Recursive Concurrent Stochastic Games

Kousha Etessami[1] and Mihalis Yannakakis[2]

[1] LFCS, School of Informatics, University of Edinburgh
[2] Department of Computer Science, Columbia University

Abstract. We study Recursive Concurrent Stochastic Games (RCSGs), extending our recent analysis of recursive simple stochastic games [14, 15] to a concurrent setting where the two players choose moves simultaneously and independently at each state. For multi-exit games, our earlier work already showed undecidability for basic questions like termination, thus we focus on the important case of single-exit RCSGs (1-RCSGs).
We first characterize the value of a 1-RCSG termination game as the least fixed point solution of a system of nonlinear minimax functional equations, and use it to show PSPACE decidability for the quantitative termination problem. We then give a strategy improvement technique, which we use to show that player 1 (maximizer) has ϵ-optimal randomized Stackless & Memoryless (r-SM) strategies, while player 2 (minimizer) has optimal r-SM strategies. Thus, such games are r-SM-determined. These results mirror and generalize in a strong sense the randomized memoryless determinacy results for finite stochastic games, and extend the classic Hoffman-Karp [19] strategy improvement approach from the finite to an infinite state setting. The proofs in our infinite-state setting are very different however.
We show that our upper bounds, even for qualitative termination, can not be improved without a major breakthrough, by giving two reductions: first a P-time reduction from the long-standing square-root sum problem to the quantitative termination decision problem for *finite* concurrent stochastic games, and then a P-time reduction from the latter problem to the qualitative termination problem for 1-RCSGs.

1 Introduction

In recent work we have studied Recursive Markov Decision Processes (RMDPs) and turn-based Recursive Simple Stochastic Games (RSSGs) ([14, 15]), providing a number of strong upper and lower bounds for their analysis. These define infinite-state (perfect information) stochastic games that extend Recursive Markov Chains (RMCs) ([12, 13]) with nonprobabilistic actions controlled by players. Here we extend our study to Recursive Concurrent Stochastic Games (RCSGs), where the two players choose moves simultaneously and independently at each state, unlike RSSGs where only one player can move at each state. RCSGs define a class of infinite-state zero-sum (imperfect information) stochastic games that can naturally model probabilistic procedural programs and other systems involving both recursive and probabilistic behavior, as well as concurrent interactions between the system and the environment. Informally, all such recursive models consist of a finite collection of finite state component models (of the same type) that can call each other in a potentially recursive manner. For multi-exit RMDPs and RSSGs, our earlier work already showed that basic questions such as qualitative (i.e. almost sure) termination are already undecidable, whereas we gave strong upper bounds for the important special case of *single-exit* RMDPs and RSSGs (called 1-RMDPs and 1-RSSGs).

Our focus is thus on single-exit RCSGs (1-RCSGs). These models correspond to a concurrent game version of multi-type Branching Processes and Stochastic Context-Free Grammars, both of which are important and extensively studied stochastic processes with many applications including in population genetics, nuclear chain reactions, computational biology, and natural language processing (see, e.g., [18, 20] and other references in [12, 14]). It is very natural to consider game extensions to these stochastic models. Branching processes model the growth of a population of entities of distinct types. In each generation each entity of a given type gives rise, according to a probability distribution, to a multi-set of entities of distinct types. A branching process can be mapped to a 1-RMC such that the probability of eventual extinction of a species is equal to the probability of termination in the 1-RMC. Modeling the process in a context where external agents can influence the evolution to bias it towards extinction or towards survival leads naturally to a game. A 1-RCSG models the process where the evolution of some types is affected by the concurrent actions of external favorable and unfavorable agents (forces).

In [14], we showed that for the 1-RSSG termination game, where the goal of player 1 (2) is to maximize (minimize) the probability of termination starting at a given vertex (in the empty calling context), we can decide in PSPACE whether the value of the game is $\geq p$ for a given probability p, and we can approximate this value (which can be irrational) to within given precision with the same complexity. We also showed that both players have optimal *deterministic Stackless and Memoryless* (SM) strategies in the 1-RSSG termination game; these are strategies that depend neither on the history of the game nor on the call stack at the current state. Thus from each vertex belonging to the player, such a strategy deterministically picks one of the outgoing transitions.

Already for finite-state concurrent stochastic games (CSGs), even under the simple termination objective, the situation is rather different. Memoryless strategies do suffice for both players, but randomization of strategies is necessary, meaning we can't hope for deterministic ϵ-optimal strategies for either player. Moreover, player 1 (the maximizer) can only attain ϵ-optimal strategies, for $\epsilon > 0$, whereas player 2 (the minimizer) does have optimal randomized memoryless strategies (see, e.g., [16, 10]). Another important result for finite CSGs is the classic Hoffman-Karp [19] strategy improvement method, which provides, via simple local improvements, a sequence of randomized memoryless strategies which yield payoffs that converge to the value of the game.

Here we generalize all these results to the infinite-state setting of 1-RCSG termination games. We first characterize values of the 1-RCSG termination game as the least fixed point solution of a system of nonlinear minimax functional equations. We use this to show PSPACE decidability for the *quantitative termination problem* (is the value of the game $\geq r$ for given rational r), as well as PSPACE algorithms for approximating the termination probabilities of 1-RCSGs to within a given number of bits of precision, via results for the existential theory of reals.

We then proceed to our technically most involved result, a strategy improvement technique for 1-RCSG termination games. We use this to show that in these games player 1 (maximizer) has ϵ-optimal randomized-Stackless & Memoryless (r-SM for short) strategies, whereas player 2 (minimizer) has optimal r-SM strategies. Thus, such games are r-SM-determined. These results mirror and generalize in a very strong sense the randomized memoryless determinacy results known for finite stochastic games. Our technique extends Hoffman-Karp's strategy improvement method for finite CSGs to an infinite state setting. However, the proofs in our infinite-state setting are very different. We rely on subtle analytic properties of certain power series that arise from studying 1-RCSGs.

