varepsilon, \ldots, \varepsilon)$. For a vector $\boldsymbol{x}$ (resp. vector of sets Z) and a scalar ε, define $\boldsymbol{x} + \varepsilon$ by $[\boldsymbol{x} + \varepsilon]_s = x_s + \varepsilon$ (resp. $[Z + \varepsilon]_s \stackrel{\text{def}}{=} Z_s + \varepsilon$) for all components s of $\boldsymbol{x}$ (resp. Z), where, for a set X, let $X + \varepsilon \stackrel{\text{def}}{=} \{\boldsymbol{x} + \varepsilon \mid \boldsymbol{x} \in X\}$. For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x} \cdot \boldsymbol{y}$ denotes their dot-product, and $\boldsymbol{x} \bullet \boldsymbol{y}$ denotes component-wise multiplication.

Stochastic Games. We consider turn-based action-labelled stochastic two-player games (henceforth simply called *games*), which distinguish two types of nondeterminism, each controlled by a separate player. $\mathsf{Player}\ \Diamond$ represents the controllable part for which we want to synthesise a strategy, while $\mathsf{Player}\ \Box$ represents the uncontrollable environment.

Definition 1. *A* game G *is a tuple* $\langle S, (S_\Diamond, S_\Box), \varsigma_0, \mathcal{A}, \longrightarrow\rangle$*, where* S *is a finite set of* states *partitioned into* $\mathsf{Player}\ \Diamond$ states $S_\Diamond$ *and* $\mathsf{Player}\ \Box$ states $S_\Box$*;* $\varsigma_0 \in S$ *is an* initial state*;* $\mathcal{A}$ *is a finite set of* actions*; and* $\longrightarrow \subseteq S \times (\mathcal{A} \cup \{\tau\}) \times \mathcal{D}(S)$ *is a* transition relation*, such that, for all* s*,* $\{(s, a, \mu) \in \longrightarrow\}$ *is finite.*

We write $s \xrightarrow{a} \mu$ for a *transition* $(s, a, \mu) \in \longrightarrow$. The action labels $\mathcal{A}$ on transitions model observable behaviours, whereas τ can be seen as internal: it cannot be used in winning conditions and is not synchronised in the composition. We denote the set of *moves* (also called *stochastic states*) by $S_\bigcirc \stackrel{\text{def}}{=} \{(a, \mu) \in \mathcal{A} \times \mathcal{D}(S) \mid \exists s \in S \,.\, s \xrightarrow{a} \mu\}$, and let $\overline{S} = S \cup S_\bigcirc$. Let the set of *successors* of $s \in \overline{S}$ be $\mathsf{succ}(s) \stackrel{\text{def}}{=} \{(a, \mu) \in S_\bigcirc \mid s \xrightarrow{a} \mu\} \cup \{t \in S \mid \mu(t) > 0 \text{ with } s = (a, \mu)\}$. A *probabilistic automaton* (PA, [17]) is a game with $S_\Diamond = \emptyset$, and a *discrete-time Markov chain* (DTMC) is a PA with $|\mathsf{succ}(s)| = 1$ for all $s \in S$.

A finite (infinite) *path* $\lambda = s_0(a_0, \mu_0)s_1(a_1, \mu_1)s_2 \ldots$ is a finite (infinite) sequence of alternating states and moves, such that for all $i \geq 0$, $s_i \xrightarrow{a_i} \mu_i$ and $\mu_i(s_{i+1}) > 0$. A finite path λ ends in a state, denoted $\mathsf{last}(\lambda)$. A finite (infinite) *trace* is a finite (infinite) sequence of actions. Given a path, its trace is the sequence of actions along λ, with τ projected out. Formally, $\mathsf{trace}(\lambda) \stackrel{\text{def}}{=} \textsc{proj}_{\{\tau\}}(a_0 a_1 \ldots)$, where, for $\alpha \subseteq \mathcal{A} \cup \{\tau\}$, $\textsc{proj}_\alpha$ is the morphism defined by $\textsc{proj}_\alpha(a) = a$ if $a \notin \alpha$, and ϵ (the empty trace) otherwise.

Strategies. Nondeterminism for each player is resolved by a strategy, which maps finite paths to distributions over moves. For PAs, we do not speak of player strategies, and implicitly consider strategies of Player $\Box$. Here we use an alternative, equivalent formulation of strategies using stochastic memory update [5].

Definition 2. *A Player* $\Diamond$ strategy π *is a tuple* $\langle \mathfrak{M}, \pi_u, \pi_c, \alpha \rangle$*, where* $\mathfrak{M}$ *is a countable set of* memory elements*;* $\pi_u \colon \mathfrak{M} \times S \to \mathcal{D}(\mathfrak{M})$ *is a* memory update function*;* $\pi_c \colon S_\Diamond \times \mathfrak{M} \to \mathcal{D}(S)$ *is a* next move function *s.t.* $\pi_c(s, m)(t) > 0$ *only if* $t \in \mathsf{succ}(s)$*; and* $\alpha \colon S \to \mathcal{D}(\mathfrak{M})$ *defines for each state of* G *an* initial memory distribution. *A Player* $\Box$ *strategy* σ *is defined in an analogous manner.*

A strategy is *finite-memory* if $|\mathfrak{M}|$ is finite. Applying a strategy pair (π, σ) to a game G yields an *induced DTMC* $G^{\pi,\sigma}$ [8]; an induced DTMC contains only reachable states and moves, but retains the entire action alphabet of G.

Probability Measures and Expectations. The *cylinder set* of a finite path λ (resp. finite trace $w \in \mathcal{A}^*$) is the set of infinite paths (resp. traces) with prefix λ (resp. w). For a finite path $\lambda = s_0(a_0, \mu_0)s_1(a_1, \mu_1) \dots s_n$ in a DTMC D we define $\mathrm{Pr}_{D,s_0}(\lambda)$, the measure of its cylinder set, by $\mathrm{Pr}_{D,s_0}(\lambda) \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \mu_i(s_{i+1})$, and write $\mathrm{Pr}^{\pi,\sigma}_{G,s}$ for $\mathrm{Pr}_{G^{\pi,\sigma},s}$. For a finite trace w, $\mathsf{paths}(w)$ denotes the set of minimal finite paths with trace w, i.e. $\lambda \in \mathsf{paths}(w)$ if $\mathsf{trace}(\lambda) = w$ and there is no path $\lambda' \neq \lambda$ with $\mathsf{trace}(\lambda') = w$ and λ' being a prefix of λ. The measure of the cylinder set of w is $\tilde{\mathrm{Pr}}_{D,s}(w) \stackrel{\text{def}}{=} \sum_{\lambda \in \mathsf{paths}(w)} \mathrm{Pr}_{D,s}(\lambda)$, and we call $\tilde{\mathrm{Pr}}_{D,s}$ the *trace distribution* of D. The measures uniquely extend to infinite paths due to Carathéodory's extension theorem. We denote the set of infinite paths of D starting at s by $\Omega_{D,s}$. The *expectation* of a function $\boldsymbol{\rho} : \Omega_{D,s} \to \mathbb{R}^n_{\pm\infty}$ over infinite paths in a DTMC D is $\mathbb{E}_{D,s}[\boldsymbol{\rho}] \stackrel{\text{def}}{=} \int_{\lambda \in \Omega_{D,s}} \boldsymbol{\rho}(\lambda) d\mathrm{Pr}_{D,s}(\lambda)$.

Rewards. A *reward structure* (with n-dimensions) of a game is a partial function $r : \overline{S} \to \mathbb{R}$ ($\boldsymbol{r} : \overline{S} \to \mathbb{R}^n$). A reward structure r is *defined on actions* $\mathcal{A}_r$ if $r(a, \mu) = r(a, \mu')$ for all moves $(a, \mu), (a, \mu') \in S_\bigcirc$ such that $a \in \mathcal{A}_r$, and $r(s) = 0$ otherwise; and if the context is clear we consider it as a total function $r : \mathcal{A}_r \to \mathbb{R}$ for $\mathcal{A}_r \subseteq \mathcal{A}$. Given an n-dimensional reward structure $\boldsymbol{r} : \overline{S} \mapsto \mathbb{R}^n$, and a vector $\boldsymbol{v} \in \mathbb{R}^n$, define the reward structure $\boldsymbol{r} - \boldsymbol{v}$ by $[\boldsymbol{r} - \boldsymbol{v}]_s \stackrel{\text{def}}{=} \boldsymbol{r}(s) - \boldsymbol{v}$ for all $s \in \overline{S}$. For a path $\lambda = s_0 s_1 \dots$ and a reward structure r we define $\mathrm{rew}^N(r)(\lambda) \stackrel{\text{def}}{=} \sum_{i=0}^{N} r(s_i)$, for $N \geq 0$; the *average reward* is $\mathrm{mp}(r)(\lambda) \stackrel{\text{def}}{=} \liminf_{N \to \infty} \frac{1}{N+1} \mathrm{rew}^N(r)(\lambda)$; given a reward structure c such that, for all $s \in \overline{S}$, $c(s) \geq 0$ and, for all bottom strongly connected components (BSCCs) $\mathcal{B}$ of D, there is a state s in $\mathcal{B}$ such that $c(s) > 0$, the *ratio reward* is $\mathrm{ratio}(r/c)(w) \stackrel{\text{def}}{=} \liminf_{N \to \infty} \mathrm{rew}^N(r)(w)/(1 + \mathrm{rew}^N(c)(w))$. If D has finite state space, the lim inf of the above rewards can be replaced by the true limit in the expectation, as it is almost surely defined (see Appendix B.2). Further, the above rewards straightforwardly extend to multiple dimensions using vectors.

Specifications and Objectives. A *specification* φ is a predicate on path distributions, and we write $D \models \varphi$ if $\varphi(\mathrm{Pr}_{D,\varsigma_0})$ holds. We say that a Player $\Diamond$ strategy π *wins* for a specification φ in a game G, written $\pi \models \varphi$, if, for all Player $\Box$ strategies σ, $G^{\pi,\sigma} \models \varphi$, and say that φ is *achievable* if such a winning strategy exists. A specification φ is *defined on traces of* $\mathcal{A}$ if $\varphi(\tilde{\mathrm{Pr}}_{D,\varsigma_0}) = \varphi(\tilde{\mathrm{Pr}}_{D',\varsigma'_0})$ for all DTMCs D, D' such that $\tilde{\mathrm{Pr}}_{D,\varsigma_0}(w) = \tilde{\mathrm{Pr}}_{D',\varsigma'_0}(w)$ for all traces $w \in \mathcal{A}^*$.

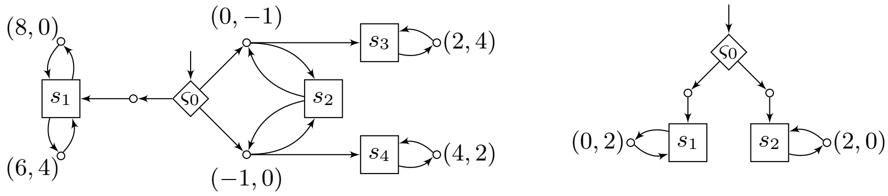


Fig. 1: Example games. Moves and states for Player $\Diamond$ and Player $\Box$ are shown as $\circ$, $\Diamond$ and $\Box$ resp.; two-dimensional rewards shown where non-zero.

A DTMC D satisfies an *expected energy* specification $\mathrm{EE}_s(\boldsymbol{r})$ if there exists $\boldsymbol{v}_0$ such that $\mathbb{E}_{D,s}[\mathrm{rew}^N(\boldsymbol{r})] \geq \boldsymbol{v}_0$ for all $N \geq 0$; D satisfies $\mathrm{EE}(\boldsymbol{r})$ if, for every state s of D, D satisfies $\mathrm{EE}_s(\boldsymbol{r})$. An *almost sure average* (resp. *ratio*) *reward objective* for target v is $\mathrm{Pmp}_s(\boldsymbol{r})(\boldsymbol{v}) \equiv \mathrm{Pr}_{D,s}(\mathrm{mp}(\boldsymbol{r}) \geq \boldsymbol{v}) = 1$ (resp. $\mathrm{Pratio}_s(\boldsymbol{r})(\boldsymbol{v}) \equiv \mathrm{Pr}_{D,s}(\mathrm{ratio}(\boldsymbol{r}/\boldsymbol{c}) \geq \boldsymbol{v}) = 1$). If the rewards $\boldsymbol{r}$ and $\boldsymbol{c}$ are understood, we omit them and write just $\mathrm{Pmp}_s(\boldsymbol{v})$ and $\mathrm{Pratio}_s(\boldsymbol{v})$. By using n-dimensional reward structures, we require that a strategy achieves the *conjunction* of the objectives defined on the individual dimensions. Minimisation is supported by inverting signs of rewards. Given an objective φ with target vector $\boldsymbol{v}$, denote by $\varphi[\boldsymbol{x}]$ the objective φ with $\boldsymbol{v}$ substituted by $\boldsymbol{x}$. A target $\boldsymbol{v} \in \mathbb{R}^n$ is a *Pareto vector* if $\varphi[\boldsymbol{v} - \varepsilon]$ is achievable for all $\varepsilon > 0$, and $\varphi[\boldsymbol{v} + \varepsilon]$ is not achievable for any $\varepsilon > 0$. The downward closure of the set of all such vectors is called a *Pareto set*.

Example. Consider the game in Figure 1 (left), showing a stochastic game with a two-dimensional reward structure. Player $\Diamond$ can achieve $\mathrm{Pmp}_{\varsigma_0}(3, 0)$ if going left at ς_0, and $\mathrm{Pmp}_{\varsigma_0}(1, 1)$ if choosing either move to the right, since then s_3 and s_4 are almost surely reached. Furthermore, achieving an expected mean-payoff does not guarantee achieving almost-sure satisfaction in general: the Player $\Diamond$ strategy going up right from ς_0 achieves an expected mean-payoff of at least $(1, 1.5)$, which by the above argument cannot be achieved almost surely. Also, synthesis in MDPs [5,20] can utilise the fact that the strategy controls reachability of end-components; e.g., if all states in the game of Figure 1 (left) are controlled by Player $\Diamond$, $(3, 2)$ is almost surely achievable.

3 Strategy Synthesis for Average Rewards

We consider the problem of computing ε-optimal strategies for almost sure average reward objectives $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{v})$. Note that, for any $\boldsymbol{v} \geq 0$, the objective $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v})$ is equivalent to $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r} - \boldsymbol{v})(\boldsymbol{0})$, i.e. with the rewards shifted by $-\boldsymbol{v}$. Hence, from now on we assume w.l.o.g. that the objectives have target $\boldsymbol{0}$.

