Journal of Computer and System Sciences 79 (2013) 640–657

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

# Strategy improvement for concurrent reachability and turn-based stochastic safety games [☆], [☆☆]

Krishnendu Chatterjee [a,*], Luca de Alfaro [b], Thomas A. Henzinger [a]

[a] *IST Austria (Institute of Science and Technology Austria), Austria*
[b] *University of California, Santa Cruz, United States*

ARTICLE INFO

*Article history:*
Received 20 April 2010
Received in revised form 3 January 2012
Accepted 12 December 2012
Available online 19 December 2012

*Keywords:*
Game theory
Stochastic games
Concurrent games
Reachability and safety objectives
Strategy-improvement algorithms

ABSTRACT

We consider concurrent games played on graphs. At every round of a game, each player simultaneously and independently selects a move; the moves jointly determine the transition to a successor state. Two basic objectives are the safety objective to stay forever in a given set of states, and its dual, the reachability objective to reach a given set of states. First, we present a simple proof of the fact that in concurrent reachability games, for all $\varepsilon > 0$, memoryless $\varepsilon$-optimal strategies exist. A memoryless strategy is independent of the history of plays, and an $\varepsilon$-optimal strategy achieves the objective with probability within $\varepsilon$ of the value of the game. In contrast to previous proofs of this fact, our proof is more elementary and more combinatorial. Second, we present a strategy-improvement (a.k.a. policy-iteration) algorithm for concurrent games with reachability objectives. Finally, we present a strategy-improvement algorithm for turn-based stochastic games (where each player selects moves in turns) with safety objectives. Our algorithms yield sequences of player-1 strategies which ensure probabilities of winning that converge monotonically (from below) to the value of the game.

## 1. Introduction

We consider games played between two players on graphs. At every round of the game, each of the two players selects a move; the moves of the players then determine the transition to the successor state. A play of the game gives rise to a path in the graph. We consider the two basic objectives for the players: *reachability* and *safety*. The reachability goal asks player 1 to reach a given set of target states or, if randomization is needed to play the game, to maximize the probability of reaching the target set. The safety goal asks player 2 to ensure that a given set of safe states is never left or, if randomization is required, to minimize the probability of leaving the target set. The two objectives are dual, and the games are determined: the supremum probability with which player 1 can reach the target set is equal to one minus the supremum probability with which player 2 can confine the game to the complement of the target set [14].

These games on graphs can be divided into two classes: *turn-based* and *concurrent*. In turn-based games, only one player has a choice of moves at each state; in concurrent games, at each state both players choose a move, simultaneously and

[☆] This work was partially supported in part by the NSF grants CCR-0132780, CNS-0720884, CCR-0225610, by the Swiss National Science Foundation, ERC Start Grant Graph Games (Project No. 279307), FWF NFN Grant S11407-N23 (RiSE), and a Microsoft faculty fellowship.

[☆☆] This paper is an improved version of Chatterjee et al. (2006, 2009) [4,3]. There is a serious and irreparable error in Theorem 4.3 of Chatterjee et al. (2009) [3] regarding the convergence property of the improvement algorithm for concurrent safety games. This is illustrated in an example of Chatterjee et al. (2012) [2]. In the present version the result is presented for turn-based stochastic safety games. We thank anonymous reviewers for many insightful comments that helped us immensely, and warmly acknowledge their help.

[*] Corresponding author.
*E-mail addresses:* krish.chat@gmail.com, krish.chat@ist.ac.at (K. Chatterjee).

0022-0000/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcss.2012.12.001

independently, from a set of available moves. For turn-based games, the solution of games with reachability and safety objectives has long been known. If each move determines a unique successor state, then the games are P-complete and can be solved in linear time in the size of the game graph. If, more generally, each move determines a probability distribution on possible successor states (called turn-based stochastic games or simple stochastic games), then the problem of deciding whether a turn-based game can be won with probability greater than a given threshold $p \in [0,1]$ is in NP ∩ co-NP [5], and the exact value of the game can be computed by a strategy-improvement algorithm for reachability objectives [6], which works well in practice. These results all depend on the fact that in turn-based reachability and safety games, both players have optimal deterministic (i.e., no randomization is required), memoryless strategies. These strategies are functions from states to moves, so they are finite in number, and this guarantees the termination of the strategy-improvement algorithm for reachability objectives.

The situation is very different for concurrent games. The player-1 *value* of the game is defined, as usual, as the sup–inf value: the supremum, over all strategies of player 1, of the infimum, over all strategies of player 2, of the probability of achieving the reachability or safety goal. In concurrent reachability games, player 1 is guaranteed only the existence of $\varepsilon$-optimal strategies, which ensure that the value of the game is achieved within a specified tolerance $\varepsilon > 0$ [14]. Moreover, while these strategies (which depend on $\varepsilon$) are memoryless, in general they require randomization [14] (even in the special case in which the transition function is deterministic). For player 2 (the safety player), *optimal* memoryless strategies exist [22], which again require randomization (even when the transition function is deterministic). All of these strategies are functions from states to probability distributions on moves. The question of deciding whether a concurrent game can be won with probability greater than $p$ is in PSPACE; this is shown by reduction to the theory of the real-closed fields [13].

