Institute of Science and Technology

The Patience of Concurrent Stochastic Games with Safety and Reachability Objectives

Krishnendu Chatterjee and Rasmus Ibsen-Jensen and Kristoffer Arnsfelt Hansen

Technical Report No. IST-2015-322-v1+1
Deposited at 19 Feb 2015 10:14
http://repository.ist.ac.at/322/1/safetygames.pdf

IST Austria (Institute of Science and Technology Austria)
Am Campus 1
A-3400 Klosterneuburg, Austria

The Patience of Concurrent Stochastic Games with Safety and Reachability Objectives

Krishnendu Chatterjee Kristoffer Arnsfelt Hansen Rasmus Ibsen-Jensen

***Abstract*—We consider finite-state concurrent stochastic games, played by $k \geq 2$ players for an infinite number of rounds, where in every round, each player simultaneously and independently of the other players chooses an action, whereafter the successor state is determined by a probability distribution given by the current state and the chosen actions. We consider reachability objectives that given a target set of states require that some state in the target set is visited, and the dual safety objectives that given a target set require that only states in the target set are visited. We are interested in the complexity of stationary strategies measured by their *patience*, which is defined as the inverse of the smallest non-zero probability employed.**

Our main results are as follows: We show that in two-player zero-sum concurrent stochastic games (with reachability objective for one player and the complementary safety objective for the other player): (i) the optimal bound on the patience of optimal and ϵ-optimal strategies, for both players is doubly exponential; and (ii) even in games with a single non-absorbing state exponential (in the number of actions) patience is necessary. In general we study the class of non-zero-sum games admitting ε-Nash equilibria. We show that if there is at least one player with reachability objective, then doubly-exponential patience is needed in general for ε-Nash equilibrium strategies, whereas in contrast if all players have safety objectives, then the optimal bound on patience for ε-Nash equilibrium strategies is only exponential.

I. Introduction

Concurrent stochastic games. Concurrent stochastic games are played on finite-state graphs by k players for an infinite number of rounds. In every round, each player simultaneously and independently of the other players chooses moves (or actions). The current state and the chosen moves of the players determine a probability distribution over the successor state. The result of playing the game (or a *play*) is an infinite sequence of states and action vectors. These games with two players were introduced in a seminal work by Shapley [31], and have been one of the most fundamental and well-studied game models in stochastic graph games. Matrix games (or normal form games) can model a wide range problems with diverse applications, when there is a finite number of interactions [28], [34]. Concurrent stochastic games can be viewed as a finite set of matrix games, such that the choices made in the current game determine which game is played next, and is the appropriate model for many applications [16]. Moreover, in analysis of reactive systems, concurrent games provide the appropriate model for reactive systems with components that interact synchronously [11], [12], [2].

Objectives. An objective for a player defines the set of desired plays for the player, i.e., if a play belongs to the objective of the player, then the player wins and gets payoff 1, otherwise the player looses and gets payoff 0. The most basic objectives for concurrent games are the *reachability* and the *safety* objectives. Given a set F of states, a reachability objective with target set F requires that some state in F is visited at least once, whereas the dual safety objective with target set F requires that only states in F are visited. In this paper, we will only consider reachability and safety objectives. A zero-sum game consists of two players (player 1 and player 2), and the objectives of the players are complementary, i.e., a reachability objective with target set F for one player and a safety objective with target set complement of F for the other player. In this work, when we refer to zero-sum games we will imply that one player has reachability objective, and the other player has the complementary safety objective. Concurrent zero-sum games are relevant for analysis of synchronous reactive systems [11], [12], [13] as well as they can model many other interesting problems, such as two-player poker games [27].

Properties of strategies in zero-sum games. Given a zero-sum concurrent stochastic game, the player-1 *value* $v_1(s)$ of the game at a state s is the limit probability with which he can guarantee his objective against all strategies of player 2. The player-2 *value* $v_2(s)$ is analogously the limit probability with which player 2 can ensure his own objective against all strategies of player 1. Concurrent zero-sum games are determined [15], i.e., for each state s we have $v_1(s) + v_2(s) = 1$. A *strategy* for a player, given a history (i.e., finite prefix of a play) specifies a probability distribution over the actions. A *stationary* strategy does not depend on the history, but only on the current state. For $\epsilon \geq 0$, a strategy is ϵ-optimal for a state s for player i if it ensures his own objective with probability at least $v_i(s) - \epsilon$ against all strategies of the opponent. A 0-optimal strategy is an *optimal* strategy. In zero-sum concurrent stochastic games, there exist stationary optimal strategies for the player with safety objectives [29], [22]; whereas in contrast, for the player with reachability objectives optimal strategies do not exist in general, however, for every $\epsilon > 0$ there exists stationary ϵ-optimal strategies [15].

The significance of patience and roundedness of strategies. The basic decision problem is as follows: given a zero-sum concurrent stochastic game and a rational threshold λ, decide whether $v_1(s) \geq \lambda$. The basic decision problem is in PSPACEand is *square-root sum* hard [14][1]. Given the hardness of the basic decision problem, the next most relevant computational problem is to compute an approximation of the value. The computational complexity of the approximation

[1]The square-root sum problem is an important problem from computational geometry, where given a set of natural numbers $n_1, n_2, \ldots, n_k$, the question is whether the sum of the square roots exceed an integer b. The problem is not known to be in NP.

problem is closely related to the size of the description of ϵ-optimal strategies. Even for special cases of zero-sum concurrent stochastic games, namely *turn-based* stochastic games, where in each state at most one player can choose between multiple moves, the best known complexity results are obtained by guessing an optimal strategy and computing the value in the game obtained after fixing the guessed strategy. A strategy has patience p if p is the inverse of the smallest non-zero probability used by a distribution describing the strategy. A rational valued strategy has roundedness q if q is the greatest denominator of the probabilities used by the distributions describing the strategy. Note that if a strategy has roundedness q, then it also has patience at most q. The description complexity of a stationary strategy can be bounded by the roundedness. A stationary strategy with exponential roundedness, can be described using polynomially many bits, whereas the explicit description of stationary strategies with doubly-exponential patience is not polynomial. Thus obtaining upper bounds on the roundedness and lower bounds on the patience is at the heart of the computational complexity analysis of concurrent stochastic games.

Strategies in non-zero-sum games and roundedness. In non-zero-sum games, the most well-studied notion of equilibrium is *Nash equilibrium* [25], which is a strategy vector (one for each player), such that no player has an incentive of unilateral deviation (i.e., if the strategies of all other players are fixed, then a player cannot switch strategy and improve his own payoff). The existence of Nash equilibrium in non-zero-sum concurrent stochastic games where all players have safety objectives has been established in [30]. It follows from the strategy characterization of the result of [30] and our Lemma 41 that if such strategies have exponential roundness and forms an ϵ-Nash equilibrium, for a constant or even logarithmic number of players, for $\epsilon > 0$, then there will be polynomial-size witness for those strategies (and the approximation of a Nash equilibrium can be achieved in TFNP, see Remark 44). Thus again the notion of roundedness is at the core of the computational complexity of non-zero-sum games.

Previous results and our contributions. In this work we consider concurrent stochastic games (both zero-sum and non-zero-sum) where the objectives of the players are either reachability or safety. We first describe the relevant previous results and then our contributions.

Previous results. For zero-sum concurrent stochastic games, the optimal bound on patience and roundedness for ϵ-optimal strategies for reachability objectives, for $\epsilon > 0$, is doubly exponential [21], [19]. The doubly-exponential lower bound is obtained by presenting a family of games (namely, Purgatory) where the reachability player requires doubly-exponential patience (however, in this game the patience of the safety player is 1) [21], [19]; whereas the doubly-exponential upper bound is obtained by expressing the values in the existential theory of reals [21], [19]. In contrast to reachability objectives that in general do not admit optimal strategies, similar to safety objectives there are two related classes of concurrent stochastic games that admit optimal stationary strategies, namely, discounted-sum objectives, and ergodic concurrent games. For both these classes the optimal bound on patience and roundedness for ϵ-optimal strategies, for $\epsilon > 0$, is exponential [10], [23]. The optimal bound on patience and roundedness for optimal and ϵ-optimal strategies, for $\epsilon > 0$, for safety objectives has been an open problem.

Our contributions. Our main results are as follows:

1) *Lower bound: general.* We show that in zero-sum concurrent stochastic games, a lower bound on patience of optimal and ϵ-optimal strategies, for $\epsilon > 0$, for safety objectives is doubly exponential (in contrast to the above mentioned related classes of games that admit stationary optimal strategies and require only exponential patience). We present a family of games (namely, Purgatory Duel) where the optimal and ϵ-optimal strategies, for $\epsilon > 0$, for both players require doubly-exponential patience.
2) *Lower bound: three states.* We show that even in zero-sum concurrent stochastic games with three states of which two are absorbing (sink states with only self-loop transitions) the patience required for optimal and ϵ-optimal strategies, for $\epsilon > 0$, is exponential (in the number of actions). An optimal (resp., ϵ-optimal, for $\epsilon > 0$) strategy in a game with three states (with two absorbing states) is basically an optimal (resp., ϵ-optimal) strategy of a matrix game, where some entries of the matrix game depends on the value of the non-absorbing state (as some transitions of the non-absorbing state can lead to itself). In standard matrix games, the patience for ϵ-optimal strategies, for $\epsilon > 0$, is only logarithmic [26]; and perhaps surprisingly in contrast we show that the patience for ϵ-optimal strategies in zero-sum concurrent stochastic games with three states is exponential (i.e., there is a doubly-exponential increase from logarithmic to exponential).
3) *Upper bound.* We show that in zero-sum concurrent stochastic games, an upper bound on the patience of optimal strategies and an upper bound on the patience and roundedness of ε-optimal strategies, for $\varepsilon > 0$, is as follows: (a) doubly exponential in general; and (b) exponential for the safety player if the number of value classes (i.e., the number of different values in the game) is constant. Hence our upper bounds on roundedness match our lower bound results for patience. Our results also imply that if the number of value classes is constant, then the basic decision problem is in coNP(resp., NP) if player 1 has reachability (resp., safety) objective.
4) *Non-zero-sum games.* We consider non-zero-sum concurrent stochastic games with reachability and safety objectives. First, we show that it easily follows from our example family of Purgatory Duel that if there are at least two players and there is at least one player with reachability objective, then a lower bound on patience for ϵ-Nash equilibrium is doubly exponential, for $\epsilon > 0$, for *all* players. In contrast, we show that if all players have safety objectives, then the optimal bound on patience of strategies for ϵ-Nash equilibrium is exponential, for $\epsilon > 0$

(i.e., for upper bound we show that there always exists an ϵ-Nash equilibrium where the strategy of each player requires at most exponential roundedness; and there exists a family of games, where for any ϵ-Nash equilibrium the strategies of all players require at least exponential patience).

In summary, we present a complete picture of the patience and roundedness required in zero-sum concurrent stochastic games, and non-zero-sum concurrent stochastic games with safety objectives for all players. Also see Section VII for a discussion on important technical aspects of our results.

Distinguishing aspects of safety and reachability. While the optimal bound on patience and roundedness we establish in zero-sum concurrent stochastic games for the safety player matches that for the reachability player, there are many distinguishing aspects for safety as compared to reachability in terms of the number of value classes (as shown in Table I). For the reachability player, if there is one value class, then the patience and roundedness required is linear: it follows from the results of [6] that if there is one value class then all the values must be either 1 or 0; and if all states have value 0, then any strategy is optimal, and if all states have value 1, then it follows from [13], [7] that there is an almost-sure winning strategy (that ensures the objective with probability 1) from all states and the optimal bound on patience and roundedness is linear. The family of game graphs defined by Purgatory has two value classes, and the reachability player requires doubly exponential patience and roundedness, even for two value classes. In contrast, if there are (at most) two value classes, then again the values are 1 and 0; and in value class 1, the safety player has an optimal strategy that is stationary and deterministic (i.e., a positional strategy) and has patience and roundedness 1 [13], and in value class 0 any strategy is optimal. While for two value classes, the patience and roundedness is 1 for the safety player, we show that for three value classes (even for three states) the patience and roundedness is exponential, and in general the patience and roundedness is doubly exponential (and such a finer characterization does not exist for reachability objectives). Finally, for non-zero-sum games (as we establish), if there are at least two players, then even in the presence of one reachability player, the patience required is at least doubly exponential, whereas if all players have safety objectives, the patience required is only exponential.

# Value classes	Reachability	Safety
1	Linear	One
2	Double-exponential	One
3	Double-exponential	**Exponential LB, Theorem 29**
Constant	Double-exponential	**Exponential UB, Corollary 34**
General	Double-exponential	**Double-exponential LB, Theorem 20 UB, Corollary 34**

TABLE I

STRATEGY COMPLEXITY (I.E., PATIENCE AND ROUNDEDNESS OF ϵ-OPTIMAL STRATEGIES, FOR $\epsilon > 0$) OF REACHABILITY VS SAFETY OBJECTIVES DEPENDING ON THE NUMBER OF VALUE CLASSES. OUR RESULTS ARE BOLD FACED, AND LB (RESP., UB) DENOTES LOWER (RESP., UPPER) BOUND ON PATIENCE (RESP., ROUNDEDNESS).

Our main ideas. Our most interesting results are the doubly-exponential and exponential lower bound on the patience and roundedness in zero-sum games. We now present a brief overview about the lower bound example.

The game of *Purgatory* [21], [19] is a concurrent reachability game [13] that was defined as an example showing that the *reachability* player must, in order to play near optimally, use a strategy with non-zero probabilities that are *doubly exponentially* small in the number of states of the game (i.e., the patience is doubly exponential).

In this paper we present another example of a reachability game where this is the case for the *safety* player as well. The game Purgatory consists of a (potentially infinite) sequence of *escape attempts*. In an escape attempt one player is given the role of the *escapee* and the other player is given the role as the *guard*. An escape attempt consists of at most N rounds. In each round, the guard selects and hides a number between 1 and m, and the escapee must try to guess the number. If the escapee successfully guesses the number N times, the game ends with the escapee as the winner. If the escapee incorrectly guesses a number which is strictly larger than the hidden number, the game ends with the guard as the winner. Otherwise, if the escapee incorrectly guesses a number which is strictly smaller than the hidden number, the escape attempt is over and the game continues.

The game of Purgatory is such that the reachability player is always given the role of the escapee, and the safety player is always given the role of the guard. If neither player wins during an escape attempt (meaning there is an infinite number of escape attempts) the safety player wins. Purgatory may be modelled as a concurrent reachability game consisting of N non-absorbing positions in which each player has m actions. The value of each non-absorbing position is 1. This means that the reachability player has, for any $\varepsilon > 0$, a stationary strategy that wins from each non-absorbing position with probability at least $1 - \varepsilon$ [15], but such strategies must have doubly-exponential patience. In fact for N sufficiently large and $m \geq 2$, such strategies must have patience at least $2^{m^{N/3}}$ for $\varepsilon = 1 - 4m^{-N/2}$ [19]. For the safety player however, the situation is simple: *any* strategy is optimal.

We introduce a game we call the *Purgatory Duel* in which the safety player must also use strategies of doubly-exponential patience to play near optimally. The main idea of the game is that it forces the safety player to behave as a reachability player. We can describe the new game as a variation on the above description of the Purgatory game. The Purgatory Duel consists also of a (potentially infinite) sequence of escape attempts. But now, before each escape attempt the role of the escapee is given to each player with probability $\frac{1}{2}$, and in each escape attempt the rules are as described above. The game remains asymmetric in the sense that if neither player wins during an escape attempt, the safety player wins.

The Purgatory Duel may be modelled as a concurrent reachability game consisting of $2N + 1$ non-absorbing positions, in which each player has m actions, except for a single position

where the players each have just a single action.

The key non-trivial aspects of our proof are as follows: first, is to come up with the family of games, namely, Purgatory Duel, where the ϵ-optimal strategies, for $\epsilon \geq 0$, for the players are symmetric, even though the objectives are complementary; and then the precise analysis of the game needs to combine and extend several ideas, such as refined analysis of matrix games, and analysis of perturbed Markov decision processes (MDPs) which are one-player stochastic games.

Related work. We have already discussed the relevant related works such as [29], [22], [15], [14], [21], [19], [13] on zero-sum games. We discuss relevant related works for non-zero-sum games. The computational complexity of *constrained* Nash equilibrium, which asks the existence of Nash (or ϵ-Nash, for $\epsilon > 0$) equilibrium that guarantees at least a payoff vector has been studied. The constrained Nash equilibrium problem is undecidable even for turn-based stochastic games, or concurrent deterministic games with randomized strategies [32]. The complexity of constrained Nash equilibrium in concurrent deterministic games with pure strategies has been studied in [4], [5]. In contrast, we study the complexity of computing some Nash equilibrium in randomized strategies in concurrent stochastic games, and our result on patience implies that with safety objectives for all players the approximation of some Nash equilibrium can be achieved in TFNP.

II. Definitions

Other number. Given a number $i \in \{1, 2\}$ let $\widehat{i}$ be the other number, i.e., if $i = 1$, then $\widehat{i} = 2$ and if $i = 2$, then $\widehat{i} = 1$.

Probability distributions. A *probability distribution* d over a finite set Z, is a map $d : Z \to [0, 1]$, such that $\sum_{z \in Z} d(z) = 1$. Fix a probability distribution d over a set Z. The distribution d is *pure (Dirac)* if $d(z) = 1$ for some $z \in Z$ and for convenience we overload the notation and let $d = z$. The *support* $\mathrm{Supp}(d)$ is the subset Z' of Z, such that $z \in Z'$ if and only if $d(z) > 0$. The distribution d is *totally mixed* if $\mathrm{Supp}(d) = Z$. The *patience* of d is $\max_{z \in \mathrm{Supp}(d)} \frac{1}{d(z)}$, i.e., the inverse of the minimum non-zero probability. The *roundedness* of d, if $d(z)$ is a rational number for all $z \in Z$, is the greatest denominator of $d(z)$. Note that roundness of d is always at least the patience of d. Given two elements $z, z' \in Z$, the probability distribution $d = \mathsf{U}(z, z')$ over Z is such that $d(z) = d(z') = \frac{1}{2}$. Let $\Delta(Z)$ be the set of all probability distributions over Z.

Concurrent game structure. A concurrent game structure for k players, consists of (1) a finite set of *states* S, of size N; and (2) for each state $s \in S$ and each player i a set A^i_s of *actions* (and $A^i = \bigcup_s A^i_s$ is the set of all actions for player i, for each i; and $A = \bigcup_i A^i$ is the set of all actions) such that A^i_s consists of at most m actions; and (3) a stochastic *transition function* $\delta : S \times A^1 \times A^2 \times \cdots \times A^k \to \Delta(S)$. Also, a state s is *deterministic* if $\delta(s, a_1, a_2, \ldots, a_k)$ is pure (deterministic), for all $a_i \in A^i_s$ and for all i. A state s is called *absorbing* if $A^i_s = \{a\}$ for all i and $\delta(s, a, a, \ldots, a) = s$. The number $\delta_{\min}$ is

$$\min_{s, a_1, \ldots, a_k, s' \in \mathrm{Supp}(\delta(s, a_1, a_2, \ldots, a_k))} (\delta(s, a_1, a_2, \ldots, a_k)(s')) \ ,$$

i.e., the smallest non-zero transition probability.

Safety and reachability objectives. Each player i, who has a safety or reachability objective, is identified by a pair (t_i, S^i), where $t_i \in \{\mathsf{Reach}, \mathsf{Safety}\}$ and $S^i \subseteq S$.