3.1 Expected Energy Objectives

We show how synthesis for almost sure average reward objectives reduces to synthesis for expected energy objectives. Applying finite-memory strategies to games results in finite induced DTMCs. Infinite memory may be required for winning strategies of Player $\Diamond$ [5]; here we synthesise only finite-memory strategies for Player $\Diamond$, in which case only finite memory for Player $\Box$ is sufficient:

Lemma 1. *A finite-memory Player $\Diamond$ strategy is winning for the objective* $\mathrm{EE}(\boldsymbol{r})$ *(resp.* $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v})$*) if it wins against all finite-memory Player $\Box$ strategies.*

We now state our key reduction lemma to show that almost sure average reward objectives can be ε-approximated by considering EE objectives.

Lemma 2. *Given a finite-memory strategy π for Player $\Diamond$, the following hold:*

(i) if π satisfies $\mathrm{EE}(\boldsymbol{r})$*, then π satisfies* $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\mathbf{0})$*; and*
(ii) if π satisfies $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\mathbf{0})$*, then, for all $\varepsilon > 0$, π satisfies* $\mathrm{EE}(\boldsymbol{r}+\boldsymbol{\varepsilon})$*.*

Our method described in Theorem 2 below allows us to compute $\mathrm{EE}(\boldsymbol{r}+\boldsymbol{\varepsilon})$, and hence, by virtue of Lemma 2(*i*), derive ε-optimal strategies for $\mathrm{Pmp}_{\varsigma_0}(\mathbf{0})$. Item (*ii*) of Lemma 2 guarantees completeness of our method, in the sense that, for any vector $\boldsymbol{v}$ such that $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v})$ is achievable, we compute an ε-optimal strategy; however, if $\boldsymbol{v}$ is not achievable, our algorithm does not terminate.

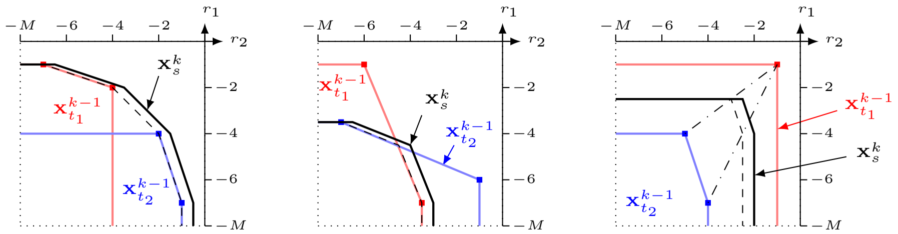
3.2 Strategy Construction

We define a value iteration method that in k iterations computes the sets X_s^k of shortfall vectors at state s, so that for any $\boldsymbol{v}_0 \in X_s^k$, Player $\Diamond$ can keep the expected energy above $\boldsymbol{v}_0$ during k steps of the game. Moreover, if successive sets X_s^{k+1} and X_s^k satisfy $X_s^k \sqsubseteq X_s^{k+1} + \varepsilon$, where $A \sqsubseteq B \Leftrightarrow \mathsf{dwc}(A) \subseteq \mathsf{dwc}(B)$, then we can construct a finite-memory strategy for $\mathrm{EE}(\boldsymbol{r}+\varepsilon)$ using Theorem 1.

Value Iteration. Let $\mathrm{Box}_M \stackrel{\text{def}}{=} [-M, 0]^n$. The *$M$-downward closure* of a set X is $\mathrm{Box}_M \cap \mathsf{dwc}(X)$. Let $\mathcal{P}_c^M(X)$ be the set of convex closed M-downward-closed subsets of X. Let $\mathcal{L}_M \stackrel{\text{def}}{=} (\mathcal{P}_c^M(\mathrm{Box}_M))^{|\overline{S}|}$, endow it with the partial order $X \subseteq Y \Leftrightarrow \forall s \in \overline{S} \,.\, X_s \subseteq Y_s$, and add the *top element* $\top \stackrel{\text{def}}{=} \mathrm{Box}_M^{|\overline{S}|}$. For a fixed M, define the operator $F_M : \mathcal{L}_M \to \mathcal{L}_M$ by $[F_M(X)]_s \stackrel{\text{def}}{=} \mathrm{Box}_M \cap \mathsf{dwc}(Y_s)$, where

$$Y_s \stackrel{\text{def}}{=} \boldsymbol{r}(s) + \begin{cases} \mathsf{conv}(\bigcup_{t \in \mathsf{succ}(s)} X_t) & \text{if } s \in S_\Diamond \\ \bigcap_{t \in \mathsf{succ}(s)} X_t & \text{if } s \in S_\Box \\ \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X_t & \text{if } s = (a, \mu) \in S_\bigcirc . \end{cases}$$

The operator F_M reflects what Player $\Diamond$ can achieve in the respective state types. In $s \in S_\Diamond$, Player $\Diamond$ can achieve the values in successors (union), and can randomise between them (convex hull). In $s \in S_\Box$, Player $\Diamond$ can achieve only values that are in all successors (intersection), since Player $\Box$ can pick arbitrarily. Lastly, in $s \in S_\bigcirc$, Player $\Diamond$ can achieve values with the prescribed distribution. F_M is closely related to our operator for expected total rewards in [7], but here we cut off values above zero with Box_M, similarly to the controllable predecessor operator of [6] for computing energy in non-stochastic games. Box_M ensures that the



(a) Player $\Diamond$ state s. (b) Player $\Box$ state s. (c) Move $s \in S_{\bigcirc}$.

Fig. 2: Value iteration and strategy construction, for state s with successors t_1, t_2, and reward $r_1(s) = 0.5$, $r_2(s) = 0$. The Pareto set under-approximation X_s^k is computed from $X_{t_1}^{k-1}$ and $X_{t_2}^{k-1}$. To achieve a point $\boldsymbol{p} \in C_s^k$, the strategy updates its memory as follows: for $s \in S_\Box$, for all $t \in \mathsf{succ}(s)$, $\boldsymbol{p} - \boldsymbol{r}(s) \in \mathsf{conv}(C_t^{k-1})$; for $s \in S_\Diamond \cup S_\bigcirc$, there exist successors $t \in \mathsf{succ}(s)$ and a distribution α s.t. $\boldsymbol{p} - \boldsymbol{r}(s) \in \sum_t \alpha(t) \times \mathsf{conv}(C_t^k)$, where, for $s = (a, \mu) \in S_\bigcirc$, we fix $\alpha = \mu$. As F is order preserving, it is sufficient to use X_t^l instead of X_t^k for any $l \geq k$.

strategy we construct in Theorem 1 below never allows the energy to diverge in any reachable state. For example, in Figure 1 (right), for $\boldsymbol{v} = (\frac{1}{2}, \frac{1}{2})$, $\mathrm{EE}_{\varsigma_0}(\boldsymbol{r} - \boldsymbol{v})$ is achievable while, for the states $s \in \{s_1, s_2\}$, $\mathrm{EE}_s(\boldsymbol{r} - \boldsymbol{v})$ is not. Since one of s_1 or s_2 must be reached, $\mathrm{EE}(\boldsymbol{r} - \boldsymbol{v})$ is not achievable, disallowing the use of Lemma 2(i); and indeed, $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{v})$ is not achievable. Bounding with M allows us to use a geometric argument in Lemma 3 below, replacing the finite lattice arguments of [6], since our theory is more involved as it reflects the continuous essence of randomisation.

We show in the following proposition that F_M defines a monotonic fixpoint computation and that it converges to the greatest fixpoint of F_M. Its proof relies on Scott-continuity of F_M, and invokes the Kleene fixpoint theorem.

Proposition 1. *F_M is order-preserving, $\top \supseteq F_M(\top) \supseteq F_M^2(\top) \supseteq \cdots$, and the greatest fixpoint $fix(F_M)$ exists and is equal to $\lim_{k\to\infty} F_M^k(\top) = \cap_{k\geq 0} F_M^k(\top)$.*

Further, we use F_M to compute the set of shortfall vectors required for Player $\Diamond$ to win for $\mathrm{EE}_s(\boldsymbol{r})$ via a value iteration with relative stopping criterion defined using ε, see Lemma 3 below. Denote $X^k \stackrel{\text{def}}{=} F_M^k(\top)$. The value iteration is illustrated in Figure 2: at iteration k, the set X_s^k of possible shortfalls until k steps is computed from the corresponding sets X_t^{k-1} for successors $t \in \mathsf{succ}(s)$ of s at iteration $k-1$. The values are restricted to be within Box_M, so that obtaining an empty set at a state s in the value iteration is an indicator of divergence at s. Any state that must be avoided by Player $\Diamond$ yields an empty set. For instance, in Figure 1 (left), with target $(1, 1)$ the value iteration diverges at s_1 for any $M \geq 0$, but at ς_0, Player $\Diamond$ can go to the right to avoid accessing s_1. The following proposition ensures completeness of our method, stated in Theorem 2 below.

Proposition 2. *If $\mathrm{EE}(\boldsymbol{r})$ is achievable then $[fix(F_M)]_{\varsigma_0} \neq \emptyset$ for some $M \geq 0$.*

Proof (Sketch). First, we consider the expected enrgy of finite DTMCs, where, at every step, we cut off the positive values. This entails that the sequence of the resulting truncated non-positive expected energies decreases and converges toward a limit vector $\boldsymbol{u}$ whose coordinates are finite if $\mathrm{EE}(\boldsymbol{r})$ is satisfied. We show that, when $\mathrm{EE}(\boldsymbol{r})$ is satisfied by a strategy π, there is a global lower bound $-M$ on every coordinate of the limit vector $\boldsymbol{u}$ for the DTMC $G^{\pi,\sigma}$ induced by any Player $\Box$ strategy σ. We show that, for this choice of M, the fixpoint of F_M for the game G is non-empty in every state reachable under π. We conclude that $[fix(F_M)]_{\varsigma_0} \neq \emptyset$ for some $M \geq 0$ whenever $\mathrm{EE}(\boldsymbol{r})$ is achievable.

Lemma 3. *Given M and ε, for every non-increasing sequence (X^i) of elements of $\mathcal{L}_M$ there exists $k \leq k^{**} \stackrel{\text{def}}{=} \left[2n((\lceil \frac{M}{\varepsilon} \rceil + 2)^2 + 2)\right]^{|\overline{S}|}$ such that $X^k \sqsubseteq X^{k+1} + \varepsilon$.*

Proof (Sketch). We first consider a single state s, and construct a graph with vertices from the sequence of sets (X^i), and edges indicating dimensions where the distance is at least ε. Interpreting each dimension as a colour, we use a Ramseyan argument to find the bound $k^* \stackrel{\text{def}}{=} n \cdot ((\lceil \frac{M}{\varepsilon} \rceil + 2)^2 + 2)$ for a single state. To find the bound $k^{**} \stackrel{\text{def}}{=} (2k^*)^{|\overline{S}|}$, which is for *all* states, we extract successive subsequences of $\{1, 2, \ldots, k^{**}\} \stackrel{\text{def}}{=} I_0 \supseteq I_1 \supseteq \cdots \supseteq I_{|\overline{S}|}$, where going from I_i to I_{i+1} means that one additional state has the desired property, and such that the invariant $|I_{i+1}| \geq |I_i|/(2k^*)$ is satisfied. At the end $I_{|\overline{S}|}$ contains at least one index $k \leq k^{**}$ for which all states have the desired property.

Strategy Construction. The strategies are constructed so that their memory corresponds to the extreme points of the sets computed by $F_M^k(\top)$. The strategies stochastically update their memory, and so the expectation of their memory elements corresponds to an expectation over such extreme points.

Let C_s^k be the set of *extreme points* of $\mathsf{dwc}(X_s^k)$, for all $k \geq 0$ (since $X^k \in \mathcal{L}_M$, the sets X_s^k are closed). For any point $\boldsymbol{p} \in X_s^k$, there is some $\boldsymbol{q} \geq \boldsymbol{p}$ that can be obtained by a convex combination of points in C_s^k, and so the strategy we construct uses C_s^k as memory, randomising to attain the convex combination $\boldsymbol{q}$. Note that the sets C_s^k are finite, yielding finite-memory strategies.

If $X_{\varsigma_0}^{k+1} \neq \emptyset$ and $X^k \sqsubseteq X^{k+1} + \varepsilon$ for some $k \in \mathbb{N}$ and $\varepsilon \geq 0$, we can construct a Player $\Diamond$ strategy π for $\mathrm{EE}(\boldsymbol{r} + \varepsilon)$. Denote by $\overline{T} \subseteq \overline{S}$ the set of states s for which $X_s^{k+1} \neq \emptyset$. For $l \geq 1$, define the *standard l-simplex* by $\Delta^l \stackrel{\text{def}}{=} \{B \in [0,1]^l \mid \sum_{\beta \in B} \beta = 1\}$. The memory $\mathfrak{M} \stackrel{\text{def}}{=} \bigcup_{s \in \overline{T}} \{(s, \boldsymbol{p}) \mid \boldsymbol{p} \in C_s^k\}$ is initialised according to α, defined by $\alpha(s) \stackrel{\text{def}}{=} [(s, \boldsymbol{q}_0^s) \mapsto \beta_0^s, \ldots, (s, \boldsymbol{q}_n^s) \mapsto \beta_n^s]$, where $\boldsymbol{\beta}^s \in \Delta^n$, and, for all $1 \leq i \leq n$, $\boldsymbol{q}_i^s \in C_s^k$. The update π_u and next move function π_c are defined as follows: at state s with memory $(s, \boldsymbol{p})$, for all $t \in \mathsf{succ}(s)$, pick n vectors $\boldsymbol{q}_i^t \in C_t^k$ for $1 \leq i \leq n$, with coefficients $\boldsymbol{\beta}^t \in \Delta^n$, such that

- **for** $s \in S_\Diamond$**,** there is $\boldsymbol{\gamma} \in \Delta^{|\mathsf{succ}(s) \cap \overline{T}|}$, such that $\sum_t \gamma_t \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p} - \boldsymbol{r}(s) - \varepsilon$;
- **for** $s \in S_\Box$**,** for all $t \in \mathsf{succ}(s)$, $\sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p} - \boldsymbol{r}(s) - \varepsilon$; and
- **for** $s = (a, \mu) \in S_\bigcirc$, we have $\sum_{t \in \mathsf{supp}(\mu)} \mu(t) \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p} - \boldsymbol{r}(s) - \varepsilon$;

Algorithm 1 PMP Strategy Synthesis

1: **function** $\textsc{SynthPMP}(G,\ \boldsymbol{r},\ \boldsymbol{v},\ \varepsilon)$
2: Set the reward structure to $\boldsymbol{r} - \boldsymbol{v} + \frac{\varepsilon}{2}$; let $k \leftarrow 0$; $M \leftarrow 2$; $X^0 \leftarrow \top$;
3: **while** true **do**
4: **while** $X^k \not\sqsubseteq X^{k+1} + \frac{\varepsilon}{2}$ **do**
5: $k \leftarrow k + 1$; $X^{k+1} \leftarrow F_M(X^k)$;
6: **if** $X^k_{\varsigma_0} \neq \emptyset$ **then**
7: Construct π for $\frac{\varepsilon}{2}$ and any $\boldsymbol{v}_0 \in C^k_{\varsigma_0}$ using Theorem 1; **return** π
8: **else**
9: $k \leftarrow 0$; $M \leftarrow M^2$;

and, for all $t \in \mathsf{succ}(s)$, let $\pi_u((s, \boldsymbol{p}), t)(t, \boldsymbol{q}^t_i) \stackrel{\text{def}}{=} \beta^t_i$ for all i, and $\pi_c(s, (s, \boldsymbol{p}))(t) \stackrel{\text{def}}{=} \gamma_t$ if $s \in S_\Diamond$.