To summarize: while strategy-improvement algorithms are available for turn-based stochastic reachability games [6], so far no strategy-improvement algorithms were known for concurrent reachability games. For turn-based stochastic safety games, one could apply the strategy-improvement algorithm for turn-based stochastic reachability games, however there were no strategy-improvement algorithm for turn-based stochastic safety games that converges from below to the value of the game and yields a sequence of improving strategies that converges to an optimal strategy.

**Our results for concurrent reachability games.** Concurrent reachability games belong to the family of stochastic games [24,14], and they have been studied more specifically in [10,9,11]. Our contributions for concurrent reachability games are two-fold. First, we present a simple and combinatorial proof of the existence of memoryless $\varepsilon$-optimal strategies for concurrent games with reachability objectives, for all $\varepsilon > 0$. Second, using the proof techniques we developed for proving existence of memoryless $\varepsilon$-optimal strategies, for $\varepsilon > 0$, we obtain a strategy-improvement (a.k.a. policy-iteration) algorithm for concurrent reachability games. Unlike in the special case of turn-based games the algorithm need not terminate in finitely many iterations.

It has long been known that optimal strategies need not exist for concurrent reachability games, and for all $\varepsilon > 0$, there exist $\varepsilon$-optimal strategies that are memoryless [14]. A proof of this fact can be obtained by considering limit of discounted games. The proof considers *discounted* versions of reachability games, where a play that reaches the target in $k$ steps is assigned a value of $\alpha^k$, for some discount factor $0 < \alpha \leqslant 1$. It is possible to show that, for $0 < \alpha < 1$, memoryless optimal strategies exist. The result for the undiscounted ($\alpha = 1$) case followed from an analysis of the limit behavior of such optimal strategies for $\alpha \to 1$. The limit behavior is studied with the help of results from the field of real Puiseux series [21]. This proof idea works not only for reachability games, but also for total-reward games with nonnegative rewards (see [15, Chapter 5] for details). A more recent result [13] establishes the existence of memoryless $\varepsilon$-optimal strategies for certain infinite-state (recursive) concurrent games, but again the proof relies on results from analysis and properties of solutions of certain polynomial functions. Another proof of existence of memoryless $\varepsilon$-optimal strategies for reachability objectives follows from the result of [14] and the proof uses induction on the number of states of the game. We show the existence of memoryless $\varepsilon$-optimal strategies for concurrent reachability games by more combinatorial and elementary means. Our proof relies only on combinatorial techniques and on simple properties of Markov decision processes [1,8]. As our proof is more combinatorial, we believe that the proof techniques will find future applications in game theory.

Our proof of the existence of memoryless $\varepsilon$-optimal strategies, for all $\varepsilon > 0$, is built upon a value-iteration scheme that converges to the value of the game [11]. The value-iteration scheme computes a sequence $u_0, u_1, u_2, \ldots$ of valuations, where for $i = 0, 1, 2, \ldots$ each valuation $u_i$ associates with each state $s$ of the game a lower bound $u_i(s)$ on the value of the game, such that $\lim_{i\to\infty} u_i(s)$ converges to the value of the game at $s$. The convergence is monotonic from below, but no rate of convergence was known. From each valuation $u_i$, we can extract a memoryless, randomized player-1 strategy, by considering the (randomized) choice of moves for player 1 that achieves the maximal one-step expectation of $u_i$. In general, a strategy $\pi_i$ obtained in this fashion is not guaranteed to achieve the value $u_i$. We show that $\pi_i$ is guaranteed to achieve the value $u_i$ if it is *proper*, that is, if regardless of the strategy adopted by player 2, the play reaches with probability 1 states that are either in the target, or that have no path leading to the target. Next, we show how to extract from the sequence of valuations $u_0, u_1, u_2, \ldots$ a sequence of memoryless randomized player-1 strategies $\pi_0, \pi_1, \pi_2, \ldots$ that are guaranteed to be proper, and thus achieve the values $u_0, u_1, u_2, \ldots$. This proves the existence of memoryless $\varepsilon$-optimal strategies for all $\varepsilon > 0$. Our proof is completely different as compared to the proof of [14]: the proof of [14] uses induction on the number of states, whereas our proof is based on the notion of ranking function obtained from the value-iteration algorithm.

We then apply the techniques developed for the above proof to design a *strategy-improvement* algorithm for concurrent reachability games. Strategy-improvement algorithms, also known as *policy-iteration* algorithms in the context of Markov