Concurrent Stochastic Games with Stateful-Discounted and Parity Objectives: Complexity and Algorithms

Ali Asadi
Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Krishnendu Chatterjee
Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Raimundo Saona
Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Jakub Svoboda
Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Abstract

We study two-player zero-sum concurrent stochastic games with finite state and action space played for an infinite number of steps. In every step, the two players simultaneously and independently choose an action. Given the current state and the chosen actions, the next state is obtained according to a stochastic transition function. An objective is a measurable function on plays (or infinite trajectories) of the game, and the value for an objective is the maximal expectation that the player can guarantee against the adversarial player. We consider: (a) stateful-discounted objectives, which are similar to the classic discounted-sum objectives, but states are associated with different discount factors rather than a single discount factor; and (b) parity objectives, which are a canonical representation for ω-regular objectives. For stateful-discounted objectives, given an ordering of the discount factors, the limit value is the limit of the value of the stateful-discounted objectives, as the discount factors approach zero according to the given order.

The computational problem we consider is the approximation of the value within an arbitrary additive error. The above problem is known to be in EXPSPACE for the limit value of stateful-discounted objectives and in PSPACE for parity objectives. The best-known algorithms for both the above problems are at least exponential time, with an exponential dependence on the number of states and actions. Our main results for the value approximation problem for the limit value of stateful-discounted objectives and parity objectives are as follows: (a) we establish TFNP[NP] complexity; and (b) we present algorithms that improve the dependency on the number of actions in the exponent from linear to logarithmic. In particular, if the number of states is constant, our algorithms run in polynomial time.

2012 ACM Subject Classification Theory of computation → Logic and verification

Keywords and phrases Concurrent Stochastic Games, Parity Objectives, Discounted-sum Objectives

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2024.5

Related Version *Full Version*: `https://arxiv.org/abs/2405.02486`

Funding This research was partially supported by ERC CoG 863818 (ForM-SMArt), Austrian Science Fund (FWF) 10.55776/COE12, and French Agence Nationale de la Recherche (ANR) ANR-21-CE40-0020 (CONVERGENCE project).

1 Introduction

In this work, we present improved complexity results and algorithms for the value approximation of concurrent stochastic games with two classic objectives. Below we present the model of concurrent stochastic games, the relevant objectives, the computational problems, previous results, and finally our contributions.

44th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2024).
Editors: Siddharth Barman and Sławomir Lasota; Article No. 5; pp. 5:1–5:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Concurrent stochastic games. Concurrent stochastic games are two-player zero-sum games played on finite-state graphs for an infinite number of steps. These games were introduced in the seminal work of Shapley [26] and are a fundamental model in game theory. In each step, both players simultaneously and independently of the other player choose an action. Given the current state and the chosen actions, the next state is obtained according to a stochastic transition function. An infinite number of such steps results in a play which is an infinite sequence of states and actions. Concurrent stochastic games have been widely studied in the literature from the mathematical perspective [26, 15, 16, 24], and from the algorithmic and computational complexity perspective, including: complexity for reachability objectives [9, 14, 17, 22], algorithms for limit-average objectives [21, 25], complexity for qualitative solutions for omega-regular objectives [8], complexity for quantitative solutions for omega-regular objectives [6, 13], and in the context of temporal logic [1]. In particular, in the analysis of reactive systems, concurrent games provide the appropriate model for reactive systems with components that interact synchronously [1, 11, 12]. Hence, concurrent parity games are relevant for the verification of synchronous reactive systems.

Objectives. An objective is a measurable function that assigns to every play a real-valued reward. The classic discounted-sum objective is as follows: every transition is assigned a reward and the objective assigns to a play the discounted-sum of the rewards. While the classic objective has a single discount factor, the stateful-discounted objective has multiple discount factors. In the stateful-discounted objective, each state is associated with a discount factor, and, in the objective, the discount at a step depends on the current state. We also consider the boolean parity objectives, which are a canonical form to express all ω-regular objectives [27], which provide a robust specification for all properties that arise in verification. For example, all LTL formulas can be converted to deterministic parity automata. In parity objectives, every state is associated with an integer priority, and a play is winning for (or satisfies) the objective if the minimum priority visited infinitely often is even.

Strategies and values. Strategies are recipes that define the choice of actions of the players. They are functions that, given a game history, return a distribution over actions. Given a concurrent stochastic game and an objective, the value of Player 1 at a state is the maximal expectation that the player can guarantee for the objective against all strategies of Player 2. For stateful-discounted objectives, given an ordering of the discount factors, the limit value at a state is the limit of the value function for the discounted objective as the discount factors approach zero in the given order.

Computational problems. Given a concurrent stochastic game, the main computational problems are: (a) the *value-decision* problem, given a state and a threshold α, asks whether the value at the state is at least α; and (b) the *value-approximation* problem, given a state and an error $\varepsilon > 0$, asks to compute an approximation of the value for the state within an additive error of ε. We consider the above problems for the limit value of stateful-discounted objectives and the value for parity objectives.

Motivation. The motivation to study the limit of the stateful-discounted objective is as follows. First, this limit generalizes the classic limit-average objectives. Second, it characterizes the value for the parity objectives in concurrent stochastic games [18, 10], where the order of limit corresponds to the order of importance of priorities in parity objectives. Third, the limit value has been shown to correspond to the value for other objectives such as priority mean-payoff for various subclasses of concurrent stochastic games [19]. The motivation to study the value-approximation problem as opposed to the value-decision is that for concurrent games, even for special classes of objectives such as reachability and safety, values can be irrational, and the decision problem related to exact value is SQRT-SUM hard [14] as explained below. Hence, approximation of values is a natural problem to study from an algorithmic and computational complexity perspective.

Previous results. For a single discount factor, the limit value corresponds to the value of the well-studied mean-payoff or long-run average objectives [24], and, for parity objectives, the computational problems admit a linear reduction to the limit value of stateful-discounted objectives [18, 10]. The value-decision problem for concurrent stochastic games is SQRT-SUM hard [14]: this result holds for reachability objectives, and hence also for parity objectives and the limit value for even a single discount factor. The SQRT-SUM problem is a classic problem in computational geometry, and whether SQRT-SUM belongs to NP has been a long-standing open problem. The complexity upper bounds for the value-approximation problem of concurrent stochastic games are as follows: (a) EXPSPACE for the limit value of stateful-discounted objectives; and (b) PSPACE for parity objectives [5, 6]. The above result for the limit value follows from a reduction to the theory of reals, where the number of discount factors corresponds to the number of quantifier alternation. For the special class of reachability objectives, the complexity upper bound of TFNP[NP] for the value-approximation problem has been established in [17], where TFNP[NP] is the total functional form of the second level of the polynomial hierarchy. The result of [17] has been recently extended to limit-average objectives (which correspond to the limit-value of single discount factor) [4]. To the best of our knowledge, the above complexity upper bounds are the best bounds for limit value of general stateful-discounted objectives and parity objectives. The best-known algorithms for the value-approximation problem are as follows: (a) double exponential time for the limit value of stateful-discounted objectives; (b) exponential time for parity objectives, where the exponent is a product that depends at least linearly on the number of states and actions [6, 5] (see Section 3 for further details). While iterative approaches are desirable, they neither exist for parity objectives nor guarantee efficiency even in special cases. For example, for reachability and safety objectives, iterative approaches like value-iteration or strategy-iteration have a double-exponential lower bound [20]; and, for parity objectives, iterative approaches like strategy iteration are not known as strategies require infinite-memory [8].

Our contributions. In this work, our main contributions are as follows: (a) we establish TFNP[NP] upper bounds for the value-approximation problem for concurrent stochastic games, both for the limit value of stateful-discounted and parity objectives; and (b) we present algorithms which are exponential time and improve the dependency on the number of actions in the exponent from linear to logarithmic. In particular, if the number of states is constant, our algorithms run in polynomial time. The comparison of previous results and our results is summarized in Tables 1 and 2.