Note that our PSPACE upper bounds for the quantitative termination problem for 1-RCSGs can not be improved to NP without a major breakthrough, since already for 1-RMCs we showed in [12] that the quantitative termination problem is at least as hard as the square-root sum problem (see [12]). In fact, here we show that even the qualitative termination problem for 1-RCSGs, where the problem is to decide whether the value of the game is exactly 1, is already as hard as the square-root sum problem, and moreover, so is the quantitative termination decision problem for *finite* CSGs. We do this via two reductions: we give a P-time reduction from the square-root sum problem to the quantitative termination decision problem for *finite* CSGs, and a P-time reduction from the quantitative finite CSG termination problem to the qualitative 1-RCSG termination problem. Note that this is despite the fact that in recent work Chatterjee et. al. ([6]) have shown that the *approximate* quantitative problems for finite CSGs, including for termination and for more general parity winning conditions, are in NP∩coNP. In other words, we show that quantitative decision problems for finite CSGs will require surmounting significant new difficulties that don't arise for approximation of game values.

We note that, as is known already for finite concurrent games ([5]), probabilistic nodes do not add any power to these games, because the stochastic nature of all the games we consider can in fact be simulated by concurrency alone. The same is true for 1-RCSGs. Specifically, given a finite CSG (or 1-RCSG), G, there is a P-time reduction to a finite concurrent game (or 1-RCG, respectively) $F(G)$, without any probabilistic vertices, such that the value of the game G is exactly the same as the value of the game $F(G)$.

Related work. Stochastic games go back to Shapley [24], who considered finite concurrent stochastic games with (discounted) rewards. See, e.g., [16] for a recent book on stochastic games. Turn-based "simple" finite stochastic games were studied by Condon [8]. As mentioned, we studied RMDPs and (turn-based) RSSGs and their quantitative and qualitative termination problems in [14, 15]. In [15] we showed that the qualitative termination problem for finite 1-RMDPs is in P, and for 1-RSSGs is in NP∩coNP. Our earlier work [12, 13] developed theory and algorithms for Recursive Markov Chains (RMCs), and [11, 3] have studied probabilistic Pushdown Systems which are essentially equivalent to RMCs.

Finite-state concurrent stochastic games have been studied extensively in recent CS literature (see, e.g., [6, 10, 9]). In particular, [6] have shown that for finite CSGs the *approximate* reachability problem and *approximate* parity game problem are in NP∩coNP; however, their results do not resolve the decision problem, which asks whether the value of the game is $\geq r$. (Their approximation theorem (Thm 3.3, part 1.) in its current form is slightly misstated in a way that would actually imply that the decision problem is also in NP∩coNP, but this will be corrected in a journal version of their paper ([5]).) Indeed, we show here that the quantitative decision problem for finite CSGs, as well as the qualitative problem for 1-RCSGs, are as hard as the square-root sum problem, for which containment even in NP is a long standing open problem. Thus our upper bound here, even for the qualitative termination problem for 1-RCSGs, can not be improved to NP without a major breakthrough. Unlike for 1-RCSGs, the qualitative termination problem for finite CSGs is known to be decidable in P-time ([9]). We note that in recent work Allender et. al. [1] have shown that the square-root sum problem is in (the 4th level of) the "Counting Hierarchy" CH, which is inside PSPACE, but it remains a major open problem to bring this complexity down to NP.

2 Basics

Let Γ_1 and Γ_2 be finite sets constituting the *move alphabet* of players 1 and 2, respectively. A *Recursive Concurrent Stochastic Game (RCSG)* is a tuple $A = (A_1, \ldots, A_k)$, where each *component* $A_i = (N_i, B_i, Y_i, En_i, Ex_i, \mathtt{pl}_i, \delta_i)$ consists of:

1. A set N_i of *nodes*, with a distinguished subset En_i of *entry* nodes and a (disjoint) subset Ex_i of *exit* nodes.

2. A set B_i of *boxes*, and a mapping $Y_i : B_i \mapsto \{1, \ldots, k\}$ that assigns to every box (the index of) a component. To each box $b \in B_i$, we associate a set of *call ports*, $Call_b = \{(b, en) \mid en \in En_{Y(b)}\}$, and a set of *return ports*, $Return_b = \{(b, ex) \mid ex \in Ex_{Y(b)}\}$. Let $Call^i = \cup_{b \in B_i} Call_b$, $Return^i = \cup_{b \in B_i} Return_b$, and let $Q_i = N_i \cup Call^i \cup Return^i$ be the set of all nodes, call ports and return ports; we refer to these as the *vertices* of component A_i.

3. A mapping $\mathtt{pl}_i : Q_i \mapsto \{0, play\}$ that assigns to every vertex u a type describing how the next transition is chosen: if $\mathtt{pl}_i(u) = 0$ it is chosen probabilistically and if $\mathtt{pl}_i(u) = play$ it is determined by moves of the two players. Vertices $u \in (Ex_i \cup Call^i)$ have no outgoing transitions; for them we let $\mathtt{pl}_i(u) = 0$.

4. A transition relation $\delta_i \subseteq (Q_i \times (\mathbb{R} \cup (\Gamma_1 \times \Gamma_2)) \times Q_i)$, where for each tuple $(u, x, v) \in \delta_i$, the source $u \in (N_i \setminus Ex_i) \cup Return^i$, the destination $v \in (N_i \setminus En_i) \cup Call^i$, where if $\mathtt{pl}(u) = 0$ then x is a real number $p_{u,v} \in [0, 1]$ (the transition probability), and if $\mathtt{pl}(u) = play$ then $x = (\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_2$. We assume that each vertex $u \in Q_i$ has associated with it a set $\Gamma_1^u \subseteq \Gamma_1$ and a set $\Gamma_2^u \subseteq \Gamma_2$, which constitute player 1 and 2's *legal moves* at vertex u. Thus, if $(u, x, v) \in \delta_i$ and $x = (\gamma_1, \gamma_2)$ then $(\gamma_1, \gamma_2) \in \Gamma_1^u \times \Gamma_2^u$. Additionally, for each vertex u and each $x \in \Gamma_1^u \times \Gamma_2^u$, we assume there is exactly 1 transition of the form (u, x, v) in δ_i. For

computational purposes we assume that the given probabilities $p_{u,v}$ are rational. Furthermore they must satisfy the consistency property: for every $u \in \mathtt{pl}^{-1}(0)$, $\sum_{\{v' | (u, p_{u,v'}, v') \in \delta_i\}} p_{u,v'} = 1$, unless u is a call port or exit node, neither of which have outgoing transitions, in which case by default $\sum_{v'} p_{u,v'} = 0$.