Concurrent games and how to play them. Fix a number k of players. A concurrent game consists of a concurrent game structure for k players and for each player i a pair (t_i, S^i), identifying the type of that player. The game G, starting in state s, is played as follows: initially a pebble is placed on $v_0 := s$. In each time step $T \geq 0$, the pebble is on some state v_T and each player selects (simultaneously and independently of the other players, like in the game rock-paper-scissors) an action $a^i_{T+1} \in A^i_{v_T}$. Then, the game selects v_{T+1} according to the probability distribution $\delta(v_T, a^1_{T+1}, a^2_{T+1}, \ldots, a^k_{T+1})$ and moves the pebble onto v_{T+1}. The game then continues with time step $T+1$ (i.e., the game consists of infinitely many time steps). For a round T, let a_{T+1} be the vector of choices of the actions for the players, i.e., $(a_{T+1})_i$ is the choice of player i, for each i. Round 0 is identified by v_0 and round $T > 0$ is then identified by the pair (a_T, v_T). A *play* P_s, starting in state $v_0 = s$, is then a sequence of rounds

$$(v_0, (a_1, v_1), (a_2, v_2), \ldots, (a_T, v_T), \ldots) \ ,$$

and for each ℓ a prefix of P^ℓ_s of length ℓ is then

$$(v_0, (a_1, v_1), (a_2, v_2), \ldots, (a_T, v_T), \ldots, (a_\ell, v_\ell)) \ ,$$

and we say that P^ℓ_s *ends in* v_ℓ. For each i, player i wins in the play P_s, if $t_i = \mathsf{Safety}$ and $v_T \in S_i$ for all $T \geq 0$; or if $t_i = \mathsf{Reach}$ and $v_T \in S_i$, for some $T \geq 0$. Otherwise, player i loses. For each i, player i tries to maximize the probability that he wins.

Strategies. Fix a player i. A strategy is a recipe to choose a probability distribution over actions given a finite prefix of a play. Formally, a strategy σ_i for player i is a map from P^ℓ_s, for a play P_s of length ℓ starting at state s, to a distribution over $A^i_{v_\ell}$. Player i *follows* a strategy σ_i, if given the current prefix of a play is P^ℓ_s, he selects $a_{\ell+1}$ according to $\sigma_i(P^\ell_s)$, for all plays P_s starting at s and all lengths ℓ. A strategy σ_i for player i, is *stationary*, if for all ℓ and ℓ', and all pair of plays P_s and $P'_{s'}$, starting at states s and s' respectively, such that P^ℓ_s and $(P')^{\ell'}_{s'}$ ends in the same state t, we have that $\sigma_i(P^\ell_s) = \sigma_i((P')^{\ell'}_{s'})$; and we write $\sigma_i(t)$ for the unique distribution used for prefix of plays ending in t. The *patience* (resp., *roundedness*) of a strategy σ_i is the supremum of the patience (resp., roundedness) of the distribution $\sigma_i(P^\ell_s)$, over all plays P_s starting at state s, and all lengths ℓ. Also, a strategy σ_i is *pure* (resp., *totally mixed*) if $\sigma_i(P^\ell_s)$ is pure (resp., totally mixed), for all plays P_s starting at s and all lengths ℓ. A strategy is *positional* if it is pure and stationary. For each player i, let Σ^i be the set of all strategies for the respective player.

Strategy profiles and Nash equilibria. A *strategy profile* $\sigma = (\sigma_i)_i$ is a vector of strategies, one for each player. A strategy profile σ defines a unique probability measure on plays, denoted $\Pr_\sigma$, when the players follow their respective strategies [33]. Let $u(G, s, \sigma, i)$ be the probability that player i wins the game G when the players follow σ and the play starts in s (i.e., the utility or payoff for player i). Given a strategy profile $\sigma = (\sigma_i)_i$ and a strategy σ'_i for player i, the strategy profile $\sigma[\sigma'_i]$ is the strategy profile where the strategy for player i is σ'_i and the strategy for player j is σ_j for $j \neq i$. Fix a state s and $\varepsilon \geq 0$. A strategy profile σ forms an *ε-Nash equilibrium from state s* if for all i and all strategies σ'_i for player i, we have that $u(G, s, \sigma, i) \geq u(G, s, \sigma[\sigma'_i], i) - \varepsilon$. A strategy profile σ forms an *ε-Nash equilibrium* if it forms an ε-Nash equilibrium from all states s. Also a strategy profile forms a *Nash equilibrium* (resp., from state s, for some s) if it forms a 0-Nash equilibrium (resp., from state s). We say that a strategy profile has a property (e.g., is stationary) if each of the strategies in the profile has that property.

A. Zero-sum concurrent stochastic games

A zero-sum game consists of two players with complementary objectives. Since we only consider reachability and safety objectives, a zero-sum concurrent stochastic game consists of a two-player concurrent stochastic game with reachability objective for player 1 and the complementary safety objective for player 2 (such a game is also referred to as concurrent reachability game).

Concurrent reachability game. A concurrent reachability game is a concurrent game with two players, identified by (Reach, S^1) and $(\mathsf{Safety}, S \setminus S^1)$. Observe that in such games, exactly one player wins each play (this implies that the games are zero-sum). Note that for all strategy profiles σ we have $u(G, s, \sigma, 1) + u(G, s, \sigma, 2) = 1$. For ease of notation and tradition, we write $u(G, s, \sigma_1, \sigma_2)$ for $u(G, s, \sigma_1, \sigma_2, 1)$, for all concurrent reachability games G, states s, and strategy profiles $\sigma = (\sigma_1, \sigma_2)$. Also if the game G is clear from context we drop it from the notation.

Values of concurrent reachability games. Given a concurrent reachability game G, the *upper value* of G starting in s is

$$\overline{\mathrm{val}}(G, s) = \sup_{\sigma_1 \in \Sigma^1} \inf_{\sigma_2 \in \Sigma^2} u(G, s, \sigma_1, \sigma_2) \ ;$$

and the *lower value* of G starting in s is

$$\underline{\mathrm{val}}(G, s) = \inf_{\sigma_2 \in \Sigma^2} \sup_{\sigma_1 \in \Sigma^1} u(G, s, \sigma_1, \sigma_2) \ .$$

As shown by [15] we have that

$$\mathrm{val}(G, s) := \overline{\mathrm{val}}(G, s) = \underline{\mathrm{val}}(G, s) \ ;$$

and this common number is called the *value* of s. We will sometimes write $\mathrm{val}(s)$ for $\mathrm{val}(G, s)$ if G is clear from the context. We will also write val for the vector where $\mathrm{val}_s = \mathrm{val}(s)$.

(ε-)optimal strategies for concurrent reachability games. For an $\varepsilon \geq 0$, a strategy σ_1 for player 1 (resp., σ_2 for player 2) is called *ε-optimal* if for each state s we have that $\mathrm{val}(s) - \varepsilon \leq \inf_{\sigma_2 \in \Sigma^2} u(s, \sigma_1, \sigma_2)$ (resp., $\mathrm{val}(s) + \varepsilon \geq \sup_{\sigma_1 \in \Sigma^1} u(s, \sigma_1, \sigma_2)$). For each i, a strategy σ_i for player i is called *optimal* if it is 0-optimal. There exist concurrent reachability games in which player 1 does not have optimal strategies, see [15] for an example[2]. On the other hand in all games G player 1 has a stationary ε-optimal strategy for each $\varepsilon > 0$. In all games player 2 has an optimal stationary strategy (thus also an ε-optimal stationary strategy for all $\varepsilon > 0$) [29], [22]. Also, given a stationary strategy σ_1 for player 1 we have that there exists a positional strategy σ_2, such that $u(s, \sigma_1, \sigma_2) = \inf_{\sigma'_2 \in \Sigma^2} u(s, \sigma_1, \sigma'_2)$, i.e., we only need to consider positional strategies for player 2. Similarly, we only need to consider positional strategies for player 1, if we are given a stationary strategy for player 2.

(ε-)optimal strategies compared to (ε-)Nash equilibria. It is well-known and easy to see that for concurrent reachability games, a strategy profile $\sigma = (\sigma_1, \sigma_2)$ is optimal if and only if σ forms a Nash equilibrium. Also, if σ_1 is ε-optimal and σ_2 is ε'-optimal, for some ε and ε', then $\sigma = (\sigma_1, \sigma_2)$ forms an $(\varepsilon + \varepsilon')$-Nash equilibrium. Furthermore, if $\sigma = (\sigma_1, \sigma_2)$ forms an ε-Nash equilibrium, for some ε, then σ_1 and σ_2 are ε-optimal[3].

Markov decision processes and Markov chains. For each player i, a *Markov decision process (MDP) for player i* is a concurrent game where the size of A^j_s is 1 for all s and $j \neq i$. A *Markov chain* is an MDP for each player (that is the size of A^j_s is 1 for all s and j). A *closed recurrent set* of a Markov chain G is a maximal (i.e., no closed recurrent set is a subset of another) set $S' \subseteq S$ such that for all pairs of states $s, s' \in S$, the play starting at s reaches state s' eventually with probability 1 (note that it does not depend on the choices of the players as we have a Markov chain). For all starting states, eventually a closed recurrent set is reached with probability 1, and then plays stay in the closed recurrent set. Observe that fixing a stationary strategy for all but one player in a concurrent game, the resulting game is an MDP for the remaining player. Hence, fixing a stationary strategy for each player gives a Markov chain.

B. Matrix games and the value iteration algorithm

A (two-player, zero-sum) matrix game consists of a matrix $M \in \mathbb{R}^{r \times c}$. We will typically let M refer to both the matrix game and the matrix and it should be clear from the context what it means. A matrix game M is played as follows: player 1 selects a row a_1 and at the same time, without knowing which row was selected by player 1, player 2 selects a column a_2. The *outcome* is then M_{a_1, a_2}. Player 1 then tries to maximize the outcome and player 2 tries to minimize it.

Strategies in matrix games. A *strategy* σ_1 (resp., σ_2) for player 1 (resp., player 2) is a probability distribution over the rows (resp., columns) of M. A strategy profile $\sigma = (\sigma_1, \sigma_2)$ is

[2]note that it is not because that we require the strategy to be optimal for each start state, since if there was one for each start state separately then there would be one for all, since this is not just for stationary strategies

[3]observe that the two latter properties implies the former, but all are included to make it clear that there is a strong connection

a pair of strategies, one for each player. Given a strategy profile $\sigma = (\sigma_1, \sigma_2)$ the payoff $u(M, \sigma_1, \sigma_2)$ under those strategies is the expected outcome if player 1 picks row a_1 with probability $\sigma_1(a_1)$ and player 2 picks column a_2 with probability $\sigma_2(a_2)$ for each a_1 and a_2, i.e.,

$$u(M, \sigma_1, \sigma_2) = \sum_{a_1} \sum_{a_2} M_{a_1,a_2} \cdot \sigma_1(a_1) \cdot \sigma_2(a_2) \ .$$

Values in matrix games. The *upper value* of a matrix game is $\overline{\text{val}}(M) = \sup_{\sigma_1} \inf_{\sigma_2} u(M, \sigma_1, \sigma_2)$. The *lower value* of a matrix game is $\underline{\text{val}}(M) = \inf_{\sigma_2} \sup_{\sigma_1} \sum_{a_1} u(M, \sigma_1, \sigma_2)$. One of the most fundamental results in game theory, as shown by [34], is that $\text{val}(M) := \overline{\text{val}}(M) = \underline{\text{val}}(M)$. This common number is called the *value*.

(ε-)optimal strategies in matrix games. A strategy σ_1 for player 1 is *ε-optimal*, for some number $\varepsilon \geq 0$ if $\text{val}(M) - \varepsilon \leq \inf_{\sigma_2} u(M, \sigma_1, \sigma_2)$. Similarly, a strategy σ_2 for player 2 is *ε-optimal*, for some number $\varepsilon \geq 0$ if $\text{val}(M) + \varepsilon \geq \sup_{\sigma_1} u(M, \sigma_1, \sigma_2)$. A strategy is *optimal* if it is 0-optimal. There exists an optimal strategy for each player in all matrix games [34]. Given an optimal strategy σ_1 for player 1, consider the vector $\overline{v}$, such that $\overline{v}_j = u(M, \sigma_1, j)$ for each column j. Then we have that $\overline{v}_j = \text{val}(M)$ for each j such that there exists an optimal strategy σ_2 for player 2, where $\sigma_2(j) > 0$. Similar analysis holds for optimal strategies of player 2. This also shows that given an optimal strategy σ_1 for player 1 we have that $u(M, \sigma_1, \sigma_2)$ is minimized for some pure strategy σ_2 and similarly for optimal strategies σ_2 for player 2. Given a matrix game M, an optimal strategy for each player and the value of M can be computed in polynomial time using linear programming.

The matrix game $A^s[\overline{v}]$ and A^s. Fix a concurrent reachability game G. Given a vector $\overline{v}$ in $\mathbb{R}^S$ and a state s (in G), the matrix game $A^s[\overline{v}] = [a_{i,j}]$ is the matrix game where $a_{i,j} = \sum_{s' \in S} \delta(s, i, j)(s') \cdot \overline{v}_{s'}$. Given a state s, the matrix game A^s is the matrix game $A^s[\text{val}]$. As shown by [29], [22], each optimal stationary strategy σ_2 for player 2 in G is such that for each state s the distribution $\sigma_2(s)$ is an optimal strategy in the matrix game A^s. Also, conversely, if $\sigma_2(s)$ is an optimal strategy in A^s for each s, then σ_2 is an optimal stationary strategy in G. Furthermore, also as shown by [29], [22], we have that $\text{val}(s) = \text{val}(A^s)$ for each state s.

The value iteration algorithm. The conceptually simplest algorithm for concurrent reachability games is the *value iteration* algorithm, which is an iterative approximation algorithm. The idea is as follows: Given a concurrent reachability game G, consider the game G^t where a *time-limit* t (some non-negative integer) has been introduced. The game G^t is then played as G, except that player 2 wins if the time-limit is exceeded (i.e., he wins after round t unless a state in S^1 has been reached before that). (The game G^t has a value like in the above definition of matrix games since the game only has a finite number of pure strategies and thus can be reduced to a matrix game). The value of G^t starting in state s then converges to the value of G starting in s as t goes to infinity as shown by [15]. More precisely, the algorithm is defined on a vector $\overline{v}^t$ which is the vector where $\overline{v}^t_s$ is the value of G^t starting in s. We can compute $\overline{v}^t_s$ recursively for increasing t as follows

$$\overline{v}^t_s = \begin{cases} 1 & \text{if } s \in S^1 \\ 0 & \text{if } s \notin S^1 \text{ and } t = 0 \\ \text{val}(A^s[\overline{v}^{t-1}]) & \text{if } s \notin S^1 \text{ and } t \geq 1 \ . \end{cases}$$

We have that $\overline{v}^t_s \leq \overline{v}^{t+1}_s \leq \text{val}(s)$ for all t and s, and for all s we have $\lim_{t\to\infty} \overline{v}^t_s = \text{val}(s)$, as shown by [15]. As shown by [19], [20] the smallest time-limit t such that $\overline{v}^t_s \geq \text{val}(s) - \varepsilon$ can be as large as $\varepsilon^{-m^{\Omega(N)}}$ for some games (of N states and at most m actions in each state for each player) and s, for $\varepsilon > 0$. On the other hand it is also at most $\varepsilon^{-m^{O(N^2)}}$ as shown by [19].

III. Zero-sum Concurrent Stochastic Games: Patience Lower Bound

In this section we will establish the doubly-exponential lower bound on patience for zero-sum concurrent stochastic games. First we define the game family, namely, *Purgatory Duel* and we also recall the family *Purgatory* that will be used in our proofs. We split our proof about the patience in Purgatory Duel in three parts. First we present some refined analysis of matrix games, and use the analysis to first prove the lower bound for optimal strategies, and then for ε-optimal strategies, for $\varepsilon > 0$.

The Purgatory Duel. In this paper we specifically focus on the following concurrent reachability game, the *Purgatory Duel*[4], defined on a pair of parameters (n, m). The game consists of $N = 2n + 3$ states, namely $\{v^1_1, v^1_2, \ldots, v^1_n, v^2_1, v^2_2, \ldots, v^2_n, v_s, \top, \bot\}$ and all but v_s are deterministic. To simplify the definition of the game, let $v^1_0 = v^2_{n+1} = \bot$ and $v^2_0 = v^1_{n+1} = \top$. The states $\top$ and $\bot$ are absorbing. For each $i \in \{1, 2\}$ and $j \in \{1, \ldots, n\}$, the state v^i_j is such that $A^1_{v^i_j} = A^2_{v^i_j} = \{1, 2, \ldots, m\}$ and for each a_1, a_2 we have that

$$\delta(v^i_j, a_1, a_2) = \begin{cases} v_s & \text{if } a_1 > a_2 \\ v^i_0 & \text{if } a_1 < a_2 \\ v^i_{j+1} & \text{if } a_1 = a_2 \ . \end{cases}$$

Finally, $A^1_{v_s} = A^2_{v_s} = \{a\}$ and $\delta(v_s, a, a) = \mathsf{U}(v^1_1, v^2_1)$. Furthermore, $S^1 = \{\top\}$. There is an illustration of the Purgatory Duel with $m = n = 2$ in Figure 1.

The game Purgatory. We will also use the game *Purgatory* as defined by [19] (and also in [21] for the case of $m = 2$). Purgatory is similar to the Purgatory Duel and hence the similarity in names. Purgatory is also defined on a pair of parameters (n, m). The game consists of $N = n + 2$ states, namely, $\{v_1, v_2, \ldots, v_n, \top, \bot\}$ and each state is deterministic. To simplify the definition of the game, let $v_{n+1} = \top$. For

[4]To allow a more compact notation, we have here exchanged the criterias for when the safety player wins as a guard and when the escape attempt ends, as compared to the textual description of the game given in the introduction.

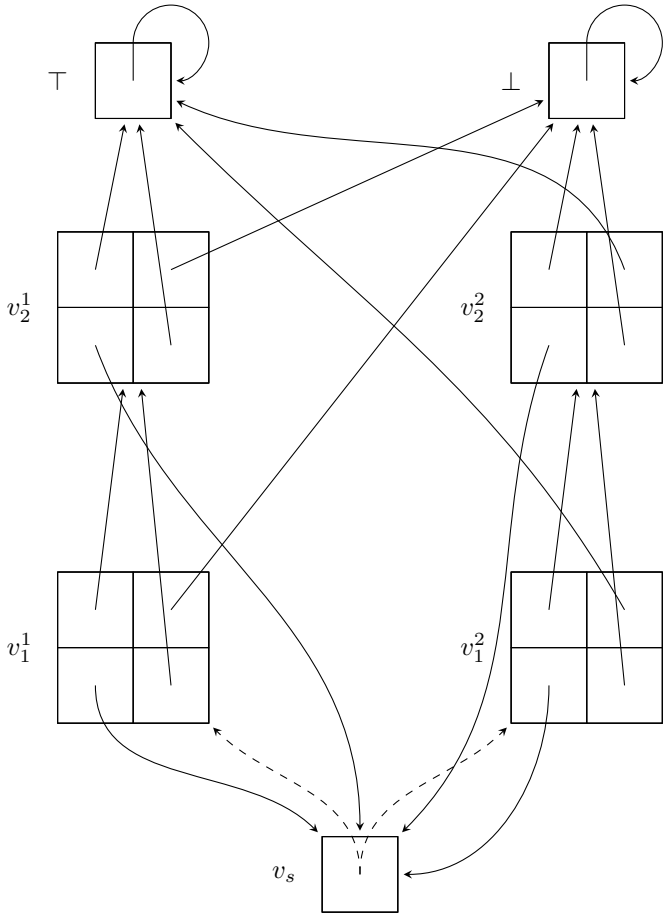

Fig. 1. An illustration of the Purgatory Duel with $m = n = 2$. The two dashed edges have probability $\frac{1}{2}$ each.

each $j \in \{1, \ldots, n\}$, the state v_j is such that $A^1_{v_j} = A^2_{v_j} = \{1, 2, \ldots, m\}$ and for each a_1, a_2 we have that

$$\delta(v_j, a_1, a_2) = \begin{cases} v_1 & \text{if } a_1 > a_2 \\ \bot & \text{if } a_1 < a_2 \\ v_{j+1} & \text{if } a_1 = a_2 \ . \end{cases}$$

The states $\top$ and $\bot$ are absorbing. Furthermore, $S^1 = \{\top\}$. There is an illustration of Purgatory with $m = n = 2$ in Figure 2.

