Strategy Complexity of Concurrent Safety Games*

Krishnendu Chatterjee[1], Kristoffer Arnsfelt Hansen[2], and Rasmus Ibsen-Jensen[3]

1 **IST Austria, Klosterneuburg, Austria**
`krish@ist.ac.at`
2 **Aarhus University, Aarhus, Denmark**
`arnsfelt@cs.au.dk`
3 **IST Austria, Klosterneuburg, Austria**
`ribsen@ist.ac.at`

Abstract

We consider two player, zero-sum, finite-state concurrent reachability games, played 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. Player 1 wins iff a designated goal state is eventually 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: (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.

1998 ACM Subject Classification I.2.1 Games

Keywords and phrases Concurrent games, Reachability and safety, Patience of strategies

Digital Object Identifier 10.4230/LIPIcs.MFCS.2017.55

1 Introduction

Concurrent reachability games

Concurrent reachability games[8] are played on finite-state graphs by 2 players for an infinite number of rounds. In every round, each player simultaneously and independently of the other player 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 actions. The play starts in a designated *start state*. Player 1 wins the play iff the play ever enters a designated *goal state*. We say that player 1 is the *reachability player* and player 2 the *safety player*. These games were introduced in a seminal work by Shapley [23], 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 of problems with diverse applications, when there is a finite number of interactions [19, 26]. Concurrent reachability 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 [11]. Moreover, in analysis of reactive systems, concurrent games provide the appropriate model for reactive systems with components that interact synchronously [6, 7, 1].

* Some proofs are missing. See full version `https://arxiv.org/abs/1506.02434`, [4]

© Krishnendu Chatterjee, Rasmus Ibsen-Jensen, and Kristoffer Arnsfelt Hansen;
licensed under Creative Commons License CC-BY
42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017).
Editors: Kim G. Larsen, Hans L. Bodlaender, and Jean-Francois Raskin; Article No. 55; pp. 55:1–55:13
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Relevance

Concurrent reachability games are relevant in many applications. For example, the synthesis problem in control theory (e.g., discrete-event systems as considered in [22]) corresponds to reactive synthesis of [21]. The synthesis problem for synchronous reactive systems is appropriately modeled as concurrent games [6, 7, 8]. Other than control theory, concurrent reachability games also provide the appropriate model to study several other interesting problems, such as two-player poker games [18].

Properties of strategies

Given a concurrent reachability games, the player-1 *value* $v_1(s)$ of the game at a state s is the limit probability with which he can guarantee that the play will eventually enter the goal state 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 reachability games are determined [10], 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 $\varepsilon \geq 0$, a strategy is ε-optimal for a state s for player i if it ensures his own objective with probability at least $v_i(s) - \varepsilon$ against all strategies of the opponent. A 0-optimal strategy is an *optimal* strategy. In concurrent reachability games, there exist stationary optimal strategies for the safety player [20, 14]; whereas in contrast, for the reachability player, optimal strategies do not exist in general, however, for every $\varepsilon > 0$ there exists stationary ε-optimal strategies [10].

The significance of patience and roundedness of strategies

The basic decision problem is as follows: given a concurrent reachability game and a rational threshold λ, decide whether $v_1(s) \geq \lambda$. The basic decision problem is in PSPACE and is *square-root sum* hard [9][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 problem is closely related to the size of the description of ε-optimal strategies. Even for special cases of concurrent reachability game, namely *turn-based* reachability 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 reachability games. Also see [27, 28, 24] for the significance of computing strategies in concurrent stochastic games.

---

[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.

Previous results and our contributions

In this work we consider concurrent reachability games. We first describe the relevant previous results and then our contributions.

