Deterministic Sub-exponential Algorithm for Discounted-sum Games with Unary Weights

Ali Asadi
ISTA
Klosterneuburg, Austria
ali.asadi@ist.ac.at

Krishnendu Chatterjee
ISTA
Klosterneuburg, Austria
krishnendu.chatterjee@ist.ac.at

Raimundo Saona
ISTA
Klosterneuburg, Austria
raimundojulian.saonaurmeneta@ist.ac.at

Jakub Svoboda
ISTA
Klosterneuburg, Austria
jakub.svoboda@ist.ac.at

ABSTRACT

Turn-based discounted-sum games are two-player zero-sum games played on finite directed graphs. The vertices of the graph are partitioned between player 1 and player 2. Plays are infinite walks on the graph where the next vertex is decided by a player that owns the current vertex. Each edge is assigned an integer weight and the payoff of a play is the discounted-sum of the weights of the play. The goal of player 1 is to maximize the discounted-sum payoff against the adversarial player 2. These games lie in NP ∩ coNP and are among the rare combinatorial problems that belong to this complexity class and the existence of a polynomial-time algorithm is a major open question. Since breaking the general exponential barrier has been a challenging problem, faster parameterized algorithms have been considered. If the discount factor is expressed in unary, then discounted-sum games can be solved in polynomial time. However, if the discount factor is arbitrary (or expressed in binary), but the weights are in unary, none of the existing approaches yield a sub-exponential bound. Our main result is a new analysis technique for a classical algorithm (namely, the strategy iteration algorithm) that present a new runtime bound which is $n^{O\left(W^{1/4}\sqrt{n}\right)}$, for game graphs with n vertices and absolute weights of at most W. In particular, our result yields a deterministic sub-exponential bound for games with weights that are constant or represented in unary.

ACM Reference Format:
Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Jakub Svoboda. 2024. Deterministic Sub-exponential Algorithm for Discounted-sum Games with Unary Weights. In *39th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS '24), July 8–11, 2024, Tallinn, Estonia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3661814.3662080

LICS '24, July 8–11, 2024, Tallinn, Estonia

ACM ISBN 979-8-4007-0660-8/24/07...$15.00
https://doi.org/10.1145/3661814.3662080

1 INTRODUCTION

Turn-based discounted-sum games. Turn-based graph games [8, 23] are two-player infinite-duration zero-sum games played on a finite directed graph. The vertex set is partitioned into player-1 and player-2 vertices. At player-1 and player-2 vertices, the respective player chooses a successor vertex. Given an initial vertex, the repeated interaction between the players generates an infinite walk (called *play*) in the graph. Strategies (or policies) for players provide the successor vertex choice at every player vertex. A payoff function assigns a real value to every play. We consider a classical and well-studied function: the discounted-sum payoff function [20, 31, 33]. Every edge of the graph is assigned an integer weight, and the payoff of a play is the discounted-sum of the weights of the play.

Value problem: complexity and algorithm. The value at a vertex of a discounted-sum game is the maximal payoff that player 1 can ensure irrespective of the strategy choice of player 2. The decision problem associated with the value (i.e., whether the value of a vertex is at least a given threshold) lies in NP ∩ coNP (even UP ∩ coUP) [9, 14, 35]. This is among the rare combinatorial problems that lie in NP ∩ coNP and the existence of a polynomial-time algorithm is a major and long-standing open problem. The classical algorithmic approaches to compute the values of a discounted-sum game are: (a) strategy iteration (or policy improvement) [20]; and (b) value iteration [4, 35]. The best-known worst-case bounds for these algorithms are exponential time. There is a randomized sub-exponential time algorithm to compute values [30] with the expected running time of $2^{\widetilde{O}(\sqrt{n})}$, where n is the number of vertices, and $\widetilde{O}(\cdot)$ hides polylogarithmic factors.