Technical contributions. We first present a bound on the roots of multi-variate polynomials with integer coefficients (Section 4.2). Given the bounds on roots of polynomials, we establish new characterizations for the limit and stateful-discounted values (Section 4.3 and Section 4.4), which lead to an approximation of the limit value by the stateful-discounted value when the discount factors are double-exponentially small (Section 4.5). Given this connection, we establish the improved complexities and algorithms for the value-approximation for the limit value of stateful-discounted objectives and parity objectives in Section 5 and Section 6. Proofs omitted due to space restrictions are provided in the Full version.

Table 1 Complexity upper bounds of the value-approximation in concurrent stochastic games for the limit value of stateful-discounted objectives and parity objectives.

	Complexity	
	Previous	**Ours**
Limit	EXPSPACE (Theory of reals)	TFNP[NP] (Theorem 5-Item 1, Theorem 6-Item 1)
Parity	PSPACE [6, 5]	

Table 2 Algorithmic upper bounds of the value-approximation in concurrent stochastic games for the limit value of stateful-discounted objectives and parity objectives, where n is the number of states, m is the number of actions, d is the number of discount factors/parity index, B is the bit-size of numbers in the input, ε is the additive error, and exp is the exponential function.

	Algorithms	
	Previous	**Ours**
Limit	$\exp\left(\mathcal{O}(2^d m^2 n + \log(1/\varepsilon) + \log(\mathrm{B}))\right)$ (Theory of reals)	$\exp\left(\mathcal{O}\left(\begin{matrix} nd\log(m) + \log(\mathrm{B}) \\ + \log(\log(1/\varepsilon))\end{matrix}\right)\right)$ (Theorem 5-Item 2, Theorem 6-Item 2)
Parity	$\exp\left(\mathcal{O}\left(\begin{matrix} mn + d\log(n) + \log(\mathrm{B}) \\ + \log(\log(1/\varepsilon))\end{matrix}\right)\right)$ [6, 5]	

2 Preliminaries

We present standard definitions related to concurrent stochastic games.

Basic Notations. Given a finite set $\mathcal{X}$, a probability distribution over $\mathcal{X}$ is a function $\mu\colon \mathcal{X} \to [0,1]$ such that $\sum_{x\in\mathcal{X}} \mu(x) = 1$. The set of all probability distributions over $\mathcal{X}$ is denoted by $\Delta(\mathcal{X})$. For $\mu \in \Delta(\mathcal{X})$, the support of μ is defined as $\mathrm{supp}(\mu) \coloneqq \{x \mid \mu(x) > 0\}$. For a positive integer k, the set of positive integers smaller than or equal to k is defined as $[k] \coloneqq \{1, \ldots, k\}$. Given a real x, we denote 2^x by $\exp(x)$.

Concurrent stochastic games. A concurrent stochastic game (CSG) is a two-player finite game $G = (\mathcal{S}, \mathcal{A}, \mathcal{B}, \delta)$ consisting of

- the set of states $\mathcal{S}$, of size n;
- the sets of actions for each player $\mathcal{A}$ and $\mathcal{B}$, with at most m actions; and
- the stochastic transition function $\delta\colon \mathcal{S} \times \mathcal{A} \times \mathcal{B} \to \Delta(\mathcal{S})$.

Steps. Given an initial state $s \in \mathcal{S}$, the game proceeds as follows. In each step, both players choose an action simultaneously, $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Based on both actions (a, b) and current state s, the next state is drawn according to the probability distribution $\delta(s, a, b)$.

Histories and plays. At step k of CSGs, each player possesses information in the form of the finite sequence of the states visited and the actions chosen by both players. A k-history $\omega^{(k)} = \langle s_0, a_0, b_0, s_1, a_1, b_1, \ldots, s_k\rangle$ is a finite sequence of states and actions such that, for all steps $0 \le t < k$, we have $s_{t+1} \in \mathrm{supp}(\delta(s_t, a_t, b_t))$. The set of all k-histories is denoted by $\Omega^{(k)}$. Similarly, a play $\omega = \langle s_0, a_0, b_0, s_1, a_1, b_1, \ldots\rangle$ is an infinite sequence of states and actions such that, for all steps $t \ge 0$, we have $s_{t+1} \in \mathrm{supp}(\delta(s_t, a_t, b_t))$. The set of all plays is denoted by Ω. For any state s, the set of all plays starting at s, i.e., $\omega = \langle s_0, a_0, b_0, \ldots\rangle$ where $s_0 = s$, is denoted by Ω_s.

Objectives. An objective is a measurable function that assigns a real number to all plays. Player 1 aims to maximize the expectation of the objective, while Player 2 minimizes it.

- *Parity objective.* Given a priority function $p\colon \mathcal{S} \to \{0, \dots, d\}$ with d as its index, the *parity* objective is an indicator of the even parity condition on minimal priority visited infinitely often in plays. More formally, we define $\mathrm{Parity}_p\colon \Omega \to \{0,1\}$ as

$$\mathrm{Parity}_p(\omega) \coloneqq \begin{cases} 1 & \min\{p(s) \mid \forall i \geq 0\ \exists j \geq i\ s_j = s\} \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

- *Stateful-discounted objective.* Consider d discount factors $\lambda_1, \cdots, \lambda_d \in (0,1]$. Given an assignment function $\chi\colon \mathcal{S} \to [d]$, we define the discount function $\Lambda\colon \mathcal{S} \to \{\lambda_1, \cdots, \lambda_d\}$ as $\Lambda(s) \coloneqq \lambda_{\chi(s)}$ for all states $s \in \mathcal{S}$. Given a reward function $r\colon \mathcal{S} \times \mathcal{A} \times \mathcal{B} \to [0,1]$ that assigns a reward value $r(s,a,b)$ for all (s,a,b), the stateful-discounted objective $\mathrm{Disc}_\Lambda\colon \Omega \to [0,1]$ is defined as, for all $\omega = \langle s_0, a_0, b_0, \cdots \rangle$,

$$\mathrm{Disc}_\Lambda(\omega) \coloneqq \sum_{i \geq 0} \left(r(s_i, a_i, b_i)\Lambda(s_i) \prod_{j<i} 1 - \Lambda(s_j) \right) .$$

Strategies. A strategy is a function that assigns a probability distribution over actions to every finite history and is denoted by $\sigma\colon \bigcup_k \Omega^{(k)} \to \Delta(\mathcal{A})$ for Player 1 (resp. $\tau\colon \bigcup_k \Omega^{(k)} \to \Delta(\mathcal{B})$ for Player 2). Given strategies σ and τ, the game proceeds as follows. At step k, the current history is some $\omega^{(k)} \in \Omega^{(k)}$. Player 1 (resp. Player 2) chooses an action according to the distribution $\sigma\left(\omega^{(k)}\right)$ (resp. $\tau\left(\omega^{(k)}\right)$). The set of all strategies for Player 1 and Player 2 is denoted by Σ and Γ respectively. A *stationary* strategy depends on the past observations only through the current state. A stationary strategy for Player 1 (resp. Player 2) is denoted by $\sigma\colon \mathcal{S} \to \Delta(\mathcal{A})$ (resp. $\tau\colon \mathcal{S} \to \Delta(\mathcal{B})$). The set of all stationary strategies for Player 1 and Player 2 is denoted by Σ^S and Γ^S respectively. A *pure stationary* strategy $\sigma\colon \mathcal{S} \to \mathcal{A}$ (resp. $\tau\colon \mathcal{S} \to \mathcal{B}$) for Player 1 (resp. Player 2) is a stationary strategy that maps to Dirac distributions only. The set of all pure stationary strategies for Player 1 and Player 2 is denoted by Σ^{PS} and Γ^{PS} respectively.