A. Analysis of matrix games

In this section we present some refined analysis of some simple matrix games, which we use in the later sections to find optimal strategies for the players and the values of the states in the Purgatory Duel.

Definition 1. *Given a positive integer m and reals x, y and z, let $M^{x,y,z,m}$ be the $(m \times m)$-matrix with x below the diagonal, y in the diagonal and z above the diagonal, i.e.,*

$$M^{x,y,z,m} = \begin{pmatrix} y & z & z & \ldots & z \\ x & y & z & \ldots & z \\ \vdots & x & \ddots & \ddots & \vdots \\ x & \vdots & \ddots & y & z \\ x & x & \ldots & x & y \end{pmatrix} .$$

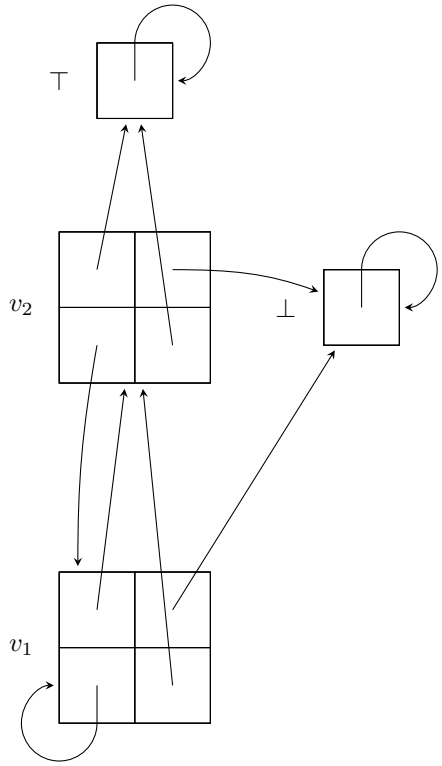

Fig. 2. An illustration of Purgatory with $m = n = 2$.

We first explain the significance of the matrix game $M^{x,y,z,m}$ in relation to Purgatory Duel. Consider the Purgatory Duel defined on parameters (n, m), for some n. We will later establish that for any j, let v (resp., v') be state v_j^1 (resp., v_j^2) of the Purgatory Duel, then we have that $A^v = M^{0,\text{val}(v_{j+1}^1),\text{val}(v_s),m}$ (resp., $A^{v'} = M^{1,\text{val}(v_{j+1}^2),\text{val}(v_s),m}$). In this section we show that for $0 < z < y$ we have that $M = M^{0,y,z,m}$ is such that $\text{val}(M) > z$ and each optimal strategy for either player is totally mixed. Similarly, for $1 > z' > y'$ we show that $M' = M^{1,y',z',m}$ is such that $\text{val}(M') < z$ and each optimal strategy for either player is totally mixed. We also compute the value and the patience of each optimal strategy in the matrix game $M^{0,\frac{1}{2}+\varepsilon,\frac{1}{2},m}$ (since we will establish in the next section, using the results of this section, that $\text{val}(v_s) = \frac{1}{2}$ and $\text{val}(v_j^1) > \text{val}(s)$ for all j).

Lemma 2. *For all positive integers m and reals y and z such that $0 < z < y$, the matrix game $M = M^{0,y,z,m}$ has value strictly above z.*

Proof. Let $\varepsilon > 0$ be some number to be defined later. Consider the probability distribution σ_1^ε given by

$$\sigma_1^\varepsilon(a) = \begin{cases} \varepsilon^{a-1} - \varepsilon^a & \text{if } 1 \le a \le m-1 \\ \varepsilon^{m-1} & \text{if } a = m \ . \end{cases}$$

If player 2 plays column a against σ_1, for $a \le m-1$, then the payoff $u(M, \sigma_1, a)$ is $y \cdot (\varepsilon^{a-1} - \varepsilon^a) + y \cdot (1 - \varepsilon^{a-1})$; and if player 2 plays column m, then the payoff $u(M, \sigma_1, m)$ is $y \cdot (\varepsilon^{m-1}) + z \cdot (1 - \varepsilon^{m-1})$. For any ε such that $y \cdot (1 - \varepsilon) > z$, the payoff is strictly greater than z implying that the value of M is strictly greater than z. □

Lemma 3. *For all positive integers m and reals y and z such that $0 < z < y$, each optimal strategy for player 1 in the matrix game $M^{0,y,z,m}$ is totally mixed.*

Proof. Consider some strategy σ_1 for player 1 in $M^{0,y,z,m}$ which is not totally mixed. Thus there exists some row a, where $\sigma_1(a) = 0$. Consider the pure strategy σ_2 that plays column a with probability 1. Playing σ_1 against σ_2 ensures that each outcome is either z or 0, i.e., the payoff is at most z which is strictly less than the value by Lemma 2. □

Lemma 4. *For all positive integers m and reals y and z such that $0 < z < y$, each optimal strategy for player 2 in the matrix game $M = M^{0,y,z,m}$ is totally mixed.*

Proof. Given a strategy σ_1 for player 1 and two rows a' and a'', let the strategy $\sigma_1[a' \to a'']$ be the strategy where the probability mass on a' is moved to a'', i.e.,

$$\sigma_1[a' \to a''](a) = \begin{cases} \sigma_1(a) & \text{if } a' \neq a \neq a'' \\ 0 & \text{if } a = a' \\ \sigma_1(a') + \sigma_1(a'') & \text{if } a = a'' \ . \end{cases}$$

Consider some optimal totally mixed strategy σ_1 for player 1, which exists by Lemma 3 and let v be the value of M. Consider some strategy σ_2 for player 2 such that $u(M, \sigma_1, \sigma_2) = v$, but σ_2 is not totally mixed. We will argue that σ_2 is not optimal. This shows that any optimal strategy σ_2^* is totally mixed, since any optimal strategy σ_2 is such that $u(M, \sigma_1, \sigma_2) = v$.

Let b' be the first column such that $\sigma_2'(b) = 0$. There are two cases, either $b' = 1$ or $b' > 1$. If $b' = 1$ let b'' be the first action such that $\sigma_2(b'') > 0$. Let $\sigma_1' = \sigma_1[b' \to b'']$. The payoff $u(M, \sigma_1', \sigma_2)$ of playing σ_1' against σ_2 is strictly more than the payoff $u(M, \sigma_1, \sigma_2)$ of playing σ_1 against σ_2. This is because the payoff $u(M, \sigma_1', b'')$ is such that

$$\begin{aligned} u(M, \sigma_1', b'') &= \sigma_1'(b'') \cdot y + z \cdot \sum_{a=1}^{b''-1} \sigma_1'(a) \\ &= \sigma_1'(b'') \cdot y + z \cdot \sum_{a=2}^{b''-1} \sigma_1'(a) \\ &= (\sigma_1(b'') + \sigma_1(1)) \cdot y + z \cdot \sum_{a=2}^{b''-1} \sigma_1'(a) \\ &> \sigma_1(b'') \cdot y + z \cdot \sum_{a=1}^{b''-1} \sigma_1(a) \\ &= u(M, \sigma_1, b'') \ , \end{aligned}$$

where the second equality comes from that $\sigma_1'(1) = 0$. The inequality comes from that $y > z$. Also, the payoff $u(M, \sigma_1', b)$, for $b > b''$ is such that

$$\begin{aligned} u(M, \sigma_1', b) &= \sigma_1'(b) \cdot y + z \cdot \sum_{a=1}^{b-1} \sigma_1'(a) \\ &= \sigma_1(b) \cdot y + z \cdot \sum_{a=1}^{b-1} \sigma_1(a) = u(M, \sigma_1, b) \ , \end{aligned}$$

because σ_1' is not different from σ_1 on those actions. We can then find the payoff $u(M, \sigma_1', \sigma_2)$ as follows

$$\begin{aligned} u(M, \sigma_1', \sigma_2) &= \sum_{b=1}^{m} \sigma_2(b) \cdot u(M, \sigma_1', b) \\ &= \sum_{b=b''}^{m} \sigma_2(b) \cdot u(M, \sigma_1', b) \\ &= \sigma_2(b'') \cdot u(M, \sigma_1', b'') + \sum_{b=b''+1}^{m} \sigma_2(b) \cdot u(M, \sigma_1', b) \\ &> \sigma_2(b'') \cdot u(M, \sigma_1, b'') + \sum_{b=b''+1}^{m} \sigma_2(b) \cdot u(M, \sigma_1, b) \\ &= u(M, \sigma_1, \sigma_2) \ , \end{aligned}$$

where the second equality comes from that b'' is the first action σ_2 plays with positive probability. Since the payoff $u(M, \sigma_1, \sigma_2)$ is the value, by definition of σ_2, and the payoff $u(M, \sigma_1', \sigma_2)$ is strictly more, the strategy σ_2 cannot be optimal. This completes the case where $b' = 1$.

The case where $b' \neq 1$ follows similarly but considers $\sigma_1'' = \sigma_1[b' \to 1]$ instead of σ_1'. □

Lemma 5. *For all positive integers m and $0 < \varepsilon \leq \frac{1}{2}$, the matrix game $M = M^{0,\frac{1}{2}+\varepsilon,\frac{1}{2},m}$ has the following properties:*

- **Property 1.** *The patience of any optimal strategy is (i) at least $(2\varepsilon)^{-m+1}$ and (ii) decreasing in ε.*
- **Property 2.** *The value is (i) at most $\frac{1}{2} + \varepsilon \cdot (2\varepsilon)^{m-1}$ and (ii) increasing in ε.*
- **Property 3.** *Any optimal strategy σ_1 for player 1 (resp., σ_2 for player 2) is such that $\sigma_1(1) > \frac{1}{2}$ (resp., $\sigma_2(m) > \frac{1}{2}$).*
- **Property 4.** *For $\varepsilon = \frac{1}{2}$, the value is $\text{val}(M) = \frac{1}{2} + \frac{1}{2^{m+1}-2}$ and the patience of any optimal strategy is $2^m - 1$.*

Proof. Let σ_i be an optimal strategy for player i in M, for each i. By Lemma 3 and Lemma 4 the strategy σ_i is totally mixed for each i. We can therefore consider the vector $\overline{v}$. Recall that $\overline{v}_j = u(M, \sigma_1, j)$ and that for each j such that $\sigma_2(j) > 0$ we have that $\overline{v}_j = \text{val}(M)$. Hence, since σ_2 is totally mixed, all entries of $\overline{M}$ are $\text{val}(M)$. For any row $a' < m$, that $\overline{v}_{a'} = \overline{v}_{a'+1}$ implies that

$$\begin{aligned} (\frac{1}{2} + \varepsilon) \cdot \sigma_1(a') + \frac{1}{2} \cdot \sum_{a=1}^{a'-1} \sigma_1(a) \\ = (\frac{1}{2} + \varepsilon) \cdot \sigma_1(a'+1) + \frac{1}{2} \cdot \sum_{a=1}^{a'} \sigma_1(a) \Rightarrow \\ \varepsilon \cdot \sigma_1(a') = (\frac{1}{2} + \varepsilon) \cdot \sigma_1(a'+1) \Rightarrow \\ \sigma_1(a') = \frac{\frac{1}{2} + \varepsilon}{\varepsilon} \cdot \sigma_1(a'+1) \ , \end{aligned}$$

indicating that $\sigma_1(a') > \sigma_1(a'+1)$ and thus the patience is

$1/\sigma_1(m)$. Also, since σ_1 is a probability distribution

$$\begin{aligned}1 &= \sum_{a=1}^{m} \sigma_1(a) \\ &= \sigma_1(m) \cdot \sum_{a=1}^{m} \left(\frac{\frac{1}{2}+\varepsilon}{\varepsilon}\right)^{m-a}\end{aligned}$$

We then get that

$$\sigma_1(m) = \frac{1}{\sum_{a=1}^{m} \left(\frac{\frac{1}{2}+\varepsilon}{\varepsilon}\right)^{m-a}}$$

We have that $\frac{\frac{1}{2}+\varepsilon}{\varepsilon} = 1 + \frac{1}{2\varepsilon}$ is decreasing in ε. This indicates that $\sigma_1(m)$ is increasing in ε and thus the patience is decreasing in ε. This shows (ii) of Property 1 for player 1. We also have that $\text{val}(M) = \overline{v}_m$ indicating that

$$\begin{aligned}\text{val}(M) &= \overline{v}_m \\ &= \sigma_1(m) \cdot (\frac{1}{2}+\varepsilon) + \frac{1}{2} \cdot \sum_{a=1}^{m-1} \sigma_1(a) \\ &= \varepsilon \cdot \sigma_1(m) + \frac{1}{2}\end{aligned}$$

and thus, the value is increasing in ε (because ε and $\sigma_1(m)$ both are). This shows (ii) of Property 2.

Also, we get that,

$$\begin{aligned}\sigma_1(m) &= \frac{1}{\sum_{a=1}^{m} \left(\frac{\frac{1}{2}+\varepsilon}{\varepsilon}\right)^{m-a}} \\ &= \frac{\varepsilon^{m-1}}{\sum_{a=1}^{m} (\frac{1}{2}+\varepsilon)^{m-a} \cdot \varepsilon^{a-1}} \\ &= \frac{\varepsilon^{m-1}}{\left(\frac{1}{2}\right)^{m-1} + \varepsilon \cdot p(\varepsilon)} \quad ,\end{aligned}$$

where p is some polynomial of degree $m-1$ in which all terms have a positive sign (p is found by multiplying out $\sum_{a=1}^{m} (\frac{1}{2}+\varepsilon)^{m-a} \cdot \varepsilon^{a-1}$). Hence, we have that $\sigma_1(m)$ is at most

$$\sigma_1(m) = \frac{\varepsilon^{m-1}}{\left(\frac{1}{2}\right)^{m-1} + \varepsilon \cdot p(\varepsilon)} < (2\varepsilon)^{m-1} \quad .$$

Thus, the patience is at least $(2\varepsilon)^{-m+1}$. This shows (i) of Property 1 for player 1. Using that $\text{val}(M) = \varepsilon \cdot \sigma_1(m) + \frac{1}{2}$ from above, we get that $\text{val}(M) < \frac{1}{2} + \varepsilon \cdot (2\varepsilon)^{m-1}$. This shows (i) of Property 2.

Furthermore, we can also consider the vector $\overline{v'}$ such that $\overline{v'}_j = u(M, j, \sigma_2)$ for all j (which like $\overline{v}$ has all entries equal to $\text{val}(M)$). Since the expression, when σ_2 is taken to be an unknown vector, for the j'th entry of $\overline{v'}$ is the same as for the $m+1-j$'th entry of $\overline{v}$, when σ_1 is taken to be an unknown vector, we see that $\sigma_1(a) = \sigma_2(m+1-a)$, implying that the patience of player 2's optimal strategies is also at least $(2\varepsilon)^{-m+1}$ and that it is decreasing in ε. This shows Property 1 for player 2.

Observe that since the value is above $\frac{1}{2}$, by Lemma 2, we have that $\sigma_1(1) > \frac{1}{2}$ (because otherwise, if player 2 plays 1 with probability 1, the payoff will not be above $\frac{1}{2}$) and thus also $\sigma_2(m) > \frac{1}{2}$. This shows Property 3.

Also, for $\varepsilon = \frac{1}{2}$ we see that

$$\begin{aligned}\sigma_1(m) &= \frac{1}{\sum_{a=1}^{m} \left(\frac{\frac{1}{2}+\varepsilon}{\varepsilon}\right)^{m-a}} \\ &= \frac{1}{\sum_{a=1}^{m} 2^{m-a}} \\ &= \frac{1}{2^m - 1} \quad .\end{aligned}$$

Similarly to above, we also get that $\sigma_2(m) = \frac{1}{2^m-1}$ and that $\text{val}(M) = \frac{1}{2} + \frac{1}{2^{m+1}-2}$. This shows Property 4 and completes the proof. □

Lemma 6. *Given a positive integer m and reals y and z such that $1 > z > y$, the matrix game $M = M^{1,y,z,m}$ has the following properties:*

- *The value $\text{val}(M) < z$.*
- *Each optimal strategy σ_i for player i is such that there exists an optimal strategy $\sigma_{\widehat{i}}$ for player $\widehat{i}$ in $M^{0,1-y,1-z,m}$ where $\sigma_i(j) = \sigma_{\widehat{i}}(m-j+1)$.*

Proof. Let a positive integer m and reals y and z such that $1 > z > y$ be given. Consider M and let v be the value of M. Exchange the roles of the players by exchanging the rows and columns and multiply the matrix by -1. We get the matrix

$$M^1 = \begin{pmatrix} -y & -1 & -1 & \dots & -1 \\ -z & -y & -1 & \dots & -1 \\ \vdots & -z & \ddots & \ddots & \vdots \\ -z & \vdots & \ddots & -y & -1 \\ -z & -z & \dots & -z & -y \end{pmatrix} \quad .$$

We then have that each optimal strategy σ_1 in M is an optimal strategy for player 2 in M^1 and similarly, each optimal strategy σ_2 for player 2 in M is an optimal strategy for player 1 in M^1 (and vice versa). Also, the value v_1 of M^1 is $v_1 := -v$.

Let M^2 be the matrix where $M^2_{a,b} = M^1_{m+1-a,m+1-b}$, i.e.,

$$M^2 = \begin{pmatrix} -y & -z & -z & \dots & -z \\ -1 & -y & -z & \dots & -z \\ \vdots & -1 & \ddots & \ddots & \vdots \\ -1 & \vdots & \ddots & -y & -z \\ -1 & -1 & \dots & -1 & -y \end{pmatrix} \quad .$$

For each i, and for any optimal strategy σ_i for player i in M^1 the strategy σ'_i is optimal for player i in M^2, where $\sigma'_i(a) = \sigma_i(m+1-a)$ for each a (and vice versa). Also, the value v_2 of M^2 is $v_2 := v_1 = -v$.

Next, let M^3 be the matrix M^2 where we add 1 to each entry, i.e.,

$$M^3 = \begin{pmatrix} 1-y & 1-z & 1-z & \dots & 1-z \\ 0 & 1-y & 1-z & \dots & 1-z \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & \vdots & \ddots & 1-y & 1-z \\ 0 & 0 & \dots & 0 & 1-y \end{pmatrix} \quad .$$

For each i, it is clear that an optimal strategy in σ_i for player i in M^2 is an optimal strategy for player i in M^3 and that the value v_3 is $v_3 := 1 + v_2 = 1 - v$. Also, we see that $M^3 = M^{0,1-y,1-z,m}$ and that $0 < 1 - z < 1 - y$.

We then get that $1 - v > 1 - z$ from Lemma 2 and thus $v < z$. □

B. The patience of optimal strategies

In this section we present an approximation of the values of the states and the patience of the optimal strategies in the Purgatory Duel. We first show that the values of the states (besides $\top$ and $\bot$) are strictly between 0 and 1.