We use the symbols $(N, B, Q, \delta$, etc.) without a subscript, to denote the union over all components. Thus, eg. $N = \cup_{i=1}^{k} N_i$ is the set of all nodes of A, $\delta = \cup_{i=1}^{k} \delta_i$ the set of all transitions, etc.

An RCSG A defines a global denumerable stochastic game $M_A = (V, \Delta, \mathtt{pl})$ as follows. The global *states* $V \subseteq B^* \times Q$ of M_A are pairs of the form $\langle \beta, u \rangle$, where $\beta \in B^*$ is a (possibly empty) sequence of boxes and $u \in Q$ is a *vertex* of A. More precisely, the states $V \subseteq B^* \times Q$ and transitions Δ are defined inductively as follows: 1. $\langle \epsilon, u \rangle \in V$, for $u \in Q$ (ϵ denotes the empty string.); 2. if $\langle \beta, u \rangle \in V$ & $(u, x, v) \in \delta$, then $\langle \beta, v \rangle \in V$ and $(\langle \beta, u \rangle, x, \langle \beta, v \rangle) \in \Delta$; 3. if $\langle \beta, (b, en) \rangle \in V$, with $(b, en) \in \mathit{Call}_b$, then $\langle \beta b, en \rangle \in V$ & $(\langle \beta, (b, en) \rangle, 1, \langle \beta b, en \rangle) \in \Delta$; 4. if $\langle \beta b, ex \rangle \in V$, & $(b, ex) \in \mathit{Return}_b$, then $\langle \beta, (b, ex) \rangle \in V$ & $(\langle \beta b, ex \rangle, 1, \langle \beta, (b, ex) \rangle) \in \Delta$. Item 1. corresponds to the possible initial states, item 2. corresponds to control staying within a component, item 3. is when a new component is entered via a box, item 4. is when control exits a box and returns to the calling component. The mapping $\mathtt{pl} : V \mapsto \{0, \mathit{play}\}$ is given by $\mathtt{pl}(\langle \beta, u \rangle) = \mathtt{pl}(u)$. The set of vertices V is partitioned into V_0, V_{play}, where $V_0 = \mathtt{pl}^{-1}(0)$ and $V_{play} = \mathtt{pl}^{-1}(\mathit{play})$.

We consider M_A with various *initial states* of the form $\langle \epsilon, u \rangle$, denoting this by M_A^u. Some states of M_A are *terminating states* and have no outgoing transitions. These are states $\langle \epsilon, ex \rangle$, where ex is an exit node. If we wish to view M_A as a non-terminating CSG, we can consider the terminating states as absorbing states of M_A, with a self-loop of probability 1.

An RCSG where $|\Gamma_2| = 1$ (i.e., where player 2 has only one action) is called a maximizing *Recursive Markov Decision Process* (RMDP), likewise, when $|\Gamma_1| = 1$ is a minimizing RMDP. An RSSG where $|\Gamma_1| = |\Gamma_2| = 1$ is essentially a *Recursive Markov Chain* ([12, 13]).

Our goal is to answer termination questions for RCSGs of the form: *"Does player 1 have a strategy to force the game to terminate (i.e., reach node $\langle \epsilon, ex \rangle$), starting at $\langle \epsilon, u \rangle$, with probability $\geq p$, regardless of how player 2 plays?"*.

First, some definitions: a *strategy* σ for player i, $i \in \{1, 2\}$, is a function $\sigma : V^* V_{play} \mapsto \mathcal{D}(\Gamma_i)$, where $\mathcal{D}(\Gamma_i)$ denotes the set of probability distributions on the finite set of moves Γ_i. In other words, given a history $ws \in V^* V_{play}$, and a strategy σ for, say, player 1, $\sigma(ws)(\gamma)$ defines the probability with which player 1 will play move γ. Moreover, we require that the function σ has the property that for any global state $s = \langle \beta, u \rangle$, with $\mathtt{pl}(u) = \mathit{play}$, $\sigma(ws) \in \mathcal{D}(\Gamma_i^u)$. In other words, the distribution has support only over eligible moves at vertex u.

Let Ψ_i denote the set of all strategies for player i. Given a history $ws \in V^* V_{play}$ of play so far, and given a strategy $\sigma \in \Psi_1$ for player 1, and a strategy $\tau \in \Psi_2$ for player 2, the strategies determine a distribution on the next move of play to a new global state, namely, the transition $(s, (\gamma_1, \gamma_2), s') \in \Delta$ has probability $\sigma(ws)(\gamma_1) * \tau(ws)(\gamma_2)$. This way, given a start node u, a strategy $\sigma \in \Psi_1$, and a strategy $\tau \in \Psi_2$, we define a new Markov chain (with initial state

u) $M_A^{u,\sigma,\tau} = (\mathcal{S}, \Delta')$. The states $\mathcal{S} \subseteq \langle \epsilon, u \rangle V^*$ of $M_A^{u,\sigma,\tau}$ are non-empty sequences of states of M_A, which must begin with $\langle \epsilon, u \rangle$. Inductively, if $ws \in \mathcal{S}$, then: (0) if $s \in V_0$ and $(s, p_{s,s'}, s') \in \Delta$ then $wss' \in \mathcal{S}$ and $(ws, p_{s,s'}, wss') \in \Delta'$; (1) if $s \in V_{play}$, where $(s, (\gamma_1, \gamma_2), s') \in \Delta$, then if $\sigma(ws)(\gamma_1) > 0$ and $\tau(ws)(\gamma_2) > 0$ then $wss' \in \mathcal{S}$ and $(ws, p, wss') \in \Delta'$, where $p = \sigma(ws)(\gamma_1) * \tau(ws)(\gamma_2)$.