Probability space. An initial state s and a pair of strategies (σ, τ) induce a unique probability over Ω_s, endowed with the sigma-algebra generated by the cylinders corresponding to finite histories. We denote by $\mathbb{P}_s^{\sigma,\tau}$ and $\mathbb{E}_s^{\sigma,\tau}$ the probability and the expectation respectively.

We state the determinacy for CSGs with stateful-discounted and parity objectives.

▶ **Theorem 1** (Parity determinacy [23]). *For all CSGs, states s, and priority functions p,*

$$\sup_{\sigma \in \Sigma} \inf_{\tau \in \Gamma} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Parity}_p] = \inf_{\tau \in \Gamma} \sup_{\sigma \in \Sigma} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Parity}_p] .$$

▶ **Theorem 2** (Stateful-discounted determinacy [26]). *For all CSGs, states s, reward functions, and discount functions Λ, we have*

$$\sup_{\sigma \in \Sigma^S} \inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda] = \inf_{\tau \in \Gamma^S} \sup_{\sigma \in \Sigma^S} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda] .$$

Values. The above determinacy results imply that switching the quantification order of strategies do not make a difference and leads to the unique notion of value. The stateful-discounted value for a state s is defined as $\mathrm{val}_\Lambda(s) \coloneqq \sup_{\sigma \in \Sigma^S} \inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda]$. We define the parity value $\mathrm{val}_p(s)$ for a state s analogously. The limit value for a state s is defined as $\mathrm{val}_\chi(s) \coloneqq \lim_{\lambda_1 \to 0^+} \cdots \lim_{\lambda_d \to 0^+} \mathrm{val}_\Lambda(s)$.

ε-optimal strategies. Given $\varepsilon \geq 0$, a strategy σ for Player 1 is *ε-optimal* for the stateful-discounted objective if, for all states $s \in \mathcal{S}$, we have $\inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda] \geq \mathrm{val}_\Lambda(s) - \varepsilon$. We say the strategy is *optimal* if $\varepsilon = 0$. The notion of ε-optimal strategies for Player 2 is defined analogously. Similarly, we define ε-optimal strategies for the parity objectives.

Approximate value problems. We consider two value problems stated as follows.

> LimitValue. Consider a CSG G, a state s, a reward function r, an assignment function $\chi \colon \mathcal{S} \to [d]$, and an additive error ε. The transition function δ and the reward function r are represented by rational numbers using at most B bits. Compute an approximation v of the limit value at state s such that $|v - \mathrm{val}_\chi(s)| \leq \varepsilon$.

> ParityValue. Consider a CSG G, a state s, a priority function p with index d, and an additive error ε. The transition function δ is represented by rational numbers using at most B bits. Compute an approximation v of the parity value at state s such that $|v - \mathrm{val}_p(s)| \leq \varepsilon$.

3 Overview of Results

We first discuss the previous results in the literature, and then, we show our contributions.

Previous results. We discuss the previous works on computing the approximation of limit and parity values in CSGs. A natural approach for these computational problems is via the theory of reals. We first recall the main computational result of the theory of reals, which is a specialization of [3, Theorem 1].

▶ **Theorem 3** ([3, Theorem 1]). *Consider ℓ variables $x_1, \cdots, x_\ell$ and the set of polynomials $\mathcal{P} = \{P_1, \cdots, P_k\}$, where, for all $i \in [k]$, we have P_i is a polynomial in $x_1, \cdots, x_\ell$ of degree at most D with integer coefficients of bit-size at most* B*. Let $X_1, \cdots, X_d$ be a partition of $x_1, \cdots, x_\ell$ into d subsets such that X_i has size ℓ_i. Let $\Phi = (Q_d X_d) \cdots (Q_1 X_1)\ \phi(P_1, \cdots, P_k)$ be a sentence with d alternating quantifiers $Q_i \in \{\exists, \forall\}$ such that $Q_{i+1} \neq Q_i$, and $\phi(P_1, \cdots, P_k)$ is a quantifier-free formula with atomic formulas of the form $P_i \bowtie 0$ where $\bowtie \in \{<, >, =\}$. Then, there exists an algorithm to decide the truth of Φ in time*

$$k^{\prod_i \mathcal{O}(\ell_i + 1)} \cdot D^{\prod_i \mathcal{O}(\ell_i)} \cdot \mathcal{O}(\mathrm{len}(\phi) \mathrm{B}^2)\,,$$

where $\mathrm{len}(\phi)$ *is the length of the quantifier-free formula* ϕ.

Along with the above algorithmic result, the following complexity result also follows from [3]: if there is constant number of quantifier alternations, then complexity is PSPACE, and in general the complexity is EXPSPACE. We now discuss the algorithms and complexity results from the literature for the limit value of stateful discounted-sum objectives. The basic computational approach is via the theory of reals. For a single discount factor, the reduction to the theory of reals and dealing with its limit (which corresponds to limit-average objectives) was presented in [7]. In the general case (d discount factors), each limit can be considered as the quantification $\exists \varepsilon_{i'}\ \forall \varepsilon_i \leq \varepsilon_{i'}$ in the theory of reals. Thus, concurrent stochastic games with the limit value of stateful-discounted objectives can be reduced to the theory of reals with quantifier alternation. This reduction gives a theory of reals sentence with the following parameters:

$$\ell = \mathcal{O}(m^2 n), \quad k = \mathcal{O}(m^2 n), \quad D = 4, \quad \prod_i (\ell_i + 1) = \mathcal{O}(2^d m^2 n)\,,$$

Applying Theorem 3 to the reduction we obtain the following result.

▶ **Theorem 4** (LimitValue: Previous Result). *For the* LimitValue *problem, the following assertions hold.*

1. *The problem is in* EXPSPACE*; and*
2. *the problem can be solved in time* $\exp\left(\mathcal{O}\left(2^d m^2 n + \log(1/\varepsilon) + \log(\mathrm{B})\right)\right)$.

For parity objectives, the result of [18, 10] reduces CSGs with parity objectives to CSGs with the limit value of stateful-discounted objectives. The reduction is achieved as follows. Consider the formula $R(a_0, a_1, \dots, a_{2n-1})$ from [10], which is a formula with multiple discount factors. Since for stateful-discounted objectives the mapping is contractive, the fixpoints are unique (least and greatest fixpoints coincide). The last sentence of [10, Theorem 4] states that the limit of $R(a_0, a_1, \dots, a_{2n-1})$ corresponds to the value for parity objectives. The Pre operator of the formula corresponds to the Bellman-operator for stateful-discounted objectives, which establish the connection to stateful-discounted games. This connection is made more explicit in the construction provided in [18, Section 2.2]. This linear reduction and the above theorem lead to similar results for parity objectives.

Besides this reduction to the theory of reals, there are two other approaches for the ParityValue problem. First, we can consider the nested fixpoint characterization as provided in [13] and a reduction to the theory of reals with quantifier alternation. However, this does not lead to a better complexity. Second, a different approach is presented in [6, 5, Chapter 8]. This approach has the following components: (a) it enumerates over all possible subsets of actions for every state; (b) for each of the enumeration, it requires a solution of a qualitative value problem (or limit-sure winning) in concurrent stochastic games with parity objectives, and the value-approximation for concurrent stochastic games with reachability objectives. This approach gives PSPACE complexity and the algorithmic complexity is $\exp\left(\mathcal{O}\left(mn + d\log(n) + \log(\log(1/\varepsilon)) + \log(\mathrm{B})\right)\right)$.

Our contributions. Our main results are the following.