Theorem 1. *If $X^{k+1}_{\varsigma_0} \neq \emptyset$ and $X^k \sqsubseteq X^{k+1} + \varepsilon$ for some $k \in \mathbb{N}$ and $\varepsilon \geq 0$, then the Player $\Diamond$ strategy constructed above is finite-memory and wins for $\mathrm{EE}(\boldsymbol{r} + \boldsymbol{\varepsilon})$.*

Proof (Sketch). We show the strategy is well-defined, i.e. the relevant extreme points and coefficients exist, which is a consequence of $X^k \sqsubseteq X^{k+1} + \varepsilon$. We then show that, when entering a state s_o with a memory $\boldsymbol{p}_o$, the expected memory from this state after N steps is above $\boldsymbol{p}_o - \mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r})] - N\boldsymbol{\varepsilon}$. As the memory is always non-positive, this implies that $\mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r} + \boldsymbol{\varepsilon})] \geq \boldsymbol{p}_o \geq -M$ for every state s_o with memory $\boldsymbol{p}_o$, for every N. We conclude that $\mathrm{EE}(\boldsymbol{r} + \boldsymbol{\varepsilon})$ holds.

3.3 Strategy Synthesis Algorithm

Given a game G, a reward structure $\boldsymbol{r}$ with target vector $\boldsymbol{v}$, and $\varepsilon > 0$, the semi-algorithm given in Algorithm 1 computes a strategy winning for $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v} - \varepsilon)$.

Theorem 2. *Whenever $\boldsymbol{v}$ is in the Pareto set of $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})$, then Algorithm 1 terminates with a finite-memory ε-optimal strategy.*

Proof (Sketch). Since $\boldsymbol{v}$ is in the Pareto set of the almost sure average reward objective, by Lemma 2(ii) the objective $\mathrm{EE}(\boldsymbol{r} - \boldsymbol{v} + \frac{\boldsymbol{\varepsilon}}{2})$ is achievable, and, by Proposition 2, there exists an M such that $fix(F_M)$ is nonempty. The condition in Line 6 is then satisfied as $\emptyset \neq [fix(F_M)]_{\varsigma_0} \subseteq X^k_{\varsigma_0}$. Further, due to the bound M on the size of the box Box_M in the value iteration, the inner loop terminates after a finite number of steps, as shown in Lemma 3. Then, by Theorem 1, the strategy constructed in Line 7 (with degradation factor $\frac{\varepsilon}{2}$ for the reward $\boldsymbol{r} - \boldsymbol{v} + \frac{\boldsymbol{\varepsilon}}{2}$) satisfies $\mathrm{EE}(\boldsymbol{r} - \boldsymbol{v} + \boldsymbol{\varepsilon})$, and hence, using Lemma 2(i), $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v} - \boldsymbol{\varepsilon})$.

4 Compositional Synthesis

In order to synthesise strategies compositionally, we introduced in [4] a composition of games, and showed that assume-guarantee rules for PAs can be applied in synthesis for games: whenever there is a PA verification rule, the corresponding game synthesis rule has the same form and side-conditions (Theorem 1 of [4]). We present a PA assume-guarantee rule for ratio rewards. The PA rules in [13] only support total expected rewards, while our rule works with any specification defined on traces, and in particular with ratio rewards (Proposition 4).

Ratio Rewards. Ratio rewards $\mathrm{ratio}(\boldsymbol{r}/\boldsymbol{c})$ generalise average rewards $\mathrm{mp}(\boldsymbol{r})$, since, to express the latter, we let $\boldsymbol{c}(s) = 1$ for all $s \in \overline{S}$. The following proposition states that to solve $\mathrm{Pratio}_{\leqslant 0}(\boldsymbol{r}/\boldsymbol{c})(\boldsymbol{v})$ it suffices to solve $\mathrm{Pmp}_{\leqslant 0}(\boldsymbol{r})(\boldsymbol{v} \bullet \boldsymbol{c})$.

Proposition 3. *A finite-memory Player $\Diamond$ strategy π satisfies* $\mathrm{Pratio}_{\leqslant 0}(\boldsymbol{r}/\boldsymbol{c})(\boldsymbol{v})$ *if and only if it satisfies* $\mathrm{Pmp}_{\leqslant 0}(\boldsymbol{r})(\boldsymbol{v} \bullet \boldsymbol{c})$.

Fairness. Given a composed PA $\mathcal{M} = \ \|_{i \in I} M^i$, a strategy σ is *fair* if at least one action of each component $\mathcal{M}_i$ is chosen infinitely often with probability 1. We write $\mathcal{M} \models^f \varphi$ if, for all fair strategies σ, $\mathcal{M}^\sigma \models \varphi$.

Theorem 3. *Given compatible PAs $\mathcal{M}_1$ and $\mathcal{M}_2$, specifications φ^{G_1} and φ^{G_2} defined on traces of $\mathcal{A}_{G_i} \subseteq \mathcal{A}_i$ for $i \in \{1, 2\}$, then the following is sound:*

$$\frac{\mathcal{M}_1 \models^f \varphi^{G_1} \qquad \mathcal{M}_2 \models^f \varphi^{G_2}}{\mathcal{M}_1 \parallel \mathcal{M}_2 \models^f \varphi^{G_1} \wedge \varphi^{G_2}}.$$

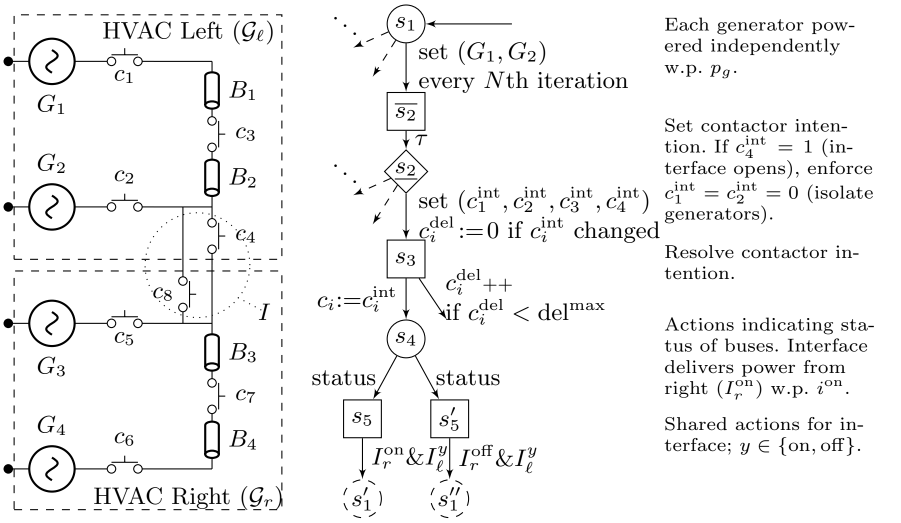
To use Theorem 3, we show that objectives using total or ratio rewards are defined on traces over some subset of actions.

Proposition 4. *If n-dimensional reward structures $\boldsymbol{r}$ and $\boldsymbol{c}$ are defined on actions $\mathcal{A}_{\boldsymbol{r}}$ and $\mathcal{A}_{\boldsymbol{c}}$, respectively, then objectives using ratio rewards* $\mathrm{ratio}(\boldsymbol{r}/\boldsymbol{c})$ *are defined on traces of $\mathcal{A}_{\boldsymbol{r}} \cup \mathcal{A}_{\boldsymbol{c}}$.*

Note that average rewards are not defined over traces in general, since its divisor counts the transitions, irrespective of whether the specification takes them into account. In particular, when composing systems, the additional transitions in between those originally counted skew the value of the average rewards. Moreover, τ-transitions are counted, but do not appear in the traces.

5 A Case Study: Aircraft Power Distribution

We demonstrate our synthesis methods on a case study for the control of the electrical power system of a more-electric aircraft [15], see Figure 3(a). Power is to be routed from generators to buses (and loads attached to them) by controlling the contactors (i.e. controllable switches) connecting the network nodes. Our models are based on a game-theoretic study of the same control problem in [21], where the control objective is to ensure the buses are powered, while avoiding unsafe configurations. The controllers have to take into account that contactors have delays, and the generators available in the system may be reconfigured, or even exhibit failures. We show that, by incorporating stochasticity in the models derived from the reliability statistics of the generators, controllers synthesised from ratio rewards achieve better uptimes compared to those reported in [21].



(a) Single-line diagram. (b) HVAC Left ($\mathcal{G}_\ell$).

Fig. 3: Aircraft electric power system, adapted from a Honeywell, Inc. patent [15]. The single-line diagram of the full power system (a) shows how power from the generators (G_i) can be routed to the buses (B_i) through the contactors (c_i). The left HVAC subsystem model $\mathcal{G}_\ell$ is shown in (b), and $\mathcal{G}_r$ is symmetric. I_ℓ^x and I_r^y is the interface status on the left and right side, resp., where x, y stand for either "on" or "off". One iteration of the reactive loop goes from s_1 to s_5 and starts again at s_1, potentially with some variables changed, indicated as s_1' or s_1''.

5.1 Model

The system comprises several components, each consisting of buses and generators, and we consider the high-voltage AC (HVAC) subsystem, shown in Figure 3(a), where the dashed boxes represent the components set out in [15]. These components are physically separated for reliability, and hence allow limited interaction and communication. Since the system is reactive, i.e. the aircraft is to be controlled continually, we use long-run properties to specify correctness.

The game models and control objectives in [21] are specified using LTL properties. We extend their models to stochastic games with quantitative specifications, where the contactors are controlled by Player $\Diamond$ and the contactor dynamics and the interfaces are controlled by Player $\Box$, and compose them by means of the synchronising parallel composition of [4]. The advantage of stochasticity is that the reliability specifications desired in [21] can be faithfully encoded. Further, games allow us to model truly adversarial behaviour (e.g. uncontrollable contactor dynamics), as well as nondeterministic interleaving in the composition.

Contactors, Buses and Generators. We derive the models based on the LTL description of [21]: the status of the buses and generators are kept in

Boolean variables $B_1, \ldots, B_4$ and $G_1, \ldots, G_4$ resp., and their truth value represents whether the bus or generator is powered; the contactor status is kept in Boolean variables $c_1, \ldots, c_8$, and their truth value represents if the corresponding contactor lets the current flow. For instance, if in $\mathcal{G}_\ell$ the generator G_1 is on but G_2 is off, the controller needs to switch the contactors c_1 and c_3 on, in order to power both buses B_1 and B_2. At the same time, short circuits from connecting generators to each other must be avoided, e.g. contactors c_1, c_2 and c_3 cannot be on at the same time, as this configuration connects G_1 and G_2. The contactors are, for example, solid state power controllers [19], which typically have non-negligible reaction times with respect to the times the buses should be powered. Hence, as in [21], we model that Player $\Diamond$ can only set the *intent* c_i^{int} of contactor i, and only after some delay is the contactor status c_i set to this intent. For the purposes of this demonstration, we only model a delayed turn-off time, as it is typically larger than the turn-on time (e.g. 40 ms, the turn-off time reported in [9]). Whether or not a contactor is delayed is controlled by Player $\Box$.

Interface. The components can deliver power to each other via the interface I, see Figure 3(a), which is bidirectional, i.e. power can flow both ways. The original design in [15] does not include connector c_8, and so c_4 has to ensure that no short circuits occur over the interface: if B_3 is powered, c_4 may only connect if B_2 is unpowered, and vice versa; hence, c_4 can only be on if both B_2 and B_3 are unpowered. By adding c_8, we break this cyclic dependence.

Actions shared between components model transmission of power. The actions I_r^x and I_ℓ^y for $x, y \in \{\text{on}, \text{off}\}$ model whether power is delivered via the interface from the right or left, respectively, or not. Hence, power flows from left to right via c_8, and from right to left via c_4; and we ensure via the contactors that power cannot flow in the other direction, preventing short circuits.

Reactive Loop. We model each component as an infinite loop of Player $\Box$ and Player $\Diamond$ actions. One iteration of the loop, called *time step*, represents one time unit T, and the system steps through several stages, corresponding to the states in $\mathcal{G}_\ell$ (and $\mathcal{G}_r$): in s_1 the status of the generators is set every Nth time step; in $\underline{s_2}$ the controller sets the contactors; in s_3 the delay is chosen nondeterministically; in s_4 actions specify whether both buses are powered, and whether a failure occurs; and in s_5 information is transmitted over the interface. The τ-labelled Dirac transitions precede all Player $\Diamond$ states to enable composition [4].

Generator Assumptions. We assume that the generator status remains the same for N time steps, i.e. after 0, N, $2N$, ... steps the status may change, with the generators each powered with probability p_g, independently from each other. N and p_g can be obtained from the mean-time-to-failure of the generators. This is in contrast to [21], where, due to non-probabilistic modelling, the strongest assumption is that generators do not fail at the same time.

5.2 Specifications and Results

The main objective is to maximise uptime of the buses, while avoiding failures due to short circuits, as in [21]. Hence, the controller has to react to the gener-

Table 1: Performance statistics, for various choices of b (bus uptime), f (failure rate), i^{on} (interface uptime), and model and algorithm parameters. A minus (−) for i^{on} means the interface is not used. The Pareto and Strategy columns show the times for EE Pareto set computation and strategy construction, respectively.