Previous results

For concurrent reachability game, the optimal bound on patience and roundedness for ε-optimal strategies for the reachability player, for $\varepsilon > 0$, is doubly exponential [13, 12]. 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) [13, 12]; whereas the doubly-exponential upper bound is obtained by expressing the values in the existential theory of reals [13, 12]. In contrast to the reachability player that in general do not have optimal strategies, similar to the safety player there are two related classes of concurrent stochastic games that admit optimal stationary strategies, namely, discounted-sum, and ergodic concurrent games. For both these classes the optimal bound on patience and roundedness for ε-optimal strategies, for $\varepsilon > 0$, is exponential [5, 15]. The optimal bound on patience and roundedness for optimal and ε-optimal strategies, for $\varepsilon > 0$, for the safety player has been an open problem.

Our contributions

Our main results are as follows:

1. *Lower bound: general.* We show that in concurrent reachability games, a lower bound on patience of optimal and ε-optimal strategies, for $\varepsilon > 0$, for the safety player is doubly exponential (in contrast to the above mentioned related classes of games that only require exponential patience). We present a family of games (namely, Purgatory Duel) where optimal and ε-optimal strategies, for $\varepsilon > 0$, for both players require doubly-exponential patience.
2. *Lower bound: three states.* We show that even in concurrent reachability 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 $\varepsilon > 0$, is exponential (in the number of actions). An optimal (resp., ε-optimal, for $\varepsilon > 0$) 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 depend 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 $\varepsilon > 0$, is only logarithmic [17]; and perhaps surprisingly in contrast we show that the patience for ε-optimal strategies in concurrent reachability games with only three states is exponential (i.e., there is a doubly-exponential increase from logarithmic to exponential).
3. *Upper bound.* We show that in concurrent reachability 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.

In summary, we present a complete picture of the patience and roundedness required in concurrent reachability games.

Table 1 Strategy complexity (i.e., patience and roundedness of ε-optimal strategies, for $\varepsilon > 0$) of the reachability vs safety player 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).

# Value classes	Reachability	Safety
1	Linear	One
2	Double-exponential	One
3	Double-exponential	**Exponential** **LB, Theorem 13**
Constant	Double-exponential	**Exponential** **UB, Theorem 14**
General	Double-exponential	**Double-exponential** **LB, Theorem 12** **UB, Theorem 14**

Distinguishing aspects of safety and reachability

While the optimal bound on patience and roundedness we establish in concurrent reachability 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 1). For the reachability player, if there is one value class, then the patience and roundedness required is linear: it follows from the results of [2] 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 [8, 3] 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 [8], 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 the reachability player).

Our main ideas

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

The game of *Purgatory* [13, 12] is a concurrent reachability game 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 modeled 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 $1 - \varepsilon$ [10], 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 $\varepsilon = 1 - 4m^{-N/2}$ [12]. 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 modeled 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.

Technical contributions

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 $\varepsilon \geq 0$, 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.

Highlights

We highlight two features of our results, namely, the surprising aspects and the significance (see Section Discussion and Conclusion of the full version for further details).

1. *Surprising aspects.* The first surprising aspect of our result is the doubly-exponential lower bound for the safety player in concurrent reachability games. The properties of strategies for the safety player in concurrent reachability games resemble concurrent discounted games, as in both cases optimal stationary strategies exist, and locally optimal strategies are optimal. We show that in contrast to concurrent discounted games where exponential patience suffices for the safety player in concurrent reachability games doubly-exponential patience is necessary. The second surprising aspect is the lower bound example itself. The lower bound example is obtained as follows: (i) given Purgatory we first obtain

simplified Purgatory by changing the start state such that it deterministically goes to the next state; (ii) we then consider its dual where the roles of the players are exchanged; and (iii) Purgatory duel is obtained by merging the start states of simplified Purgatory and its dual. Both in simplified Purgatory and its dual, there are only two value classes, and positional optimal strategies exist for the safety player. Surprisingly we show that a simple merge operation gives a game with linear number of value classes and the patience increases from 1 to doubly-exponential. Finally, the properties of strategies for the reachability- and safety-player in concurrent reachability games differ substantially. An important aspect of our lower bound example is that we show how to modify an example for the reachability player to obtain the result for safety player.