Lemma 7. *Each state*

$$v \in \{v_1^1, v_2^1, \ldots, v_n^1, v_1^2, v_2^2, \ldots, v_2^1, v_s\}$$

is such that $\text{val}(v) \in [\frac{1}{m^{n+2}}, 1 - \frac{1}{m^{n+2}}]$

Proof. Fix $v \in \{v_1^1, v_2^1, \ldots, v_n^1, v_1^2, v_2^2, \ldots, v_2^1, v_s\}$. The fact that $\text{val}(v) \geq \frac{1}{m^{n+2}}$ follows from that if player 1 plays uniformly at random all actions in every state v_j^i for all i, j, then against all strategies for player 2 there is a probability of at least $\frac{1}{m}$ to go (1) from v_j^1 to v_{j+1}^1, for all j; and (2) from v_s to v_1^1; and (3) from v_j^2 to v_s, for all j. By following such steps for at most $n + 2$ steps, the state $v_{n+1}^1 = \top$ is reached. Similarly that $\text{val}(v) \leq 1 - \frac{1}{m^{n+2}}$ follows from player 2 playing uniformly at random all actions in every state v_j^i for all i, j (and using that $\top$ cannot be reached from $\bot$). □

Next we show that every optimal stationary strategy for player 2 must be totally mixed.

Lemma 8. *Let σ_2 be an optimal stationary strategy for player 2. The distribution $\sigma_2(v_j^i)$ is totally mixed and* $\text{val}(v_j^1) > \text{val}(v_s) > \text{val}(v_j^2)$*, for all i, j.*

Proof. Let $v = v_j^i$ for some i, j. We will use that $\text{val}(v) = \text{val}(A^v)$. For $i = 1$ we have that $A^v = M^{0,\text{val}(v_{j+1}^1),\text{val}(v_s),m}$ and for $i = 2$ we have that $A^v = M^{1,\text{val}(v_{j+1}^2),\text{val}(v_s),m}$.

Consider first $i = 1$. We will show using induction in j (with base case $j = n$ and proceeding downwards), that $\text{val}(v_j^1) > \text{val}(v_s)$ and that the distribution $\sigma_2(v_j^1)$ is totally mixed.

Base case, $j = n$: We have that $A^v = M^{0,1,\text{val}(v_s),m}$. By Lemma 7 we have that $1 > \text{val}(v_s) > 0$ and thus, that $\text{val}(v) > \text{val}(v_s)$ follows from Lemma 2. That $\sigma_2(v)$ is totally mixed follows from Lemma 4.

Induction case, $j \leq n - 1$: We have that $A^v = M^{0,\text{val}(v_{j+1}^1),\text{val}(v_s),m}$. By Lemma 7 we have that $\text{val}(v_s) > 0$ and by induction we have that $\text{val}(v_{j+1}^1) > \text{val}(v_s)$ and thus, that $\text{val}(v) > \text{val}(v_s)$ follows from Lemma 2. That $\sigma_2(v)$ is totally mixed follows from Lemma 4.

The argument for $i = 2$ is similar but uses Lemma 6 together with Lemma 3, instead of Lemma 4 and Lemma 2. □

Next, we show that if either player follows a stationary strategy that is totally mixed on at least one side (that is, if there is an i', such that for each j the stationary strategy plays totally mixed in $v_j^{i'}$), then eventually either $\top$ or $\bot$ is reached with probability 1.

Lemma 9. *For any i and i', let σ_i be a stationary strategy for player i, such that $\sigma_i(v_j^{i'})$ is totally mixed for all j. Let $\sigma_{\widehat{i}}$ be some positional strategy for the other player. Then, each closed recurrent set in the Markov chain defined by the game and σ_i and $\sigma_{\widehat{i}}$ consists of only the state $\top$ or only the state $\bot$.*

Proof. In the Markov chain defined by the game and σ_i and $\sigma_{\widehat{i}}$, we have that there are at most two closed recurrent sets, namely, the one consisting of only $\top$ and the one consisting of only $\bot$. The reasoning is as follows: If either $\top$ or $\bot$ is reached, then the respective state will not be left. Also, for each j, since σ_i is totally mixed there is a positive probability to go to either $v_0^{i'}$ or $v_{j+1}^{i'}$ from $v_j^{i'}$ (the remaining probability goes to v_s). The probability to go from v_s to $v_1^{i'}$ in one step is $\frac{1}{2}$. Also if neither $\top$ nor $\bot$ has been reached, then v_s is visited after at most $n + 1$ steps. Hence, in every $n + 1$ steps there is a positive probability that in the next $n + 1$ steps either $\top$ or $\bot$ is reached (i.e., from v_s there is a positive probability that the next states are either (i) $v_1^{i'}, \ldots, v_j^{i'}, v_0^{i'}$; or (ii) $v_1^{i'}, \ldots, v_n^{i'}, v_{n+1}^{i'}$). This shows that eventually either $\top$ or $\bot$ is reached with probability 1. □

Remark 10. *Note that Lemma 9 only requires that the strategy σ_i is totally mixed on one "side" of the Purgatory Duel. For the purpose of this section, we do not use that it only requires one side to be totally mixed, since we only use the result for optimal strategies for player 2, which are totally mixed by Lemma 8. However the lemma will be reused in the next section, where the one sidedness property will be useful.*

The following definition basically *"mirrors"* a strategy σ_i for player i, for each i and gives it to the other player. We show (in Lemma 12) that if σ_2 is optimal for player 2, then the mirror strategy is optimal for player 1. We also show that if σ_2 is an ε-optimal strategy for player 2, for $0 < \varepsilon < \frac{1}{3}$, then so is the mirror strategy for player 1 (in Lemma 16).

Definition 11 (Mirror strategy)**.** *Given a stationary strategy σ_i for player i, for either i, let the mirror strategy $\sigma_{\widehat{i}}^{\sigma_i}$ for player $\widehat{i}$ be the stationary strategy where $\sigma_{\widehat{i}}^{\sigma_i}(v_j^{\widehat{i'}}) = \sigma_i(v_j^{i'})$ for each i' and j.*

We next show that player 1 has optimal stationary strategies in the Purgatory Duel and give expressions for the values of states.

Lemma 12. *Let σ_2 be some optimal stationary strategy for player 2. Then the mirror strategy $\sigma_1^{\sigma_2}$ is optimal for player 1. We have* $\text{val}(v_s) = \frac{1}{2}$ *and* $\text{val}(v_j^i) = 1 - \text{val}(v_j^{\widehat{i}})$*, for all i, j.*

Proof. Consider some optimal stationary strategy σ_2 for player 2. It is thus totally mixed, by Lemma 8. Let $\sigma_1 = \sigma_1^{\sigma_2}$ be the mirror strategy for player 1.

Playing σ_1 against σ_2 and starting in v_s we see that we have probability $\frac{1}{2}$ to reach $\top$ and probability $\frac{1}{2}$ to reach $\bot$, by symmetry and Lemma 9. This shows that the value is at least $\frac{1}{2}$ because σ_2 is optimal. On the other hand, consider some stationary strategy σ_1' for player 1, and the mirror strategy

$\sigma_2' = \sigma_2^{\sigma_1'}$ for player 2. If player 2 plays σ_2' against σ_1', then the probability to eventually reach $\bot$ is equal to the probability to eventually reach $\top$ and then there is some probability p (perhaps 0) that neither will be reached. The payoff $u(v_s, \sigma_1', \sigma_2', 1)$ is then $\frac{1-p}{2} \leq \frac{1}{2}$. This shows that player 1 cannot ensure value strictly more than $\frac{1}{2}$, which is then the value of v_s. Finally, we argue that σ_1 is optimal. If not, then consider σ_2^* such that $u(v_s, \sigma_1, \sigma_2^*, 1) < 1/2$, and then the mirror strategy $\sigma_1^* = \sigma_1^{\sigma_2^*}$ ensures that $u(v_s, \sigma_1^*, \sigma_2, 1) > 1/2$ contradicting optimality of σ_2.

Similarly, for any i, j, playing σ_1 against σ_2 and starting in v_j^i we see that the probability with which we reach $\top$ is equal to the probability of reaching $\bot$ starting in $v_j^{\widehat{i}}$ and vice versa, by symmetry. Also, by Lemma 9 the probability to eventually reach either $\bot$ or $\top$ is 1. Observe that the probability to reach $\bot$ starting in $v_j^{\widehat{i}}$ is at least $1 - \text{val}(v_j^{\widehat{i}})$, by optimality of σ_2 and that with probability 1 either $\bot$ is reached or $\top$ is reached. Also, again because σ_2 is optimal, the probability to reach $\top$ starting in v_j^i is at most $\text{val}(v_j^i)$. This shows that $\text{val}(v_j^i) \geq 1 - \text{val}(v_j^{\widehat{i}})$. Using an argument like the one above, we obtain that $\text{val}(v_j^i) = 1 - \text{val}(v_j^{\widehat{i}})$ and that σ_1 is optimal if the play starts in v_j^i. □

Finally, we give an approximation of the values of states in the Purgatory Duel and a lower bound on the patience of any optimal strategy of $2^{(m-1)^2 m^{n-2}}$.

Theorem 13. *For each j in $\{1, \ldots, n\}$, the value of state v_j^1 in the Purgatory Duel is less than $\frac{1}{2} + 2^{(1-m)\cdot m^{n-j}-1}$ and for any optimal stationary strategy σ_i for either player i, the patience of $\sigma_i(v_j^1)$ is at least $2^{(m-1)^2 m^{n-j-1}}$.*

Proof. Consider some optimal stationary strategy σ_2 for player 2. We will show using induction in j that $\text{val}(v_j^1)$ is less than $\frac{1}{2} + 2^{(1-m)\cdot m^{n-j}-1}$ and that the patience of $\sigma_2(v_j^1)$ is at least $2^{(m-1)^2 m^{n-j-1}}$. Note that using Lemma 12, a similar result holds for optimal strategies for player 1. Let $v = v_j^i$.

Base case, $j = n$: We see that the matrix A^v is $M^{0,1,\frac{1}{2},m}$ and thus, by Lemma 5 (Property 1 and 2) we have that the value

$$\begin{aligned}\text{val}(v) &= \text{val}(A^v) \\ &= \frac{1}{2} + \frac{1}{2^{m+1} - 2} \\ &< \frac{1}{2} + 2^{-m} \\ &= \frac{1}{2} + 2^{(1-m)\cdot m^0 - 1} \quad ,\end{aligned}$$

and $\sigma_2(v)$ has patience $2^m - 1 > 2^{(m-1)^2 \cdot m^{-1}}$.

Induction case, $j \leq n-1$: We see that the matrix A^v is $M = M^{0,\text{val}(v_{j+1}^i),\frac{1}{2},m}$. By induction we have that $\text{val}(v_{j+1}^i) < \frac{1}{2} + 2^{(1-m)\cdot m^{n-j-1}-1}$. Let $\varepsilon = 2^{(1-m)\cdot m^{n-j-1}-1}$ and consider $M' = M^{0,\frac{1}{2}+\varepsilon,\frac{1}{2},m}$. By Lemma 5 (Property 1 and 2) we get that $\text{val}(M') \geq \text{val}(M)$ and that the patience of M' is smaller than the one for M. Also, we get that

$$\begin{aligned}\text{val}(M') &< \frac{1}{2} + \varepsilon \cdot (2\varepsilon)^{m-1} \\ &= \frac{1}{2} + 2^{m-1} \cdot 2^{(1-m)\cdot m^{n-j} - m} \\ &= \frac{1}{2} + 2^{(1-m)\cdot m^{n-j} - 1} \quad ,\end{aligned}$$

and that the patience of M' (and thus M) is at least

$$\begin{aligned}(2\varepsilon)^{-m+1} &= 2^{m-1} \cdot 2^{(1-m)^2 \cdot m^{n-j-1} - m + 1} \\ &= 2^{(1-m)^2 \cdot m^{n-j-1}} \quad .\end{aligned}$$

This completes the proof. □

Remark 14. *It can be seen using induction that the value of each state in the Purgatory Duel is a rational number. First notice that v_n^1 and v_n^2 are the value of a matrix game with numbers in $\{0, \frac{1}{2}, 1\}$ and hence are rational. Similarly, using induction in i, we see that for $j \in \{1, 2\}$ the number v_i^j is rational, since it is the value of a matrix game with numbers in $\{v_0^j, \frac{1}{2}, v_{i+1}^j\}$ (recall that $v_0^1 = 0$ and $v_0^2 = 1$).*

C. The patience of ε-optimal strategies

In this section we consider the patience of ε-optimal strategies for $0 < \varepsilon < \frac{1}{3}$. First we argue that each such strategy for player 2 is totally mixed on one side.

Lemma 15. *For all $0 < \varepsilon < \frac{1}{2}$, each ε-optimal stationary strategy σ_2 for player 2 is such that $\sigma_2(v_j^2)$ is totally mixed, for all j.*

Proof. Fix $0 < \varepsilon < \frac{1}{2}$ and fix some stationary strategy σ_2 such that there exists j such that $\sigma_2(v_j^2)$ is not totally mixed. We will show that σ_2 is not ε-optimal.

Let η be such that $0 < \eta < \frac{1}{2} - \varepsilon$. Let a be an action such that $\sigma_2(v_j^2)(a) = 0$. Let σ_1^η be an η-optimal strategy in *Purgatory* (not the Purgatory Duel) (with the same parameters n and m). Let σ_1 be the strategy such that (i) $\sigma_1(v_{j'}^2)(1) = 1$ for each j'; and (ii) $\sigma_1(v_j^2)(a) = 1$; and (iii) $\sigma_1(v_j^1) = \sigma_1^\eta(v_j)$. Consider a play starting in v_s. Whenever the play is in state $v_{j'}^2$, for some $j' \neq j$ in each step there is a probability of either going back to v_s or going to $v_{j'+1}^2$. Thus, the play either reaches v_j^2 or has gone back to v_s. If it reaches v_j^2, then the next state is either v_s or $\top$ (i.e., v_{j+1}^2 cannot be reached). If the play is in v_1^1, then there is a positive probability to reach $\top$ before going back to v_s, which is at least $\frac{1-\eta}{\eta}$ times the probability to reach $\bot$ before going back to v_s, since σ_1 follows an η-optimal strategy in Purgatory. Hence, the probability to eventually reach $\top$ is at least $1 - \eta > \frac{1}{2} + \varepsilon$ and thus σ_2 is not ε-optimal, since the value of v_s is $\frac{1}{2}$ by Lemma 7. □

We now show that if we mirror an ε-optimal strategy, then we get an ε-optimal strategy.

Lemma 16. *For all $0 < \varepsilon < \frac{1}{3}$, each ε-optimal stationary strategy σ_2 for player 2 in the Purgatory Duel, is such that the mirror strategy $\sigma_1^{\sigma_2}$ is ε-optimal for player 1.*

Proof. Fix $0 < \varepsilon < \frac{1}{3}$ and let σ_2 be some ε-optimal stationary strategy for player 2. Also, let $\sigma_1 = \sigma_1^{\sigma_2}$ be the mirror strategy.

By Lemma 15 the strategy σ_2 is such that $\sigma_2(v_j^2)$ is totally mixed, for all j. We can then apply Lemma 9 and get that either $\top$ or $\bot$ is reached with probability 1. Hence, since σ_2 is ε-optimal we reach $\bot$ with probability at least $1-\text{val}(v)-\varepsilon$ starting in v against all strategies for player 1, for each v. It is clear that any play P of σ_2 against any given strategy σ_1' for player 1 starting in v corresponds, by symmetry, to a play P' of $\sigma_2^{\sigma_1'}$ against σ_1 starting in $f(v)$, where

$$f(v) = \begin{cases} v_s & \text{if } v = v_s \\ \widehat{v_j^i} & \text{if } v = v_j^i \\ \bot & \text{if } v = \top \\ \top & \text{if } v = \bot \ , \end{cases}$$

such that in round i we have that $P_i = f(P_i')$ and the plays are equally likely. Thus, the probability to reach $f(\bot) = \top$, starting in state $f(v)$, for each v is at least $1 - \text{val}(v) - \varepsilon = \text{val}(f(v)) - \varepsilon$, where the equality follows from Lemma 12. Hence, σ_1 is ε-optimal for player 1. □

Next we give a definition and a lemma, which is similar to Lemma 6 in [24]. The purpose of the lemma is to identify certain cases where one can change the transition function of an MDP in a specific way and obtain a new MDP with larger values. We cannot simply obtain the result from Lemma 6 in [24], since the direction is opposite (i.e., Lemma 6 in [24] considers some cases where one can change the transition function and obtain a new MDP with *smaller* values) and our lemma is also for a slightly more general class of MDPs.

Definition 17. *Let G be an MDP with safety objectives. A* replacement set *is a set of triples of states, actions and distributions over the states $Q = \{(s_1, a_1, \delta_1), \ldots, (s_\ell, a_\ell, \delta_\ell)\}$. Given the replacement set Q, the MDP $G[Q]$ is an MDP over the same states as G and with the same set of safe states, but where the transition function δ' is*

$$\delta'(s,a) = \begin{cases} \delta_i & \text{if } s = s_i \text{ and } a = a_i \text{ for some } i \\ \delta(s,a) & \text{otherwise} \end{cases}$$

Lemma 18. *Let G be an MDP with safety objectives. Consider some replacement set*

$$Q = \{(s_1, a_1, \delta_1), \ldots, (s_\ell, a_\ell, \delta_\ell)\} \ ,$$

such that for all t and i we have that

$$\sum_{s \in S} (\delta(s_i, a_i)(s) \cdot \overline{v}_s^t) \leq \sum_{s \in S} (\delta_i(s) \cdot \overline{v}_s^t) \ .$$

Let $\overline{v'}^t$ be the value vector for $G[Q]$ with finite horizon t. (1) For all states s and time limits t we have that

$$\overline{v}_s^t \leq \overline{v'}_s^t \ .$$

(2) For all states s, we have that

$$\text{val}(G, s) \leq \text{val}(G[Q], s) \ .$$

Proof. We first present the proof of first item. We will show, using induction in t, that $\overline{v}_s^t \leq \overline{v'}_s^t$ for all s. Let δ' be the transition function for $G[Q]$.

Base case, $t = 0$: Consider some state s. Clearly we have that $\overline{v}_s^t = \overline{v'}_s^t$ because we have not changed the safe states.

Induction case, $t \geq 1$: The induction hypothesis state that $\overline{v}_s^{t-1} \leq \overline{v'}_s^{t-1}$ for all s. Consider some state s. Consider any action a' such that there is an i such that $s = s_i$ and $a = a_i$. We have that

$$\sum_{s'} (\delta(s, a')(s') \cdot \overline{v}_{s'}^{t-1}) \leq \sum_{s'} (\delta'(s, a')(s') \cdot \overline{v}_{s'}^{t-1})$$

by definition for such a' (the statement is true for all time limits and thus also for $t-1$). For all other actions a'' we have that

$$\sum_{s'} (\delta(s, a'')(s') \cdot \overline{v}_{s'}^{t-1}) = \sum_{s'} (\delta'(s, a'')(s') \cdot \overline{v}_{s'}^{t-1}) \ ,$$

since $\delta(s, a'') = \delta'(s, a'')$. Hence,

$$\min_a \sum_{s'} (\delta(s, a)(s') \cdot \overline{v}_{s'}^{t-1}) \leq \min_a \sum_{s'} (\delta'(s, a)(s') \cdot \overline{v}_{s'}^{t-1})$$

We then have, using the recursive definition of $\overline{v}_s^t$, that

$$\begin{aligned} \overline{v}_s^t &= \min_a \sum_{s'} (\delta(s, a)(s') \cdot \overline{v}_{s'}^{t-1}) \\ &\leq \min_a \sum_{s'} (\delta'(s, a)(s') \cdot \overline{v}_{s'}^{t-1}) \\ &\leq \min_a \sum_{s'} (\delta'(s, a)(s') \cdot \overline{v'}_{s'}^{t-1}) \\ &= \overline{v'}_s^t \ . \end{aligned}$$

where we just argued the first inequality; and the second inequality comes from the induction hypothesis and that each factor is positive. (Note that the optimal strategy for player 2 in a matrix game $A^s[\overline{v}^{t-1}]$ of 1 row is to pick one of the columns with the smallest entry with probability 1 and thus $\overline{v}_s^t = \text{val}(A^s[\overline{v}^{t-1}]) = \min_a \sum_{s'} (\delta(s, a)(s') \cdot \overline{v}_{s'}^{t-1})$ and similarly for $\overline{v'}_s^t$). This completes the proof of the first item. The second item follows from the first item and since the value of a time limited game goes to the value of the game without the time limit as the time limit grows to ∞, as shown by [15]. □

We next show that for player 1, the patience of ε-optimal strategies is high.