Given initial vertex u, and final exit ex in the same component, and given strategies $\sigma \in \Psi_1$ and $\tau \in \Psi_2$, for $k \geq 0$, let $q^{k,\sigma,\tau}_{(u,ex)}$ be the probability that, in $M_A^{u,\sigma,\tau}$, starting at initial state $\langle \epsilon, u \rangle$, we will reach a state $w\langle \epsilon, ex \rangle$ in at most k "steps" (i.e., where $|w| \leq k$). Let $q^{*,\sigma,\tau}_{(u,ex)} = \lim_{k\to\infty} q^{k,\sigma,\tau}_{(u,ex)}$ be the probability of ever terminating at ex, i.e., reaching $\langle \epsilon, ex \rangle$. (Note, the limit exists: it is a monotonically non-decreasing sequence bounded by 1). Let $q^k_{(u,ex)} = \sup_{\sigma\in\Psi_1} \inf_{\tau\in\Psi_2} q^{k,\sigma,\tau}_{(u,ex)}$ and let $q^*_{(u,ex)} = \sup_{\sigma\in\Psi_1} \inf_{\tau\in\Psi_2} q^{*,\sigma,\tau}_{(u,ex)}$. For a strategy $\sigma \in \Psi_1$, let $q^{k,\sigma}_{(u,ex)} = \inf_{\tau\in\Psi_2} q^{k,\sigma,\tau}_{(u,ex)}$, and let $q^{*,\sigma}_{(u,ex)} = \inf_{\tau\in\Psi_2} q^{*,\sigma,\tau}_{(u,ex)}$. Lastly, given a strategy $\tau \in \Psi_2$, let $q^{k,\cdot,\tau}_{(u,ex)} = \sup_{\sigma\in\Psi_1} q^{k,\sigma,\tau}_{(u,ex)}$, and let $q^{*,\cdot,\tau}_{(u,ex)} = \sup_{\sigma\in\Psi_1} q^{*,\sigma,\tau}_{(u,ex)}$.

From, general determinacy results (e.g., "Blackwell determinacy" [22] which applies to all Borel two-player zero-sum stochastic games with countable state spaces; see also [21]) it follows that the games M_A are *determined*, meaning: $\sup_{\sigma\in\Psi_1} \inf_{\tau\in\Psi_2} q^{*,\sigma,\tau}_{(u,ex)} = \inf_{\tau\in\Psi_2} \sup_{\sigma\in\Psi_1} q^{*,\sigma,\tau}_{(u,ex)}$.

We call a strategy σ for either player a (randomized) *Stackless and Memoryless* (*r-SM*) strategy if it neither depends on the history of the game, nor on the current call stack. In other words, a r-SM strategy σ for player i is given by a function $\sigma : Q \mapsto \mathcal{D}(\Gamma_i)$, which maps each vertex u of the RCSG to a probability distribution $\sigma(u) \in \mathcal{D}(\Gamma_i^u)$ on the moves available to player i at vertex u.

We are interested in the following computational problems.

(1) The *qualitative* termination problem: Is $q^*_{(u,ex)} = 1$?
(2) The *quantitative* termination (decision) problem: given $r \in [0, 1]$, is $q^*_{(u,ex)} \geq r$? The *approximate* version: approximate $q^*_{(u,ex)}$ to within desired precision.

As mentioned, for multi-exit RCSGs these are all undecidable. Thus we focus on *single-exit* RCSGs (*1-RCSGs*), where every component has one exit. Since for 1-RCSGs it is always clear which exit we wish to terminate at starting at vertex u (there is only one exit in u's component), we abbreviate $q^*_{(u,ex)}, q^{*,\sigma}_{(u,ex)}$, etc., as $q^*_u, q^{*,\sigma}_u$, etc., and we likewise abbreviate other subscripts.

3 Nonlinear minimax equations for 1-RCSGs

In ([14]) we defined a monotone system S_A of nonlinear min-& -max equations for 1-RSSGs, and showed that its *least fixed point* solution yields the desired probabilities q^*_u. Here we generalize these to nonlinear minimax systems for 1-RCSGs. Let us use a variable x_u for each unknown q^*_u, and let x be the vector of all x_u , $u \in Q$. The system S_A has one equation of the form $x_u = P_u(\mathbf{x})$ for each vertex u. Suppose that u is in component A_i with (unique) exit ex. There are 4 cases based on the "*Type*" of u.

1. $u \in Type_1$: $u = ex$. In this case: $x_u = 1$.

2. $u \in Type_{rand}$: $\texttt{pl}(u) = 0$ & $u \in (N_i \backslash \{ex\}) \cup Return^i$: $x_u = \sum_{\{v | (u,p_{u,v},v) \in \delta\}} p_{u,v} x_v$. (If u has no outgoing transitions, this equation is by definition $x_u = 0$.)
3. $u \in Type_{call}$: $u = (b, en)$ is a call port: $x_{(b,en)} = x_{en} \cdot x_{(b,ex')}$, where $ex' \in Ex_{Y(b)}$ is the unique exit of $A_{Y(b)}$.
4. $u \in Type_{play}$: $x_u = \mathrm{Val}(A_u(x))$.
We have to define this case. Given a value vector x, and a play vertex u, consider the zero-sum matrix game given by matrix $A_u(x)$, whose rows are indexed by player 1's moves Γ_1^u from node u, and whose columns are indexed by player 2's moves Γ_2^u. The payoff to player 1 under the pair of deterministic moves $\gamma_1 \in \Gamma_1^u$, and $\gamma_2 \in \Gamma_2^u$, is given by $(A_u(x))_{\gamma_1,\gamma_2} := x_v$, where $(u, (\gamma_1, \gamma_2), v) \in \delta$. Let $\mathrm{Val}(A_u(x))$ be the value of this zero-sum matrix game. By von Neumann's minmax theorem, the value and optimal mixed strategies exist, and they can be obtained by solving a set of linear inequality constraints with coefficients given by the x_i's.