2. *Significance.* Our most important results are the lower bounds, and the main significance is threefold. First, the most well-studied way to obtain computational complexity result in games is to explicitly guess strategies, and then verify the game obtained fixing the strategy. The lower bound for the reachability player by itself did not rule out that better complexity results can be obtained through better strategy complexity for the safety player (indeed, for constant number of value classes, we obtain a better complexity result than known before due to the exponential bound on roundedness). Our doubly-exponential lower bound shows that in general the method of explicitly guessing strategies would require exponential space, and would not yield NP or coNP upper bounds. Second, one of the most well-studied algorithm for games is the strategy-iteration algorithm. Our result implies that any natural variant of the strategy-iteration algorithm for the safety player that explicitly compute strategies require exponential space in the worst-case. Finally, in games, strategies that are witness to the values and specify how to play the game, are as important as values, and our results establish the precise strategy complexity (matching upper bound of roundedness with lower bounds of patience).

Full-version: Proofs and non-zero-sum games.

In the full version [4], we give full proofs of all our lemmas and also consider non-zero-sum and non-two-player concurrent games, but where each player has either a reachability or safety objective (concurrent reachability games is then the special case of 1 player with a reachability objective and 1 player with the complementary safety objective).

2 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 vice-versa.

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* $\text{Supp}(d)$ 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 $\text{Supp}(d) = Z$. The *patience* of d is $\max_{z \in \text{Supp}(d)} \frac{1}{d(z)}$, 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 reachability games

A concurrent reachability game, 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_s^i of *actions* (and $A^i = \bigcup_s A_s^i$ 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_s^i consists of at most m actions; and (3) a stochastic *transition function* $\delta : S \times A^1 \times A^2 \to \Delta(S)$; and (4) a designated *goal state* $g \in S$. A state s is *deterministic* if $\delta(s, a_1, a_2)$ is pure (deterministic), for all $a_i \in A_s^i$ and for all i. A state s is called *absorbing* if $A_s^i = \{a\}$ for all i and $\delta(s, a, a) = s$. The number $\delta_{\min}$ is the smallest non-zero transition probability.

How to play a concurrent reachability game

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 $a_{T+1}^i \in A_{v_T}^i$. Then, the game selects v_{T+1} according to the probability distribution $\delta(v_T, a_{T+1}^1, a_{T+1}^2)$ 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 pair 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 $(v_0, (a_1, v_1), (a_2, v_2), \dots, (a_T, v_T), \dots)$, and for each ℓ a prefix of P_s^ℓ of length ℓ is then $(v_0, (a_1, v_1), (a_2, v_2), \dots, (a_T, v_T), \dots, (a_\ell, v_\ell))$, and we say that P_s^ℓ *ends in* v_ℓ. Player 1 wins a play P_s iff $v_T = g$ for some T. Similarly, player 2 wins a play P_s iff $v_T \neq g$ for *all* T We refer to player 1 as the reachability player and player 2 as the safety player.

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_s^ℓ, for a play P_s of length ℓ starting at state s, to a distribution over $A_{v_\ell}^i$. Player i *follows* a strategy σ_i, if given the current prefix of a play is P_s^ℓ, he selects $a_{\ell+1}$ according to $\sigma_i(P_s^\ell)$, 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_s^ℓ and $(P')_{s'}^{\ell'}$ ends in the same state t, we have that $\sigma_i(P_s^\ell) = \sigma_i((P')_{s'}^{\ell'})$; 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 $\sigma_i(P_s^\ell)$, over all plays P_s starting at state s, and all lengths ℓ. Also, a strategy σ_i is *pure* (resp., *totally mixed*) if $\sigma_i(P_s^\ell)$ 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. Let Σ^i be the set of all strategies for player i.