Related game problems on directed graphs. There are interesting related problems for games on directed graphs, namely, *parity games* [18] and *mean-payoff games* [17, 24, 35]. Parity games and mean-payoff games also lie in NP ∩ coNP [18] (even UP ∩ coUP [26]), and the existence of polynomial-time algorithms are major open problems. Parity games admit linear-time reduction to mean-payoff games [26] and mean-payoff games admit linear-time reduction to discounted-sum games [35]. However, reductions in the converse direction are not known. For parity games, several important algorithmic improvements have been achieved. In particular, deterministic sub-exponential time algorithm [28] and deterministic quasi-polynomial time algorithm [7], that break the existing exponential-time barrier. No similar algorithmic improvements have been achieved for mean-payoff games and discounted-sum games. Indeed, no deterministic sub-exponential time algorithm is known for mean-payoff or discounted-sum games.

Parameterized algorithms and open problem. Given that the algorithmic improvements for discounted-sum games have been rare, it is natural to consider faster parameterized algorithms. The natural restriction to consider is the representation of the numbers related to discounted-sum games. There are two sources of numbers related to discounted-sum games: (a) the weights; and (b) the discount factor. The results of [25] establish, that if the discount factor is constant, then the strategy iteration algorithm runs in polynomial time (this result holds even in games with stochastic transitions [25]). However, when the discount factor is arbitrary but the weight function is expressed in unary, no better bound than the exponential bound is known, and the existence of a deterministic sub-exponential algorithm is an important open problem.

Motivation. While the problem of discounted-sum games with unary weights is theoretically interesting, there are practical motivations as well. Game graphs are models of reactive systems, where vertices represent states of the system, edges represent transitions, and players represent agents controlling different transitions. In analysis of reactive systems, small/constant weights are natural, e.g., when there are good and bad events and weights represent the relative importance of the good and bad events [10, 12]. Since discounted-sum objectives are studied in reactive systems analysis [15], improved algorithm for such games are of practical relevance along with their theoretical importance.

Our result. In this work, our main result answers the above open question. We present an improved analysis of the strategy iteration algorithm and show that, given a discounted-sum game with n vertices and absolute weights at most W, the running time is $n^{O\left(W^{1/4}\sqrt{n}\right)}$. Hence, if the weights are constant or represented in unary, then the algorithm is a deterministic sub-exponential time algorithm.

Technical contributions. We first present new bounds on the roots of polynomials with bounded integer coefficients. We show a sub-exponential lower bound and two upper bounds for the roots of polynomials: (i) a non-constructive sub-exponential upper bound; and (ii) an explicit quasi-polynomial upper bound. Our key insight is to associate all strategy profiles in discounted-sum games with rational polynomial functions with bounded integer coefficients. This main technical contribution establishes a connection between the lower bound for the polynomials and the running time analysis of the strategy iteration algorithm. To the best of our knowledge, such a connection has not been established before. Given the connection and our bounds on the roots of polynomials, we establish an improved running time analysis of the strategy iteration algorithm for discounted-sum games. Section 4 presents the results on bounds on the roots of polynomials and Section 5 presents the key insights of analysis of strategy iteration algorithm using results of Section 4.

Related works. The algorithmic study of discounted-sum, mean-payoff, and parity games have received significant attention. Below we summarize some related works.

- *Parity games.* There has been a significant progress in the study of parity games. While the classical algorithms [18, 34] has exponential worst-case complexity, faster exponential-time algorithms were achieved [27, 32], and then deterministic sub-exponential [28] and deterministic quasi-polynomial time algorithms [7] were obtained. However, extending the algorithmic bounds from parity games to mean-payoff or discounted-sum games has been a major open question.
- *Mean-payoff games.* The algorithmic aspects of mean-payoff games has also been studied in several works [6, 16, 17, 24, 35]. All these algorithms have an exponential worst-case complexity. However, the classical value iteration algorithm is pseudo-polynomial and runs in polynomial time if the weights are expressed in unary.
- *Discounted-sum games.* The value iteration algorithm for discounted-sum games has been studied in [4, 35], and the strategy iteration algorithm has been studied in [25]. While the worst-case bound for the algorithms is exponential, if the discount-factor is constant, these algorithms run in polynomial time [4, 25]. The value iteration and strategy iteration algorithms have an explicit dependence on the discount factor. An algorithm that does not depend on the discount factor is presented in [29], which is inspired by the algorithm of [16] for mean-payoff games. This algorithm has a complexity of $O\left((2+\sqrt{2})^n\right)$, and is exponential without any dependence on the weights or discount factor.
- *Stochastic games.* Discounted-sum games admit a linear reduction to stochastic games with reachability objectives [35], which are games with stochastic transitions. The algorithmic study of stochastic games has been considered in several works [13, 14, 30]. However, even stochastic games with 0 and 1 weights are as hard as stochastic games with general weights [2, 14], i.e., parameterization by the weights is not useful.