Lemma 19. *For all $0 < \varepsilon < \frac{1}{3}$, each ε-optimal stationary strategy σ_1 for player 1 in the Purgatory Duel has patience at least $2^{m^{\Omega(n)}}$. For $N = 5$ the patience is $2^{\Omega(m)}$.*

Proof. Consider some ε-optimal stationary strategy σ_1 for player 1 in the Purgatory Duel. Fixing σ_1 for player 1 in the Purgatory Duel we obtain an MDP G' for player 2. Let $\overline{v}^t$ be the value vector for G' with finite horizon (time-limit) t and

let δ be the transition function for G'. For each i, let

$$\delta_i(s) = \begin{cases} \delta(v_n^2, i)(s) & \text{if } v_s \neq s \neq \bot \\ \delta(v_n^2, i)(\bot) + \delta(v_n^2, i)(v_s) & \text{if } v_s = s \\ 0 & \text{if } \bot = s \end{cases}$$

(Note that δ_i is the same probability distribution as $\delta(v_n^2, i)$, except that the probability mass on $\bot$ is moved to v_s.) Consider the replacement set $Q = \{(v_n^2, 1, \delta_1), \ldots, (v_n^2, m, \delta_m)\}$ and the MDP $G'[Q]$. We have for all t and i that

$$\sum_{s \in S} (\delta(v_n^2, i)(s) \cdot \overline{v}_s^t) \leq \sum_{s \in S} (\delta_i(s) \cdot \overline{v}_s^t)$$

because

$$\overline{v}_\bot^t = \overline{v}_{v_{n+1}^2}^t = 0 \leq \overline{v}_{v_s}^t$$

for all t and the only difference between $\delta(v_n^2, i)$ and δ_i is that the probability mass on $\bot$ is moved to v_s. We then get from Lemma 18(2) that $\text{val}(G', v_s) \leq \text{val}(G'[Q], v_s)$. Let σ_2 be an optimal positional strategy in $G'[Q]$. It is easy to see that σ_2 plays action 1 in v_j^2 for all j, because the best player 2 can hope for is to get back to v_s since $\bot$ cannot be reached from v_j^2 in $G'[Q]$ for any j and if he plays some action which is not 1, then there is a positive probability that $\top$ will be reached in one step. Thus, the MDP $G'[Q]$ corresponds to the MDP one gets by fixing the strategy σ_1' where $\sigma_1'(v_i) = \sigma_1(v_i^1)$ for player 1 in Purgatory. But the probability to reach $\top$ in $G'[Q]$ is at least $\frac{1}{2} - \varepsilon$ and hence σ_1' is $(\frac{1}{2} + \varepsilon)$-optimal in Purgatory (note that this is Purgatory and not Purgatory Duel). As shown by [19] any such strategy requires patience $2^{m^{\Omega(n)}}$. Thus, any ε-optimal stationary strategy for player 1 in the Purgatory Duel requires patience $2^{m^{\Omega(n)}}$.

It was shown by [19] that the patience of ε-optimal strategies for Purgatory with $n = 1$ Purgatory state is $2^{\Omega(m)}$, and thus similarly for the Purgatory Duel with $N = 5$. □

We are now ready to prove the main theorem of this section.

Theorem 20. *For all $0 < \varepsilon < \frac{1}{3}$, every ε-optimal stationary strategy, for either player, in the Purgatory Duel (that has $N = 2n + 3$ states and at most m actions for each player at all states) has patience $2^{m^{\Omega(n)}}$. For $N = 5$ the patience is $2^{\Omega(m)}$.*

Proof. The statement for strategies for player 1 follows from Lemma 19. By Lemma 16, for each ε-optimal strategy for player 2, there is an ε-optimal strategy for player 1 (i.e., the mirror strategy) with the same patience. Thus the result follows for strategies for player 2. □

IV. Zero-sum Concurrent Stochastic Games: Patience Lower Bound for Three States

In this section we show that the patience of all ε-optimal strategies, for all $0 < \varepsilon < \frac{1}{3}$, for both players in a concurrent reachability game G with three states of which two are absorbing, and the non-absorbing state has m actions for each player, can be as large as $2^{\Omega(m)}$. The proof consists of two phases, first we show the lower bound in a game with at most m^2 actions for each player; and second, we show that all but $2m - 1$ actions can be removed for both players in the game without changing the patience.

The first game, the *3-state Purgatory Duel*, is intuitively speaking the Purgatory Duel for $N = 5$, where we replace the states v_1^1, v_1^2 and v_s with a state v_s' while in essence keeping the same set of ε-optimal strategies. The idea is to ensure that one step in the 3-state Purgatory Duel corresponds to two steps in the Purgatory Duel with $N = 5$, by having the players pick all the actions they might use in the next two steps at once. The game is formally defined as follows:

The 3-state Purgatory Duel consists of $N = 3$ states, named v_s', $\top'$ and $\bot'$ respectively. The states $\top'$ and $\bot'$ are absorbing. The state v_s' is such that

$$A_{v_s'}^1 = A_{v_s'}^2 = \{(i, j) \mid 1 \leq i, j \leq m\} \ .$$

Also, let δ' be the transition function for the Purgatory Duel with $N = 5$. Let p be the function that given a state in $\{v_s, \bot, \top\}$ in the Purgatory Duel for $i = 1$ outputs the primed state (which is then a state in the 3-state Purgatory Duel). Recall that $\mathsf{U}(s, s')$ is the uniform distribution over s and s'. Observe that the deterministic distributions $\delta'(v_1^1, a_1, a_2)$ and $\delta'(v_1^2, a_1, a_2)$ are in $\{v_s, \top, \bot\}$ for all a_1 and a_2. For each pair of actions $(a_1^1, a_1^2) \in A_{v_s'}^1$ and $(a_2^1, a_2^2) \in A_{v_s'}^2$ in the 3-state Purgatory Duel, we have that

$$\begin{aligned} &\delta(v_s', (a_1^1, a_1^2), (a_2^1, a_2^2)) = \\ &\mathsf{U}(p(\delta'(v_1^1, a_1^1, a_2^1)), p(\delta'(v_1^2, a_1^2, a_2^2))) \ . \end{aligned}$$

To make the game easier to understand on its own, we now give a more elaborate description of the transition function δ without using the transition function for the Purgatory Duel. To make the pattern as clear as possible we write $\mathsf{U}(s, s)$ instead of s for all s.

$$\delta(v_s', (a_1^1, a_1^2), (a_2^1, a_2^2)) = \begin{cases} \mathsf{U}(\bot', \top') & \text{if } a_1^1 > a_2^1 \text{ and } a_1^2 > a_2^2 \\ \mathsf{U}(\bot', \bot') & \text{if } a_1^1 > a_2^1 \text{ and } a_1^2 = a_2^2 \\ \mathsf{U}(\bot', v_s') & \text{if } a_1^1 > a_2^1 \text{ and } a_1^2 < a_2^2 \\ \mathsf{U}(\top', \top') & \text{if } a_1^1 = a_2^1 \text{ and } a_1^2 > a_2^2 \\ \mathsf{U}(\top', \bot') & \text{if } a_1^1 = a_2^1 \text{ and } a_1^2 = a_2^2 \\ \mathsf{U}(\top', v_s') & \text{if } a_1^1 = a_2^1 \text{ and } a_1^2 < a_2^2 \\ \mathsf{U}(v_s', \top') & \text{if } a_1^1 < a_2^1 \text{ and } a_1^2 > a_2^2 \\ \mathsf{U}(v_s', \bot') & \text{if } a_1^1 < a_2^1 \text{ and } a_1^2 = a_2^2 \\ \mathsf{U}(v_s', v_s') & \text{if } a_1^1 < a_2^1 \text{ and } a_1^2 < a_2^2 \ . \end{cases}$$

Furthermore, $S^1 = \{\top'\}$. We will use τ_i for strategies in the 3-state Purgatory Duel to distinguish them from strategies in the Purgatory Duel. There is an illustration of the Purgatory Duel with $N = 5$ and $m = 2$ in Figure 3 and the corresponding 3-state Purgatory Duel in Figure 4.

Given a strategy τ_i for player i in the 3-state Purgatory Duel we define the strategy σ_i in the Purgatory Duel with $N = 5$ which is the projection of τ_i and vice versa (note that the other direction maps to a set of strategies).

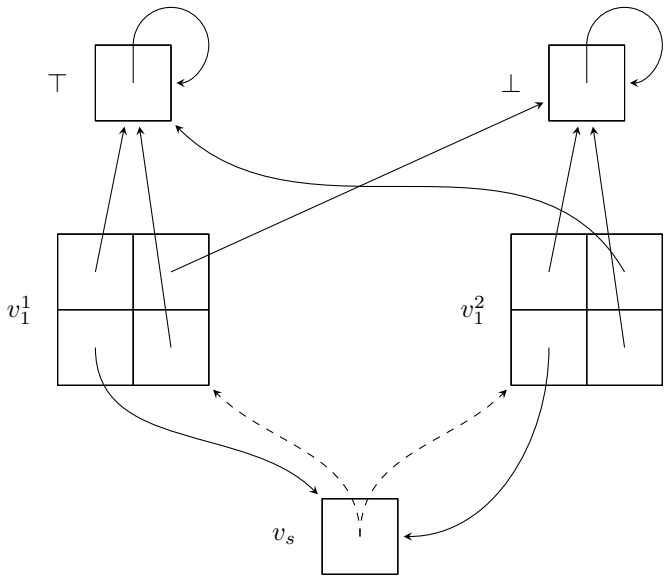

Fig. 3. An illustration of the Purgatory Duel with $N = 5$ and $m = 2$. The two dashed edge have probability $\frac{1}{2}$ each.

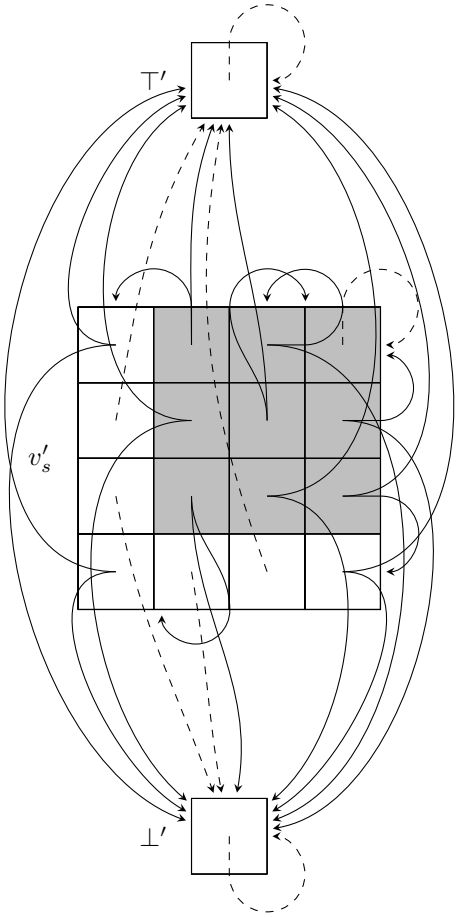

Fig. 4. An illustration of the 3-state Purgatory Duel $m = 2$. The *non-dashed* edges have probability $\frac{1}{2}$ each. The order of the actions is $(1,1),(1,2),(2,1),(2,2)$. The actions (i.e., $(2,2)$ for player 1 and $(1,1)$ for player 2) with white background cannot be played in a restricted strategy.

Definition 21. *Given a strategy τ_i for player i in the 3-state Purgatory Duel, let $\sigma_i^{\tau_i}$ be the stationary strategy for player i in the Purgatory Duel with $N = 5$ where*

$$\sigma_i^{\tau_i}(v_1^1)(a_1^1) = \sum_{a_1^2} \tau_i(v_s')(a_1^1, a_1^2)$$

and

$$\sigma_i^{\tau_i}(v_1^2)(a_1^2) = \sum_{a_1^1} \tau_i(v_s')(a_1^1, a_1^2) \ .$$

Also, for any stationary strategy σ_i in the Purgatory Duel with $N = 5$, let $\mathcal{T}_i^{\sigma_i}$ be the set of stationary strategies in the 3-state Purgatory Duel such that $\tau_i \in \mathcal{T}_i^{\sigma_i}$ implies that $\sigma_i^{\tau_i} = \sigma_i$.

Lemma 22. *Consider any $\varepsilon \geq 0$. Let G be the Purgatory Duel with $N = 5$ and G' be the 3-state Purgatory Duel. For any ε-optimal stationary strategy τ_i for player i in G', we have that $\sigma_i^{\tau_i}$ is ε-optimal starting in v_s in G. Similarly, for any ε-optimal stationary strategy σ_i in G starting in v_s each strategy in $\mathcal{T}_i^{\sigma_i}$ is ε-optimal in G'. Also,* $\text{val}(v_s') = \frac{1}{2}$.

Proof. Consider some pair of strategies τ_i and $\sigma_i^{\tau_i}$ for player i in G' and G, respectively. Fixing τ_i and $\sigma_i^{\tau_i}$ as the strategy for player i we get two MDPs H' and H, respectively. We will argue that $\text{val}(H', v_s') = \text{val}(H, v_s)$. Let $\overline{v'}^t$ and $\overline{v}^t$ be the vector of values for the value iteration algorithm in iteration t when run on H' and H respectively (i.e., the values of H' and H with time limit t). We have that $\overline{v}_{v_s}^{2t} = \overline{v'}_{v_s'}^{t}$ by definition of the value-iteration algorithm and the transition function in the 3-state Purgatory Duel. Hence, since $\overline{v}_{v_s}^{2t}$ and $\overline{v'}_{v_s'}^{t}$ converges to the value of state v_s and v_s' in H and H' respectively, they have the same value. We know that the value of v_s is $\frac{1}{2}$ and thus that is also the value of v_s'. □

Corollary 23. *The patience of ε-optimal stationary strategies for both players, for $0 < \varepsilon < \frac{1}{3}$, in the 3-state Purgatory Duel is at least $2^{\Omega(m)}$, where m^2 is the number of actions in state v_s.*

Proof. The patience of ε-optimal strategies, for $0 < \varepsilon < \frac{1}{3}$, in the Purgatory Duel with $N = 5$ is $2^{\Omega(m)}$ from Theorem 20. Thus, by Lemma 22, the patience of the 3-state Purgatory Duel is $2^{\Omega(m)}$. □

The restricted 3-state Purgatory Duel. The above corollary only shows that the for the 3-state Purgatory Duel, in which one state have m^2 actions and others have 1, the patience is at least $2^{\Omega(m)}$. We now show how to decrease the number of actions from quadratic down to linear, while keeping the same patience.

From Lemma 5 and Lemma 6 we see that for any optimal strategy σ_1 for player 1 (resp., σ_2 for player 2) in the Purgatory Duel with $N = 5$, we have that $\sigma_1(v_1^1)(1) > \frac{1}{2}$ and that $\sigma_1(v_1^2)(1) > \frac{1}{2}$ (resp., $\sigma_2(v_1^1)(m) > \frac{1}{2}$ and that $\sigma_2(v_1^2)(m) > \frac{1}{2}$). Hence, there exists an optimal strategy for player 1 in the 3-state Purgatory Duel that only plays actions on the form $(1, a_1^2)$ and $(a_1^1, 1)$ with positive probability. More precisely, the strategy τ_1 where (1) $\tau_1(v_s)((1, a_1^2)) = \sigma_1(v_1^2)(a_1^2)$; and (2) $\tau_1(v_s)((a_1^1, 1)) = \sigma_1(v_1^1)(a_1^1)$; and (3) has the remaining

probability mass on $(1,1)$ is optimal in the 3-state Purgatory Duel, since $\sigma_1^{\tau_1}$ is σ_1. Similarly for player 2 and the actions (m, a_2^2) and (a_2^1, m). Let

$$R_1 = \{(i,j) \mid i = 1 \vee j = 1, 1 \leq i,j \leq m\}$$

and

$$R_2 = \{(i,j) \mid i = m \vee j = m, 1 \leq i,j \leq m\} \ .$$

Observe that $|R_1| = |R_2| = 2m - 1$. We say that a strategy for player i, for each i, is *restricted* if the strategy uses only actions in R_i. The sub-matrix corresponding to the restricted 3-state Purgatory Duel for $m = 2$ is depicted as the grey sub-matrix in Figure 4. This suggests the definition of the *restricted 3-state Purgatory Duel*, which is like the 3-state Purgatory Duel, except that the strategies for the players are restricted. We next show that ε-optimal strategies in the restricted 3-state Purgatory Duel also have high patience (note, that while this is perhaps not surprising, it does not follow directly from the similar result for the 3-state Purgatory Duel, since it is possible that the restriction removes the optimal best reply to some strategy which would otherwise not be ε-optimal). The key idea of the proof is as follows: (i) we show that the patience of player i in the 3-state Purgatory Duel remains unchanged even if only the opponent is enforced to use restricted strategies; and (ii) each player has a restricted strategy that is optimal in the 3-state Purgatory Duel as well as in the restricted 3-state Purgatory Duel.

Lemma 24. *The value of state v'_s in the restricted 3-state Purgatory Duel is $\frac{1}{2}$*

Proof. Each player has a restricted strategy which is optimal in the 3-state Purgatory Duel and ensures value $\frac{1}{2}$. Thus, these strategies must still be optimal in the restricted 3-state Purgatory Duel and still ensure value $\frac{1}{2}$. □

The next lemma is conceptually similar to Lemma 15 for $N = 5$ (however, it does not follow from Lemma 15, since the strategies for player 1 are restricted here).

Lemma 25. *Let τ_2 be an ε-optimal stationary strategy for player 2 in the restricted 3-state Purgatory Duel, for $0 < \varepsilon < \frac{1}{2}$. Then, $\sum_{i=1}^{m} \tau_2(v'_s)(i,j) > 0$, for each j.*

Proof. Fix $0 < \varepsilon < \frac{1}{2}$. Let τ_2 be a stationary strategy in the 3-state Purgatory Duel (note, we do not require that τ_2 is restricted), such that there exists an a_2 for which $\sum_{a_1} \tau_2(v'_s)((a_1,a_2)) = 0$. Let a' be smallest such a_2.

Fix $0 < \eta < \frac{1}{2} - \varepsilon$. We show that there exists a restricted stationary strategy τ_1 for player 1, ensuring that the payoff is at least $1 - \eta > \frac{1}{2} + \varepsilon$. There are two cases. Either (i) $a' = 1$ or (ii) not.

In case (i), let $\sigma_1(v'_s)$ be an η-optimal strategy for player 1 in the *Purgatory* with parameters $(3, m)$. Then consider the strategy $\tau_1(v'_s)$, where $\tau_1(v'_s)((a,1)) = \sigma_1(v'_s)(a)$, for each a. Observe that τ_1 is a restricted strategy. Consider what happens if τ_1 is played against τ_2: In each round i, as long as $v_i = v'_s$, the next state is either defined by the first or the second component of the actions of the players. If it is defined by the second component, then the next state v_{i+1} is always v'_s, because player 1's first component is 1 and player 2's first component greater than 1. Consider the rounds where the next state is defined by the first component. In such rounds $\top$ is reached with probability $(1-\eta)\cdot p$, for some $p > 0$ and $\bot$ is reached with probability at most $\eta \cdot p$, because player 1 follows an η-optimal strategy in Purgatory on the first component. But in expectation, in every second round the first component is used and thus $\top$ is reached with probability at least $1 - \eta$, which shows that σ_2 is not ε-optimal.