Target			Model Params.				Algorithm Params.		Runtime [s]	
b	f	i^{on}	N	$\text{del}^{\max}$	p_g	$\lvert S\rvert$	ε	k	Pareto	Strategy
0.90	0.01	−	0	0	0.8	1152	0.001	20	25	0.29
0.85	0.01	−	3	1	0.8	15200	0.001	65	1100	2.9
0.90	0.01	−	3	1	0.8	15200	0.001	118	2100	2.1
0.90	0.01	0.6	0	0	0.8	2432	0.01	15	52	0.53
0.95	0.01	0.6	0	0	0.8	2432	0.01	15	49	0.46
0.90	0.01	0.6	2	1	0.8	24744	0.01	80	4300	4.80

ator status, and cannot just leave all contactors connected. The properties are specified as ratio rewards, since we are interested in the proportion of time the buses are powered. To use Theorem 3, we attach all rewards to the status actions or the synchronised actions I_ℓ^x and I_r^y. Moreover, every time step, the reward structure t attaches T to these actions to measure the progress of time.

The reward structure "buses$_\ell$" (resp. "buses$_r$") assigns T for each time unit both buses of $\mathcal{G}_\ell$ (resp. $\mathcal{G}_r$) are powered; and the reward structure "fail$_\ell$" (resp. "fail$_r$") assigns 1 for every time unit a short circuit occurs in $\mathcal{G}_\ell$ (resp. $\mathcal{G}_r$). Since the synchronised actions I_r^{on} and I_ℓ^{on} are taken whenever power is delivered over the interface, we attach reward structures, with the same name, assigning T whenever the corresponding action is taken. For each component $x \in \{\ell, r\}$, the objectives are to keep the uptime of the buses above b, i.e. $P_x^{\text{bus}} \equiv \text{Pratio}_{\leqslant 0}(\text{buses}_x/t)(b)$; to keep the failure rate below f, i.e. $P_x^{\text{safe}} \equiv \text{Pratio}_{\leqslant 0}(-\text{fail}_x/t)(-f)$, where minimisation is expressed using negation; and, if used, to keep the interface uptime above i^{on}, i.e. $P_x^{\text{int}} \equiv \text{Pratio}_{\leqslant 0}(I_x^{\text{on}}/t)(i^{\text{on}})$. We hence consider the specification $P_x^{\text{bus}} \wedge P_x^{\text{safe}} \wedge P_x^{\text{int}}$, for $x \in \{\ell, r\}$. Using the rule from Theorem 3 in Theorem 1 of [4], we obtain the strategy composed of the individual strategies to control the full system, satisfying $P_\ell^{\text{bus}} \wedge P_\ell^{\text{safe}} \wedge P_r^{\text{bus}} \wedge P_r^{\text{safe}}$, i.e. both components are safe and the buses are powered.

Strategy Synthesis. We implement the algorithms of this paper as an extension of our multi-objective strategy synthesis tool of [8], using a compact representation of the polyhedra $F_M^k(\top)$. Table 1 shows, for several parameter choices, the experimental results, which were obtained on a 2.8 GHz PC with 32 GB RAM. In [21], the uptime objective was encoded in LTL by requiring that buses are powered at least every Kth time step, yielding an uptime for the buses of $1/K$, which translates to an uptime of 20% (by letting $K = 5$). In contrast, using stochastic games we can utilise the statistics of the generator reliability, and obtain bus uptimes of up to 95% for generator health $p_g = 0.8$. For the models without delay, the synthesised strategies approximate memoryless deterministic strategies but when adding delay, randomisation is introduced in the memory updates. The model will be included in a forthcoming release of our tool.

6 Conclusion

We synthesise strategies for almost sure satisfaction of multi-dimensional average and ratio objectives, and demonstrate their application to assume-guarantee controller synthesis. It would be interesting to study the complexity class of the problem considered here. Satisfaction for arbitrary thresholds is subject to further research. Solutions involving an oracle computing the almost-sure winning region [12] would need to be adapted to handle our ε-approximations. Moreover, we are interested in strategies for disjunctions of satisfaction objectives.

Acknowledgements. Part of this work was sponsored by ERC AdG-246967 VERIWARE, and AFOSR grant FA9550-12-1-0302, ONR grant N000141310778.

References

1. C. Baier, C. Dubslaff, S. Klüppelholz, and L. Leuschner. Energy-utility analysis for resilient systems using probabilistic model checking. In *PETRI NETS*, vol. 8489 of LNCS, pages 20–39. Springer, 2014.
2. C. Baier and J.P. Katoen. *Principles of Model Checking.* MIT press, 2008.
3. N. Basset, M. Kwiatkowska, U. Topcu, and C. Wiltsche. Strategy synthesis for stochastic games with multiple long-run objectives. Technical Report RR-14-10, University of Oxford, 2014.
4. N. Basset, M. Kwiatkowska, and C. Wiltsche. Compositional controller synthesis for stochastic games. In *CONCUR*, vol. 8704 of LNCS, pages 173–187. Springer, 2014.
5. T. Brázdil, V. Brozek, K. Chatterjee, V. Forejt, and A. Kucera. Two views on multiple mean-payoff objectives in Markov decision processes. *LMCS*, 10(1), 2014.
6. K. Chatterjee, M. Randour, and J.F. Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Inf.*, 51(3–4):129–163, 2014.
7. T. Chen, V. Forejt, M. Kwiatkowska, A. Simaitis, and C. Wiltsche. On stochastic games with multiple objectives. In *MFCS*, vol. 8087 of LNCS, pages 266–277. Springer, 2013.
8. T. Chen, M. Kwiatkowska, A. Simaitis, and C. Wiltsche. Synthesis for multi-objective stochastic games: An application to autonomous urban driving. In *QEST*, vol. 8054 of LNCS, pages 322–337. Springer, 2013.
9. Automation Direct. Part number AD-SSR610-AC-280A, Relays and Timers, Book 2 (14.1), eRL-45, 2014.
10. M. Droste, W. Kuich, and H. Vogler. *Handbook of weighted automata.* Springer, 2009.
11. H. Gimbert. Pure stationary optimal strategies in Markov decision processes. In *STACS*, vol. 4393 of LNCS, pages 200–211. Springer, 2007.
12. H. Gimbert and F. Horn. Solving simple stochastic tail games. In *SODA*, pages 847–862. SIAM, 2010.
13. M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Compositional probabilistic verification through multi-objective model checking. *I&C*, 232:38–65, 2013.
14. D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov chains and mixing times.* American Mathematical Soc., 2009.
15. R.G. Michalko. Electrical starting, generation, conversion and distribution system architecture for a more electric vehicle, 2008. US Patent 7,439,634.

16. O. Sankur. *Robustness in Timed Automata: Analysis, Synthesis, Implementation.* Thèse de doctorat, Laboratoire Spécification et Vérification, ENS Cachan, France, 2013.
17. R. Segala. *Modelling and Verification of Randomized Distributed Real Time Systems.* PhD thesis, Massachusetts Institute of Technology, 1995.
18. Lloyd S Shapley. Stochastic games. *Proc. Natl. Acad. Sci. USA*, 39(10):1095, 1953.
19. M. Sinnett. 787 no-bleed systems: saving fuel and enhancing operational efficiencies. *Aero Quarterly*, pages 6–11, 2007.
20. C. von Essen. *Quantitative Verification and Synthesis.* PhD thesis, VERIMAG, 2014.
21. H. Xu, U. Topcu, and R.M. Murray. Reactive protocols for aircraft electric power distribution. In *CDC*. IEEE, 2012.

A Note on Induced PAs and DTMCs

In the following proofs we use the definition from [4] for the PAs and DTMCs induced by strategy application.

Firstly, note that a DTMC according to the definition in [4] is also a PA, but with only one move per state, which allows us to label the transitions with actions. Considering a state s together with its unique move (a, μ) as a state (s, a, μ) in itself, allows us to view a DTMC induced according to [4] as a DTMC according to the classical definition, which contains only distributions without action labels. We apply this transformation implicitly in the proofs below.

Secondly, induced PAs and DTMCs according to the definition in [4] are always infinite objects, irrespective of the memory size of the applied strategies. When dealing with finite-memory DU strategies, however, we can fold the paths of the induced PAs and DTMCs to obtain finite objects. This folding is not possible in general for SU strategies, but using the construction in [8] we can always obtain a finite induced DTMC $G^{\pi,\sigma}$ from a finite-memory SU strategy pair (π, σ).

B Energy Reduction

B.1 Proof of Lemma 1

Proof. Fix a game G, and a finite-memory Player $\Diamond$ strategy π. We show the contrapositive: we suppose that there is an infinite-memory strategy σ of Player $\Box$ such that $G^{\pi,\sigma}$ does not satisfy $\mathrm{EE}(\boldsymbol{r})$ (resp. $\mathrm{Pmp}_s(\boldsymbol{r})(\boldsymbol{v})$) and show that a finite-memory strategy for Player $\Box$ exists, also falsifying the property $\mathrm{EE}_s(\boldsymbol{r})$ (resp. $\mathrm{Pmp}_s(\boldsymbol{r})(\boldsymbol{v})$).

For $\mathrm{EE}(\boldsymbol{r})$**:** By assumption (of the contrapositive argument) there exists a state s and $N \geq 0$ such that $\mathbb{E}_{G^{\pi,\sigma},s}[rew^N(\boldsymbol{r})] < \boldsymbol{v}_0$. Now consider a finite-memory strategy $\tilde{\sigma}$ that agrees with σ on paths of length less than or equal to N, that is $\tilde{\sigma}(\lambda) = \sigma(\lambda)$ for $|\lambda| \leq N$ and such that s appears as a state of $G^{\pi,\tilde{\sigma}}$. Then, it holds that $\mathbb{E}_{G^{\pi,\sigma},s}[rew^N(\boldsymbol{r})] = \mathbb{E}_{G^{\pi,\tilde{\sigma}},s}[rew^N(\boldsymbol{r})] < \boldsymbol{v}_0$, and thus we are done, as the finite memory strategy $\tilde{\sigma}$ is such that $G^{\pi,\tilde{\sigma}}$ does not satisfy $\mathrm{EE}(\boldsymbol{r})$.

For $\mathrm{Pmp}_s(\boldsymbol{r})(\boldsymbol{v})$**:** If $G^{\pi,\sigma} \not\models \mathrm{Pmp}_s(\boldsymbol{r})(\boldsymbol{v})$, then there exists an index i such that $\mathrm{Pr}^{\pi,\sigma}_{G,s}(\mathrm{mp}(r_i) \geq v_i) < 1$, which is a single mean-payoff objective for Player $\Box$ in the induced PA G^{π}. By Proposition 1 of [11], mean-payoff specifications are "submixing", and by Theorem 1 of the same reference, for such submixing specifications finite-memory (and even memoryless) strategies suffice to win. □

B.2 Proof of Lemma 2

We first show an intermediate result describing the asymptotic behaviour of average rewards on paths of a DTMC. We denote by $\mu_{\mathcal{B}}$ the *stationary distribution* of a BSCC $\mathcal{B}$, denote by $P_{D,s}(\Diamond\mathcal{B})$ the probability to eventually reach $\mathcal{B}$ from s; and define $\mathrm{rew}(\boldsymbol{r})(\mathcal{B}) \stackrel{\text{def}}{=} \sum_{s\in\mathcal{B}} \boldsymbol{r}(s)\mu_{\mathcal{B}}(s)$, where $\boldsymbol{r}$ is the reward structure on the DTMC (this is the mean-payoff in the BSCC $\mathcal{B}$, which is the same at every state).

Lemma 4. *Given a finite-state DTMC D and a state $s \in S_D$, and a reward structure $\boldsymbol{r}$, the limit $\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda)$ almost surely exists for $\lambda \in \Omega_{D,s}$, and takes values $\boldsymbol{x}$ in the finite set $\{\mathrm{rew}(\boldsymbol{r})(\mathcal{B}) \mid \mathcal{B}$ is a BSCC of $D\}$ with probability*

$$\sum_{\mathcal{B} \text{ s.t. } \mathrm{rew}(\boldsymbol{r})(\mathcal{B})=\boldsymbol{x}} \mathrm{Pr}_{D,s}(\Diamond\mathcal{B}).$$

Consequently, $D \models \mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{v})$ if and only if $\mathrm{rew}(\boldsymbol{r})(\mathcal{B}) \geq \boldsymbol{v}$ for every BSCC $\mathcal{B}$ that is reached with positive probability.

Proof (Lemma 4). Note first that for every path λ, $\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda)$ converges if and only if for every suffix λ' of λ, $\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda')$ converges to the same limit. For every recurrent state t of $\mathcal{B}$, we denote by W_t the set of paths λ such that t is the first recurrent state along λ.

Paths $\lambda \in W_t$ have suffixes λ' distributed according to $\mathrm{Pr}_{D,t}$. Due to a classical ergodic theorem for irreducible Markov chains (see e.g. Theorem 4.16 of [14]), $\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda')$ almost surely converges to $\sum_{t'\in\mathcal{B}} \mu_{\mathcal{B}}(t')r(t')$. Thus, with probability $\mathrm{Pr}_{D,s}(\Diamond\mathcal{B}) = \sum_{t\in\mathcal{B}} P_{D,s}(W_t)$, the sequence $\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda)$ converges to $\mathrm{rew}(\boldsymbol{r})(\mathcal{B})$. To conclude, it suffices to recall that $\sum_{\mathcal{B}\in\mathrm{BSCCs}} P_{D,s}(\Diamond\mathcal{B}) = 1$ and thus the result holds almost surely. □

Proof (Lemma 2). Instead of proving $\forall\sigma\,.\,G^{\pi,\sigma} \models \psi \Rightarrow \forall\sigma\,.\,G^{\pi,\sigma} \models \varphi$, we prove the stronger statement $\forall\sigma\,.\,(G^{\pi,\sigma} \models \psi \Rightarrow G^{\pi,\sigma} \models \varphi)$. Hence, fix finite-memory strategies π and σ, and so the induced DTMC $D = \mathcal{G}^{\pi,\sigma}$ is finite. By Lemma 4, the limit $\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})$ almost surely exists. Denote by S_D the set of states of D.

For every N and path λ, it holds that $|\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})(\lambda)| \leq \max_{s\in S_D} |\boldsymbol{r}(s)|$, where the maximum is taken componentwise, and thus, by the Lebesgue dominated convergence theorem,

$$\mathbb{E}_{D,s}\left[\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] = \lim_{N\to\infty} \mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right].$$

By Lemma 4, this is equal to $\sum_{BSCC\mathcal{B}} \mathrm{Pr}_{D,s}(\Diamond\mathcal{B})\mathrm{rew}(\mathcal{B})$.