In summary, none of the existing approaches break the long-standing exponential barrier for discounted-sum games with weights in unary and we present the first deterministic sub-exponential bound.

2 PRELIMINARIES

We present standard notations and definitions related to turn-based games, similar to [11, 34].

Turn-based games. A turn-based game (TBG) is a two-player finite game $G = (V = V_1 \sqcup V_2, E)$ consisting of a finite graph with

- the set of vertices V, of size n, partitioned into player-1 vertices V_1 and player-2 vertices V_2; and
- the set of edges $E \subseteq V \times V$, of size m, such that for all $v \in V$, the set $E(v) \coloneqq \{u \mid (v,u) \in E\}$ is non-empty.

Steps and plays. Given an initial vertex $v_0 \in V$, the game proceeds as follows. In each step, the player who owns the current vertex v chooses the next vertex from the set $E(v)$. A *play* is an infinite sequence of vertices $\omega = \langle v_0, v_1, \ldots \rangle$ such that, for every step $t \geq 0$, the vertex $v_{t+1} \in E(v_t)$. We denote by Ω the set of all plays, and by Ω_v the set of all plays $\omega = \langle v_0, v_1, \ldots \rangle$ where $v_0 = v$.

Discounted-payoff objectives. We consider TBGs with a weight or reward function $r \colon E \to \mathbb{Z}$ that assigns a reward value $r(v,u)$ for all edges $(v,u) \in E$. We denote the largest absolute reward by $W := \max\{|r(v,u)| \mid (v,u) \in E\}$. For a play $\omega = \langle v_0, v_1, \ldots\rangle$ and a discount factor $\lambda \in [0,1)$, the discounted-payoff (or simply payoff) is denoted by $\mathrm{Disc}_\lambda(\omega) := \sum_{i\geq 0} \lambda^i r(v_i, v_{i+1})$. The objective of player 1 is to maximize the payoff, while player 2 minimizes the payoff.

Positional strategies. Strategies are recipes that specify how to choose the next vertex. A *positional* strategy $\sigma\colon V_1 \to V$ for player 1 (resp. $\tau\colon V_2 \to V$ for player 2) is a strategy which chooses a vertex $\sigma(v) \in E(v)$ whenever the play visits vertex v. A strategy profile (σ, τ) is a pair of strategies for both players. We denote by Σ^P and Γ^P the set of all positional strategies for player 1 and player 2, respectively. In general, strategies can depend on past history and not only the current vertex. However, for discounted-sum objective, positional strategies are as powerful as general strategies [14]. Hence, in the sequel, every strategy is positional.

Lasso-shaped plays given strategies in TBGs. We define G^σ as the restricted game where player 1 follows the strategy σ. We define G^τ and $G^{\sigma,\tau}$ similarly. Note that once both players have fixed their strategies we obtain a graph where each vertex has exactly one outgoing edge. Given an initial vertex v, we obtain a play $G_v^{\sigma,\tau} = \langle v_0, v_1, \ldots\rangle$ such that $v_0 = v$, and for any step $t \geq 0$, $v_{t+1} = \sigma(v_t)$ if $v_t \in V_1$; and $v_{t+1} = \tau(v_t)$ otherwise. In other words, given strategies σ and τ, the obtained play $G_v^{\sigma,\tau}$ is a *lasso-shaped* play that consists in a finite cycle-free path $\mathcal{P} := \langle v_0, \ldots, v_{p-1}\rangle$ followed by a simple cycle $C := \langle v_p, \ldots, v_{p+c-1}\rangle$ repeated forever.