In case (ii), consider the strategy τ_1, such that $\tau_1(v'_s)((1,a')) = 1$. Observe that τ_1 is a restricted strategy. Consider what happens if τ_1 is played against τ_2: In each round i, as long as $v_i = v'_s$, the next state is either defined by the first or the second component of the players choice. If it is defined by the first component, then the next state v_{i+1} is always v'_s or $\top$, because the choice of player 1 is 1. Consider the rounds where the next state is defined by the second component. In each such round either $\top$ or v'_s is reached and $\top$ is reached with positive probability, since player 1 plays $a' > 1$ and player 2 always plays something else and 1 with positive probability. But in expectation, in every second round the second component is used and hence $\top$ is reached with probability 1 eventually, which shows that σ_2 is not ε-optimal. □

We will now define how to mirror strategies in the restricted 3-state Purgatory Duel.

Definition 26. *Given a stationary strategy τ_i for player i in the restricted 3-state Purgatory Duel, for either i, let $\tau_{\widehat{i}}^{\tau_i}$ be the stationary strategy for player $\widehat{i}$ (referred to as the mirror strategy of τ_i) in the restricted 3-state Purgatory Duel where $\tau_{\widehat{i}}^{\tau_i}(v'_s)((a_1,a_2)) = \tau_i(v'_s)((a_2,a_1))$ for each a_1 and a_2.*

We next show that each ε-optimal stationary strategy for player 2 can be mirrored to an ε-optimal stationary for player 1. The statement and the proof idea are similar to Lemma 16, but since the strategies for the players are restricted here, there are some differences.

Lemma 27. *For all $0 < \varepsilon < \frac{1}{2}$, each ε-optimal stationary strategy τ_2 for player 2 in the restricted 3-state Purgatory Duel is such that the mirror strategy $\tau_1^{\tau_2}$ is ε-optimal for player 1 in the restricted 3-state Purgatory Duel.*

Proof. Fix ε, such that $0 < \varepsilon < \frac{1}{2}$. Consider some ε-optimal stationary strategy τ_2^* for player 2 in the restricted 3-state Purgatory Duel. Let $\tau_1^* = \tau_1^{\tau_2^*}$ be the mirror strategy for player 1 given τ_2^* and let τ_2 be an optimal best reply to τ_1^*. Let $\tau_1 = \tau_1^{\tau_2}$ be the mirror strategy for player 1 given τ_2. Observe that eventually either $\top$ or $\bot$ is reached with probability 1, when playing τ_1^* against τ_2, by Lemma 25 and the construction of the game (since there is a positive probability that the second component matches in every round in which the play is in v'_s). We have that $u(v'_s, \tau_1, \tau_2^*) \leq \frac{1}{2} + \varepsilon$, since τ_2^* is ε-optimal. This indicates that $\top'$ is reached with probability at most $\frac{1}{2} + \varepsilon$ when playing τ_1 against τ_2^*. Hence, by symmetry

$\bot'$ is reached with probability at most $\frac{1}{2}+\varepsilon$ when playing τ_1^* against τ_2. Thus, since $\bot'$ or $\top'$ is reached with probability 1, we have that $u(v'_s, \tau_1^*, \tau_2) \geq \frac{1}{2} - \varepsilon$, showing that τ_1^* is ε-optimal. □

We next show that ε-optimal stationary strategies for player 1 requires high (exponential) patience. The statement and the proof idea are similar to Lemma 19, but since the players strategies are restricted here, there are some differences.

Lemma 28. *For all $0 < \varepsilon < \frac{1}{3}$, each ε-optimal stationary strategy σ_1 for player 1 in the restricted 3-state Purgatory Duel has patience $2^{\Omega(m)}$.*

Proof. Fix some $0 < \varepsilon < \frac{1}{3}$ and some ε-optimal stationary strategy σ_1 for player 1 in the restricted 3-state Purgatory Duel. The restricted 3-state Purgatory Duel then turns into an MDP M for player 2 and we can apply Lemma 18(2). We have that $p = \sum_{a_1^1} \sigma_1(v'_s)(a_1^1, a_2^2)/2$ is the probability that player 1 plays an action with second component a_2^2 and the next state is defined by the second component. Let $d(a_1^2, a_2^2)$ be the probability distribution over successors if player 2 plays (a_1^2, a_2^2) in v'_s. Observe that the play would go to $\bot$ if both players played a_2^2 and the next state is defined by the second component and thus

$$d(a_1^2, a_2^2)(\bot) - p \geq 0 \ .$$

Let

$$d'(a_1^2, a_2^2)(v) = \begin{cases} d(a_1^2, a_2^2)(v'_s) + p & \text{if } v = v'_s \\ d(a_1^2, a_2^2)(\bot) - p & \text{if } v = \bot \\ d(a_1^2, a_2^2)(\top) & \text{if } v = \top \ . \end{cases}$$

Consider the MDP M', which is equal to M, except that it uses the distribution $d'(a_1^2, a_2^2)$ instead of $d(a_1^2, a_2^2)$. By Lemma 18(2) we have that

$$\text{val}(M') \geq \text{val}(M) \geq \frac{1}{2} - \varepsilon \geq \frac{1}{6} \ .$$

It is clear that player 2 has an optimal positional strategy in M' that plays (a_1^2, m) for some a_1^2 (this strategy is restricted), since playing (a_1^2, a_2^2), for some $a_2^2 < m$, just increases the probability to reach $\top$ in one step (because player 1 might play some action $a_2^1 > a_2^2$ and otherwise the play will go back to v'_s). But M' corresponds to the MDP obtained by playing σ_1 in the Purgatory with $N = 3$ (where v'_s corresponds to v_1), except that with probability $\frac{1}{2}$ the play goes from v'_s back to v'_s in the restricted 3-state Purgatory Duel no matter the choice of the players. This difference clearly does not change the value. Hence, σ_1 ensures payoff at least $\frac{1}{6}$ in the Purgatory with $N = 3$ and hence has patience $2^{\Omega(m)}$ by [19]. □

We are now ready for the main result of this section.

Theorem 29. *For all $0 < \varepsilon < \frac{1}{3}$, every ε-optimal stationary strategy, for either player, in the restricted 3-state Purgatory Duel (that has three states, two of which are absorbing, and the non-absorbing state has $O(m)$ actions for each player) has patience $2^{\Omega(m)}$.*

Proof. By Lemma 28, the statement is true for every ε-optimal stationary strategy for player 1. By Lemma 27, every ε-optimal stationary strategy for player 2 corresponds to an ε-optimal stationary strategy for player 1, with the same patience, and thus every ε-optimal stationary strategy for player 2 has patience $2^{\Omega(m)}$. □

V. Zero-sum Concurrent Stochastic Games: Patience Upper Bound

In this section we give upper bounds on the patience of optimal and ε-optimal stationary strategies in a zero-sum concurrent reachability game G for the safety player. Our exposition here makes heavy use of the setup of Hansen et al. [20] and will for that reason not be fully self-contained. We assume for concreteness that the player 1 is the reachability player and player 2 the safety player.

Hansen et al. showed [20, Corollary 42] for the more general class of Everett's recursive games [15] that each player has an ε-optimal stationary strategy of doubly-exponential patience. More precisely, if all probabilities have bit-size at most τ, then each player has an ε-optimal strategy of patience bounded by $(\frac{1}{\varepsilon})^{\tau m^{O(N)}}$. For zero-sum concurrent reachability games the safety player is guaranteed to have an optimal stationary strategy [29], [22]. Using this fact one may use directly the results of Hansen et al. to show that the safety player has an optimal strategy of patience bounded by $(\frac{1}{\varepsilon})^{\tau m^{O(N^2)}}$. We shall below refine this latter upper bound in terms of the number of value classes of the game. The overall approach in deriving this is the same, namely we use the general machinery of real algebraic geometry and semi-algebraic geometry [3] to derive our bounds. In order to do this we derive a formula in the first order theory of the real numbers that uniquely defines the value of the game, and from the value of the game we can express the optimal strategies. The improved bound is obtained by presenting a formula where the number of variables depend only on the number of value classes rather than the number of states.

Let below N denote the number of non-absorbing states, and $m \geq 2$ the maximum number of actions in a state for either player. Assume that all probabilities are rational numbers with numerators and denominators of bit-size at most τ, where the bit-size of a positive integer n is given by $\lfloor \lg n \rfloor + 1$. We let K denote the number of value classes. We number the non-absorbing states $1, \ldots, N$ and assume that both players have the actions $\{1, \ldots, m\}$ in each of these states. For a non-negative integer z, define $\text{bit}(z) = \lceil \lg z \rceil$.

Given valuations $v_1, \ldots, v_N$ for the non-absorbing states, we define for each state k a $m \times m$ matrix game $A^k(v)$ letting entry (i, j) be $s_{ij}^k + \sum_{\ell=1}^N p_{ij}^{k\ell} v_\ell$, where $p_{ij}^{k\ell} = \delta(k, i, j)(\ell)$ and s_{ij}^k is the probability of a transition to a state where the reachability player wins, given actions i and j in state k. The *value mapping* operator $M : \mathbb{R}^N \to \mathbb{R}^N$ is given by $M(v) = \big(\text{val}(A^1(v), \ldots, \text{val}(A^N(v)))\big)$. Everett showed that the value vector of his recursive games are given by the unique *critical vector*, which in turn is defined using the value mapping. We will instead for concurrent reachability games

use the characterization of the value vector as the coordinate-wise least fixpoint of the value mapping. The value vector v is thus characterized by the formula

$$M(v) = v \wedge (\forall v' : M(v') = v' \Rightarrow v \leq v') \quad . \tag{1}$$

Similarly to [20, proof of Theorem 13] we obtain the following statement.

Lemma 30. *There is a quantifier free formula with N variables v that expresses $M(v) = v$. The formula uses at most $N(m+2)4^m$ different polynomials, each of degree at most $m+2$ and having coefficients of bit-size at most $2(N+1)(m+2)^2 \operatorname{bit}(m)\tau$.*

Now, if we instead introduce a variable for each value class, we can express $M(v) = v$ using only K free variables, by identifying variables of the same value class. For $w \in \mathbb{R}^K$, let $v(w) \in \mathbb{R}^N$ denote the vector obtained by letting the coordinates corresponding to value class j be assigned w_j. We thus simply express $M(v(w)) = v(w)$ instead. Combining this with (1) we obtain the final formula.

Corollary 31. *There is a quantified formula with K free variables that describes whether the vector $v(w)$ is the value vector of G. The formula has a single block of quantifiers over K variables. Furthermore the formula uses at most $2N(m+2)4^m + K$ different polynomials, each of degree at most $m+2$ and having coefficients of bit-size at most $2(N+1)(m+2)^2 \operatorname{bit}(m)\tau$.*

We shall now apply the *quantifier elimination* [3, Theorem 14.16] and *sampling* [3, Theorem 13.11] procedures to the formula of Corollary 31.

First we use Theorem 14.16 of Basu, Pollack, and Roy [3] obtaining a quantifier free formula with K variables, expressing that $w(v)$ is the value of G. Next we use Theorem 13.11 of [3] to obtain a univariate representation of w such that $v(w)$ is the value vector of G. That is, we obtain univariate real polynomials $f, g_0, \ldots, g_K$, where f and g_0 are coprime, such that $w = (g_1(t)/g_0(t), \ldots, g_K(t)/g_0(t))$, where t is a root of f. These polynomial are of degree $m^{O(K^2)}$ and their coefficients have bit-size $\tau m^{O(K^2)}$. Our next task is to recover from w an optimal strategy for the safety player. For this we just need to select optimal strategies for the column player in each of the matrix games $A^k(v(w))$. Such optimal strategies correspond to basic feasible solutions of standard linear programs for computing the value and optimal strategies of matrix games (cf. [20, Lemma 3]). This means that there exists $(m+1) \times (m+1)$ matrices $M^1(w), \ldots, M^N(w)$, such that $(q_1^k(w), \ldots, q_m^k(w))$ is an optimal strategy for the column player in $A^k(v(w))$ where $q_i^k(w) = \det((M^k(w))_i)/\det(M^k(w))$, where $(M^k(w))_i$ denotes the matrix obtained from $M^k(w)$ by replacing column i with the $(m+1)$th unit vector e_{m+1}. As the matrices $M^1(w), \ldots, M^k(w)$ are obtained from the matrix games $A^1(v(w)), \ldots, A^N(v(w))$, the entries are degree 1 polynomial in w and having rational coefficients with numerators and denominators of bit-size at most τ as well. Using a simple bound on determinants [3, Proposition 8.12], and substituting the expression $g_j(t)/g_0(t)$ for w_j for each j, we obtain a univariate representation of $(q_1^k(w), \ldots, q_m^k(w))$ for each k given by polynomials of degree $m^{O(K^2)}$ and their coefficients have bit-size $\tau m^{O(K^2)}$. Substituting the root t using resultants (cf. [20, Lemma 15]) we finally obtain the following result.

Theorem 32. *Let G be a zero-sum concurrent reachability game with N non-absorbing states, at most $m \geq 2$ actions for each player in every non-absorbing state, and where all probabilities are rational numbers with numerators and denominators of bit-size at most τ. Assume further that G has at most K value classes. Then there is an optimal strategy for the safety player where each probability is a real algebraic number, defined by a polynomial of degree $m^{O(K^2)}$ and maximum coefficient bit-size $\tau m^{O(K^2)}$.*

By a standard root separation bounds (e.g. [35, Chapter 6, equation (5)]) we obtain a patience upper bound.

Corollary 33. *Let G be as in Theorem 32. Then there is an optimal strategy for the safety player of patience at most $2^{\tau m^{O(K^2)}}$.*

In general the probabilities of this optimal strategy will be irrational numbers. However we may employ the rounding scheme as explained in Lemma 14 and Theorem 15 of Hansen, Koucký, and Miltersen [21] to obtain a rational ε-optimal strategy. Letting $\varepsilon = 2^{-\ell}$ we may round each probability, except the largest, upwards to $L = \lg \frac{1}{\varepsilon} + \lg\lg \frac{1}{\varepsilon} + N\tau m^{O(K^2)}$ binary digits, and then rounding the largest probability down by the total amount the rest were rounded up. Here we use that by fixing the above strategy of patience at most $2^{\tau m^{O(K^2)}}$ for the safety player and an pure strategy for the reachability player one obtains a Markov chain where each non-zero transition probability is at least $(2^{\tau m^{O(K^2)}})^{-1}$. We thus have the following.

Corollary 34. *Let G be as in Theorem 32. Then there is an ε-optimal strategy for the safety player where each probability is a rational number with a common denominator of magnitude at most $\frac{1}{\varepsilon} \lg \frac{1}{\varepsilon} 2^{N\tau m^{O(K^2)}}$.*

We now address the basic decision problem. Let s be a state and let λ be a rational number with numerator and denominator of bit-size at most κ, and consider the task of deciding whether $v_2(s) \geq \lambda$. An equivalent task is to decide whether $v_2(s) - \lambda \geq 0$. Since $v_2(s)$ is a real algebraic number defined by a polynomial of degree $m^{O(K^2)}$ and maximum coefficient bit-size $\tau m^{O(K^2)}$ it follows that $v_2(s) - \lambda$ is a real algebraic number defined by a polynomial of degree $m^{O(K^2)}$ and maximum coefficient bit-size $(\kappa + \tau)m^{O(K^2)}$. This can be seen by subtracting λ from the univariate representation of $v_2(s)$ and substituting for the root t using a resultant. By standard root separation bounds this means that either is $v_2(s) - \lambda = 0$ or $|v_2(s) - \lambda| > \eta$, for some η of the form $d = 2^{-(\kappa+\tau)m^{O(K^2)}}$. Given an $\eta/2$-optimal strategy σ_2 for the safety player, by fixing the strategy σ_2 we obtain an MDP for

player 1, where we can find the value $\widetilde{v}_2(s)$ of state s using linear programming, and the computed estimate $\widetilde{v}_2(s)$ for $v_2(s)$ is within $\eta/2$ of the true value. Thus if $\widetilde{v}_2(s) \geq \lambda - \eta/2$ we conclude that $v_2(s) \geq \lambda$ (and similarly if $\widetilde{v}_2(s) \geq \lambda + \eta/2$ we conclude that $v_2(s) > \lambda$). Now, if we fix K to be a constant and consider the promise problem that G has at most K value classes, then a rational $\eta/2$-optimal strategy σ_2 exists with numerators and denominators of polynomial bit-size by Corollary 34. Now, by simply guessing non-deterministically the strategy σ_2 and verifying as above we have the following result.

Theorem 35. *For a fixed constant K, the promise problem of deciding whether $v_1(s) \geq \lambda$ given a zero-sum concurrent stochastic game with at most K value classes is in* coNP *if player 1 has reachability objective and in* NP *if player 1 has safety objective.*

Note that interestingly it does not follow similarly that the promise problem is in $(\mathsf{coNP} \cap \mathsf{NP})$, because the games are not symmetric.

Remark 36 (Complexity of approximation for constant value classes)**.** *As a direct consequence we have that for a game G promised to have at most K value classes, the value of a state can be approximated in $\mathsf{FP}^{\mathsf{NP}}$. This improves on the $\mathsf{FNP}^{\mathsf{NP}}$ bound of Frederiksen and Miltersen [17] (that holds in general with no restriction on the number of value classes).*

VI. Non-Zero-sum Concurrent Stochastic Games: Bounds on Patience and Roundedness

In this section we consider non-zero-sum concurrent stochastic games where each player has either a reachability or a safety objective. We first present a remark on the lower bound in the presence of even a single player with reachability objective, and then for the rest of the section focus on non-zero-sum games where all players have safety objectives.

Remark 37. *In non-zero-sum concurrent stochastic games, with at least two players, even if there is one player with reachability objectives, then at least doubly-exponential patience is required for ε-Nash equilibrium strategies. We have the property if $k = 2$ and one player is a reachability player and the other is a safety player, from Section III-C. It is also easy to see that Lemma 9 together with Lemma 15 implies that if player 1 is identified with the objective $(\mathsf{Reach}, \{\top\})$ and player 2 is identified with the objective $(\mathsf{Reach}, \{\bot\})$ and they are playing the Purgatory Duel, then each strategy profile σ, that forms a ε-Nash equilibrium, for any $0 < \varepsilon < \frac{1}{3}$, in the Purgatory Duel, has patience $2^{m^{\Omega(n)}}$. This is because player 2 has a harder objective (a subset of the plays satisfies it) than in Section III-C, but can still ensure the same payoff (by using an optimal strategy for player 2 in the concurrent reachability variant, which ensures that $\bot$ is reached with probability at least $\frac{1}{2}$). In this case, we say that a strategy is optimal (resp., ε-optimal) for a player, if it is optimal (resp., ε-optimal) for the corresponding player in the concurrent reachability version. It is clear that only if both strategies are optimal (resp., ε-optimal), then the strategies forms a Nash equilibrium (resp., ε-Nash equilibrium). Thus the doubly-exponential lower bound follows even for non-zero-sum games with two reachability players. The key idea to extend to more players, of which at least one is a reachability player, is as follows: Consider some reachability player i. The game for which the lower bound holds can be described as follows. First player i picks another player j and they then proceed to play the Purgatory Duel with parameters n, m against each other. This can be captured by a game with $k(2n+1)+3$ states, where each matrix has size at most $\max(m, k)$. Each player must then use doubly-expoential patience in every strategy profile that forms an ε-Nash equilibrium, for sufficently small $\varepsilon > 0$. First consider a player j that is different from i, and a strategy for player j with low patience. It follows that player i would then simply play against player j and win with good probability. Second, consider a strategy for player i with low patience and there are two cases. Either player i gets a payoff close to $\frac{1}{2}$ or not. If he gets a payoff close to $\frac{1}{2}$, then the player he is most likely to play against can deviate to an optimal strategy and increase his payoff by an amount close to $\frac{1}{2k}$, which player i loses. On the other hand, if player i gets a payoff far from $\frac{1}{2}$, then he can deviate to an optimal strategy and then he gets payoff $\frac{1}{2}$.*

The rest of the section is devoted to non-zero-sum concurrent stochastic games with safety objectives for all players, and first we establish an exponential upper bound on patience and then an exponential lower bound for ε-Nash equilibrium strategies, for $\varepsilon > 0$.