Proof of (i): We assume that $\mathrm{EE}(\boldsymbol{r})$ is satisfied, and show that $\mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{0})$ is also satisfied. There exists a finite vector $\boldsymbol{v}_0$ such that for all $s \in S_D$ and all N it holds that $\mathbb{E}_{D,s}\left[\mathrm{rew}^N(\boldsymbol{r})\right] \geq \boldsymbol{v}_0$. Dividing both sides of this inequality by $N+1$, we obtain $\mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] \geq \frac{\boldsymbol{v}_0}{N+1}$, and taking the limit, we obtain $\lim_{N\to\infty} \mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] \geq 0$, i.e. we get $\mathbb{E}_{D,s}[\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})] \geq \boldsymbol{0}$ for all $s \in S_D$. The left hand side equals $\mathrm{rew}(\boldsymbol{r})(\mathcal{B})$ whenever s is in a BSCC $\mathcal{B}$ of D. We have $\mathrm{rew}(\boldsymbol{r})(\mathcal{B}) \geq \boldsymbol{0}$ for every BSCC $\mathcal{B}$, which concludes (i) by virtue of Lemma 4.

Proof of (ii): Assume that $D \models \mathrm{Pmp}_{\varsigma_0}(\boldsymbol{r})(\boldsymbol{0})$, and hence that $\mathrm{rew}(\boldsymbol{r})(\mathcal{B}) \geq \boldsymbol{0}$ for every BSCC $\mathcal{B}$. This implies that $\lim_{N\to\infty} \mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] \geq \boldsymbol{0}$ for every s. Fix $\varepsilon > 0$. For every s, there exists $N_{\varepsilon,s} \in \mathbb{N}$ such that for all $N \geq N_{\varepsilon,s}$ it holds that $\mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] \geq -\boldsymbol{\varepsilon}$. Hence for all $N \geq 0$, for every s, it holds that $\mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(\boldsymbol{r})\right] \geq -(N+1)\cdot\boldsymbol{\varepsilon} + \boldsymbol{v}_0^s$ with $v_{0,i}^s = \min_{N\leq N_{\varepsilon,s}} \mathbb{E}_{D,s}\left[\frac{1}{N+1}\mathrm{rew}^N(r_i)\right]$ for all i. Therefore, we have that $\mathrm{rew}^N(\boldsymbol{r}+\boldsymbol{\varepsilon})(\lambda) = \mathrm{rew}^N(\boldsymbol{r})(\lambda) + (N+1)\boldsymbol{\varepsilon} \geq \boldsymbol{v}_0^s$, and hence D satisfies $\mathrm{EE}(\boldsymbol{r}+\boldsymbol{\varepsilon})$. □

C Strategy Synthesis

C.1 Proof of Proposition 1

Below we use (after recalling them) quite standard results on fixpoint theory and refer the reader to Chapter 2 of [10].

Scott Continuity and Kleene Fixpoint Theorem. For $D \subseteq \mathcal{L}_M$, the infimum $\inf D$ is defined via $[\inf\{X \in D\}]_s \stackrel{\text{def}}{=} \bigcap_{X\in D} X_s$ for all $s \in \overline{S}$. A linearly ordered subset $D \subseteq \mathcal{L}_M$ is a *chain*. For any chain $D \in \mathcal{L}_M$, it holds that $\inf D \in \mathcal{L}_M$, as the intersection of convex topologically-closed M-downward-closed sets is itself convex, topologically-closed, and M-downward-closed. Hence, $\mathcal{L}_M$ endowed with $\subseteq$ is a *complete partial order* (cpo). A function $f : \mathcal{L}_M \to \mathcal{L}_M$ on a cpo $(\mathcal{L}_M, \subseteq)$ is *(Scott) continuous* if for all (countable) nonempty chains $D \subseteq \mathcal{L}_M$ we have that $f(\inf D) = \inf f(D)$, cf. Definition 2.3 in [10]. We recall that, according to the *Kleene fixpoint theorem*, the greatest fixpoint of a Scott continuous function $f : \mathcal{L}_M \to \mathcal{L}_M$ over a cpo $(\mathcal{L}_M, \subseteq)$ exists and is equal to the infimum of the $\subseteq$-descending chain $\top \supseteq f(\top) \supseteq f^2(\top)\cdots$, c.f. Corollary 2.6 in [10]. We note that this chain is countable.

Proof (of Proposition 1). The properties claimed in the proposition are consequences of Scott-continuity of F_M and the Kleene fixpoint theorem.

To show Scott-continuity, it is sufficient to show that for every (countable) $\subseteq$-descending chain D, we have $F_M(\inf D) = \inf F_M(D)$ componentwise, i.e. it is sufficient to show that for all $s \in \overline{S}$ we have that $[F_M(\inf D)]_s = \inf([F_M(D)]_s)$. Take any countable descending chain $D = \{X^k \in \mathcal{L}_M \mid k \in \mathbb{N}\} \subseteq \mathcal{L}_M$, and any $s \in \overline{S}$. We consider three cases:

– **Case** $s \in S_\Diamond$**.** The result follows from the following chain of equalities

$$
\begin{aligned}
[F_M(\inf(D))]_s &\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} [\inf D]_t)) \\
&\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} \bigcap_{k\geq 0} X_t^k)) \\
&\stackrel{*}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{k\geq 0} \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} X_t^k)) \\
&= \bigcap_{k\geq 0} (\text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} X_t^k))) \\
&\stackrel{\text{def}}{=} \bigcap_{k\geq 0} [F_M(X^k)]_s \\
&\stackrel{\text{def}}{=} [\inf F_M(D)]_s,
\end{aligned}
$$

where we prove $*$ just below. Let $Y_s^k \stackrel{\text{def}}{=} \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} X_t^k)$ and define $Y^\infty \stackrel{\text{def}}{=} \bigcap_{k\geq 0} Y^k$. We first prove that, for every k, $Y^k \in \mathcal{L}_M$. The sets X_t^k for $t \in \mathsf{succ}(s)$ are closed and bounded, thus compact. By a corollary of Carathéodory's theorem for convex sets their convex hull Y_s^k is also compact and convex. Moreover, it is M-downward closed, so, for every k, $Y^k \in \mathcal{L}_M$.

The sets Y^k clearly form a descending chain, since the X^k form a descending chain, and thus $\bigcap_{k\geq 0} Y^k = Y^\infty \in \mathcal{L}_M$.

The equality to prove is $\mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} \bigcap_{k\geq 0} X_t^k) = Y^\infty$. For the $\subseteq$ case, take $\boldsymbol{y} \in \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} \bigcap_{k\geq 0} X_t^k)$. Then $\boldsymbol{y} = \sum \mu(t)\boldsymbol{x}_t$ for some distribution μ and some $\boldsymbol{x}_t \in \bigcap_{k\geq 0} X_t^k$. Hence, for every k, $\boldsymbol{y} \in Y^k$, and so, $\boldsymbol{y} \in Y^\infty$.

Now we prove the case $\supseteq$. Let $\boldsymbol{y}^\infty \in Y_\infty$. We note that, for every $k \geq 0$, $\boldsymbol{y}^\infty = \sum_{t\in\mathsf{succ}(s)} \mu_k(t)\boldsymbol{x}_t^k$ for some distribution μ_k and some vector of points $\boldsymbol{x}_t^k \in Y_t^k$. As the sets of distributions and Y^k are compact, one can extract a subsequence of indices i_k such that μ_{i_k} and $\boldsymbol{x}_t^{i_k}$ converge toward limits, which we respectively denote μ and $\boldsymbol{x}_t$ for every $t \in \mathsf{succ}(s)$. Moreover, $\lim_{k\to\infty} \boldsymbol{x}_t^{i_k} = \boldsymbol{x}_t \in Y_t^l$ for every $l \geq 0$ as Y^l is compact. Hence, $\boldsymbol{x}_t \in \bigcap_{k\geq 0} X_t^k$ for every t and we conclude $\boldsymbol{y}^\infty = \sum_{t\in\mathsf{succ}(s)} \mu(t)\boldsymbol{x}_t \in \mathsf{conv}(\bigcup_{t\in\mathsf{succ}(s)} \bigcap_{k\geq 0} X_t^k)$.

– **Case** $s \in S_{\Box}$**.** We have

$$\begin{aligned}
[F_M(\inf(D))]_s &\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{t \in \mathsf{succ}(s)} [\inf D]_t) \\
&\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{t \in \mathsf{succ}(s)} \bigcap_{k \geq 0} X^k) \\
&= \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{k \geq 0} \bigcap_{t \in \mathsf{succ}(s)} X^k) \\
&= \bigcap_{k \geq 0} (\text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{t \in \mathsf{succ}(s)} X^k)) \\
&\stackrel{\text{def}}{=} \bigcap_{k \geq 0} [F_M(X^k)]_s \\
&\stackrel{\text{def}}{=} [\inf F_M(D)]_s.
\end{aligned}$$

– **Case** $s = (a, \mu) \in S_{\bigcirc}$**.** The result follows from the following chain of equalities

$$\begin{aligned}
[F_M(\inf(D))]_s &\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times [\inf D]_t) \\
&\stackrel{\text{def}}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times \bigcap_{k \geq 0} X^k) \\
&\stackrel{*}{=} \text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{k \geq 0} \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X^k) \\
&= \bigcap_{k \geq 0} (\text{Box}_M \cap \mathsf{dwc}(\boldsymbol{r}(s) + \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X^k)) \\
&\stackrel{\text{def}}{=} \bigcap_{k \geq 0} [F_M(X^k)]_s \\
&\stackrel{\text{def}}{=} [\inf F_M(D)]_s,
\end{aligned}$$

where we prove $*$ just below. We first show that $\sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times \bigcap_{k \geq 0} X^k = \bigcap_{k \geq 0} \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X^k$, analogously to the case for $\mathsf{Player}\ \Diamond$. The inclusion $\sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times \bigcap_{k \geq 0} X^k \subseteq \bigcap_{k \geq 0} \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X^k$ is straightforward. For the $\supseteq$ direction, take $\boldsymbol{x} \in \bigcap_{k \geq 0} \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X_t^k$, and so, for all $k \geq 0$, there exist vectors $\boldsymbol{y}_t^k \in X_t$ for $t \in \mathsf{supp}(\mu)$, such that $\boldsymbol{x} = \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \boldsymbol{y}_t^k$. We extract a subsequence of indices i_k such that $\boldsymbol{y}_t^{i_k}$ tends to a limit $\boldsymbol{y}_t$. This limit necessarily lies in $\bigcap_{k \geq 0} X_t^k$ (recall the case for $\mathsf{Player}\ \Diamond$) and hence $\boldsymbol{x} = \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \boldsymbol{y}_t \in \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times \bigcap_{k \geq 0} X^k$.

Now, since F_M is Scott-continuous, by the Kleene fixpoint theorem, the greatest fixpoint exists and is equal to the limit $fix(F_M) = \lim_{n \to \infty} F_M^n(\top) \stackrel{\text{def}}{=} \cap_{n \in \mathbb{N}} F_M^n(\top)$. □

C.2 Proof of Proposition 2

The proof is given by a chain of implications corresponding respectively to Lemma 5–8. The Lemmas are based on relating fixpoints of several operators in order to pass from EE that deals with DTMCs to the fixpoint of the operator F_M that deals with games. The proof involves instantiating the operator F_M for different kinds of games, namely, a version $F_{G,M}$ associated with the game G and a version $F_{\mathcal{M},M}$ associated with the PA $\mathcal{M}$ (which is a particular game).

Write S_D for the state space of a DTMC D, and write $\mathsf{succ}_D(s)$ for the successors of $s \in S_D$. Write $\overline{S}_{\mathcal{M}}$ for the set of states and moves of a PA M, and write $\mathsf{succ}_{\mathcal{M}}(s)$ for the successors of $s \in \overline{S}_{\mathcal{M}}$. For two vectors $\boldsymbol{a}, \boldsymbol{b} \in (\mathbb{R} \cup -\infty)^n$, we denote by $\min(\boldsymbol{a}, \boldsymbol{b})$ the vector $\boldsymbol{c}$ such that $c_i = \min(a_i, b_i)$, i.e. the componentwise minimum.

Given a PA $\mathcal{M}$, we consider the set $([-\infty, 0]^n)^{|\overline{S}_{\mathcal{M}}|}$ with the partial order $\boldsymbol{u} \leq \boldsymbol{u}'$, which is defined to hold if and only if $u_{s,i} \leq u'_{s,i}$ for every $s \in \overline{S}_{\mathcal{M}}$ and every $i \in [1, n]$. We define an operator $F_{\mathcal{M}}$ on $([-\infty, 0]^n)^{|\overline{S}_{\mathcal{M}}|}$ as follows:

$$[\tilde{F}_{\mathcal{M}}(\boldsymbol{u})]_s = \begin{cases} \min(\boldsymbol{0}, \boldsymbol{r}(s) + \min_{t \in \mathsf{succ}_{\mathcal{M}}(s)} \boldsymbol{u}_t) & \text{if } s \in S_\Box \\ \min(\boldsymbol{0}, \boldsymbol{r}(s) + \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \boldsymbol{u}_t) & \text{if } s = (a, \mu) \in S_\bigcirc. \end{cases}$$

The sequence $(\tilde{F}^k_{\mathcal{M}}(\boldsymbol{0}))_{k \geq 0}$ is non-increasing, that is, $\tilde{F}^0_{\mathcal{M}}(\boldsymbol{0}) \geq \tilde{F}^1_{\mathcal{M}}(\boldsymbol{0}) \geq \ldots \geq \tilde{F}^k_{\mathcal{M}}(\boldsymbol{0}) \ldots$, and converges toward the greatest fixpoint of $\tilde{F}_{\mathcal{M}}$, written $fix(\tilde{F}_{\mathcal{M}})$.

Lemma 5. *If π satisfies $\mathrm{EE}(r)$, then for every σ, $fix(\tilde{F}_{G^{\pi,\sigma}})$ is lower-bounded.*

Proof. Let σ be a strategy, let $D \stackrel{\text{def}}{=} G^{\pi,\sigma}$, and let P be the transition probability matrix of the DTMC D (see e.g. [2]). Let $\tilde{\boldsymbol{E}}^* \stackrel{\text{def}}{=} fix(\tilde{F}_{G^{\pi,\sigma}})$. First we show how to reduce our case to an irreducible DTMC by applying qualitative graph analysis on the DTMC $G^{\pi,\sigma}$. We use that for every edge $(s,t) \in S_D \times S_D$ (that is, $P_{s,t} > 0$), if $\tilde{\boldsymbol{E}}^*_t$ is unbounded then so is $\tilde{\boldsymbol{E}}^*_s$ (due to the min-operator in $\tilde{\boldsymbol{E}}^*_s = \min(\boldsymbol{0}, \boldsymbol{r}(s) + \sum_{t \in \mathsf{succ}_D(s)} P_{s,t} \tilde{\boldsymbol{E}}^*_t))$. In particular, if $fix(\tilde{F}_{G^{\pi,\sigma}})$ is lower-bounded in the initial state then it is lower-bounded at every state (by definition the induced DTMC $G^{\pi,\sigma}$ contains only reachable states).