▶ **Theorem 5** (LimitValue: Complexity and Algorithm). *For the* LimitValue *problem, the following assertions hold.*

1. *The problem is in* TFNP[NP]*; and*
2. *the problem can be solved in time* $\exp\left(\mathcal{O}\left(nd\log(m) + \log(\mathrm{B}) + \log(\log(1/\varepsilon))\right)\right)$.

▶ **Theorem 6** (ParityValue: Complexity and Algorithm). *For the* ParityValue *problem, the following assertions hold.*

1. *The problem is in* TFNP[NP]*; and*
2. *the problem can be solved in time* $\exp\left(\mathcal{O}\left(nd\log(m) + \log(\mathrm{B}) + \log(\log(1/\varepsilon))\right)\right)$.

4 Mathematical Properties

In this section, we present an approach of the limit value approximation via the stateful-discounted value (Theorem 12). We follow the approach of [2] extending it step by step using similar arguments. We use this technical result to improve complexities and algorithmic bounds of computing ε-approximation of the limit and parity values. This section is organized as follows. In Section 4.1, we present some useful definitions. In Section 4.2, we present a bound on the roots of multi-variate polynomials which is used to establish a connection between the stateful-discounted and limit values. In Sections 4.3 and 4.4, we introduce new characterizations of the stateful-discounted and limit values. In Section 4.5, we show Theorem 12.

4.1 Definitions

We present some basic notations and definitions related to polynomials and matrices.

Basic notations. Given a positive integer k, we define $\mathrm{bit}(k) \coloneqq \lceil \log_2(k+2) \rceil$. For a rational k_1/k_2, we define $\mathrm{bit}(k_1/k_2) \coloneqq \mathrm{bit}(k_1) + \mathrm{bit}(k_2)$. Given a real x, the standard sign function is $\mathrm{sign}(x) = -1$ if $x < 0,\ 0$ if $x = 0,\ 1$ if $x > 0$. Moreover, we use the classic arithmetic with infinity, i.e., $x + \infty = \infty$ and $x - \infty = -\infty$.

Polynomials. A *uni-variate* polynomial P of degree D with integer coefficients of bit-size B is defined as $P(x) \coloneqq \sum_{i=0}^{D} c_i x^i$ where $|c_i| < 2^{\mathrm{B}}$. A k-variate polynomial P in $x_1, \cdots, x_k$ of degree $D_1, \cdots, D_k$ with integer coefficients of bit-size B is defined as

$$P(x_1, \cdots, x_k) \coloneqq \sum_{0 \le i_1 \le D_1} \cdots \sum_{0 \le i_k \le D_k} c_{i_1, \cdots, i_k} \prod_{j=1}^{k} x_j^{i_j} \,,$$

where $|c_{i_1, \cdots, i_k}| < 2^{\mathrm{B}}$. Polynomial P is nonzero if $c_{i_1, \cdots, i_k} \neq 0$ for some $i_1, \cdots, i_k$. We say α is a root of P if $P(\alpha) = 0$. In this work, we only consider real roots.

Matrix notations. Given a square matrix M, we denote the determinant of M by $\det(M)$ and denote the signed sum of all minors of M by $S(M)$. Given two $k \times \ell$ matrices M_1 and M_2, we denote the Hadamard product of M_1 and M_2 by $M_1 \odot M_2$. Given a positive integer k, often implicitly clear from context, we denote by $\mathbf{1}$ (resp. $\mathbf{0}$) the k-dimensional vector with all elements equal to 1 (resp. 0) and denote by Id the $k \times k$ identity matrix.

Matrix games. A matrix M defines a game played between two opponents, where rows (resp. columns) correspond to possible actions for the row- (resp. column-) player, and the entry $(M)_{i,j}$ is the reward the column-player pays the row-player when the pair of actions (i, j) is chosen. The value of a matrix game, denoted $\operatorname{val} M$, is the maximum amount the row-player can guarantee, i.e., the amount they can obtain regardless of the column-player's strategy.

4.2 Bounds on Roots of Polynomials with Integer Coefficients

We present a bound on the roots of multi-variate polynomials P with integer coefficients (Theorem 7). This result shows that there exists a region close to $\mathbf{0}$ within which P does not have a root. We use this technical result to establish a connection between the stateful-discounted value and limit value.

▶ **Lemma 7.** *Consider a nonzero polynomial P in $x_1, \cdots, x_\ell$ of degrees $D_1, \cdots, D_\ell$ with integer coefficients of bit-size* B*. Let $D \coloneqq \max(D_1, \cdots, D_\ell)$ and $\mathrm{B}_1 \coloneqq 4\ell \operatorname{bit}(D) + \mathrm{B} + 1$. Then,*

$$\forall x_1 \in (0, \exp(-\mathrm{B}_1)] \quad \forall x_2 \in \left(0, (x_1)^{D+1}\right] \quad \cdots \quad \forall x_\ell \in \left(0, (x_{\ell-1})^{D+1}\right]$$
$$|P(x_1, \cdots, x_\ell)| \geq \exp(\mathrm{B}_1 - \ell) \cdot (x_\ell)^{D+1} .$$

Proof sketch. We partition P into $P(x) = P_0 + P_1(x_1) + \cdots + P_\ell(x_1, \cdots, x_\ell)$. Consider the smallest i where $P_i \neq 0$. The proof has following key components: (a) By fixing $x_1, \cdots, x_{i-1}$, we obtain a uni-variate polynomial from P_i. By the fact that $x_i \ll x_j$ for all $j < i$, we obtain $|P_i(x_1, \cdots, x_i)| \neq 0$. (b) Due to the constraints over the variables, $x_j \ll x_i$ for all $j > i$, we have $|P_j(x_1, \cdots, x_j)| \ll |P_i(x_1, \cdots, x_i)|$. Thus, we can ignore the effect of $P_j(x_1, \cdots, x_j)$ on $|P(x_1, \cdots, x_\ell)|$ for all $j > i$. Thus, $P(x_1, \cdots, x_\ell) \approx P_i(x_1, \cdots, x_i) \neq 0$. ◀

4.3 Characterization of Stateful-discounted Value

We introduce a new characterization of the stateful-discounted value in CSGs (Theorems 8 and 9), which generalizes results of [2] from a single discount factor to multiple discount factors. In particular, Theorem 9 generalizes Theorem 1 of [2].

Stateful-discounted payoff. Consider a CSG G, a state s, a reward function r, and a discount function Λ. Given a pair of stationary strategies (σ, τ), we define the stateful-discounted payoff as $\nu^{\sigma,\tau}(s) \coloneqq \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda]$. By fixing σ and τ, we obtain a transition function $\delta^{\sigma,\tau}(s, s') \coloneqq \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \sigma(s)(a) \cdot \tau(s)(b) \cdot \delta(s, a, b)(s')$, which is described as a matrix, i.e., $\delta^{\sigma,\tau} \in \mathbb{R}^{n\times n}$. The stage reward function is defined as $r^{\sigma,\tau}(s) \coloneqq \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \sigma(s)(a) \cdot \tau(s)(b) \cdot r(s, a, b)$, which is described as a vector, i.e., $r^{\sigma,\tau} \in \mathbb{R}^n$. Therefore, the Bellman operator defined in [26] can be written as a recursive expression:

$$\nu^{\sigma,\tau} = \Lambda \odot r^{\sigma,\tau} + (\mathbf{1} - \Lambda) \odot (\delta^{\sigma,\tau} \nu^{\sigma,\tau}) .$$

The matrix $\mathrm{Id} - \left((\mathbf{1} - \Lambda)\, \mathbf{1}^\top \right) \odot \delta^{\sigma,\tau}$ is strictly diagonally dominant, and therefore, is invertible. By Cramer's rule, we have

$$\nu^{\sigma,\tau}(s) = \frac{\nabla^s_\Lambda(\sigma, \tau)}{\nabla_\Lambda(\sigma, \tau)} , \tag{1}$$

where $\nabla_\Lambda(\sigma, \tau) \coloneqq \det\left(\mathrm{Id} - \left((\mathbf{1} - \Lambda)\, \mathbf{1}^\top \right) \odot \delta^{\sigma,\tau}\right)$ and $\nabla^s_\Lambda(\sigma, \tau)$ is the determinant of a matrix derived by substituting the s-th column of the matrix $\mathrm{Id} - \left((\mathbf{1} - \Lambda)\mathbf{1}^\top\right) \odot \delta^{\sigma,\tau}$ with $\Lambda \odot r^{\sigma,\tau}$.