In vector notation, we denote the system S_A by $x = P(x)$. Given 1-exit RCSG A, we can easily construct this system. Note that the operator $P : \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}^n_{\geq 0}$ is *monotone*: for $x, y \in \mathbb{R}^n_{\geq 0}$, if $x \leq y$ then $P(x) \leq P(y)$. This follows because for two game matrices A and B of the same dimensions, if $A \leq B$ (i.e., $A_{i,j} \leq B_{i,j}$ for all i and j), then $\mathrm{Val}(A) \leq \mathrm{Val}(B)$. Note that by definition of $A_u(x)$, for $x \leq y$, $A_u(x) \leq A_u(y)$. We now identify a particular solution to $x = P(x)$, called the *Least Fixed Point* (LFP) solution, which gives precisely the termination game values. Define $P^1(x) = P(x)$, and define $P^k(x) = P(P^{k-1}(x))$, for $k > 1$. Let $q^* \in \mathbb{R}^n$ denote the n-vector $q^*_u, u \in Q$ (using the same indexing as used for x). For $k \geq 0$, let q^k denote, similarly, the n-vector $q^k_u, u \in Q$.

Theorem 1. *Let $x = P(x)$ be the system S_A associated with 1-RCSG A. Then $q^* = P(q^*)$, and for all $q' \in \mathbb{R}^n_{\geq 0}$, if $q' = P(q')$, then $q^* \leq q'$ (i.e., q^* is the* Least Fixed Point*, of $P : \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}^n_{\geq 0}$). Moreover, $\lim_{k\to\infty} P^k(\mathbf{0}) \uparrow q^*$, i.e., the "value iteration" sequence $P^k(\mathbf{0})$ converges monotonically to the LFP, q^*.*

The proof is omitted due to space constraints. We will need an important fact established in the proof: suppose for some $q' \in \mathbb{R}^n_{\geq 0}$, $q' = P(q')$. Let τ' be the r-SM strategy for player 2 that always picks, at any state $\langle \beta, u \rangle$, for vertex $u \in \texttt{pl}^{-1}(play)$, the mixed 1-step strategy which is an optimal minimax strategy in the matrix game $A_u(q')$. Then $q^{*,\cdot,\tau'} \leq q'$. In other words, τ' achieves a value $\leq q'_u$ for the game starting from every vertex u (in the empty context).

Theorem 2. *Given a 1-exit RCSG A and a rational probability p, there is a PSPACE algorithm to decide whether $\mathbf{q}^*_u \leq p$. The running time is $O(|A|^{O(n)})$ where n is the number of variables in $\mathbf{x} = P(\mathbf{x})$. We can also approximate $\mathbf{q}^*$ to within a given number of bits i of precision (i given in unary), in PSPACE and in time $O(i|A|^{O(n)})$.*

Proof. Using the system $x = P(x)$, we can express facts such as $q^*_u \leq c$ as

$$\exists x_1, \ldots, x_n \bigwedge_{i=1}^{n} (x_i = P_i(x_1, \ldots, x_n)) \wedge x_u \leq c$$

We only need to show how to express equations of the form $x_v = \mathrm{Val}(A_v(\mathbf{x}))$ in the existential theory of reals. We can then appeal to well known results for deciding that theory ([4, 23]). But this is a standard fact in game theory (see, e.g., [2, 16, 10] where it is used for finite CSGs). Namely, the minimax theorem and its LP encoding allow the predicate "$y = \mathrm{Val}(A_v(\mathbf{x}))$" to be expressed as an existential formula $\varphi(y, x)$ in the theory of reals with free variables y and $x_1, \ldots, x_n$, such that for every $x \in \mathbb{R}^n$, there exists a unique y (the game value) satisfying $\varphi(y, \mathbf{x})$. To approximate the game values within given precision we can do binary search using such queries. □

4 Strategy improvement & randomized-SM-determinacy

The proof of Theorem 1 implies the following (see discussion after Thm 1):

Corollary 1. *In every 1-RCSG termination game, player 2 (the minimizer) has an optimal r-SM strategy.*

Proof. Consider the strategy τ' in the discussion after Theorem 1, chosen not for just any fixed point $\mathbf{q}'$, but for $\mathbf{q}^*$ itself. That strategy is r-SM. □

Player 1 does not have optimal r-SM strategies, not even in finite concurrent stochastic games (see, e.g., [16, 10]). We next establish that it does have finite r-SM *ϵ-optimal strategies*, meaning that it has, for every $\epsilon > 0$, a r-SM strategy that guarantees a value of at least $\mathbf{q}^*_u - \epsilon$, starting from every vertex u in the termination game. We say that a game is *r-SM-determined* if, letting Ψ'_1 and Ψ'_2 denote the set of r-SM strategies for players 1 and 2, respectively, we have $\sup_{\sigma \in \Psi'_1} \inf_{\tau \in \Psi'_2} q_u^{*,\sigma,\tau} = \inf_{\tau \in \Psi'_2} \sup_{\sigma \in \Psi'_1} q_u^{*,\sigma,\tau}$.

Theorem 3.

1. *(Strategy Improvement) Starting at any r-SM strategy σ_0 for player 1, via local strategy improvement steps at individual vertices, we can derive a series of r-SM strategies $\sigma_0, \sigma_1, \sigma_2, \ldots$, such that for all $\epsilon > 0$, there exists $i \geq 0$ such that for all $j \geq i$, σ_j is an ϵ-optimal strategy for player 1 starting at any vertex, i.e., $q_u^{*,\sigma_j} \geq q_u^* - \epsilon$ for all vertices u.*
   Each strategy improvement step involves solving the quantitative termination problem for a corresponding 1-RMDP. Thus, for classes where this problem is known to be in P-time (such as linearly-recursive 1-RMDPs, [14]), strategy improvement steps can be carried out in polynomial time.
2. *Player 1 has ϵ-optimal r-SM strategies, for all $\epsilon > 0$, in 1-RCSG termination games.*
3. *1-RCSG termination games are r-SM-determined.*

Proof. Note that (2.) follows immediately from (1.), and (3.) follows because by Corollary 1, player 2 has an optimal r-SM strategy and thus
$\sup_{\sigma \in \Psi'_1} \inf_{\tau \in \Psi'_2} q_u^{*,\sigma,\tau} = \inf_{\tau \in \Psi'_2} \sup_{\sigma \in \Psi'_1} q_u^{*,\sigma,\tau}$.