Strategy profiles

A *strategy profile* $\sigma = (\sigma_1, \sigma_2)$ is a pair 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 [25]. We say that a strategy profile has a property (e.g., is stationary) if each of the strategies in the profile has that property.

Values

Let $u(G, s, \sigma)$ be the probability that player 1 wins the game G when the players follow σ and the play starts in s (i.e., the utility or payoff for player 1). Also if the game G is clear from context we drop it from the notation. Given a concurrent reachability game G, the *upper value* $\overline{\text{val}}(G, s)$ (resp., *lower value* $\underline{\text{val}}(G, s)$) of G starting in s is

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

As shown by [10] we have that $\text{val}(G, s) := \overline{\text{val}}(G, s) = \underline{\text{val}}(G, s)$; which is called the *value* of s. We will sometimes write $\text{val}(s)$ for $\text{val}(G, s)$ if G is clear from the context. We will also write val for the vector where $\text{val}_s = \text{val}(s)$.

(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 $\text{val}(s) - \varepsilon \leq \inf_{\sigma_2 \in \Sigma^2} u(s, \sigma_1, \sigma_2)$ (resp., $\text{val}(s) + \varepsilon \geq \sup_{\sigma_1 \in \Sigma^1} u(s, \sigma_1, \sigma_2)$). 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 [10] for an example. On the other hand in all concurrent reachability games G player 1 has a stationary ε-optimal strategy for each $\varepsilon > 0$. In all concurrent reachability games player 2 has an optimal stationary strategy (thus also an ε-optimal stationary strategy for all $\varepsilon > 0$) [20, 14]. 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.

Markov decision processes and Markov chains

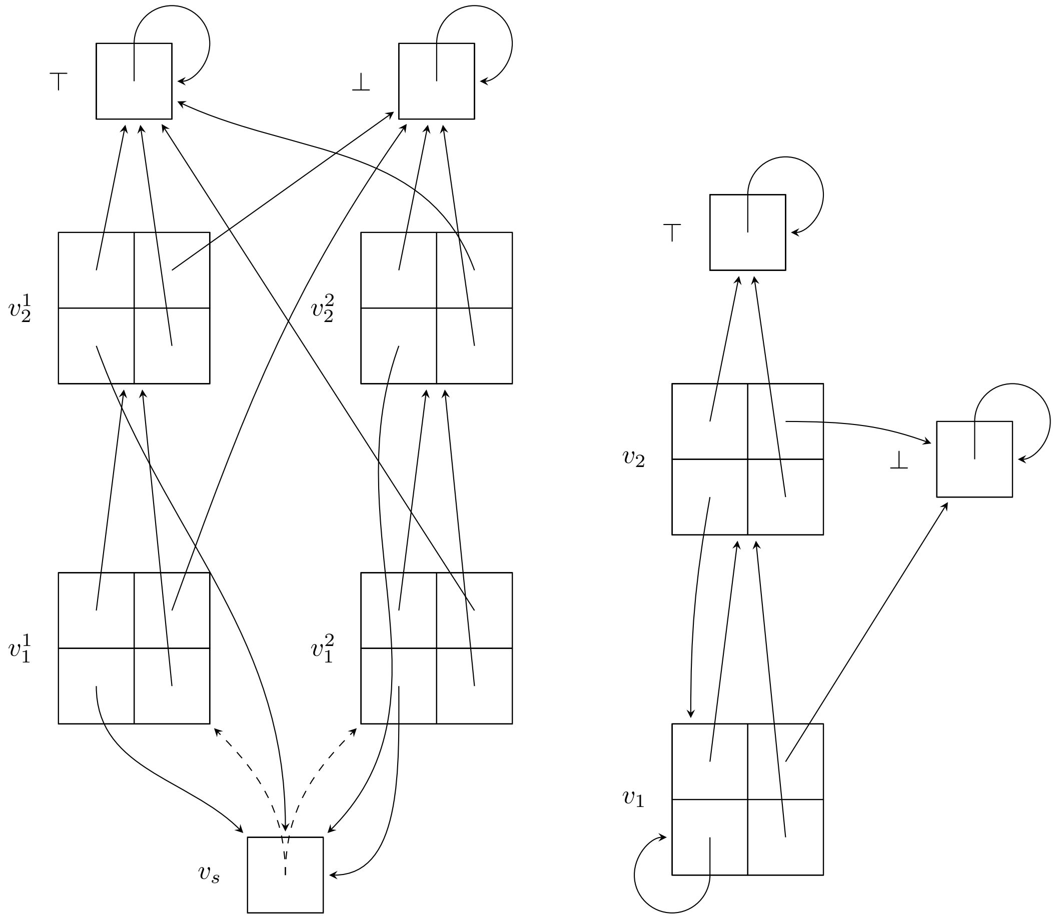
For each player i, a *Markov decision process (MDP) for player i* is a concurrent game where the size of A_s^j is 1 for all s and $j \neq i$. A *Markov chain* is an MDP for each player (that is the size of A_s^j is 1 for all s and j). A *closed recurrent set* of a Markov chain G is a maximal (i.e., no 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 reccurent 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.