A. Exponential upper bound on roundedness

In this section we consider non-zero-sum concurrent safety games, with $k \geq 2$ players, and such games are also called stay-in-a-set games, by [30]. We will argue that, for all $0 < \varepsilon < \frac{1}{4}$, in any such game, there exists a strategy profile σ that forms an ε-Nash equilibrium and have roundedness at most

$$\frac{-32 \cdot k^2 \cdot \ln(\varepsilon) \cdot n \cdot (\delta_{\min})^{-n} \cdot m}{\varepsilon} .$$

Note that the roundedness is only exponential, as compared to the doubly-exponential patience when there is at least one reachability player (Remark 37). Note that the bound is polynomial in m and k; and also polynomial in n if $\delta_{\min} = 1$.

Players already lost, and all winners. For a prefix of a play $P_s^{\ell'}$, for a starting state s, play P_s and length ℓ', let $\widehat{L}(P_s^{\ell'})$ be the set of players that have not lost already in $P_s^{\ell'}$ (note that for each i, player i has lost in a play prefix if a state not in S^i has been visited in the prefix). Let $P_s^{\ell'}$ be some prefix of a play and we define $W(P_s^{\ell'})$ as the event that each player in $\widehat{L}(P_s^{\ell'})$ wins with probability 1.

Player-stationary strategies. As shown by [30], there exists a strategy profile $\sigma = (\sigma_i)_i$ that forms a Nash equilibrium. They show that the strategy σ_i, for any player i, in the witness Nash equilibrium strategy profile has the following properties: For each set of players Π and state s, there exists a probability distribution $\widehat{\sigma}_i(\Pi, s)$, such that for each prefix of a play $P_s^{\ell'}$,

play P_s and length ℓ', if $P_s^{\ell'}$ ends in s', we have that $\sigma_i(P_s^{\ell'}) = \widehat{\sigma}_i(\widehat{L}(P_s^{\ell'}), s')$ (i.e., the strategy only depends on the players who have not lost yet and the current state). Also, there exists some positional strategy σ_i', such that $\widehat{\sigma}_i(\Pi, s) = \sigma_i'(s)$, for all $i \notin \Pi$ (i.e., players who have lost already play some fixed positional strategy). This allows them to only consider the subgame G^{Π}, which is the game in which each player i not in Π plays σ_i'. Also, if there is a strategy profile which ensures that each player in Π wins with probability 1 if the play starts in s of G^{Π}, then the probability distribution $\widehat{\sigma}_i(\Pi, s)$ is pure[5] and it ensures that the players in Π wins with probability 1. We call strategies with these properties *player-stationary strategies*.

The real number ε and the length ℓ. In the remainder of this section, fix $0 < \varepsilon < \frac{1}{4}$ and fix the length ℓ, such that

$$\ell = -n \cdot k \cdot \ln(\varepsilon/(4k)) \cdot (\delta_{\min})^{-n} \ .$$

We will, in Lemma 39, argue that any player-stationary strategy is such that with probability $1 - \varepsilon$ no player loses after ℓ steps. Also several lemmas in this section will use ℓ and ε.

The event $E(P_s^{\ell'})$. Given a play P_s, starting in state s for some s and any ℓ', let $E(P_s^{\ell'})$ be the event that either the event $(\widehat{L}(P_s^{\ell'}) \subsetneq \widehat{L}(P_s^{\ell'-1}))$ (i.e., some player lost at the ℓ'-th step) or the event $W(P_s^{\ell'})$ (i.e., the remaining players win with probability 1) happens. In [30, 2.1 Lemma] they show[6]:

Lemma 38. *Fix a player-stationary strategy profile σ. Let $T \geq 0$ denote a round (or a step of plays). Let $Y^{T,s}$ be the set of plays, where for all plays P_s in $Y^{T,s}$, either the remaining players win with probability 1 in round T (i.e., the event $W(P_s^T)$ happens) or some player loses in round T (i.e., the event $\widehat{L}(P_s^T) \subsetneq \widehat{L}(P_s^{T-1})$ happens). For a constant c and length ℓ', let $y_{c,\ell'} = \Pr_\sigma[\exists T : \ell' < T \leq \ell' + cn \wedge P_s \in Y^{T,s}]$ denote the probability that event $Y^{T,s}$ happens for some T between ℓ' and $\ell' + cn$. Then, for all constants c and length ℓ', we have that*

$$y_{c,\ell'} \geq 1 - (1 - (\delta_{\min})^n)^c \ .$$

Note that T above depends on the play P_s. It is straightforward that players can lose at most k times in any play P_s, simply because there are at most k players, and if the remaining players win with probability 1 in round T, then they also win with probability 1 in round $T + 1$, by construction of σ.

Proof overview. Our proof will proceed as follows. Consider the game, while the players play some player-stationary strategy profile that forms a Nash equilibria. First, we show that it is unlikely (low-probability event) that the players do not play positional (like they do if the event $W(P_s^{\ell'})$ has happened) after some exponential number of steps. Second, we show that if we change each of the probabilities used by an exponentially small amount as compared to the Nash equilibria, then it is unlikely that that there will be a large difference in the first exponentially many steps. This allows us to round the probabilities to exponentially small probabilities while the players only lose little.

Lemma 39. *Fix some player-stationary strategy profile σ. Consider the set P of plays P_s, under σ, such that $W(P_s^{\ell})$ does not happen. Then, the probability $\Pr_\sigma[P]$ is less than $\varepsilon/4$.*

Proof. Fix $0 < \varepsilon < \frac{1}{2}$ and a player-stationary strategy profile σ. Let $c = -\ln(\varepsilon/(4k)) \cdot (\delta_{\min})^{-n} > 1$. We will argue that the event $E(P_s^{\ell'})$ happens at least k times with probability at least $1 - \varepsilon/4$ over $c \cdot n \cdot k = \ell$ steps.

We consider two cases, either $\delta_{\min} = 1$ or $0 < \delta_{\min} < 1$. If $\delta_{\min} = 1$, the event $\exists 1 \leq T \leq n : E(P_s^{\ell'+T})$ always happens (otherwise, in case it did not in some play, then a deterministic cycle satisfying the safety objectives of all players who have not lost yet is executed, and then the players could win by playing whatever they did the last time they were in a given state). If $0 < \delta_{\min} < 1$, we see that $c \geq c' = \frac{\ln(\varepsilon/(4k))}{\ln(1-(\delta_{\min})^{-n})}$, since $1+x \leq e^x$ and that $\exists 1 \leq T \leq c' \cdot n : E(P_s^{\ell'+T})$ happens with probability at least $1 - \varepsilon/(4k)$ by Lemma 38. In either case, we have that the event $\exists 1 \leq T \leq c \cdot n : E(P_s^{\ell'+T})$ happens with probability at least $1 - \varepsilon/(4k)$.

Next, split the plays up in epochs of length $c \cdot n$ each, and we get that the event $E(P_s^T)$ happens at least once for T ranging over the steps of an epoch with probability at least $1 - \varepsilon/(4k)$ and hence happens at least once in each of the first k epochs with probability at least $1 - \varepsilon/4$ using union bound. At that point the remaining players win with probability 1. The first k epochs have length $c \cdot k \cdot n = \ell$ and the lemma follows. □

We use the above lemma to show that any strategy profile close to a Nash equilibrium ensures payoffs close to that equilibrium. To do so, we use coupling (similar to [10]).

Variation distance. The *variation distance* is a measure of the similarity between two distributions. Given a finite set Z, and two distributions d_1 and d_2 over Z, the variation distance of the distributions is

$$\mathsf{var}(d_1, d_2) = \frac{1}{2} \cdot \sum_{z \in Z} |d_1(z) - d_2(z)| \ .$$

We will extend the notion of variation distances to strategies as follows: Given two strategies σ_i and σ_i' for player i the variation distance between the strategies is

$$\mathsf{var}(\sigma_i, \sigma_i') = \sup_{P_s^\ell} \mathsf{var}(\sigma_i(P_s^\ell), \sigma_i'(P_s^\ell)) \ ;$$

i.e., it is the supremum over the variation distance of the distributions used by the strategies for finite-prefixes of plays.

Coupling and coupling lemma. Given a pair of distributions, a coupling is a probability distribution over the joint set of possible outcomes. Let Z be a finite set. For distributions d_1 and d_2 over the finite set Z, a *coupling* ω is a distribution over $Z \times Z$, such that for all $z \in Z$ we have

[5] it is not explicitly mentioned in [30] that the distributions are pure, but it follows from the fact that if all players can ensure their objectives with probability 1, then there exists a positional strategy profile ensuring so, by just considering an MDP (with all players together) with a conjunction of safety objectives

[6] they do not explicitly show that the constant is $1-(\delta_{\min})^n$, but it follows easily from an inspection of the proof

$\sum_{z' \in Z} \omega(z, z') = d_1(z)$ and also for all $z' \in Z$ we have $\sum_{z \in Z} \omega(z, z') = d_2(z')$. One of the most important properties of coupling is the coupling lemma [1] of which we only mention and use the second part:

- **(Coupling lemma).** For a pair of distributions d_1 and d_2, there exists a coupling ω of d_1 and d_2, such that for a random variable (X, Y) from the distribution ω, we have that $\mathsf{var}(d_1, d_2) = \Pr[X \neq Y]$.

Smaller support. Fix a pair of strategies σ_i and σ'_i for player i for some i. We say that σ'_i has *smaller support* than σ_i, if for all P_s^ℓ we have that

$$\mathrm{Supp}(\sigma'_i(P_s^\ell)) \subseteq \mathrm{Supp}(\sigma_i(P_s^\ell)) \ .$$

Lemma 40. *Let $\sigma = (\sigma_i)_i$ and $\sigma' = (\sigma'_i)_i$ be player-stationary strategy profiles, such that*

$$\mathsf{var}(\sigma, \sigma') \leq \frac{\varepsilon}{\ell \cdot k \cdot 4} \ ,$$

and such that σ'_i has smaller support than σ_i, for all i. Then σ' is such that

$$u(G, s, \sigma', i) \in [u(G, s, \sigma, i) - \varepsilon/2, u(G, s, \sigma, i) + \varepsilon/2]$$

for each player i and state s.

Proof. Fix σ and σ' according to the lemma statement. For any prefix of a play $P_s^{\ell'}$, for any state s and length ℓ' and player i, we have that $\mathsf{var}(\sigma_i(P_s^{\ell'}), \sigma'_i(P_s^{\ell'})) \leq \frac{\varepsilon}{\ell \cdot k \cdot 4}$ and thus, we can create a coupling $\omega = (X_i^{P_s^{\ell'}}, Y_i^{P_s^{\ell'}})$ between the two distributions $\sigma_i(P_s^{\ell'})$ and $\sigma'_i(P_s^{\ell'})$, i.e., $X_i^{P_s^{\ell'}} \sim \sigma_i(P_s^{\ell'})$ and $Y_i^{P_s^{\ell'}} \sim \sigma'_i(P_s^{\ell'})$ is such that $\Pr[X_i^{P_s^{\ell'}} \neq Y_i^{P_s^{\ell'}}] \leq \frac{\varepsilon}{\ell \cdot k \cdot 4}$. Then, consider some state s and consider a play P_s, picked using the random variables $X_i^{P_s^{\ell'}}$, and a play Q_s, picked using the random variables $Y_i^{P_s^{\ell'}}$ (where, if the players uses the same action in $P_s^{\ell'}$ and $Q_s^{\ell'}$, then the next state is also the same, using an implicit coupling). Then according to Lemma 39, the probability that $W(P_s^\ell)$ occurs is at least $1 - \varepsilon/4$. In that case, we are interested in the probability that $Q_s = P_s$. Observe that we just need to ensure that P_s^ℓ and Q_s^ℓ are the same, since at that point the players play according to the same positional strategy, because of the smaller support. For each $\ell'' \leq \ell$, if the first ℓ'' steps match, then the next step match with probability at least $1 - \frac{\varepsilon}{\ell \cdot k \cdot 4} \cdot k$, since each of the k players has a probability of $\frac{\varepsilon}{\ell \cdot k \cdot 2}$ to differ in the two plays. Hence, all ℓ steps match with probability at least $1 - \frac{\varepsilon}{\ell \cdot k \cdot 4} \cdot \ell \cdot k = 1 - \varepsilon/4$. Hence, with probability at least $1 - \varepsilon/2$ we have that P_s equals Q_s and thus, especially, the payoff for each player must be the same in that case. But observe that P_s is distributed like plays under σ and Q_s is distributed like plays under σ' and the statement follows. □

We will next show that we only need to consider deviations to player-stationary strategies for the purpose of player-stationary equilibria.

Lemma 41. *For all player-stationary strategy profiles σ and each player i, there exists a pure player-stationary strategy σ'_i for player i maximizing $u(G, s, \sigma[\sigma'_i], i)$.*

Proof. Observe first that it does not matter what player i does if he has already lost, and we can consider him to play some fixed positional strategy in that case. Also, when the remaining players play according to σ, we can view the game as being an MDP, in the games G^Π. The objective of player i is then to reach a sub-game of G^Π and a state in that sub-game, from which he cannot lose. But it is well-known that such reachability objectives have positional optimal strategies in MDPs. Hence, this strategy forms a pure player-stationary strategy in the original game. □

We will use Lemma 3 from [10]. The proof only appears in [9], where the lemma is Lemma 4.

Lemma 42. ***(Lemma 3, [10]).*** *Let Z be a set of size ℓ. Let d_1 be some distribution over Z and let $q \geq \ell$ be some integer. Then there exists some distribution d_2, such that for each $z \in Z$, there exists an integer p such that $d_2(z) = \frac{p}{q}$ and such that $|d_1(z) - d_2(z)| < \frac{1}{q}$.*

We are now ready to show the main theorem of this section.

Theorem 43. *For all concurrent stochastic games with all k safety players, for all $0 < \varepsilon < \frac{1}{4}$, there exists a player-stationary strategy profile σ that forms an ε-Nash equilibrium and has roundedness at most*

$$4n \cdot k^2 \cdot m \cdot \varepsilon^{-1} \cdot \ln(4k/\varepsilon) \cdot (\delta_{\min})^{-n} \ .$$

Proof. Fix some player-stationary strategy profile σ that forms a Nash-equilibrium and some $0 < \varepsilon < \frac{1}{4}$ and let

$$\ell := -n \cdot k \cdot \ln(\varepsilon/(4k)) \cdot (\delta_{\min})^{-n} \ .$$

Consider some distribution d_1 over some set Z. Observe that for each distribution d_2 with smaller support than d_1 and such that $|d_1(z) - d_2(z)| < \frac{1}{q}$, for each $z \in \mathrm{Supp}(d_1)$, we have $\mathsf{var}(d_1, d_2) \leq \frac{|\mathrm{Supp}(d_1)|}{q}$. Then, applying Lemma 42, for $q = \frac{\ell \cdot k \cdot 4 \cdot m}{\varepsilon}$ and $Z = \mathrm{Supp}(d)$, to each probability distribution d defining σ, we see that there exists a player-stationary strategy profile $\sigma' = (\sigma'_i)_i$, such that (1)

$$\mathsf{var}(\sigma, \sigma') \leq \frac{m}{q} = \frac{\varepsilon}{\ell \cdot k \cdot 4} \ ;$$

and (2) σ'_i has smaller support than σ_i; and (3) $\sigma'_i(P_s^\ell)$ is a fraction with denominator q. Observe that the strategy has roundedness q.

We now argue that σ' is an ε-Nash equilibrium. Consider some player i and a player-stationary strategy σ''_i maximizing the probability that player i wins when the remaining players play according to σ', which is known to exists by Lemma 41. From Lemma 40, we have that

$$u(G, s, \sigma[\sigma''_i], i) \geq u(G, s, \sigma'[\sigma''_i], i) - \varepsilon/2$$

and

$$u(G, s, \sigma, i) \leq u(G, s, \sigma', i) + \varepsilon/2 \ .$$

Thus, $u(G, s, \sigma', i) \geq u(G, s, \sigma'[\sigma_i''], i) - \varepsilon$. This completes the proof. □

Remark 44 (Finding an ε-Nash equilibria in TFNP)**.** *We explain how the results of this section imply that for non-zero-sum concurrent stochastic games with safety objectives for all players, if the number k of players is only a constant or logarithmic, then we can compute an ε-Nash equilibria in* TFNP*, where $\varepsilon > 0$ is given in binary as part of the input. Note that there is a polynomial-size witness (to guess) for a stationary strategy with exponential roundedness. Observe that a player-stationary strategy for a player is defined by $2^{k-1}+1$ stationary strategies, one used in case that the respective player has lost, and one for each subset of other players. Thus, we can guess polynomial-size witnesses of k player-stationary strategies with exponential roundedness, given that the number of players is at most logarithmic in the size of the input. Hence, according to Theorem 43, we can guess a candidate strategy profile σ that forms an ε-Nash equilibrium in non-deterministic polynomial time. For each player i, constructing the (polynomial-sized) MDP described in the proof of Lemma 41 and then solving it using linear programming gives us the payoff of playing the strategy maximizing the value for player i while the remaining players follows σ. If, for each player i, the payoff only differs at most ε from what achieved by player i when all players follows σ, then the strategy profile σ is an ε-Nash equilibrium. It follows that the approximation of some ε-Nash equilibria can be achieved in* TFNP*, given that the number of players is at most logarithmic.*

B. Exponential lower bound on patience

In this section, we show that $\Omega((\delta_{\min})^{-(n-3)/6})$ patience is required, for each strategy profile that forms an ε-Nash equilibrium, for any $0 < \varepsilon < \frac{1}{6}$, in a family of games $\{G_c^{(\delta_{\min})} \mid c \in \mathbb{N} \wedge \delta_{\min} < 6^{-3}\}$ with two safety players.