Let σ be any r-SM strategy for player 1. Consider $q^{*,\sigma}$. First, let us note that if $q^{*,\sigma} = P(q^{*,\sigma})$ then $q^{*,\sigma} = q^*$. This is so because, by Theorem 1, $q^* \leq q^{*,\sigma}$, and on the other hand, σ is just one strategy for player 1, and for every vertex u, $q^*_u = \sup_{\sigma' \in \Psi_1} \inf_{\tau \in \Psi_2} q^{*,\sigma',\tau}_u \geq \inf_{\tau \in \Psi_2} q^{*,\sigma,\tau}_u = q^{*,\sigma}_u$.

Next we claim that, for all vertices $u \notin Type_{play}$, $q^{*,\sigma}_u$ satisfies its equation in $x = P(x)$. In other words, $q^{*,\sigma}_u = P_u(q^{*,\sigma})$. To see this, note that for vertices $u \notin Type_{play}$, no choice of either player is involved, thus the equation holds by definition of $q^{*,\sigma}$. Thus, the only equations that may fail are those for $u \in Type_{play}$, of the form $x_u = \mathrm{Val}(A_u(x))$. We need the following (proof omitted).

Lemma 1. *For any r-SM strategy σ for player 1, and for any $u \in Type_{play}$, $q^{*,\sigma}_u \leq Val(A_u(q^{*,\sigma}))$.*

Now, suppose that for some $u \in Type_{play}$, $q^{*,\sigma}_u \neq Val(A_u(q^{*,\sigma}))$. Thus by the lemma $q^{*,\sigma}_u < Val(A_u(q^{*,\sigma}))$. Consider a revised r-SM strategy for player 1, σ', which is identical to σ, except that locally at vertex u the strategy is changed so that $\sigma'(u) = p^{*,u,\sigma}$, where $p^{*,u,\sigma} \in \mathcal{D}(\Gamma^u_1)$ is an optimal mixed minimax strategy for player 1 in the matrix game $A_u(q^{*,\sigma})$. We will show that switching from σ to σ' will improve player 1's payoff at vertex u, and will not reduce its payoff at any other vertex.

Consider a parameterized 1-RCSG, $A(t)$, which is identical to A, except that u is a randomizing vertex, all edges out of vertex u are removed, and replaced by a single edge labeled by probability variable t to the exit of the same component, and an edge with remaining probability $1 - t$ to a dead vertex. Fixing the value t determines an 1-RCSG, $A(t)$. Note that if we restrict the r-SM strategies σ or σ' to all vertices other than u, then they both define the same r-SM strategy for the 1-RCSG $A(t)$. For each vertex z and strategy τ of player 2, define $q^{*,\sigma,\tau,t}_z$ to be the probability of eventually terminating starting from $\langle \epsilon, z \rangle$ in the Markov chain $M^{z,\sigma,\tau}_{A(t)}$. Let $f_z(t) = \inf_{\tau \in \Psi_2} q^{*,\sigma,\tau,t}_z$. Recall that $\sigma'(u) = p^{*,u,\sigma} \in \mathcal{D}(\Gamma^u_1)$ defines a probability distribution on the actions available to player 1 at vertex u. Thus $p^{*,u,\sigma}(\gamma_1)$ is the probability of action $\gamma_1 \in \Gamma_1$. Let $\gamma_2 \in \Gamma_2$ be any action of player 2 for the 1-step zero-sum game with game matrix $A_u(q^{*,\sigma})$. Let $w(\gamma_1, \gamma_2)$ denote the vertex such that $(u, (\gamma_1, \gamma), w(\gamma_1, \gamma_2)) \in \delta$. Let $h_{\gamma_2}(t) = \sum_{\gamma_1 \in \Gamma_1} p^{*,u,\sigma}(\gamma_1) f_{w(\gamma_1,\gamma_2)}(t)$.

Lemma 2. *Fix the vertex u. Let $\varphi : \mathbb{R} \mapsto \mathbb{R}$ be any function $\varphi \in \{f_z \mid z \in Q\} \cup \{h_\gamma \mid \gamma \in \Gamma^u_2\}$. The following properties hold:*

1. *If $\varphi(t) > t$ at some point $t \geq 0$, then $\varphi(t') > t'$ for all $0 \leq t' < t$.*
2. *If $\varphi(t) < t$ at some point $t \geq 0$, then $\varphi(t') < t'$ for all $1 > t' > t$.*

Proof. First, we prove this for $\varphi = f_z$, for some vertex z.

Note that, once player 1 picks a r-SM strategy, a 1-RCSG becomes a 1-RMDP. By a result of [14], player 2 has an optimal deterministic SM response strategy. Furthermore, there is such a strategy that is optimal regardless of the starting vertex. Thus, for any value of t, player 2 has an optimal deterministic SM strategy τ_t, such that for any start vertex z, we have $\tau_t = \arg\min_{\tau \in \Psi_2} q^{*,\sigma,\tau,t}_z$. Let $g_{(z,\tau)}(t) = q^{*,\sigma,\tau,t}_z$, and let $d\Psi_2$ be the (finite) set of deterministic SM strategies of player 2. Then $f_z(t) = \min_{\tau \in d\Psi_2} g_{z,\tau}(t)$. Now, note that the function $g_{z,\tau}(t)$

is the probability of reaching an exit in an RMC starting from a particular vertex. Thus, by [12], $g_{z,\tau}(t) = (\lim_{k\to\infty} R^k(\mathbf{0}))_z$ for a polynomial system $\mathbf{x} = R(\mathbf{x})$ with non-negative coefficients, but with the additional feature that the variable t appears as one of the coefficients. Since this limit can be described by a power series in the variable t with non-negative coefficients, $g_{z,\tau}(t)$ has the following properties: it is a continuous, differentiable, and nondecreasing function of $t \in [0,1]$, with continuous and nondecreasing derivative, $g'_{z,\tau}(t)$, and since the limit defines probabilities we also know that for $t \in [0,1]$, $g_{z,\tau}(t) \in [0,1]$. Thus $g_{z,\tau}(0) \geq 0$ and $g_{z,\tau}(1) \leq 1$.