Auxiliary matrix game $W^s_\Lambda(z)$. We define a matrix game where the actions of each player are the pure stationary strategies in the stochastic game. The payoff of the game is obtained by the linearization of the quotient in Eq. (1). More formally, for all parameters $z \in \mathbb{R}$, $\widehat{\sigma} \in \Sigma^{PS}$, and $\widehat{\tau} \in \Gamma^{PS}$, we define the payoff of the matrix game as

$$W^s_\Lambda(z)[\widehat{\sigma}, \widehat{\tau}] \coloneqq \nabla^s_\Lambda(\widehat{\sigma}, \widehat{\tau}) - z \cdot \nabla_\Lambda(\widehat{\sigma}, \widehat{\tau}) .$$

The value of $W^s_\Lambda(z)$ is denoted by $\mathrm{val}\,(W^s_\Lambda(z))$.

The following results (Theorem 8 and Theorem 9) connect the stateful-discounted value with the value of the matrix game.

▶ **Lemma 8.** *Consider a CSG G, a state s, a reward function, and an assignment function $\chi\colon \mathcal{S} \to [d]$. Then, the following assertions hold.*

1. *The map $(z, \lambda_1, \cdots, \lambda_d) \mapsto \mathrm{val}(W^s_\Lambda(z))$ is continuous;*
2. *for all discount factors $\lambda_1, \cdots, \lambda_d$ and $z_1, z_2 \in \mathbb{R}$ such that $z_1 \leq z_2$, we have that $\mathrm{val}\left(W^s_\Lambda(z_1)\right) \geq \mathrm{val}\left(W^s_\Lambda(z_2)\right) + (z_2 - z_1)\left(\min_i \lambda_i\right)^n$, in particular, $z \mapsto \mathrm{val}(W^s_\Lambda(z))$ is strictly decreasing; and*
3. *for all discount factors $\lambda_1, \cdots, \lambda_d$, we have $\mathrm{val}\left(W^s_\Lambda(\mathrm{val}_\Lambda(s))\right) = 0$.*

Proof sketch. We present the proof sketch for each item as follows.

1. The entries of the matrix game depend continuously on the parameters $z, \lambda_1, \cdots, \lambda_d$. Thus, its value is also continuous in the parameters.
2. For all $z_1 < z_2$, we show that all of the entries of $W^s_\Lambda(z_1)$ are strictly larger than $W^s_\Lambda(z_2)$. By quantifying this difference in each entry of the matrix, we obtain the desired inequality.
3. We show that there exist optimal strategies in the matrix game $W^s_\Lambda(\mathrm{val}_\Lambda(s))$ which are derived from optimal strategies in G. These strategies guarantee that the value of the matrix game is 0. ◀

▶ **Corollary 9.** *Consider a CSG G, a state s, a reward function, and a discount function Λ. Then, $\mathrm{val}_\Lambda(s)$ is the unique $z^* \in \mathbb{R}$ such that $\mathrm{val}\left(W^s_\Lambda(z^*)\right) = 0$.*

Proof. By Theorem 8-Item 2, the mapping $z \mapsto \mathrm{val}\left(W^s_\Lambda(z)\right)$ is strictly decreasing. By Theorem 8-Item 3, we know that $\mathrm{val}\left(W^s_\Lambda(\mathrm{val}_\Lambda(s))\right) = 0$. Hence, there exists the unique $z^* = \mathrm{val}_\Lambda(s) \in \mathbb{R}$ such that $\mathrm{val}\left(W^s_\Lambda(z^*)\right) = 0$, which yields the result. ◀

4.4 Characterization of Limit Value

We introduce a new characterization of the limit value in CSGs (Theorems 10 and 11), which generalizes results presented in [2] from a single discount factor to multiple discount factors. In particular, Theorem 11 generalizes Theorem 2 of [2].

Limit function. Given a CSG G, a reward function, and an assignment function $\chi\colon \mathcal{S} \to [d]$, we define the limit function as

$$F^s_\chi(z) := \lim_{\lambda_1 \to 0^+} \cdots \lim_{\lambda_d \to 0^+} \frac{\mathrm{val}\left(W^s_\Lambda(z)\right)}{(\lambda_d)^n}\,.$$

The following statements (Theorem 10 and Theorem 11) connect the limit value with the limit function.

▶ **Lemma 10.** *Consider a CSG G, a state s, a reward function, and an assignment function $\chi\colon \mathcal{S} \to [d]$. Then, the following assertions hold.*

1. *For all $z \in \mathbb{R}$, the limit $F^s_\chi(z)$ exists in $\mathbb{R} \cup \{-\infty, +\infty\}$; and*
2. *There exists $z_1, z_2 \in \mathbb{R}$ such that $F^s_\chi(z_2) \leq 0 \leq F^s_\chi(z_1)$.*

Proof sketch. We present the proof sketch for each item as follows.

1. By Theorem 7, we show that given a fixed parameter z, there exist two multi-variate polynomials P and Q in $\lambda_1, \cdots, \lambda_d$ such that for all small enough $\lambda_1, \cdots, \lambda_d$, we have $\mathrm{val}(W^s_\Lambda(z)) = \frac{P(\lambda_1, \cdots, \lambda_d)}{Q(\lambda_1, \cdots, \lambda_d)}$, which implies the existence of the limit.
2. We provide two numbers z_1 and z_2 such that $\mathrm{sign}(F^s_\chi(z_1)) \neq \mathrm{sign}(F^s_\chi(z_2))$. ◀

▶ **Corollary 11.** *Consider a CSG G, a state s, a reward function, and an assignment function $\chi\colon \mathcal{S} \to [d]$. Then, $\mathrm{val}_\chi(s)$ is the unique $z^* \in \mathbb{R}$ such that*

$$\forall z > z^* \quad F^s_\chi(z) < 0 \qquad \text{and} \qquad \forall z < z^* \quad F^s_\chi(z) > 0\,.$$

Proof sketch. By Theorem 8-Item 2, F^s_χ is decreasing. By Theorem 10-Item 2, there exists a unique sign-changing point. By the definition of the limit value and F^s_χ, we show that this sign-changing point is $\mathrm{val}_\chi(s)$. ◀

4.5 Approximation of Limit Value

We introduce an approach for the approximation of the limit value via the stateful-discounted value (Theorem 12).