We now show that if $fix(\tilde{F}_{G^{\pi,\sigma}})$ is lower-bounded in all BSCCs then it is also lower-bounded in the transient states. We first give a linear over-approximation of $\tilde{\boldsymbol{E}}^* = \tilde{F}_{G^{\pi,\sigma}}(\tilde{\boldsymbol{E}}^*)$ in the following inequality:

$$\tilde{\boldsymbol{E}}^* \geq \boldsymbol{a} + P \tilde{\boldsymbol{E}}^*, \tag{1}$$

where $\boldsymbol{a}$ is defined by $\boldsymbol{a}_s = \min\{\boldsymbol{0}, \boldsymbol{r}(s)\}$. We denote by P_{TT} (resp. $P_{T\perp}$) the transition probability matrix P of D restricted to probabilities between transient states (resp. to probabilities of going from transient to recurrent states). We denote by $\tilde{\boldsymbol{E}}^*_T$ (resp. $\tilde{\boldsymbol{E}}^*_\perp$) the sub-vector of $\tilde{\boldsymbol{E}}^*$ corresponding to the transient states (resp. the recurrent states). Our purpose is to show that $\tilde{\boldsymbol{E}}^*_T$ is a vector with finite entries (that is, $> -\infty$) knowing that $\tilde{\boldsymbol{E}}^*_\bot$ has finite entries. Applying (1) to transient states yields

$$\tilde{\boldsymbol{E}}^*_T \geq \boldsymbol{a}_T + P_{TT}\tilde{\boldsymbol{E}}^*_T + P_{T\bot}\tilde{\boldsymbol{E}}^*_\bot,$$

and hence

$$(I - P_{TT})\tilde{\boldsymbol{E}}^*_T \geq \boldsymbol{a}_T + P_{T\bot}\tilde{\boldsymbol{E}}^*_\bot.$$

In particular, $(I - P_{TT})\tilde{\boldsymbol{E}}^*_T$ is a vector with finite entries. The matrix $(I - P_{TT})$ is invertible, since $(I - P_{TT})^{-1} = \sum_{k=0}^{+\infty} P^k_{TT}$ is well defined (see e.g. Remark 10.20 in [2]), and hence $\tilde{\boldsymbol{E}}^*_T$ has also finite entries.

Note that, if $G^{\pi,\sigma}$ satisfies EE($\boldsymbol{r}$), then it satisfies EE($\boldsymbol{r}$) in every BSCC. Thus it suffices to show the statement of the Lemma for an arbitrary single BSCC.

Justified by the above argument, we suppose for the rest of this proof that the DTMC $D = G^{\pi,\sigma}$ considered is irreducible. We use $\boldsymbol{E}^N_s$ as a shorthand for $\mathbb{E}_{D,s}[\mathrm{rew}^N(\boldsymbol{r})]$. $\boldsymbol{E}^N_s$ can be expressed via a recurrence as follows: $\boldsymbol{E}^0_s \stackrel{\text{def}}{=} \boldsymbol{0}$, and for $N \geq 0$,

$$\boldsymbol{E}^{N+1}_s \stackrel{\text{def}}{=} \boldsymbol{r}(s) + \sum_{t \in \mathsf{succ}(s)} P_{s,t}\boldsymbol{E}^N_t.$$

Writing this in matrix form for $N \geq 0$ and $j \geq 1$ yields

$$\boldsymbol{E}^{N+j} = \sum_{i=0}^{j-1} P^i \boldsymbol{r} + \boldsymbol{E}^N. \tag{2}$$

We want to show that the sequence $(\boldsymbol{E}^N)$ is not lower-bounded whenever the sequence $(\tilde{\boldsymbol{E}}^N)$ is not, where $\tilde{\boldsymbol{E}}^N \stackrel{\text{def}}{=} \tilde{F}^N_D(\boldsymbol{0}))$ for all $N \geq 0$. It suffices to prove the result for one arbitrary dimension of the reward structure $\boldsymbol{r}$, which we simply denote r. We use the lower case $\tilde{\boldsymbol{e}}$ (resp. $\boldsymbol{e}$) instead of $\tilde{\boldsymbol{E}}$ (resp. $\boldsymbol{E}$) to emphasize that we work with single-dimensional reward, but note that these are still vectors, with one component for each state s of D; when looking at $\tilde{\boldsymbol{e}}$ or $\boldsymbol{e}$ at a specific state, we write e_s and $\tilde{e}_s$ (note that e_s and $\tilde{e}_s$ are scalars specific to the state s and to the same dimension as r). We show the following facts, noting that D is assumed to be irreducible:

- F1 For every $s \in \overline{S}_D$ and $k \geq 0$, $|\tilde{e}^k_s - \tilde{e}^{k+1}_s| \leq \max_s r(s)$;
- F2 For all $s, t \in \overline{S}_D$ and $k \geq 0$, it holds that: $\tilde{e}^k_s \leq p^{|\overline{S}_D|}_{\min}\tilde{e}^k_t + 2|\bar{S}| \max_s r(s)$, where $p_{\min}$ is the smallest probability on any transition in D, and thus, if $(\tilde{\boldsymbol{e}}^k)_{k\in\mathbb{N}}$ is unbounded, then $(\max_s(\tilde{e}^k_s))_{k\in\mathbb{N}}$ is likewise unbounded.
- F3 If $(\tilde{\boldsymbol{e}}^k)_{k\in\mathbb{N}}$ is unbounded then, for every M, there exist $k, m \in \mathbb{N}$ such that $\boldsymbol{e}^{k+m} - P^m\boldsymbol{e}^k \not\geq -2\boldsymbol{M}$.
- F4 For every $k, m, M \in \mathbb{N}$, if $\boldsymbol{e}^k$ is lower-bounded by $-M$ then $\boldsymbol{e}^{k+m} - P^m\boldsymbol{e}^k \geq -\boldsymbol{M}$.

Proof of F1. We show this by induction on k. The base case is $|\tilde{e}_s^0 - \tilde{e}_s^1| = |\tilde{e}_s^1| = \min(0, |r(s)|) \leq \max_s r(s)$. Now assume the result holds for $k-1$. Then $\tilde{e}_s^k - \tilde{e}_s^{k+1} = \min(0, r(s) + \sum_t P_{s,t}\tilde{e}_t^{k-1}) - \min(0, r(s) + \sum_t P_{s,t}\tilde{e}_t^k)$. If both $\tilde{e}_s^k$ and e_s^{k+1} are zero, then the inequality is trivial. If both $\tilde{e}_s^k$ and e_s^{k+1} are negative then

$$|\tilde{e}_s^k - \tilde{e}_s^{k+1}| = \left|\sum_t P_{s,t}(\tilde{e}_t^{k-1} - \tilde{e}_t^k)\right| \leq \sum_t P_{s,t}|\tilde{e}_t^{k-1} - \tilde{e}_t^k|$$
$$\leq \sum_t P_{s,t} \max_s r(s) \leq \max_s r(s).$$

If $\tilde{e}_s^k = 0$, then, from the induction hypothesis, $|\tilde{e}_s^{k-1}| \leq \max_s r(s)$, and so $0 \leq r(s) + \sum_t P_{s,t}\tilde{e}_t^{k-1} \leq r(s) + \sum_t P_{s,t}(\tilde{e}_t^k + \max_s r(s))$. Hence $|\tilde{e}_s^k - \tilde{e}_s^{k+1}| = |\tilde{e}_s^{k+1}| = |\min(0, r(s) + \sum_t P_{s,t}\tilde{e}_t^k)| \leq \max_s r(s)$.

Proof of F2. If t is a successor of s, then

$$\tilde{e}_s^{k+1} \leq r(s) + \sum_t P_{s,t}\tilde{e}_t^k \leq r(s) + P_{s,t}\tilde{e}_t^k \leq r(s) + p_{\min}\tilde{e}_t^k.$$

If t is reachable from s in $l \leq |S|$ steps, then, by iterating,

$$\tilde{e}_s^{k+l} \leq \sum_{i=0}^{l-1} p_{\min}^i \max_s r(s) + p_{\min}^l \tilde{e}_t^k \leq l \max_s r(s) + p_{\min}^l \tilde{e}_t^k.$$

The result follows by applying fact F1 l times, to obtain

$$\tilde{e}_s^k \leq l \max_s r(s) + l \max_s r(s) + p_{\min}^l \tilde{e}_t^k \leq 2|S| \max_s r(s) + p_{\min}^{|S|} \tilde{e}_t^k.$$

Proof of F3. Let l be such that $\tilde{e}_s^l \leq B$ for every state $s \in \overline{S}_D$, where $B \stackrel{\text{def}}{=} -2M\left(p_{\min}^{-|\bar{S}|}|\bar{S}| + 2|\bar{S}|\max(r)\right)$. For every s the sequence $(\tilde{e}_s^{l-m})_{m \leq l}$ is zero for $m = l$, and so there is a first index m_s (potentially different for every state), such that $\tilde{e}_s^{l-m_s} \geq -C$. We let s^* be the state for which m_{s^*} is minimal amongst the m_s, and so that $\tilde{e}_{s^*}^{l-m_{s^*}}$ is maximal amongst the $\tilde{e}_s^{l-m_s}$. By definition of s^*, every $k < m_{s^*}$, and for every s it holds that $\tilde{e}_s^{l-k} < -C$. This implies that $\tilde{e}_s^{l-m_{s^*}} \leq \tilde{e}_s^{l-m_{s^*}-1} + C < 0$. This yields that, for every $k \leq m_{s^*}$, and every s, we have $\tilde{e}_s^{l-k} < 0$. Hence there is no cut-off between indices $l-m$ and l, enabling us to use the recurrence (2), to obtain

$$\tilde{\boldsymbol{e}}^{l-m+j} = \sum_{i=0}^{j-1} P^i \boldsymbol{r} + P^j \tilde{\boldsymbol{e}}^{l-m} \tag{3}$$

Hence $\boldsymbol{e}^l - \tilde{\boldsymbol{e}}^l = P^m \boldsymbol{e}^{l-m} - P^m \tilde{\boldsymbol{e}}^{l-m}$, and $\boldsymbol{e}^l - P^m \boldsymbol{e}^{l-m} = \tilde{\boldsymbol{e}}^l - P^m \tilde{\boldsymbol{e}}^{l-m}$.

We further have, for every state $t \in \overline{S}_D$, that

$$[-P^m \tilde{\boldsymbol{e}}^{l-m}]_s = \sum_{t \in \bar{S}} [P^m]_{st}(-\tilde{e}_t^{l-m_{s^*}}) \leq -\tilde{e}_{t^*}^{l-m_{s^*}},$$

where t^* is such that $\tilde{e}_{t^*}^{l-m_{s^*}}$ is minimal. We therefore have

$$[\tilde{\boldsymbol{e}}^l - P^{m_{s^*}}\tilde{\boldsymbol{e}}^{l-m_{s^*}}]_{t^*} \leq \tilde{e}_{t^*}^l - \tilde{e}_{t^*}^{l-m_{s^*}} \leq B - \tilde{e}_{t^*}^{l-m_{s^*}}.$$

By virtue of F2, we get

$$-\tilde{e}_{t^*}^{l-m_{s^*}} \leq -p_{\min}^{-|\bar{S}|}\tilde{e}_{s^*}^{l-m_{s^*}} - 2|\bar{S}|\max_s r(s),$$

which is by definition of s^* less than $-p_{\min}^{-|\bar{S}|}|\bar{S}| - 2|\bar{S}|\max_s r(s) = \frac{1}{2M}B$. Putting the previous statements together, we obtain the desired result:

$$[\boldsymbol{e}^l - P^{m_{s^*}}\boldsymbol{e}^{l-m_{s^*}}]_{t^*} = [\tilde{\boldsymbol{e}}_{t^*}^l - P^{m_{s^*}}\tilde{\boldsymbol{e}}^{l-m_{s^*}}]_{t^*} \leq \left(1 + \frac{1}{2M}\right)B \leq -2M.$$

Proof of F4. We use recurrence (2) to obtain $\boldsymbol{e}^{k+m} - P^m\boldsymbol{e}^k = \sum_{i=0}^{m-1}P^i\boldsymbol{r} = \boldsymbol{e}^m \geq -M$.

We are now ready to complete the proof. If $(\tilde{\boldsymbol{e}}^k)_{k\in\mathbb{N}}$ is unbounded, then, using F3, for every M there exist $k, m \in \mathbb{N}$ such that $\boldsymbol{e}^{k+m} - P^m\boldsymbol{e}^k \not> -M$ and in particular $\boldsymbol{e}^{k+m} - P^m\boldsymbol{e}^k \not\geq -2M$. By F4, this implies that $(\boldsymbol{e}^k)_{k\in\mathbb{N}}$ is not lower-bounded by $-M$. We have shown that $(\boldsymbol{e}^k)_{k\in\mathbb{N}}$ is not lower-bounded when $(\tilde{\boldsymbol{e}}^k)_{k\in\mathbb{N}}$ is not; this concludes the proof. □

Lemma 6. *Given a PA $\mathcal{M}$, it holds that $fix(\tilde{F}_{\mathcal{M}^\sigma})$ is lower-bounded for every strategy σ if and only if $fix(\tilde{F}_{\mathcal{M}})$ is lower-bounded.*

Proof. A straightforward induction allows us to get that for every $k \geq 0$ and every state λ of $\mathcal{M}^\sigma$ (which is by definition of the induced DTMC a path of $\mathcal{M}$) it holds that $[\tilde{F}_{\mathcal{M}^\sigma}^k(\mathbf{0})]_\lambda \geq [\tilde{F}_{\mathcal{M}}^k(\mathbf{0})]_{\mathsf{last}(\lambda)}$. Hence $[fix(\tilde{F}_{\mathcal{M}^\sigma})]_\lambda \geq [fix(\tilde{F}_{\mathcal{M}})]_{\mathsf{last}(\lambda)}$ for every σ. This implies the "only if" direction.