Notation. For simplicity, we denote by $\mathrm{Disc}_\lambda(G^{\sigma,\tau})$ a vector whose v-th coordinate is the discounted payoff for vertex v, i.e., $\mathrm{Disc}_\lambda(G_v^{\sigma,\tau})$.

We recall the fundamental determinacy in positional strategies for TBGs with discounted-payoff objectives.

Theorem 2.1 ([14]). *For all TBGs G, vertices v, reward functions, and discount factors $\lambda \in [0,1)$, we have*

$$\max_{\sigma\in\Sigma^P} \min_{\tau\in\Gamma^P} Disc_\lambda(G_v^{\sigma,\tau}) = \min_{\tau\in\Gamma^P} \max_{\sigma\in\Sigma^P} Disc_\lambda(G_v^{\sigma,\tau}) .$$

Value and optimal strategies. Theorem 2.1 implies that switching the quantification order of positional strategies does not make a difference and leads to the unique notion of value, defined for a vertex v as

$$\mathrm{val}_\lambda(v) := \max_{\sigma\in\Sigma^P} \min_{\tau\in\Gamma^P} \mathrm{Disc}_\lambda(G_v^{\sigma,\tau}) .$$

A strategy σ for player 1 is *optimal* if, for all vertices $v \in V$, we have that

$$\min_{\tau\in\Gamma^P} \mathrm{Disc}_\lambda(G_v^{\sigma,\tau}) = \mathrm{val}_\lambda(v) .$$

The notion of optimal strategies for player 2 is defined analogously. Optimal strategies are guaranteed to exist for both players (Theorem 2.1). Therefore, restricting the attention to positional strategies does not change the notion of value.

Value problem. The value problem for turn-based discounted-sum games is defined as follows

> DiscVal. Given a game G, a reward function r, and discount factor λ, compute the value function val_λ.

Three variants. There are three variants of the DiscVal problem with respect to the representation of r and λ.

- DiscVal-Bin: Both the reward function r and the discount factor λ are given in binary.
- DiscVal-Dun: The reward function r is given in binary but the discount factor λ is given in unary.
- DiscVal-Wun: The reward function r is given in unary but the discount factor λ is given in binary.

3 OVERVIEW OF RESULTS

We discuss previous results from the literature and present our main result.

Previous results. The two classical algorithms for DiscVal are: (a) Value Iteration (VI); and (b) Strategy Iteration (SI). Both algorithms are iterative algorithms and the running time is a product of two factors: (i) the number of iterations and (ii) the complexity of every iteration. For both algorithms, the running time of every iteration is polynomial: (a) $O(m)$ for VI; and (b) $O(mn^2 \log m)$ for SI [1]. The bounds on the number of iterations are as follows.

Theorem 3.1. *The following assertions hold:*

- *The* VI *algorithm solves* DiscVal *with* $O\left(\frac{\log(W)}{1-\lambda} + n\right)$ *iterations [4, 35].*
- *The* SI *algorithm solves* DiscVal *with* $O\left(\frac{m}{1-\lambda} \log \frac{n}{1-\lambda}\right)$ *iterations [25].*

Remark 3.1 (Implications). We discuss the implications of Theorem 3.1 for the variants of DiscVal.

- For DiscVal-Bin, the above running times for VI and SI are exponential.
- For DiscVal-Dun, the above result shows that VI and SI run in polynomial-time. Moreover, even for stochastic games DiscVal-Dun has polynomial-time upper bound [25].
- For DiscVal-Wun, the current bounds for the above and other existing algorithms do not break the exponential-time barrier.

Lower bounds for SI. Lower bounds for SI have been an active research topic. Exponential lower bound for SI for parity games was established in [21], which was extended to other settings (such as randomized pivoting algorithm) [22]. Moreover, some complexity hardness result has also been established for SI (e.g., the decision problem of whether SI modifies an edge is known to be PSPACE-complete) [19].