▶ **Theorem 12.** *Consider a CSG G, a state s, a reward function r, an assignment function $\chi\colon \mathcal{S} \to [d]$, and an additive error $\varepsilon > 0$. The transition function δ and the reward function r are represented by rational numbers of bit-size* B. *Fix*

$$D := \max(|\Sigma^{PS}|, |\Gamma^{PS}|), \quad \mathrm{B}_1 := 11Dn(\mathrm{B} + \mathrm{bit}(n) + \mathrm{bit}(D) + \mathrm{bit}(\varepsilon)) ,$$

and, for all $1 \le i \le d$, we set $\lambda_i^0 := \exp\left(-\mathrm{B}_1(nD+1)^{i-1}\right)$. Then, we have

$$|\,\mathrm{val}_{\Lambda^0}(s) - \mathrm{val}_\chi(s)| \le \varepsilon .$$

Proof sketch. The proof has following key components:

- We show that there exists a finite set of rational functions with bounded degrees and coefficients such that for all $z, \lambda_1, \cdots, \lambda_d$, the value of $W^s_\Lambda(z)$ corresponds to one of these functions evaluated in $z, \lambda_1, \cdots, \lambda_d$.
- We derive some insights on the asymptotic behavior of the sign of the map $(\lambda_1, \cdots, \lambda_d) \mapsto \mathrm{val}(W^s_\Lambda(z))$ as $\lambda_d, \cdots, \lambda_1$ go to 0 respectively.
- We establish the connection between the stateful-discounted and limit value by the characterizations introduced in Theorem 9 and Theorem 11, and above insights. ◀

Novelty. As mentioned previously, our result is a generalization of [2]. The key non-trivial aspect of the generalization relies on the fact that [2] considers uni-variate polynomials, whereas our result requires analysis of multi-variate polynomials. Theorem 7 is the key mathematical foundation, and the proofs require significant technical generalization.

5 Algorithms for LimitValue and ParityValue

In this section, we present algorithms for computing ε-approximation of stateful-discounted, limit, and parity values. The section is organized as follows. In Section 5.1, we present an algorithm for computing ε-approximate stateful-discounted value. In Section 5.2, we present an algorithm for computing ε-approximate limit value, and as a consequence, we also obtain an algorithm for computing ε-approximate parity value.

5.1 Algorithm for Approximate Stateful-discounted Value

In this subsection, we present an algorithm for computing ε-approximation of the stateful-discounted value in CSGs. Given a CSG G, a reward function, and a discount function Λ, Algorithm 1 runs a binary search over the stateful-discounted value of state s. At the beginning, $\underline{z}$ and $\overline{z}$ are the under and over approximation of $\mathrm{val}_\Lambda(s)$. In each step, the algorithm halves the interval $[\underline{z}, \overline{z}]$ by increasing $\underline{z}$ or decreasing $\overline{z}$ based on the sign of $\mathrm{val}\left(W^s_\Lambda\left(\frac{\underline{z}+\overline{z}}{2}\right)\right)$. After $\mathrm{bit}(\varepsilon)$ steps, the algorithm outputs the ε-approximate value $(\overline{z}+\underline{z})/2$. The correctness and the time complexity of the algorithm are shown in Theorem 13.

Algorithm 1 ApproxDiscounted.

Input: Game G, state s, reward function r, a discount function Λ, additive error ε
Output: Approximate stateful-discounted value v such that $|v - \mathrm{val}_\Lambda(s)| \le \varepsilon$

procedure ApproxDiscounted$(G, s, r, \Lambda, \varepsilon)$
    $\underline{z} \leftarrow 0$ and $\overline{z} \leftarrow 1$
    while $\overline{z} - \underline{z} > \varepsilon$ **do**
        $z \leftarrow \frac{\underline{z}+\overline{z}}{2}$
        $\nu \leftarrow \mathrm{val}(W^s_\Lambda(z))$
        if $\nu \ge 0$ **then** $\underline{z} \leftarrow z$
        else $\overline{z} \leftarrow z$
    return $\frac{\underline{z}+\overline{z}}{2}$

▶ **Lemma 13.** *Consider a CSG G, a state s, a reward function r, a discount function Λ, and an additive error $\varepsilon > 0$. The transition function δ, the reward function r, and the discount function Λ are represented by rational numbers of bit-size* B*. Then, Algorithm 1 computes the ε-approximation of the stateful-discounted value of state s. Moreover, the algorithm runs in time* $\exp\left(\mathcal{O}(n\log(m) + \log(B) + \log(\log(1/\varepsilon))\right)$.

Proof sketch. The correctness of Algorithm 1 is due to the properties established in Theorem 8. The time complexity of Algorithm 1 is due to the construction of the matrix game and computing its value. ◀

5.2 Algorithms for Approximate Limit and Parity Values

In this subsection, we present an algorithm for computing ε-approximation of the limit and parity values in CSGs. Given a CSG G, a reward function, and an assignment function χ, Algorithm 2 outputs the $\varepsilon/2$-approximate of the stateful-discounted value of state s for some Λ_0 by calling ApproxDiscounted. By Theorem 12, the stateful-discounted value is an $\varepsilon/2$-approximation of the limit value. Thus, the returned value of the algorithm is indeed an ε-approximate of the limit value. The correctness and the time complexity of the algorithm is shown in Theorem 14. Since CSGs with parity objectives have a linear-size reduction to CSGs with the limit value of stateful-discounted objectives, as a consequence of the above algorithm, we obtain an algorithm for parity value approximation.

Algorithm 2 ApproxLimit.

Input: Game G, state s, reward function r, assignment function χ, additive error ε
Output: Approximate limit value v such that $|v - \mathrm{val}(s)| \le \varepsilon$

procedure ApproxLimit$(G, s, r, \chi, \varepsilon)$
    $D \leftarrow m^n$
    $\mathrm{B}_1 \leftarrow 11Dn\left(\mathrm{B} + \mathrm{bit}(n) + \mathrm{bit}(D) + \mathrm{bit}(\varepsilon)\right)$
    for $i \leftarrow 1$ to d **do**
        $\lambda^0_i \leftarrow \exp\left(-\mathrm{B}_1(nD+1)^{i-1}\right)$
    $v \leftarrow$ ApproxDiscounted$(G, s, r, \Lambda^0, \varepsilon/2)$
    return v

▶ **Lemma 14.** *Consider a CSG G, a state s, a reward function r, an assignment function $\chi\colon \mathcal{S} \to [d]$, and an additive error $\varepsilon > 0$. The transition function δ and the reward function r are represented by rational numbers of bit-size* B*. Then, Algorithm 2 computes the ε-approximation of the limit value of state s. Moreover, the algorithm runs in time* $\exp\left(\mathcal{O}(nd\log(m) + \log(\mathrm{B}) + \log(\log(1/\varepsilon)))\right)$.

Proof sketch. The correctness of the algorithm is due to Theorem 13 and Theorem 12. The time complexity is due to the size of the representation of Λ^0 and Theorem 13. ◀

Proof of Theorem 5-Item 2. It is a direct implication of Theorem 14. ◀

▶ **Corollary 15.** *Consider a CSG G, a priority function p, a state s, and an additive error $\varepsilon > 0$. The transition function δ is represented by rational numbers of bit-size* B*. Then, there exists an algorithm that computes the ε-approximation of the parity value of state s. Moreover, the algorithm runs in time* $\exp\left(\mathcal{O}(nd\log(m) + \log(\mathrm{B}) + \log(\log(1/\varepsilon)))\right)$.

Proof. By [18, 10], there exists a linear-size reduction from the CSGs with parity objectives to the CSGs with the limit-value of stateful-discounted objectives. Therefore, the result follows from Theorem 14. ◀

Proof of Theorem 6-Item 2. It is a direct implication of Theorem 15. ◀

6 Computational Complexities of LimitValue and ParityValue

In this section, we show that the LimitValue and ParityValue problems are in TFNP[NP]. This section is organized as follows. In Section 6.1, we present some useful definitions related to Markov Chains (MCs) and Markov Decision Processes (MDPs), and floating-point representation. In Section 6.2, we present the complexity results for LimitValue and ParityValue problems.

6.1 Definitions

We present some basic notations and definitions related to Markov Chains, Markov Decision Processes, and the classic symbolic representation for numbers and probability distributions, called floating-point.