For the "if" direction, we assume that $fix(\tilde{F}_{\mathcal{M}^\sigma})$ is lower-bounded for every σ and, in particular, for every memoryless deterministic (MD) σ. We show as a sufficient condition that the inequality $[fix(\tilde{F}_{\mathcal{M}^\sigma})]_\lambda \geq [fix(\tilde{F}_{\mathcal{M}})]_{\mathsf{last}(\lambda)}$ becomes an equality for a MD strategy σ we are about to define. Indeed, it suffices to define σ that, in every state $s \in S_\Box$, takes the successor t_s for which the minimum is reached in the definition of the fixpoint $[fix(\tilde{F}_{\mathcal{M}})]_s = \min(0, \boldsymbol{r}(s) + \min_{t\in\mathsf{succ}_{\mathcal{M}}(s)}[fix(\tilde{F}_{\mathcal{M}})]_t) = \min(0, \boldsymbol{r}(s) + [fix(\tilde{F}_{\mathcal{M}})]_{t_s})$. Then $\tilde{F}_{\mathcal{M}^\sigma}$ and $\tilde{F}_{\mathcal{M}}$ have exactly the same recurrent definition and hence the same fixpoint. □

Lemma 7. *Given a PA $\mathcal{M}$ and a bound M, if $fix(\tilde{F}_{\mathcal{M}}) \geq -M$ then, for every $s \in \overline{S}_{\mathcal{M}}$, $[fix(F_{\mathcal{M},M})]_s \neq \emptyset$.*

Proof. Let $\boldsymbol{u}^* \stackrel{\text{def}}{=} fix(\tilde{F}_{\mathcal{M}})$. We show, by induction on k, that, for every k, $\boldsymbol{u}^* \in F_{\mathcal{M},M}^k(\top)$. The statement for $k = 0$ holds since $\boldsymbol{u}^* \in \mathrm{Box}_M$. For the inductive step, assume that $\boldsymbol{u}^* \in F_{\mathcal{M},M}^{k-1}(\top)$ for some $k \geq 1$. Then, for every $\Box$-state s, it holds that $\min_{t\in\mathsf{succ}_{\mathcal{M}}(s)}\boldsymbol{u}_t^* \in \cap_{t\in\mathsf{succ}_{\mathcal{M}}(s)}[F_{\mathcal{M},M}^k(\top)]_t$, since the sets resulting from $F^k_{\mathcal{M},M}$ are M-downward-closed. It also holds that, for every move (a,μ), $\sum_{t\in\mathsf{supp}(\mu)}\mu(t)\boldsymbol{u}^*_t \in \sum_{t\in\mathsf{supp}(\mu)}\mu(t)[F^k_{\mathcal{M},M}(\top)]_t$. In both cases we deduce that $\boldsymbol{u}^*_s \in [F^k_{\mathcal{M},M}(\top)]_s$, and hence the property is proved. We conclude that $\boldsymbol{u}^*_s \in \cap_{k\geq 0}[F^k_{\mathcal{M},M}(\top)]_s = [fix(F_{\mathcal{M},M})]_s$, which is thus non-empty for every s. □

Lemma 8. *Given a game G and a bound M, if there exists a strategy π such that $[fix(F_{G^\pi,M})]_{\varsigma_0} \neq \emptyset$ then $[fix(F_{G,M})]_{\varsigma_0} \neq \emptyset$.*

Proof. A straightforward induction yields that for every $k \geq 0$ and every state λ of G^π (which is by definition of the induced PA a path of G) it holds that $[F^k_{G^\pi,M}(\top)]_\lambda \subseteq [F^k_G(\top)]_{\mathsf{last}(\lambda)}$. Hence $[fix(F_{G^\pi,M})]_\lambda \subseteq [fix(F_G)]_{\mathsf{last}(\lambda)}$ for every state λ of G^π. By assumption, $[fix(F_{G^\pi,M})]_{\varsigma_0} \neq \emptyset$ and hence $[fix(F_{G^\pi,M})]_{\varsigma_0} \neq \emptyset$. □

Proof (Proposition 2). Since $\mathrm{EE}(\boldsymbol{r})$ is achievable, there is a Player $\Diamond$ strategy π satisfying $\mathrm{EE}(\boldsymbol{r})$, and so, by Lemma 5, for every Player $\Box$ strategy σ, $fix(\tilde{F}_{G^{\pi,\sigma}})$ is lower bounded by some $-M_\sigma > -\infty$. From Lemma 6, there is a lower bound $-M > -\infty$ on $fix(\tilde{F}_{G^{\pi,\sigma}})$, which holds for every Player $\Box$ strategy σ. Hence, by Lemma 7, the fixpoint $[fix(F_{G^\pi,M})]_s$ is nonempty for every state $s \in \overline{S}_{G^\pi}$ of the PA G^π induced by applying π to G. Finally, Lemma 8 states that the same bound M is also valid for the game G, and hence $[fix(F_{G,M})]_{\varsigma_0} \neq \emptyset$, concluding the proof. □

C.3 Proof of Lemma 3

Lemma 9. *Given a sequence $(Y^k_s)_{k\leq m+1}$ of elements of $\mathcal{P}^M_c(\mathrm{Box}_M)$ such that for $k \leq m$, $\emptyset \subsetneq Y^{k+1}_s \sqsubseteq Y^k_s$ and $Y^k_s \not\sqsubseteq Y^{k+1}_s + \varepsilon$. Then it must be that $m \leq k^*$ with $k^* \stackrel{\text{def}}{=} n \cdot ((\lceil \frac{M}{\varepsilon} \rceil + 2)^2 + 2)$.*

The proof of Lemma 9 makes use of the following Ramsey like theorem.

Theorem 4 (Theorem 4.5.2 of [16]). *Let $G = (V,E)$ be a linearly-ordered complete graph over m nodes given with an n-coloring of its edges. Then G contains a monochromatic directed path of length $\lfloor \sqrt{m/n - 2} \rfloor - 1$*

Proof (of Lemma 9). By assumption there exists a sequence $(\boldsymbol{x}^k)_{k\leq m} \in (Y^k_s \cap (\mathrm{Box}_M \setminus \mathsf{dwc}(Y^{k+1}_s + \varepsilon)))_{k\leq m}$ of points. We first prove that for all $j < k$, there exists a coordinate $c(j,k)$ for which $x^j_{c(j,k)} - x^k_{c(j,k)} > \varepsilon$. For $j < k$, $\boldsymbol{x}^j - \varepsilon \notin \mathsf{dwc}(\{\boldsymbol{x}^k\})$; otherwise $\boldsymbol{x}^j - \varepsilon$ would be in Y^k_s, for which we have $Y^k_s \sqsubseteq Y^{j+1}_s$. This implies that there exists a coordinate $c(j,k)$ for which $\boldsymbol{x}^j_{c(j,k)} - \boldsymbol{x}^k_{c(j,k)} > \varepsilon$.

Now consider the linearly-ordered complete graph over m nodes $V = \{1,...,m\}$ and with edges (j,k) for $1 \leq j < k \leq m$. Endow the edges of this graph with the n-colouring c given above, i.e. there is one color per dimension of the M-polyhedrals.

Assume toward a contradiction that m is greater than or equal to $k^* = n \cdot ((\frac{M}{\varepsilon} + 2)^2 + 2)$. Then, by Theorem 4, there exists a monochromatic path

$j_1 \to j_2 \to \cdots \to j_l$ of length $l \geq \lceil \frac{M}{\varepsilon} \rceil$, and thus by denoting c the colour of this path it holds that $x_c^{j_1} > x_c^{j_2} + \varepsilon > \ldots > x_c^{j_l} + l\varepsilon \geq -M + \frac{M}{\varepsilon}\varepsilon \geq 0$, a contradiction. □

Proof (Lemma 3). Our goal is to find an index $i \leq k^{**} \stackrel{\text{def}}{=} (2k^*)^{|\bar{S}|}$ such that $X_s^{i+1} \subseteq X_s^i + \varepsilon$ for all $s \in \bar{S}$. Fix an ordering on states, $s_0, \ldots, s_{|\bar{S}|-1}$.

We recursively define a finite sequence of sets of indices $(I_l)_{l \leq |\bar{S}|}$ such that $|I_l| = (2k^*)^{|\bar{S}|-l}$, $I_{l+1} \subseteq I_l$, and for all $i \in I_{l+1}$, $X_{s_l}^i \subseteq X_{s_l}^{i+1} + \varepsilon$. Let $I_0 = \{0, 1, \ldots, (2k^*)^{|\bar{S}|} - 1\}$. Assume that $I_l = \{i_0, \ldots, i_{(2k^*)^{|\bar{S}|-l}/k^*-1}\}$ is already constructed. For every $p \leq (2k^*)^{|\bar{S}|-l}/k^* - 1$, we use Lemma 9 with the sequence $X_s^{i_{k^*p}}, \ldots, X_s^{i_{k^*(p+1)-1}}$ to find an index i_{j_p} among the k^* indices in the set $\{i_{k^*p}, \ldots, i_{k^*(p+1)-1}\}$, such that $X_s^{i_{j_p}} \sqsubseteq X_s^{i_{j_p}+1} + \varepsilon$ or $X_s^{i_{j_p}+1} = \emptyset$. Note that if $X_s^{i_{j_p}+1} = \emptyset$ then $X_s^{i_{j_q}} = X_s^{i_{j_q}+1} = \emptyset$ for $q > p$. Then there can be only one index for which $X_s^{i_{j_p}} \not\sqsubseteq X_s^{i_{j_p}+1} + \varepsilon$. If such an index i_{j_q} exists, we can define the two sets $\{i_{j_p} | p < q\}$ and $\{i_{j_p} | p > q\}$, and we define I_{l+1} to be the set among these with maximal cardinality. Thus I_{l+1} contains at least $|I_l|/(2k^*) = (2k^*)^{|\bar{S}|-l-1}$ elements. Up to deleting some indices we define I_{l+1} with exactly $(2k^*)^{|\bar{S}|-l-1}$ elements.

We now have a unique index $i \in I_{|\bar{S}|}$, such that for all $s \in \bar{S}$, it holds that $X_s^i \sqsubseteq X_s^{i+1} + \varepsilon$. Further, the index $i \in I_{|\bar{S}|} \subseteq I_0$ is bounded by $k^{**} = (2k^*)^{|\bar{S}|}$, which is the bound we set out to prove. □

C.4 Proof of Theorem 1

Proof. Let $\varepsilon \geq 0$, and let k such that $X_{\varsigma_0}^{k+1} \neq \emptyset$ and $X^k \sqsubseteq X^{k+1} + \varepsilon$.[1] We define $\overline{T} \subseteq \bar{S}$ as the set of states s for which $X_s^{k+1} \neq \emptyset$ (in particular $\varsigma_0 \in \overline{T}$). Note that, if a move or a Player □ state is in $\overline{T}$, then it means that, for every $t \in \mathsf{succ}(s)$, $X_t^{k+1} + \varepsilon \sqsupseteq X_t^k \neq \emptyset$ and hence $t \in \overline{T}$.

Firstly, we show that the strategy is well-defined. We now show that the points $\boldsymbol{q}_i^t$ can be picked from X_t^k. For any state $s \in \bar{S}$, depending on the type of s (i.e. Player ♢, Player □, or move), we define an auxiliary set Y_s^k without the cut-off. We then show that we can find the required coefficients and corner points for every point in Y_s^k, and prove that for all extreme points $\boldsymbol{p}$ of X_s^k we have $\boldsymbol{p} - \varepsilon$ in Y_s^k for $k \geq 0$, allowing us to show well-definedness of the strategy. Take $s \in \overline{T}$.

- **Case** $s \in S_\Diamond$**.** Let $Y_s^k \stackrel{\text{def}}{=} \boldsymbol{r}(s) + \mathsf{conv}(\bigcup_{t \in \mathsf{succ}(s) \cap \overline{T}} X_t^k)$. Take any $\boldsymbol{p'} \in Y_s^k$. Due to the convex hull, there now exist coefficients $\boldsymbol{\gamma} \in \Delta^{|\mathsf{succ}(s) \cap \overline{T}|}$, $\boldsymbol{\beta}^t \in \Delta^n$ and points $\boldsymbol{q}_i^t \in C_t^k$ for $t \in \mathsf{succ}(s) \cap \overline{T}$, such that $\sum_{t \in \mathsf{succ}(s)} \gamma_t \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p'} - \boldsymbol{r}(s)$.

[1] Note that we allow $\varepsilon = 0$. As long as k is finite, the sets X^k can be finitely represented, yielding finite-memory Player ♢ strategies.

– **Case** $s \in S_\Box$**.** Let $Y_s^k \stackrel{\text{def}}{=} \mathsf{dwc}(\boldsymbol{r}(s) + \bigcap_{t \in \mathsf{succ}(s)} X_t^k)$. Take any $\boldsymbol{p'} \in Y_s^k$. For any $t \in \mathsf{succ}(s)$, there now exist coefficients $\boldsymbol{\beta}^t \in \Delta^n$ and points $\boldsymbol{q}_i^t \in C_t^k$ such that $\sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p'} - \boldsymbol{r}(s)$.
– **Case** $s = (a, \mu) \in S_\bigcirc$**.** Let $Y_s^k \stackrel{\text{def}}{=} \boldsymbol{r}(s) + \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \times X_t^k$. Take any $\boldsymbol{p'} \in Y_s^k$. Due to the Minkowski sum, there now exist coefficients $\boldsymbol{\beta}^t \in \Delta^n$ and points $\boldsymbol{q}_i^t \in C_t^k$ such that $\sum_{t \in \mathsf{supp}(\mu)} \mu(t) \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t \geq \boldsymbol{p'} - \boldsymbol{r}(s)$.

We have $X^k \sqsubseteq X^{k+1} + \varepsilon$, and so $\mathsf{dwc}(Y_s^k) \cap \mathrm{Box}_M = X_s^{k+1} \sqsupseteq X_s^k - \boldsymbol{\varepsilon}$. Then, for any point $\boldsymbol{p} \in C_s^k$, it holds that $\boldsymbol{p} - \boldsymbol{\varepsilon} \in \mathsf{dwc}(Y_s^k) \cap \mathrm{Box}_M$. Hence we can find for $\boldsymbol{p}' = \boldsymbol{p} - \boldsymbol{\varepsilon}$ the corresponding coefficients and extreme points to construct the strategy, as described in Section 3.2. Note that only states in $\overline{T}$ are accessed by the strategy.

Secondly, we let σ be an arbitrary Player $\Box$ strategy, let D be the DTMC induced by applying (π, σ) to $\mathcal{G}$ and let s_o be an arbitrary state of D. It has the form $s_o = (s, (s, \boldsymbol{p}_o), m)$, where s is a state of $\mathcal{G}$ and $(s, \boldsymbol{p}_o)$ is the memory of Player $\Diamond$ and m is the memory of Player $\Box$.