Hence, since $g'_{z,\tau}(t)$ is non-decreasing, if for some $t \in [0,1]$, $g_{z,\tau}(t) > t$, then for all $t' < t$, $g_{z,\tau}(t') > t'$. To see this, note that if $g_{z,\tau}(t) > t$ and $g'_{z,\tau}(t) \geq 1$, then for all $t'' > t$, $g_{z,\tau}(t'') > t''$, which contradicts the fact that $g_{z,\tau}(1) = 1$. Thus $g'_{z,\tau}(t') < 1$ for all $t' \leq t$, and since $g_{z,\tau}(t) > t$, we also have $g_{z,\tau}(t') > t'$ for all $t' < t$. Similarly, if $g_{z,\tau}(t) < t$ for some t, then $g_{z,\tau}(t'') < t''$ for all $t'' \in [t,1)$. To see this, note that if for some $t'' > t$, $t'' < 1$, $g_{z,\tau}(t'') = t''$, then since $g'_{z,\tau}$ is non-decreasing and $g_{z,\tau}(t) < t$, it must be the case that $g'_{z,\tau}(t'') > 1$. But then $g_{z,\tau}(1) > 1$, which is a contradiction.

It follows that $f_z(t)$ has the same properties, namely: if $f_z(t) > t$ at some point $t \in [0,1]$ then $g_{z,\tau}(t) > t$ for all τ, and hence for all $t' < t$ and for all $\tau \in d\Psi_2$, $g_{z,\tau}(t') > t'$, and thus $f_z(t') > t'$ for all $t' \in [0,t]$. On the other hand, if $f_z(t) < t$ at $t \in [0,1]$, then there must be some $\tau' \in d\Psi_2$ such that $g_{z,\tau'}(t) < t$. Hence $g_{z,\tau'}(t'') < t''$, for all $t'' \in [t,1)$, and hence $f_z(t'') < t''$ for all $t'' \in [t,1)$.

Next we prove the lemma for every $\varphi = h_\gamma$, where $\gamma \in \Gamma_2^u$. For every value of t, there is one SM strategy τ_t of player 2 (depending only on t) that minimizes simultaneously $g_{z,\tau}(t)$ for all nodes z. So $h_\gamma(t) = \min_\tau r_{\gamma,\tau}(t)$, where $r_{\gamma,\tau}(t) = \sum_{\gamma_1 \in \Gamma_1} p^{*,u,\sigma}(\gamma_1) g_{w(\gamma_1,\gamma),\tau}(t)$ is a convex combination (i.e., a "weighted average")of some g functions at the same point t. The function $r_{\gamma,\tau}$ (for any subscript) inherits the same properties as the g's: continuous, differentiable, nondecreasing, with continuous nondecreasing derivatives, and $r_{\gamma,\tau}$ takes value between 0 and 1. As we argued for the g functions, in the same way it follows that $r_{\gamma,\tau}$ has properties 1 and 2. Also, as we argued for f's based on the g's, it follows that h's also have the same properties, based on the r's. □

Let $t_1 = q_u^{*,\sigma}$, and let $t_2 = \mathrm{Val}(A_u(q^{*,\sigma}))$. By assumption $t_2 > t_1$. Observe that $f_z(t_1) = q_z^{*,\sigma}$ for every vertex z. Thus, $h_{\gamma_2}(t_1) = \sum_{\gamma_1 \in \Gamma_1} p^{*,u,\sigma}(\gamma_1) f_{w(\gamma_1,\gamma_2)}(t_1) = \sum_{\gamma_1} p^{*,u,\sigma}(\gamma_1) q^{*,\sigma}_{w(\gamma_1,\gamma_2)}$. But since, by definition, $p^{*,u,\sigma}$ is an optimal strategy for player 1 in the matrix game $A_u(q^{*,\sigma})$, it must be the case that for every $\gamma_2 \in \Gamma_2^u$, $h_{\gamma_2}(t_1) \geq t_2$, for otherwise player 2 could play a strategy against $p^{*,u,\sigma}$ which would force a payoff lower than the value of the game. Thus $h_{\gamma_2}(t_1) \geq t_2 > t_1$, for all γ_2. This implies that $h_{\gamma_2}(t) > t$ for all $t < t_1$ by Lemma 2, and for all $t_1 \leq t < t_2$, because h_{γ_2} is nondecreasing. Thus, $h_{\gamma_2}(t) > t$ for all $t < t_2$.

Let $t_3 = q_u^{*,\sigma'}$. Let τ' be an optimal global strategy for player 2 against σ'; by [14], we may assume τ' is a pure SM strategy. Let γ' be player 2's action in τ' at node u. Then the value of any node z under the pair of strategies σ' and τ' is $f_z(t_3)$, and thus since $h_{\gamma'}(t_3)$ is a weighted average of $f_z(t_3)$'s for some set of z's, we have $h_{\gamma'}(t_3) = t_3$. Thus, by the previous paragraph, it must be that

$t_3 \geq t_2$, and we know $t_2 > t_1$. Thus, $t_3 = q_u^{*,\sigma'} \geq \mathrm{Val}(A_u(q^{*,\sigma})) > t_1 = q_u^{*,\sigma}$. We have shown:

Lemma 3. $q_u^{*,\sigma'} \geq \mathit{Val}(A_u(q^{*,\sigma})) > q_u^{*,\sigma}$.

Note that since $t_3 > t_1$, and f_z is non-decreasing, we have $f_z(t_3) \geq f_z(t_1)$ for all vertices z. But then $q_z^{*,\sigma'} = f_z(t_3) \geq f_z(t_1) = q_z^{*,\sigma}$ for all z. Thus, $q^{*,\sigma'} \geq q^{*,\sigma}$, with strict inequality at u, i.e., $q_u^{*,\sigma'} > q_u^{*,\sigma}$. Thus, we have established that such a "strategy improvement" step does yield a strictly better payoff for player 1.

Suppose we conduct this "strategy improvement" step repeatedly, starting at an arbitrary initial r-SM strategy σ_0, as long as we can. This leads to a (possibly infinite) sequence of r-SM strategies $\sigma_0, \sigma_1, \sigma_2, \ldots$. Suppose moreover, that during these improvement steps we always "prioritize" among vertices at which to improve so that, among all those vertices $u \in Type_{play}$ which can be improved, i.e., such that $q_u^{*,\sigma_i} < \mathrm{Val}(A_u(q^{*,\sigma_i}))$, we choose the vertex which has not been improved for the longest number of steps (or one that has never been improved yet). This insures that, infinitely often, at every vertex at which the local strategy can be improved, it eventually is improved.