Game illustration

When we illustrate our games, we illustrate each state as a matrix, where the rows corresponds to the actions of the reachability player, the columns corresponds to the actions of the safety player. Thus, each entry e corresponds to an pair of actions (i, j) and a state s and we have an edge to $\delta(s, i, j)$ from e.

3 Patience Lower Bound

In this section we will establish the doubly-exponential lower bound on patience for concurrent reachability games. First we define the game family, namely, *Purgatory Duel* and we also recall the family *Purgatory*.



(a) Illustration of the Purgatory Duel with $m = n = 2$. The dashed edges have probability $\frac{1}{2}$ each.

(b) Illustration of Purgatory with $m = n = 2$.

Figure 1 Illustration of the games used for lower bounds.

The Purgatory Duel

In this paper we specifically focus on the following concurrent reachability game, the *Purgatory Duel*, defined on a pair of parameters (n, m). The game consists of $N = 2n+3$ states, namely $\{v_1^1, v_2^1, \ldots, v_n^1, v_1^2, v_2^2, \ldots, v_n^2, v_s, \top, \bot\}$ and all but v_s are deterministic. To simplify the definition of the game, let $v_0^1 = v_{n+1}^2 = \bot$ and $v_0^2 = v_{n+1}^1 = \top$. The states $\top$ and $\bot$ are absorbing. For each $i \in \{1, 2\}$ and $j \in \{1, \ldots, n\}$, the state v_j^i is such that $A^1_{v_j^i} = A^2_{v_j^i} = \{1, 2, \ldots, m\}$ and for each a_1, a_2 we have that $\delta(v_j, a_1, a_2)$ is (1) v_s if $a_1 > a_2$, (2) v_0^i if $a_1 < a_2$ and (3) v_{j+1}^i if $a_1 = a_2$. Finally, $A^1_{v_s} = A^2_{v_s} = \{a\}$ and $\delta(v_s, a, a) = \mathsf{U}(v_1^1, v_1^2)$. There is an illustration of the Purgatory Duel with $m = n = 2$ in Figure 1a.

The game Purgatory

We will also use the game *Purgatory* as defined by [12] (and also in [13] 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 each $j \in \{1, \dots, n\}$, the state v_j is such that $A^1_{v_j} = A^2_{v_j} = \{1, 2, \dots, m\}$ and for each a_1, a_2 we have that $\delta(v_j, a_1, a_2)$ is (1) v_1 if $a_1 > a_2$, (2) $\bot$ if $a_1 < a_2$ and (3) v_{j+1} if $a_1 = a_2$. The states $\top$ and $\bot$ are absorbing. Furthermore, $S^1 = \{\top\}$. For an illustration of Purgatory with $m = n = 2$ see Figure 1b.