Markov decision processes and Markov chains. For $i \in \{1, 2\}$, a Player-i Markov decision process (Player-i MDP) is a special class of CSGs where the other player has only one action and is denoted by $\mathcal{P} = (\mathcal{S}, \mathcal{A}, \delta\colon \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S}))$. A Markov chain (MC) is a special class of MDPs where both players have only one action and is denoted by $\mathcal{C} = (\mathcal{S}, \delta\colon \mathcal{S} \to \Delta(\mathcal{S}))$. In Markov chains we write $\delta(s, s')$ to denote $\delta(s)(s')$.

Absorbing MCs. We say an MC $\mathcal{C}$ is *absorbing* if there exists a subset of absorbing states $\mathcal{S}_0 \subseteq \mathcal{S}$ such that

- For all $s \in \mathcal{S}_0$, we have $\delta(s, s) = 1$; and
- For all $s_0 \in \mathcal{S} \setminus \mathcal{S}_0$, there exist states $s_1, \ldots, s_k$ such that $\delta(s_i, s_{i+1}) > 0$ and $s_k \in \mathcal{S}_0$.

States in $\mathcal{S}_0$ are called absorbing.

MDPs and MCs given stationary strategies in CSGs. Given a stationary strategy σ for Player 1 in a game G, by fixing the strategy σ, we obtain a Player-2 MDP $G_\sigma = (S, \mathcal{B}, \delta_\sigma)$ where the transition function $\delta_\sigma : \mathcal{S} \times \mathcal{B} \to \Delta(\mathcal{S})$ is given by $\delta_\sigma(s, b)(s') := \sum_{a \in \mathcal{A}} \delta(s, a, b)(s') \cdot \sigma(s)(a)$, for all $s, s' \in \mathcal{S}$ and $b \in \mathcal{B}$. Analogously, we obtain Player-1 MDP G_τ by fixing a stationary strategy τ for Player 2. Moreover, by fixing stationary strategies σ and τ for both players, we obtain an MC $G_{\sigma,\tau} = (\mathcal{S}, \delta_{\sigma,\tau})$, where the transition function $\delta_{\sigma,\tau} : \mathcal{S} \to \Delta(\mathcal{S})$ is given by $\delta_{\sigma,\tau}(s)(s') = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \delta(s, a, b)(s') \cdot \sigma(s)(a) \cdot \tau(s)(b)$, for all $s, s' \in \mathcal{S}$.

Reachability objectives in MCs. Given an MC $\mathcal{C}$ and a target set $T \subseteq \mathcal{S}$, the reachability objective is the indicator function of plays eventually reaching T. More formally, for a play $\omega = \langle s_0, s_1, \cdots \rangle$, we define $\mathrm{Reach}_T : \Omega \to \{0, 1\}$ as

$$\mathrm{Reach}_T(\omega) := \begin{cases} 1 & \exists i \in \mathbb{N}\; s_i \in T \\ 0\,. & \text{otherwise} \end{cases}$$

We define the probability of reaching the target set T from state s as $\mathrm{val}_T(s) := \mathbb{E}_s[\mathrm{Reach}_T]$.

Floating-point number representation. We define the set of floating-point numbers with precision ℓ as $\mathcal{F}(\ell) := \{ m \cdot 2^e \mid m \in \{0, \cdots, 2^\ell - 1\},\ e \in \mathbb{Z} \}$. The floating-point representation of an element $x = m \cdot 2^e \in \mathcal{F}(\ell)$ uses $\mathrm{bit}(m) + \mathrm{bit}(|e|)$ bits. We define the relative distance of two positive real numbers $x, \widetilde{x}$ as $\mathrm{rel}(x, \widetilde{x}) := \max \left\{ \frac{x}{\widetilde{x}}, \frac{\widetilde{x}}{x} \right\} - 1$. We say x is (ℓ, i)-close to $\widetilde{x}$ if $\mathrm{rel}(x, \widetilde{x}) \le (1 - 2^{1-\ell})^{-i} - 1$, where ℓ is a positive integer and i is a non-negative integer.

Floating-point probability distribution representation. We denote by $\mathcal{D}(\ell)$ the set of all floating-point probability distributions with precision ℓ. A probability distribution $\mu \in \Delta([t])$ belongs to $\mathcal{D}(\ell)$ if there exists $w_1, w_2, \cdots, w_t \in \mathcal{F}(\ell)$ such that

- For all $i \in [t]$, we have $\mu(i) = \frac{w_i}{\sum_{j \in [t]} w_j}$; and
- $\sum_{j \in [t]} w_j$ and 1 are (ℓ, t)-close.

We define the relative distance rel for probability distributions as $\mathrm{rel}(\mu, \widetilde{\mu}) := \max\{\mathrm{rel}(\mu(i), \widetilde{\mu}(i)) : i \in [t]\}$. We say μ is (ℓ, i)-close to $\widetilde{\mu}$ if $\mathrm{rel}(\mu, \widetilde{\mu}) \le (1 - 2^{1-\ell})^{-i} - 1$, where ℓ is a positive integer and i is a non-negative integer.

6.2 Complexity Results

We present algorithms for computing ε-approximate stateful-discounted value in MCs, MDPs, and CSGs, which generalize the result of [4] from a single discount factor to multiple discount factors. The algorithm for MCs (Theorem 16) is achieved by a reduction from stateful-discounted objective to reachability objectives in MCs and using the algorithm for computing reachability values in MCs presented in [17]. By our technical result on the limit value approximation via the stateful-discounted value (Theorem 12), we consequently obtain a TFNP[NP] procedure for LimitValue. Since there exists a linear-size reduction from CSGs with parity objectives to CSGs with the limit-value of stateful-discounted objectives [18, 10], ParityValue is also in TFNP[NP].

▶ **Lemma 16.** *Consider an MC $\mathcal{C}$, a reward function r, and a discount function Λ. For all $s \in \mathcal{S}$, we set*

$$\delta(s) \in D(\ell), \quad r(s) \in \mathcal{F}(\ell), \quad \Lambda(s) \in \mathcal{F}(\ell)\,,$$

where $\ell \geq 1000n^2$. Then, there exists a polynomial-time algorithm that for all states $s \in \mathcal{S}$, computes an approximation v for the stateful-discounted value such that $|v - \mathrm{val}_\Lambda(s)| \leq 104n^4 2^{-\ell}$.

Proof sketch. We construct a new MC $\widetilde{\mathcal{C}}$ with a reachability objective derived from the MC $\mathcal{C}$ such that the reachability value of $\widetilde{\mathcal{C}}$ is an approximation of the stateful-discounted value of $\mathcal{C}$. The result follows from the algorithm for computing ε-approximation of reachability value in MCs presented in [17]. ◀

▶ **Lemma 17.** *The problem of deciding if the stateful-discounted value for Player-1 MDPs is below a threshold up to an additive error is in NP where the input is a Player-1 MDP $\mathcal{P}$, a reward function r, a discount function Λ, a state s, a threshold $0 \leq \alpha \leq 1$, an additive error $\varepsilon = 2^{-\kappa}$ and a positive integer ℓ such that, for all $s' \in \mathcal{S}$ and $a \in \mathcal{A}$, we have*

$$\delta(s', a) \in D(\ell), \quad r(s', a) \in \mathcal{F}(\ell), \quad \Lambda(s') \in \mathcal{F}(\ell), \quad \ell \geq 1000n^2 + \kappa\,.$$

Note that, the numbers α and ε are represented in fixed-point binary and the NP procedure is such that

- *If $\alpha \leq \mathrm{val}_\Lambda(s) - \varepsilon$, then it outputs YES; and*
- *If $\alpha \geq \mathrm{val}_\Lambda(s) + \varepsilon$, then it outputs NO.*

Proof sketch. The procedure guesses a pure stationary strategy σ. By fixing σ, we obtain an MC $\mathcal{P}_\sigma$. The algorithm verifies whether the threshold can be approximately achieved if the player follows the strategy σ. The verification procedure is implemented by Theorem 16. ◀

Theorem 17 also holds for Player-2 MDPs by symmetric arguments. More formally, we have the following result.