Under this strategy improvement regime, we show that $\lim_{i\to\infty} q^{*,\sigma_i} = q^*$, and thus, for all $\epsilon > 0$, there exists a sufficiently large $i \geq 0$ such that σ_i is an ϵ-optimal r-SM strategy for player 1. Note that after every strategy improvement step, i, which improves at a vertex u, by Lemma 3 we will have $q_u^{*,\sigma_{i+1}} \geq \mathrm{Val}(A_u(q^{*,\sigma_i}))$. Since our prioritization assures that every vertex that can be improved at any step i will be improved eventually, for all $i \geq 0$ there exists $k \geq 0$ such that $q^{*,\sigma_i} \leq P(q^{*,\sigma_i}) \leq q^{*,\sigma_{i+k}}$. In fact, there is a uniform bound on k, namely $k \leq |Q|$, the number of vertices. This "sandwiching" property allows us to conclude that, in the limit, this sequence reaches a fixed point of $x = P(x)$. Note that since $q^{*,\sigma_i} \leq q^{*,\sigma_{i+1}}$ for all i, and since $q^{*,\sigma_i} \leq q^*$, we know that the limit $\lim_{i\to\infty} q^{*,\sigma_i}$ exists. Letting this limit be q', we have $q' \leq q^*$. Finally, we have $q' = P(q')$, because letting i go to infinity in all three parts of the "sandwiching" inequalities above, we get $q' \leq \lim_{i\to\infty} P(q^{*,\sigma_i}) \leq q'$. But note that $\lim_{i\to\infty} P(q^{*,\sigma_i}) = P(q')$, because the mapping $P(x)$ is continuous on $\mathbb{R}^n_{\geq 0}$. Thus q' is a fixed point of $x = P(x)$, and $q' \leq q^*$. But since q^* is the *least* fixed point of $x = P(x)$, we have $q' = q^*$. □

Finally, we give the following two reductions (proofs omitted due to space). Recall that the *square-root sum problem* (see, e.g., [17, 12]) is the following: given $(d_1, \ldots, d_n) \in \mathbb{N}^n$ and $k \in \mathbb{N}$, decide whether $\sum_{i=1}^{n} \sqrt{d_i} \geq k$.

Theorem 4. *There is a P-time reduction from the square-root sum problem to the quantitative termination (decision) problem for finite CSGs.*

Theorem 5. *There is a P-time reduction from the quantitative termination (decision) problem for finite CSGs to the qualitative termination problem for 1-RCSGs.*

Acknowledgement We thank Krishnendu Chatterjee for clarifying for us several results about finite CSGs obtained by himself and others. This work was partially supported by NSF grant CCF-04-30946.

References

1. E. Allender, P. Bürgisser, J. Kjeldgaard-Pedersen, and P. B. Miltersen. On the complexity of numerical analysis. In *21st IEEE Computational Complexity Conference*, 2006.
2. T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Math. Oper. Res.*, 1(3):197–208, 1976.
3. T. Brázdil, A. Kučera, and O. Stražovský. Decidability of temporal properties of probabilistic pushdown automata. In *Proc. of STACS'05*, 2005.
4. J. Canny. Some algebraic and geometric computations in PSPACE. In *Proc. of 20th ACM STOC*, pages 460–467, 1988.
5. K. Chatterjee, Personal communication.
6. K. Chatterjee, L. de Alfaro, and T. Henzinger. The complexity of quantitative concurrent parity games. In *Proc. of SODA'06*, 2006.
7. K. Chatterjee, R. Majumdar, and M. Jurdzinski. On Nash equilibria in stochastic games. In *CSL'04*, volume LNCS 3210, pages 26–40, 2004.
8. A. Condon. The complexity of stochastic games. *Inf. & Comp.*, 96(2):203–224, 1992.
9. L. de Alfaro, T. A. Henzinger, and O. Kupferman. Concurrent reachability games. In *Proc. of FOCS'98*, pages 564–575, 1998.
10. L. de Alfaro and R. Majumdar. Quantitative solution of omega-regular games. *J. Comput. Syst. Sci.*, 68(2):374–397, 2004.
11. J. Esparza, A. Kučera, and R. Mayr. Model checking probabilistic pushdown automata. In *Proc. of 19th IEEE LICS'04*, 2004.
12. K. Etessami and M. Yannakakis. Recursive markov chains, stochastic grammars, and monotone systems of nonlinear equations. In *Proc. of 22nd STACS'05.* Springer, 2005.
13. K. Etessami and M. Yannakakis. Algorithmic verification of recursive probabilistic state machines. In *Proc. 11th TACAS*, vol. 3440 of LNCS, 2005.
14. K. Etessami and M. Yannakakis. Recursive markov decision processes and recursive stochastic games. In *Proc. of 32nd Int. Coll. on Automata, Languages, and Programming (ICALP'05)*, 2005.
15. K. Etessami and M. Yannakakis. Efficient qualitative analysis of classes of recursive markov decision processes and simple stochastic games. In *Proc. of 23rd STACS'06.* Springer, 2006.
16. J. Filar and K. Vrieze. *Competitive Markov Decision Processes.* Springer, 1997.
17. M. R. Garey, R. L. Graham, and D. S. Johnson. Some NP-complete geometric problems. In *8th ACM STOC*, pages 10–22, 1976.
18. T. E. Harris. *The Theory of Branching Processes.* Springer-Verlag, 1963.
19. A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Sci.*, 12:359–370, 1966.
20. P. Jagers. *Branching Processes with Biological Applications.* Wiley, 1975.
21. A. Maitra and W. Sudderth. Finitely additive stochastic games with Borel measurable payoffs. *Internat. J. Game Theory*, 27(2):257–267, 1998.
22. D. A. Martin. Determinacy of Blackwell games. *J. Symb. Logic*, 63(4):1565–1581, 1998.
23. J. Renegar. On the computational complexity and geometry of the first-order theory of the reals, parts I-III. *J. Symb. Comp.*, 13(3):255–352, 1992.
24. L.S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci.*, 39:1095–1100, 1953.