3.1 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 1.** *Each state* $v \in \{v^1_1, v^1_2, \dots, v^1_n, v^2_1, v^2_2, \dots, v^1_2, v_s\}$ *is such that* $\mathrm{val}(v) \in [\frac{1}{m^{n+2}}, 1 - \frac{1}{m^{n+2}}]$

The proof of the above lemma is obtained by considering the strategy, for either player, that plays uniformly at random all available actions at every state. Next we show that every optimal stationary strategy for player 2 must be totally mixed.

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

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 $v^{i'}_j$), then eventually either $\top$ or $\bot$ is reached with probability 1. The proof relies on the analysis of the Markov chain obtained given the strategies.

▶ **Lemma 3.** *For any* i *and* i'*, let* σ_i *be a stationary strategy for player* i*, such that* $\sigma_i(v^{i'}_j)$ *is totally mixed for all* j*. Let* $\sigma_{\widehat{i}}$ *be some stationary strategy for the other player. Then, each closed recurrent set in the Markov chain given by the game,* σ_i*, and* $\sigma_{\widehat{i}}$*, consists of only the state* $\top$ *or only the state* $\bot$*.*

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 5) 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 8).

▶ **Definition 4** (Mirror strategy)**.** Given a stationary strategy σ_i for player i, for either i, let the mirror strategy $\sigma^{\sigma_i}_{\widehat{i}}$ for player $\widehat{i}$ be the stationary strategy where $\sigma^{\sigma_i}_{\widehat{i}}(v^{\widehat{i'}}_j) = \sigma_i(v^{i'}_j)$ 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.

▶ **Lemma 5.** *Let* σ_2 *be some optimal stationary strategy for player 2. Then the mirror strategy* $\sigma^{\sigma_2}_1$ *is optimal for player 1. We have* $\mathrm{val}(v_s) = \frac{1}{2}$ *and* $\mathrm{val}(v^i_j) = 1 - \mathrm{val}(v^{\widehat{i}}_j)$*, for all* i, j*.*

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 6.** *For each* j *in* $\{1, \dots, n\}$*, the value of state* v^1_j *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* $\sigma_i(v^1_j)$ *is at least* $2^{(m-1)^2 m^{n-j-1}}$*.*

3.2 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 7.** *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.*

The idea is that against any strategy σ_2 that does not play totally mixed in some v_j^2, player 1 can ensure that if v_1^2 is entered, then $\bot$ is not reached before v_s is entered again (by playing 1 in $v_{j'}^2$, for $j' < j$ and some action not played by σ_2 in v_j^2). This allows player 1 to play a near optimal strategy from Purgatory in the states $v_{j'}^1$, ensuring that $\top$ is eventually reached with probability close to 1 from v_s and showing that σ_2 is far from optimal. We now show that if we mirror an ε-optimal strategy, then we get an ε-optimal strategy.

▶ **Lemma 8.** *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.*

Next we give a definition and a lemma, which is similar to Lemma 6 in [16]. 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.

▶ **Definition 9.** Let G be an MDP for a safety player. A *replacement set* is a set of triples of states, actions and distributions over the states $Q = \{(s_1, a_1, \delta_1), \dots, (s_\ell, a_\ell, \delta_\ell)\}$. Given the replacement set Q, the MDP $G[Q]$ is an MDP over the same states as G, with the same set of safe states, and where the transition function δ' is similar to δ, except that $\delta'(s, a) = \delta_i$ if $s = s_i$ and $a = a_i$ for some i.

▶ **Lemma 10.** *Let G be an MDP with a safety player. Consider some replacement set $Q = \{(s_1, a_1, \delta_1), \dots, (s_\ell, a_\ell, \delta_\ell)\}$, such that for all t and i we have that $\sum_{s \in S}(\delta(s_i, a_i)(s) \cdot \overline{v}_s^t) \le \sum_{s \in S}(\delta_i(s) \cdot \overline{v}_s^t)$. 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 $\overline{v}_s^t \le \overline{v'}_s^t$. (2) For all states s, we have that $\mathrm{val}(G, s) \le \mathrm{val}(G[Q], s)$.*

The proof of the lemma is in the full version [4]. We next show that for player 1, the patience of ε-optimal strategies is high.