Game family $G_c^{\delta_{\min}}$. For a fixed number $c \geq 1$ and $0 < \delta_{\min} < 6^{-3}$, the game $G_c^{\delta_{\min}}$ is defined as follows: There are $n = 4 \cdot c + 3$ states, namely, $S = \{v_s, v_1, v_2, \top, \bot\} \cup \{v_j^\ell \mid j \in \{1,2\} \wedge \ell \in \{1, \ldots, 2 \cdot c - 1\}\}$. For player i in state v_j, for $j = 1, 2$, there are two actions, called $a_i^{j,1}$ and $a_i^{j,2}$, respectively. For each other state s and each player i, there is a single action, a. For simplicity, for each pair of states s, s' we write $d(s, s')$ for the probability distribution, where $d(s, s')(s) = 1 - \delta_{\min}$ and $d(s, s')(s') = \delta_{\min}$. Also, we define v_1^0 as $\top$ and v_2^0 as $\bot$. The states $\bot$ and $\top$ are absorbing. The state v_s is such that[7] $\delta(v_s, a, a) = \mathsf{U}(v_1, v_2)$. For each $j \in \{1, 2\}$, the transition function of state v_j is

$$\delta(v_j, a_1^{j,\ell}, a_2^{j,\ell'}) = \begin{cases} d(v_s, v_j^{c-1}) & \text{if } \ell = \ell' \\ d(v_s, v_{\widehat{j}}^{2c-1}) & \text{if } \ell < \ell' \\ v_{\widehat{j}}^0 & \text{if } \ell > \ell' \end{cases}$$

For each other state v_j^ℓ, the transition function is $\delta(v_j^\ell, a, a) = d(v_s, v_j^{\ell-1})$. The objective of player 1 is $(\mathsf{Safety}, S \setminus \{\bot\})$ and the objective of player 2 is $(\mathsf{Safety}, S \setminus \{\top\})$. See Figure 5 for an illustration of $G_2^{\delta_{\min}}$.

[7] recall that $\mathsf{U}(s, s')$ is the uniform distribution over s and s'

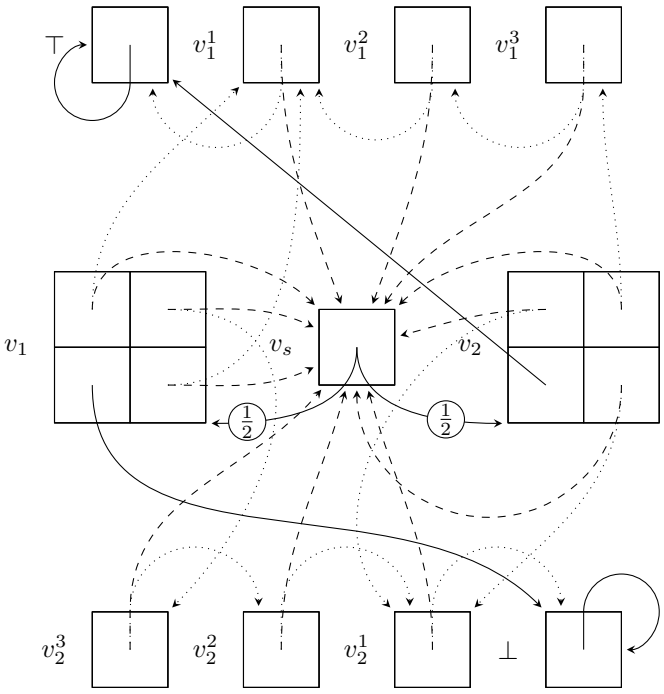

Fig. 5. An illustration of the game $G_2^{\delta_{\min}}$. The probabilities are as follows: The probability of each dashed edge is $1 - \delta_{\min}$; and the probability of each dotted edge is $\delta_{\min}$; and the probability of each solid edge is 1. The only exception is the edges from v_s, where the probability is written on each edge (it is $\frac{1}{2}$ in each case).

Near-zero-sum property. Observe that either $\bot$ or $\top$ is reached with probability 1 (and once $\top$ or $\bot$ is reached, the game stays there). The reasoning is as follows: there is a probability of at least $(\delta_{\min})^{2c}$ to reach either $\top$ or $\bot$ within the next $2c + 1$ steps from any state. If the current state is v_s, then the next state is either v_1 or v_2, and from v_1 or v_2 through v_j^ℓ for each ℓ from 1 to $2c - 1$, for some j, either $\top$ or $\bot$ is reached, and each of the steps from v_1 or v_2 onward happens with probability at least $\delta_{\min}$, no matter the choice of the players. Hence, the game is in essence zero-sum, since precisely one player wins with probability 1.

Proof overview. Our proof has two parts. We show that there is a strategy for player i, for each i, that ensures that against all strategies for the other player, the payoff is at least $\frac{1}{2}$ for player i. Also, we show that for each strategy of player i with patience at most $(\delta_{\min})^{-2/3 \cdot c}$, there is a strategy for the other player such that the payoff is less than $\frac{1}{6}$ for player i. This then allows us to show that no strategy profile that forms a $\frac{1}{6}$-Nash equilibrium has patience less than $(\delta_{\min})^{-2/3 \cdot c}$.

Lemma 45. *For each i, player i has a strategy σ_i such that*

$$\inf_{\sigma_{\widehat{i}}} u(G, v_s, \sigma_1, \sigma_2, i) = \frac{1}{2} \ .$$

Proof. Consider the stationary strategy σ_1, where

$$\sigma_1(v_1)(a_1^{1,1}) = \sigma_1(v_2)(a_1^{2,1}) = \frac{1 + (\delta_{\min})^{-c}}{2 + (\delta_{\min})^{-c} + (\delta_{\min})^{c}}$$

and

$$\sigma_1(v_1)(a_1^{1,2}) = \sigma_1(v_2)(a_1^{2,2}) = \frac{1+(\delta_{\min})^c}{2+(\delta_{\min})^{-c}+(\delta_{\min})^c} \ .$$

Observe that fixing σ_1 as the strategy for player 1, the game turns into an MDP for player 2. Such games have a positional strategy ensuring that the payoff for player 2 is as large as possible. Going through all four candidates for σ_2, one can see that $\max_{\sigma_2} u(G, v_s, \sigma_1, \sigma_2, 2) = \frac{1}{2}$. Because of the near-zero-sum property, this minimizes the payoff for player 1 (since $u(G, v_s, \sigma_1, \sigma_2, 1) + u(G, v_s, \sigma_1, \sigma_2, 2) = 1$), which is then $\inf_{\sigma_2} u(G, v_s, \sigma_1, \sigma_2, 1) = \frac{1}{2}$. The strategy for player 2 follows from σ_1 and the symmetry of the game. □

We next argue that if player i uses a low-patience strategy, then the opponent can ensure low payoff for player i.

Lemma 46. *Let σ_i be a strategy for player i with patience at most $(\delta_{\min})^{-2/3\cdot c}$. Then there exists a pure strategy $\sigma_{\widehat{i}}$ such that $u(G, v_s, \sigma_1, \sigma_2, \widehat{i}) > 1 - \frac{1}{6}$.*

Proof. Consider first player 1 (the argument for player 2 follows from symmetry). Let σ_1 be some strategy with patience at most $(\delta_{\min})^{-(n-3)/6} = (\delta_{\min})^{-2/3\cdot c}$.

The pure strategy σ_2 is defined given σ_1 as follows. For plays P_s^ℓ ending in state v_1 or v_2 we have that

$$\sigma_2(P_s^\ell) = \begin{cases} a_2^{j,j} & \text{if } \sigma_1(P_s^\ell)(a_2^{j,2}) > 0 \\ a_2^{j,\widehat{j}} & \text{if } \sigma_1(P_s^\ell) = a_2^{j,1} \end{cases} .$$

To argue that $u(G, v_s, \sigma_1, \sigma_2, 2) > 1 - \frac{1}{6}$, we consider a play P_{v_s} picked according to (σ_1, σ_2), such that either $\bot$ or $\top$ is eventually reached. This is true with probability 1. Consider the last round ℓ, such that $v_\ell = v_j$, for some $j = 1, 2$. We now consider four cases: Either we have that

1) $j = 1$ and $\sigma_1(P_s^\ell)(a_2^{j,2}) > 0$ or
2) $j = 1$ and $\sigma_1(P_s^\ell) = a_2^{j,1}$ or
3) $j = 2$ and $\sigma_1(P_s^\ell)(a_2^{j,2}) > 0$ or
4) $j = 2$ and $\sigma_1(P_s^\ell) = a_2^{j,1}$.

The probability to eventually reach $\bot$ is then at least the minimum probability to eventually reach $\bot$ in each of the four cases. In case (2) and case (4), we see that player 2 wins with probability 1. In case (1) observe that from a round ℓ' where $\sigma_1(P_s^{\ell'})(a_2^{1,2}) > 0$ player 1 wins (i.e., reaches $\top$ before entering v_s again) with probability $(1-(\delta_{\min})^{2/3\cdot c})\cdot(\delta_{\min})^c < (\delta_{\min})^c$ and player 2 wins (i.e., reaches $\bot$ before entering v_s again) with probability $(\delta_{\min})^{2/3\cdot c}$. Hence, the probability that player 1 wins if such a round is round ℓ is at most

$$\frac{(\delta_{\min})^c}{(\delta_{\min})^{2/3\cdot c}+(\delta_{\min})^c} < \frac{(\delta_{\min})^c}{(\delta_{\min})^{2/3\cdot c}} = (\delta_{\min})^{c/3} < \frac{1}{6} \ ,$$

where the last inequality comes from that $c \geq 1$ and $\delta_{\min} < 6^{-3}$. In case (3) observe that from a round ℓ' where $\sigma_1(P_s^{\ell'})(a_2^{2,2}) > 0$ player 1 wins (i.e., reaches $\top$ before entering v_s again) with probability at most $(1 - (\delta_{\min})^{2/3\cdot c}) \cdot (\delta_{\min})^{2c} < (\delta_{\min})^{2c}$ and player 2 wins (i.e., reaches $\bot$ before entering v_s again) with probability at least $(\delta_{\min})^{2/3\cdot c} \cdot (\delta_{\min})^c = (\delta_{\min})^{5/3\cdot c}$. Hence, the probability that player 1 wins if such a round is round ℓ is at most

$$\frac{(\delta_{\min})^{2\cdot c}}{(\delta_{\min})^{5/3\cdot c}+(\delta_{\min})^{2\cdot c}} < \frac{(\delta_{\min})^{2\cdot c}}{(\delta_{\min})^{5/3\cdot c}} = (\delta_{\min})^{c/3} < \frac{1}{6} \ ,$$

where the last inequality comes from that $c \geq 1$ and $\delta_{\min} < 6^{-3}$. The desired result follows. □

We now prove the main result that no strategy with patience only $(\delta_{\min})^{-2/3\cdot c}$ can be a part of a $\frac{1}{6}$-Nash equilibrium.

Theorem 47. *For all $c \in \mathbb{N}$ and all $0 < \delta_{\min} < 6^{-3}$, consider the game $G_c^{\delta_{\min}}$ (that has $n = 4c + 3$ states and at most two actions for each player at all states). Each strategy profile $\sigma = (\sigma_i)_i$ that forms an $\frac{1}{6}$-Nash equilibrium has patience at least $(\delta_{\min})^{-(n-3)/6}$.*

Proof. Fix some $c \in \mathbb{N}$ and $0 < \delta_{\min} < 6^{-3}$. The proof will be by contradiction. Consider first player 1 (the argument for player 2 follows from symmetry). Let σ_1 be some strategy with patience at most $(\delta_{\min})^{-(n-3)/6} = (\delta_{\min})^{-2/3\cdot c}$.

Consider some strategy σ_2 for player 2. We consider two cases, either

$$u(G, v_s, \sigma_1, \sigma_2, 2) \leq \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

or not. If

$$u(G, v_s, \sigma_1, \sigma_2, 2) \leq \frac{2}{3} \ ,$$

then player 2 can play a strategy σ_2', shown to exist in Lemma 46, instead and get payoff strictly above $1 - \frac{1}{6} = \frac{5}{6}$, showing that (σ_1, σ_2) is not an $\frac{1}{6}$-Nash equilibrium. On the other hand, if

$$u(G, v_s, \sigma_1, \sigma_2, 2) > \frac{2}{3} \ ,$$

then $u(G, v_s, \sigma_1, \sigma_2, 1) < \frac{1}{3}$ and player 1 can play a strategy σ_1', shown to exist in Lemma 45, for which $u(G, v_s, \sigma_1', \sigma_2, 1) \geq \frac{1}{2}$. Hence, (σ_1, σ_2) does not form an $\frac{1}{6}$-Nash equilibrium in this case either. The desired result follows. □

Remark 48. *Using ideas similar to Remark 37 we can construct a game with $k \geq 3$ safety players in which the patience is at least $(\delta_{\min})^{-(n-3)/(6k)}$ for all strategy profiles that forms an $\frac{1}{6k}$-Nash equilibrium.*

VII. Discussion

In this section, we discuss some interesting technical aspects of our results.

Remark 49 (Difference of exponential bounds)**.** *In this work we present two different exponential bound on patience. The first for zero-sum concurrent stochastic games, and the second for non-zero-sum concurrent stochastic games with safety objectives for all players. However, note that the nature of the lower bounds are very different. The first lower bound is exponential in the number of actions, and the size of the state space is constant. In contrast, for non-zero-sum concurrent stochastic games with safety objectives for all players, if the*

size of the state space is constant, then our upper bound on patience is polynomial. The second lower bound in contrast to the first lower bound is exponential in the number of states (and the upper bound is polynomial in m and also the number of players).

Remark 50 (Concurrent games with deterministic transitions)**.** *We now discuss our results for concurrent games with deterministic transitions. It follows from the results of [8] that for zero-sum games, there is a polynomial-time reduction from concurrent stochastic games to concurrent games with deterministic transitions. Hence, all our lower bound results for zero-sum games also hold for concurrent deterministic games. Observe that this is also true for our lower bound on non-zero sum games with at least one reachability player, since we reduce the problem to the zero-sum case. However, in general for non-zero-sum games polynomial-time reductions from concurrent stochastic games to concurrent deterministic games are not possible. For example, for concurrent stochastic games with safety objectives for all players we establish an exponential lower bound on patience of strategies that constitute an $1/6$-Nash equilibrium, whereas in contrast, our upper bound on patience shows that if the game is deterministic (i.e., $\delta_{\min} = 1$) and ϵ is constant, then there always exists an ε-Nash equilibrium that requires only polynomial patience.*

Remark 51 (Nature of strategies for the reachability player)**.** *Another important feature of our result is as follows: for zero-sum concurrent stochastic games, the characterization of [18] of ϵ-optimal strategies as* monomial *strategies for reachability objectives, separates the description of the strategies as a part that is a function of ϵ, and a part that is independent ϵ. The previous double-exponential lower bound on patience from [21], [19] shows that the part dependent on ϵ requires double-exponential patience, whereas the part that is independent only requires linear patience. A witness for ϵ-optimal strategies in Purgatory (as described in [13] for the value-1 problem for general zero-sum concurrent stochastic game) can be obtained as a ranking function on states and actions, such that the actions with rank 0 are played with uniform probability (linear patience); and an action of rank i at a state of rank j is played with probability roughly proportional to ϵ^{i^j}. In contrast, since we show lower bound for optimal strategies (and the strategies are symmetric) in Purgatory Duel, our lower bound implies that also the part that is independent of ϵ requires double-exponential patience in general (i.e., the probability description of ϵ-optimal strategies needs to be doubly exponentially precise).*

REFERENCES

[1] D. Aldous. Random walks on finite groups and rapidly mixing markov chains. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XVII 1981/82*, volume 986 of *Lecture Notes in Mathematics*, pages 243–297. Springer Berlin Heidelberg, 1983.

[2] R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

[3] S. Basu, R. Pollack, and M. Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2nd edition, 2006.

[4] P. Bouyer, R. Brenguier, and N. Markey. Nash equilibria for reachability objectives in multi-player timed games. In *CONCUR'10*, pages 192–206, 2010.

[5] P. Bouyer, R. Brenguier, N. Markey, and M. Ummels. Concurrent games with ordered objectives. In *FOSSACS'12*, pages 301–315, 2012.

[6] K. Chatterjee. Concurrent games with tail objectives. *Theoretical Computer Science*, 388:181–198, 2007.

[7] K. Chatterjee, L. de Alfaro, and T. Henzinger. Qualitative concurrent parity games. *ACM ToCL*, 2011.

[8] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *CoRR abs/1006.0673 (Full version)*, 2010. Conference version Proc. of MFCS, Springer, LNCS 6281, pages 246–257.

[9] K. Chatterjee and R. Ibsen-Jensen. The complexity of ergodic mean-payoff games. *CoRR*, abs/1404.5734, 2014.

[10] K. Chatterjee and R. Ibsen-Jensen. The Complexity of Ergodic Mean-payoff Games. In *ICALP 2014*, pages 122–133, 2014.

[11] L. de Alfaro, T. Henzinger, and F. Mang. The control of synchronous systems. In *CONCUR'00*, LNCS 1877, pages 458–473. Springer, 2000.

[12] L. de Alfaro, T. Henzinger, and F. Mang. The control of synchronous systems, part ii. In *CONCUR'01*, LNCS 2154, pages 566–580. Springer, 2001.

[13] L. de Alfaro, T. A. Henzinger, and O. Kupferman. Concurrent reachability games. *Theor. Comput. Sci*, 386(3):188–217, 2007.

[14] K. Etessami and M. Yannakakis. Recursive concurrent stochastic games. In *ICALP'06 (2)*, LNCS 4052, Springer, pages 324–335, 2006.

[15] H. Everett. Recursive games. In *CTG*, volume 39 of *AMS*, pages 47–78, 1957.

[16] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1997.

[17] S. K. S. Frederiksen and P. B. Miltersen. Approximating the value of a concurrent reachability game in the polynomial time hierarchy. In *ISAAC*, pages 457–467, 2013.

[18] S. K. S. Frederiksen and P. B. Miltersen. Monomial strategies for concurrent reachability games and other stochastic games. In *Reachability Problems'13*, pages 122–134, 2013.

[19] K. A. Hansen, R. Ibsen-Jensen, and P. B. Miltersen. The complexity of solving reachability games using value and strategy iteration. In *CSR*, pages 77–90, 2011.

[20] K. A. Hansen, M. Koucký, N. Lauritzen, P. B. Miltersen, and E. P. Tsigaridas. Exact algorithms for solving stochastic games: extended abstract. In *STOC*, pages 205–214. ACM, 2011. Precise references are to the full version: arXiv:1202.3898.

[21] K. A. Hansen, M. Koucký, and P. B. Miltersen. Winning concurrent reachability games requires doubly-exponential patience. In *LICS*, pages 332–341, 2009.

[22] C. J. Himmelberg, T. Parthasarathy, T. E. S. Raghavan, and F. S. V. Vleck. Existence of p-equilibrium and optimal stationary strategies in stochastic games. *Proc. Amer. Math. Soc.*, 60:245–251, 1976.

[23] R. Ibsen-Jensen. *Strategy complexity of two-player, zero-sum games*. PhD thesis, Aarhus University, 2013.

[24] R. Ibsen-Jensen and P. B. Miltersen. Solving simple stochastic games with few coin toss positions. In *ESA*, pages 636–647, 2012.

[25] J. N. Jr. Equilibrium points in n-person games. *PNAS*, 36:48–49, 1950.

[26] R. Lipton, E. Markakis, and A. Mehta. Playing large games using simple strategies. In *EC 03: Electronic Commerce*, pages 36–41. ACM Press, 2003.

[27] P. B. Miltersen and T. B. Sørensen. A near-optimal strategy for a heads-up no-limit texas hold'em poker tournament. In *AAMAS'07*, pages 191–197, 2007.

[28] G. Owen. *Game Theory*. Academic Press, 1995.

[29] T. Parthasarathy. Discounted and positive stochastic games. *Bull. Amer. Math. Soc*, 77:134–136, 1971.

[30] P. Secchi and W. D. Sudderth. Stay-in-a-set games. *International Journal of Game Theory*, 30(4):479–490, 2002.

[31] L. Shapley. Stochastic games. *PNAS*, 39:1095–1100, 1953.

[32] M. Ummels and D. Wojtczak. The complexity of Nash equilibria in stochastic multiplayer games. *Logical Methods in Computer Science*, 7(3), 2011.

[33] M. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *FOCS'85*, pages 327–338. IEEE, 1985.

[34] J. von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1947.

[35] C. K. Yap. *Fundamental Problems of Algorithmic Algebra*. Oxford University Press, New York, 2000.