Our result. In this work, in contrast to several lower bound results for SI in the literature, we present an improved running time analysis for SI. Our main result for SI for DiscVal is as follows.

> Theorem 3.2 (Main result). *The* SI *algorithm solves* DiscVal *with* $n^{O\left(W^{1/4}\sqrt{n}\right)}$ *iterations.*

Remark 3.2 (Implications). A key implication of our result is that we obtain the first deterministic sub-exponential time algorithm for DiscVal-Wun. In fact, as long as $W = O(n^{2-\varepsilon})$ for $\varepsilon > 0$, we obtain a deterministic sub-exponential algorithm for DiscVal.

Overview. Our analysis focuses on the difference of values between two lasso-shaped plays and uses a class of polynomials and the properties of their roots. In Section 4, we present the results related to upper and lower bounds on the roots of polynomials. In Section 5, we present the improved analysis of the SI algorithm by employing the results of Section 4.

4 BOUNDS ON THE ROOTS OF POLYNOMIALS WITH INTEGER COEFFICIENTS

In this section, we present some bounds on the roots of polynomials with integer coefficients. For a polynomial of degree N with integer coefficients bounded by W, we show that the roots of the polynomial, which are not equal to 1, are at most sub–exponentially close to 1 in terms of N and W. This result is achieved by Theorem 4.2, which is a generalization of [5]. We then present non-constructive and constructive upper bounds on how close to 1 a root of the polynomial can be.

Some illustrations of roots. In Figure 1a, we can observe the distribution of all roots for polynomials of degree at most 5 with integer coefficients ranging from -4 to 4. Additionally, Figure 1b illustrates the behavior of roots around 1.

Polynomials. A polynomial P of degree N with real coefficients bounded by W is defined as

$$P(x) = \sum_{i=0}^{N} a_i x^i ,$$

where $|a_i| \leq W$ and $a_N \neq 0$. We denote by $\mathcal{P}_N^W$ the set of all polynomials of degree N with integer coefficients bounded by W. We denote the degree of P by $\mathfrak{d}(P)$. We say α is a root of P if $P(\alpha) = 0$. A root α is of order k if there exists a polynomial Q with rational coefficients such that $Q(\alpha) \neq 0$ and $P(x) = Q(x)(x - \alpha)^k$. In this work, we only consider roots that are real numbers.

Problem definition. The problem of the roots of a polynomial is defined as follows.

Root-Poly. Given two positive integers N and W, provide lower and upper bounds on

$$\inf \left\{ |1 - \alpha| \mid P \in \mathcal{P}_N^W, P(\alpha) = 0, \alpha \neq 1 \right\} .$$

Previous Works. Borwein et al. [5] presents an upper and lower bound for a special variant of Root-Poly problem. The main results are summarized as follows.

Theorem 4.1 ([5]). *For a fixed non-negative integer k, the following assertions hold:*

- Lower bound. *Consider a polynomial P of degree N with $\{-1, 0, +1\}$ coefficients. If P has a root of order k at 1, then, for all roots $\alpha \neq 1$, we have*
$$|1 - \alpha| \geq \frac{4^{k+1}(k+1)!}{(N+1)^{k+2}} - O\left(\frac{c_1}{N^{k+3}}\right) ,$$
where $c_1 = c_1(k)$ is independent of N.
- Upper bound. *There exists a polynomial P of degree N with $\{-1, 0, +1\}$ coefficients and a root α of P such that*
$$|1 - \alpha| \leq \frac{2^{\frac{1}{2}(k+1)(k+4)}}{N^{k+2}} + O\left(\frac{c_2}{N^{2k+3}}\right) ,$$
where $c_2 = c_2(k)$ is independent of N.

Remark 4.1. Borwein et al. consider a class of polynomials with $\{-1, 0, +1\}$ coefficients and a root of order k at 1. They present an asymptotic upper and lower bound on this class of polynomials when k is fixed, and N grows to infinity.