▶ **Lemma 11.** *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 $2^{\Omega(m)}$.*

The proof of the lemma is in the full version

We present the main theorem of this section. The proof follows easily from the previous lemmas (and is presented in details in the full version [4]).

▶ **Theorem 12.** *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 $2^{\Omega(m)}$.*

3.3 The patience lower bound for three states

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 key steps of the proof are as follows: (1) First we consider the Purgatory duel with $n = 1$, and compress it down to 3 states by considering two steps of the Purgatory Duel in a single step. This gives us a game that has three states with one non-absorbing state (which we call 3-state Purgatory Duel) where ε-optimal strategies for the players require exponential patience in m. However, since two steps are simulated by a single step, this game increases the number of actions M from m to m^2. Hence, our patience bound for 3-state Purgatory Duel is only $2^{\Omega(\sqrt{M})}$. (2) We then show that we can restrict the above game to $2m - 1$ of the m^2 actions and still get the same patience as a function of m. We refer to this game as the restricted 3-state Purgatory Duel. Formally, we establish the following result.

▶ **Theorem 13.** *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 $2^{\Omega(m)}$.*

4 Patience Upper Bound

In this section we present the upper bounds. The values of concurrent reachability games can be expressed in the existential theory of reals. Using a refined analysis we present a formula where the number of variables depends only on the number of value classes, rather than the number of states. Using techniques similarly to [13] (such as quantifier elimination, sampling, and root separation for analysis of strategies in games), for concurrent reachability games with K value classes, we show that there is an optimal stationary strategy for the safety player where each probability is a real algebraic number, defined by a polynomial of degree $m^{O(K^2)}$ and the maximum coefficient bit-size is $\tau m^{O(K^2)}$, where τ is the bit-size of numbers in the input. We obtain the following theorem.

▶ **Theorem 14.** *For all concurrent reachability games with at most K different value-classes and probabilities that are rational numbers defined using at most τ bits, the following hold:*

1. *For all $\varepsilon > 0$, there exists an ε-optimal stationary strategy with roundedness at most $\frac{1}{\varepsilon} \lg \frac{1}{\varepsilon} 2^{N\tau m^{O(K^2)}}$.*
2. *For a fixed constant K, a state s and a number λ, given in binary, the problem of deciding whether $\mathrm{val}(s) \geq \lambda$ is in* coNP.

References

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

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

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

4 K. Chatterjee, K. A. Hansen, and R. Ibsen-Jensen. Strategy complexity of concurrent stochastic games with safety and reachability objectives. *CoRR*, abs/1506.02434, 2015.

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

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

7 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.

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

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

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

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

12 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.

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

14 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.

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

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

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

18 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.

19 G. Owen. *Game Theory*. Academic Press, 1995.

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

21 A. Pnueli and R. Rosner. On the synthesis of a reactive module. In *Proc. of POPL*, pages 179–190. ACM Press, 1989.

22 P. J. Ramadge and W. M. Wonham. Supervisory control of a class of discrete-event processes. *SIAM Journal of Control and Optimization*, 25(1):206–230, 1987.

23 L. Shapley. Stochastic games. *PNAS*, 39:1095–1100, 1953.

24 E. Solan and N. Vieille. Computing uniformly optimal strategies in two-player stochastic games. *Economic Theory*, 42(1):237–253, 2010.

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

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

27 O. Vrieze and F. Thuijsman. On equilibria in repeated games with absorbing states. *International Journal of Game Theory*, 18(3):293–310, 1989.

28 O. Vrieze and S. Tijs. Fictitious play applied to sequences of games and discounted stochastic games. *International Journal of Game Theory*, 11(2):71–85, 1982.