We show that $\mathbb{E}_{D,s_o}[\mathrm{rew}^N(r)] \geq \boldsymbol{p}_o - N\boldsymbol{\varepsilon}$. For this we show that the memory of π is always above $\boldsymbol{p}_o - \mathbb{E}_{D,s_o}[\mathrm{rew}^N(r)] - N\boldsymbol{\varepsilon}$, and since this memory is always non-positive due to the Box_M intersection in the F_M operator, we get the desired result.

Let $\lambda \in \Omega_{D,s_o}$ be a path ending in $(s, (s, \boldsymbol{y}), m)$, where $s \in \overline{S}$ is the current state, $(s, \boldsymbol{y})$ is the current memory of π, and m is the current memory of σ. Let $Y_N : \Omega_D \to \mathbb{R}^n$ be the random variable that assigns $\boldsymbol{y}$ to a path λ for which $\lambda_N = (s, (s, \boldsymbol{y}), m)$. We now show that $\mathbb{E}_{D,s_o}[Y_N] \geq \boldsymbol{p}_o - \mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r})] - N\boldsymbol{\varepsilon}$ for all $N \geq 0$ by induction on the length N of paths.

– **Base case.** For $N = 0$, we have $\mathbb{E}_{D,s_o}[Y_0] = \boldsymbol{p}_o$,
– **Inductive case.** Assume that $\mathbb{E}_{D,s_o}[Y_N] \geq \boldsymbol{p}_o - \mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r})] - N\boldsymbol{\varepsilon}$. Let W_N be the set of all finite paths of length N in the induced DTMC D. For each path $\lambda' = \lambda \cdot (s, (s, \boldsymbol{y}), m) \in W_N$, we have

$$\mathbb{E}_{D,s_o}[Y_{N+1}|\lambda'] = \begin{cases} \sum_{t \in \mathsf{succ}(s)} \pi_n(s, (s, \boldsymbol{y}))(t) \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t & \text{if } s \in S_\Diamond \\ \sum_{t \in \mathsf{succ}(s)} \sigma_n(s, m)(t) \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t & \text{if } s \in S_\Box \\ \sum_{t \in \mathsf{supp}(\mu)} \mu(t) \cdot \sum_i \beta_i^t \cdot \boldsymbol{q}_i^t & \text{if } s = (a, \mu) \in S_\bigcirc, \end{cases}$$

which, by the definition of π, is greater than or equal to $\boldsymbol{y} - \boldsymbol{r}(s) - \boldsymbol{\varepsilon}$.
We now apply the law of total probability (note that the memory of the strategies is countable), to obtain (we again denote $\lambda' = \lambda \cdot (s, (s, \boldsymbol{y}), m)$)

$$\begin{aligned} \mathbb{E}_{D,s_o}[Y_{N+1}] &= \sum_{\lambda' \in W_N} \mathbb{E}_{D,s_o}[Y_{N+1}|\lambda'] \cdot \mathrm{Pr}_{\mathcal{G}}^{\pi,\sigma}(\lambda') \\ &\geq \sum_{\lambda' \in W_N} (\boldsymbol{y} - \boldsymbol{r}(s) - \boldsymbol{\varepsilon}) \cdot \mathrm{Pr}_{\mathcal{G}}^{\pi,\sigma}(\lambda') \\ &= \mathbb{E}_{D,s_o}[Y_N] - (\mathbb{E}_{D,s_o}[\mathrm{rew}^{N+1}(\boldsymbol{r})] - \mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r})]) - \boldsymbol{\varepsilon} \\ &\geq \boldsymbol{p}_o - \mathbb{E}_{D,s_o}[\mathrm{rew}^{N+1}(\boldsymbol{r})] - (N+1)\boldsymbol{\varepsilon}, \end{aligned}$$

where the last inequality holds by the induction hypothesis.

Since $\mathbb{E}_{D,s_o}[Y_N] \leq \mathbf{0}$ for all N, we have that $\mathbb{E}_{D,s_o}[\mathrm{rew}^N(\boldsymbol{r})] - \boldsymbol{p}_o + \boldsymbol{\varepsilon} \cdot N \geq -\mathbb{E}_{D,s_o}[Y_N] \geq \mathbf{0}$, and hence D satisfies $\mathrm{EE}_{s_o}(\boldsymbol{r} + \boldsymbol{\varepsilon})$ for every state s_o of D, concluding the proof.

□

D Assume-Guarantee Rules

D.1 Proof of Proposition 3

Proof. Fix π and take any Player □ strategy σ. Fix i. Using definition of c_i and Lemma 4, we know that, with probability one, $\mathrm{mp}(c_i) > 0$. We first show that the following holds for almost every path:

$$\frac{\mathrm{mp}(r_i)(\lambda)}{\mathrm{mp}(c_i)(\lambda)} = \mathrm{ratio}\left(\frac{r_i}{c_i}\right)(\lambda).$$

Again, by Lemma 4, with probability one, a path λ ends in a BSCC $\mathcal{B}$ and the limit inferior can be replaced by the true limit in the definitions of $\mathrm{mp}(c_i)$ and $\mathrm{mp}(r_i)$. Hence,

$$\frac{\mathrm{mp}(r_i)(\lambda)}{\mathrm{mp}(c_i)(\lambda)} = \frac{\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(r_i)(\lambda)}{\lim_{N\to\infty} \frac{1}{N+1}\mathrm{rew}^N(c_i)(\lambda)} = \lim_{N\to\infty} \frac{\frac{1}{N+1}\mathrm{rew}^N(r_i)(\lambda)}{\frac{1}{N+1}\mathrm{rew}^N(c_i)(\lambda)}.$$

There is no indeterminacy for this quotient of limits, as the denominator is positive and both numerator and denominator are bounded. Simplifying the $\frac{1}{N+1}$ term yields the equality $\frac{\mathrm{mp}(r_i)(\lambda)}{\mathrm{mp}(c_i)(\lambda)} = \lim_{N\to\infty} \frac{\mathrm{rew}^N(r_i)(\lambda)}{\mathrm{rew}^N(c_i)(\lambda)}$. This is almost surely equal to $\mathrm{ratio}\left(\frac{r_i}{c_i}\right)(\lambda) = \lim_{N\to\infty} \frac{\mathrm{rew}^N(r_i)(\lambda)}{1+\mathrm{rew}^N(c_i)(\lambda)}$ since $\mathrm{rew}^N(c_i)(\lambda) \to +\infty$ almost surely. It follows straightforwardly that $\mathrm{mp}(r_i)(\lambda) - v_i \cdot \mathrm{mp}(c_i)(\lambda) \geq 0$ holds almost surely exactly when $\mathrm{ratio}\left(\frac{r_i}{c_i}\right)(\lambda) = \frac{\mathrm{mp}(r_i)(\lambda)}{\mathrm{mp}(c_i)(\lambda)} \geq v_i$ holds almost surely. Since we have shown this for all strategies π and σ, almost every path λ, and all dimensions i, the proposition follows. □

D.2 Proof of Theorem 3

Projections. Given a state $s = (s_1, s_2)$ of $\mathcal{M}_1 \parallel \mathcal{M}_2$, the projection of s onto $\mathcal{M}_i$ is $s{\upharpoonright}_{\mathcal{M}_i} = s_i$, and for a distribution μ over states of $\mathcal{M}_1 \parallel \mathcal{M}_2$ we define its projection by $\mu{\upharpoonright}_{\mathcal{M}_i}(s_i) = \sum_{s{\upharpoonright}_{\mathcal{M}_i} = s_i} \mu(s)$. Given a (finite or infinite) path λ of $\mathcal{M}_1 \parallel \mathcal{M}_2$, the projection of λ onto $\mathcal{M}_i$, denoted by $\lambda{\upharpoonright}_{\mathcal{M}_i}$, is the path obtained from λ by projecting each state and distribution, and removing all moves with actions not in the alphabet of $\mathcal{M}_i$, together with the subsequent states.

Definition 3 (Strategy Projection, [13]). *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be PAs and σ a strategy of $\mathcal{M}_1 \parallel \mathcal{M}_2$. The* projection *of σ onto $\mathcal{M}_i$ is the strategy $\sigma\upharpoonright_{\mathcal{M}_i}$, defined such that for any finite path λ_i of $\mathcal{M}_i$ and transition $\mathsf{last}(\lambda_i) \xrightarrow{a} \mu_i$ (where $\mathsf{last}(\lambda)$ is last state of λ), $\sigma\upharpoonright_{\mathcal{M}_i}(\lambda_i)(a, \mu_i)$ is defined as*

$$\frac{\sum\{\mathrm{Pr}^{\sigma}_{\mathcal{M}_1\|\mathcal{M}_2}(\lambda) \cdot \sigma(\lambda)(a,\mu) \mid \lambda \in \Omega^{+}_{(\mathcal{M}_1\|\mathcal{M}_2)^\sigma} \wedge \lambda\upharpoonright_{\mathcal{M}_i} = \lambda_i \wedge \lambda\upharpoonright_{\mathcal{M}_i} = \mu_i\}}{\mathrm{Pr}^{\sigma\upharpoonright_{\mathcal{M}_i}}_{\mathcal{M}_i}(\lambda_i).}$$

Lemma 10. *Given PAs $\mathcal{M}_1$ and $\mathcal{M}_2$ with actions $\mathcal{A}_1$ and $\mathcal{A}_2$ resp., a specification φ defined on traces of $\mathcal{A} \subseteq \mathcal{A}_i$ for some $i \in \{1, 2\}$, and a strategy σ, we have $(\mathcal{M}_1 \parallel \mathcal{M}_2)^\sigma \models \varphi \Leftrightarrow \mathcal{M}_i^{\sigma\upharpoonright_{\mathcal{M}_i}} \models \varphi$.*

Proof. From Lemma 7.2.6 in [17], for PAs $\mathcal{M}_1$ and $\mathcal{M}_2$, and a strategy σ of $\mathcal{M}_1 \parallel \mathcal{M}_2$, for any trace w over actions $\mathcal{A} \subseteq \mathcal{A}_i$ we have $\mathrm{Pr}^{\sigma}_{\mathcal{M}_1\|\mathcal{M}_2}(w) = \mathrm{Pr}^{\sigma\upharpoonright_{\mathcal{M}_i}}_{\mathcal{M}_i}(w)$. Therefore, since φ is defined on traces of $\mathcal{A} \subseteq \mathcal{A}_i$, we have that $(\mathcal{M}_1 \parallel \mathcal{M}_2)^\sigma \models \varphi \Leftrightarrow \varphi(\mathrm{Pr}^{\sigma}_{\mathcal{M}_1\|\mathcal{M}_2}) \Leftrightarrow \varphi(\mathrm{Pr}^{\sigma_i\upharpoonright_{\mathcal{M}_i}}_{\mathcal{M}_i}) \Leftrightarrow \mathcal{M}_i^{\sigma_i\upharpoonright_{\mathcal{M}_i}} \models \varphi$. □

Proof (Of Theorem 3). Take any fair strategy σ of $\mathcal{M}_1 \parallel \mathcal{M}_2$. From Lemma 2 in [13], the projections $\sigma\upharpoonright_{M_2}$ and $\sigma\upharpoonright_{M_1}$ are fair strategies. We have, for $i \in \{1, 2\}$, that $\mathcal{M}_i \models^f \varphi^{G_i}$ implies $\mathcal{M}_i^{\sigma\upharpoonright_{M_i}} \models \varphi^{G_i}$, since $\sigma\upharpoonright_{M_i}$ is fair; this in turn implies $(\mathcal{M}_1 \parallel \mathcal{M}_2)^\sigma \models \varphi^{G_i}$, by Lemma 10, since $\mathcal{A}_{G_i} \subseteq \mathcal{A}_i$. Since σ was an arbitrary fair strategy of $\mathcal{M}_1 \parallel \mathcal{M}_2$, we have that this implies $\mathcal{M}_1 \parallel \mathcal{M}_2 \models^f \varphi^{G_1} \wedge \varphi^{G_2}$. □

D.3 Proof of Proposition 4

Proof. It is sufficient to show that this holds for single-dimensional reward structures r and c, as specifications over n-dimensional reward structures are defined on traces over the union of the actions of the individual dimensions.

Let $\mathfrak{A} = \{\tau\} \cup \mathcal{A}$. Given a path λ, $\lambda_{rc} \stackrel{\text{def}}{=} \mathrm{PROJ}_{\mathfrak{A}\setminus(\mathcal{A}_r\cup\mathcal{A}_c)}(\lambda)$ is the sequence of moves s such that such that $r(s) \neq 0$ or $c(s) > 0$. Given an index N and an infinite path λ, divide it into two, its prefix and suffix $\lambda^{pre}, \lambda^{post}$ such that $|\lambda^{pre}| = N$. We have for all N and m that $\mathrm{rew}^N(r)(\lambda) = \mathrm{rew}^{N'}(r)(\lambda_{rc})$, and $\mathrm{rew}^N(c)(\lambda) = \mathrm{rew}^{N'}(c)(\lambda_{rc})$, where $N' = |\lambda^{pre}_{rc}|$. Hence

$$\mathrm{rew}^N(r)(\lambda)/\mathrm{rew}^N(c)(\lambda) = \mathrm{rew}^{N'}(r)(\lambda_{rc})/\mathrm{rew}^{N'}(c)(\lambda_{rc}).$$

If λ contains an infinite number of actions in $\mathcal{A}_r \cup \mathcal{A}_c$, then, and as $N \to \infty$, also $N' \to \infty$. Hence, we have that $\mathrm{ratio}(r/c)(\lambda)$ is equal to

$$\liminf_{N\to\infty} \frac{\mathrm{rew}^N(r)(\lambda)}{1 + \mathrm{rew}^N(c)(\lambda)} = \liminf_{N\to\infty} \frac{\mathrm{rew}^N(r)(\lambda_{rc})}{1 + \mathrm{rew}^N(c)(\lambda_{rc})} \stackrel{\text{def}}{=} \mathrm{ratio}(r/c)(\lambda_{rc}).$$

If, on the other hand, λ contains only a finite number of actions in $\mathcal{A}_r \cup \mathcal{A}_c$, then $\mathrm{ratio}(r/c)(\lambda) = \mathrm{ratio}(r/c)(\lambda_{rc}) = 0$. Then, we have for any DTMCs D and D' with $\tilde{\mathrm{Pr}}_{D,\varsigma_0}(w) = \tilde{\mathrm{Pr}}_{D',\varsigma_0'}(w)$ for traces $w \in (\mathcal{A}_r \cup \mathcal{A}_c)^*$, that $\mathrm{Pr}_{D,\varsigma_0}(\mathrm{ratio}(r/c) \geq v) = \mathrm{Pr}_{D',\varsigma_0'}(\mathrm{ratio}(r/c) \geq v)$. □