▶ **Corollary 18.** *The problem of deciding if the stateful-discounted value for Player-2 MDPs is above a threshold up to an additive error is in NP where the input is a Player-2 MDP $\mathcal{P}$, a reward function r, a discount function Λ, a state s, a threshold $0 \leq \alpha \leq 1$, an additive error $\varepsilon = 2^{-\kappa}$ and a positive integer ℓ such that, for all $s' \in \mathcal{S}$ and $a \in \mathcal{A}$, we have*

$$\delta(s', a) \in D(\ell), \quad r(s', a) \in \mathcal{F}(\ell), \quad \Lambda(s') \in \mathcal{F}(\ell), \quad \ell \geq 1000n^2 + \kappa\,.$$

▶ **Lemma 19.** *The problem of computing an ε-approximation of the stateful-discounted value for CSGs is in TFNP[NP] for inputs CSGs G, reward functions, discount functions Λ, states s, additive errors $\varepsilon = 2^{-\kappa}$, and positive integers ℓ such that, for all states $s' \in \mathcal{S}$, we have $\Lambda(s') \in \mathcal{F}(\ell)$.*

Proof sketch. The procedure guesses two stationary strategies σ and τ and an approximate value α. By fixing the strategy σ (resp. τ), we obtain a Player-1 MDP G_σ (resp. Player-2 MDP G_τ). By the NP oracles calling the procedures defined in Theorems 17 and 18, the algorithm verifies whether the guessed threshold can be approximately achieved by both of the strategies. ◀

Proof of Theorem 5-Item 1. It is a direct implication of Theorem 19 and Theorem 12. ◀

Proof of Theorem 6-Item 1. By [18, 10], there exists a linear-size reduction from Parity-Value to LimitValue. Therefore, the result follows from Theorem 5-Item 1. ◀

Concluding remarks. In this work, we present improved complexity upper bounds and algorithms for the value approximation problem for concurrent stochastic games with two classic objectives. There are several interesting directions for future work. First, whether the complexity can be further improved from TFNP[NP] to TFNP is a major open question, even for reachability objectives. Second, whether for parity objectives, the dependency on d can be improved from linear to logarithmic, retaining the logarithmic dependence on m, is another interesting open question. Finally, the study of priority mean-payoff objectives for concurrent stochastic games and their connection to stateful-discounted objectives is another interesting direction for future work.

References

1 Rajeev Alur, Thomas A Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the ACM (JACM)*, 49(5):672–713, 2002. doi:10.1145/585265.585270.

2 Luc Attia and Miquel Oliu-Barton. A formula for the value of a stochastic game. *Proceedings of the National Academy of Sciences*, 116(52):26435–26443, 2019.

3 Saugata Basu. New results on quantifier elimination over real closed fields and applications to constraint databases. *Journal of the ACM (JACM)*, 46(4):537–555, 1999. doi:10.1145/320211.320240.

4 Sougata Bose, Rasmus Ibsen-Jensen, and Patrick Totzke. Bounded-memory strategies in partial-information games. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 1–14, 2024. doi:10.1145/3661814.3662096.

5 Krishnendu Chatterjee. *Stochastic ω-regular games.* University of California, Berkeley, 2007.

6 Krishnendu Chatterjee, Luca de Alfaro, and Thomas A Henzinger. The complexity of quantitative concurrent parity games. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 678–687, 2006. URL: http://dl.acm.org/citation.cfm?id=1109557.1109631.

7 Krishnendu Chatterjee, Rupak Majumdar, and Thomas A Henzinger. Stochastic limit-average games are in EXPTIME. *International Journal of Game Theory*, 37(2):219–234, 2008. doi:10.1007/S00182-007-0110-5.

8 Luca de Alfaro and Thomas A Henzinger. Concurrent omega-regular games. In *Proceedings Fifteenth Annual IEEE Symposium on Logic in Computer Science*, pages 141–154. IEEE, 2000. doi:10.1109/LICS.2000.855763.

9 Luca de Alfaro, Thomas A Henzinger, and Orna Kupferman. Concurrent reachability games. In *Proceedings 39th Annual Symposium on Foundations of Computer Science*, pages 564–564. IEEE Computer Society, 1998.

10 Luca de Alfaro, Thomas A Henzinger, and Rupak Majumdar. Discounting the future in systems theory. In *International Colloquium on Automata, Languages, and Programming*, pages 1022–1037. Springer, 2003. doi:10.1007/3-540-45061-0_79.

11 Luca de Alfaro, Thomas A Henzinger, and Freddy YC Mang. The control of synchronous systems. In *International Conference on Concurrency Theory*, pages 458–473. Springer, 2000.

12 Luca de Alfaro, Thomas A Henzinger, and Freddy YC Mang. The control of synchronous systems, part II. In *International Conference on Concurrency Theory*, pages 566–581. Springer, 2001.

13 Luca de Alfaro and Rupak Majumdar. Quantitative solution of omega-regular games. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 675–683, 2001. doi:10.1145/380752.380871.

14 Kousha Etessami and Mihalis Yannakakis. Recursive concurrent stochastic games. *Logical Methods in Computer Science*, 4, 2008. doi:10.2168/LMCS-4(4:7)2008.

15 Hugh Everett. Recursive games. *Contributions to the Theory of Games*, 3(39):47–78, 1957.

16 Jerzy Filar and Koos Vrieze. *Competitive Markov Decision Processes.* Springer-Verlag, 1997.

17 Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen. Approximating the value of a concurrent reachability game in the polynomial time hierarchy. In *International Symposium on Algorithms and Computation*, pages 457–467. Springer, 2013. doi:10.1007/978-3-642-45030-3_43.

18 Hugo Gimbert and Wiesław Zielonka. Discounting infinite games but how and why? *Electronic Notes in Theoretical Computer Science*, 119(1):3–9, 2005. doi:10.1016/J.ENTCS.2004.07.005.

19 Hugo Gimbert and Wiesław Zielonka. Blackwell optimal strategies in priority mean-payoff games. *International Journal of Foundations of Computer Science*, 23(03):687–711, 2012. doi:10.1142/S0129054112400345.

20 Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Peter Bro Miltersen. The complexity of solving reachability games using value and strategy iteration. In *International Computer Science Symposium in Russia*, pages 77–90. Springer, 2011. doi:10.1007/978-3-642-20712-9_7.

21 Kristoffer Arnsfelt Hansen, Michal Koucky, Niels Lauritzen, Peter Bro Miltersen, and Elias P Tsigaridas. Exact algorithms for solving stochastic games. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 205–214, 2011.

22 Kristoffer Arnsfelt Hansen, Michal Koucky, and Peter Bro Miltersen. Winning concurrent reachability games requires doubly-exponential patience. In *2009 24th Annual IEEE Symposium on Logic In Computer Science*, pages 332–341. IEEE, 2009. doi:10.1109/LICS.2009.44.

23 Donald A Martin. The determinacy of Blackwell games. *The Journal of Symbolic Logic*, 63(4):1565–1581, 1998. doi:10.2307/2586667.

24 Jean-François Mertens and Abraham Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–66, 1981.

25 Miquel Oliu-Barton. New algorithms for solving zero-sum stochastic games. *Mathematics of Operations Research*, 46(1):255–267, 2021. doi:10.1287/MOOR.2020.1055.

26 Lloyd Stowell Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

27 Wolfgang Thomas. Languages, automata, and logic. In *Handbook of Formal Languages: Volume 3 Beyond Words*, pages 389–455. Springer, 1997. doi:10.1007/978-3-642-59126-6_7.