Our results. We generalize the work of Borwein et al. [5] to a class of polynomials with integer coefficients bounded by W, and our result is independent of k.

Theorem 4.2. *The following assertions hold:*

(1) Sub-exponential lower bound. *Consider a polynomial P of degree at most N with integer coefficients bounded by W. For all roots $\alpha \neq 1$ of P, we have*
$$|1 - \alpha| > \frac{\left\lfloor \frac{16}{7} W^{1/4} \sqrt{N} \right\rfloor !}{2W(N+1)^{\frac{16}{7} W^{1/4} \sqrt{N} + 6}} .$$
(2) Quasi-polynomial constructive upper bound. *For a sufficiently large positive integer N, we present an explicit polynomial P of degree N with $\{-2, -1, 0, +1, +2\}$ coefficients and a root $\alpha < 1$ such that*
$$|1 - \alpha| \leq 2 \left(\left\lfloor N^{2/5} \right\rfloor - 1 \right)^{-3/2 \left\lfloor \log\left(\lfloor N^{2/5} \rfloor - 3 \right) \right\rfloor} .$$
(3) Sub-exponential nonconstructive upper bound. *For a sufficiently large positive integer N, there exists a polynomial P of degree N with $\{-2, -1, 0, +1, +2\}$ coefficients and a root $\alpha < 1$ such that*
$$|1 - \alpha| \leq 2 \left(\left\lfloor N^{2/5} \right\rfloor - 1 \right)^{-3/2 \left\lfloor \sqrt{\frac{\lfloor N^{2/5} \rfloor - 3}{\log(N^{2/5})}} \right\rfloor} .$$

Section 4.1 proves the lower bound, and Section 4.2 proves the upper bounds given in Theorem 4.2.

4.1 Lower bound

In this section, we show the lower bound on the roots of polynomials with integer coefficients. The proof relies on two components: (a) a lower bound for polynomials with a root of order k at 1 (Lemma 4.3); and (b) an upper bound on the order of 1 as a root (Lemma 4.6). These results yield Theorem 4.2-(1).

Lemma 4.3. *Consider a polynomial P of degree at most N with integer coefficients bounded by W. If P has a root of order at most k at 1, then, for all roots $\alpha \neq 1$, we have*

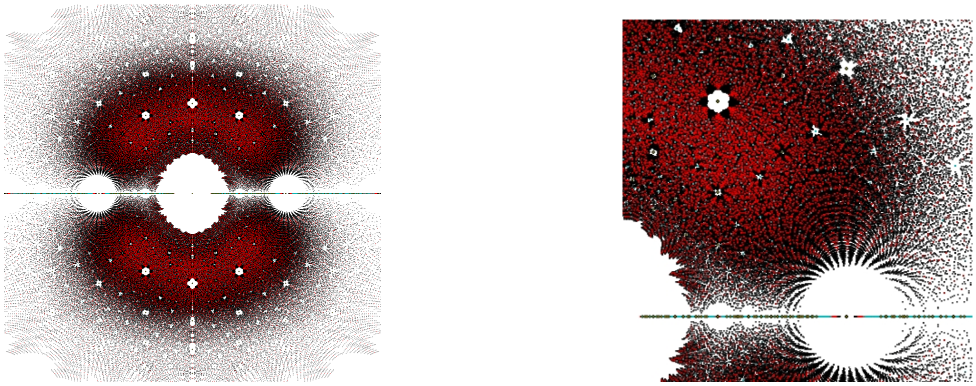
$$|1 - \alpha| > \frac{(k+1)!}{2W(N+1)^{k+2}} .$$

Proof. Consider $P = \sum_{i=0}^{N} a_i x^i$. We denote by P^j the j-th derivative of P. Note that

$$P^j(1) = \sum_{i=j}^{N} j! \binom{i}{j} a_i .$$

It follows that $P^j(1)$ is an integer divisible by $j!$, and we have

$$|P^j(1)| \leq W(N+1)^{j+1} . \quad (1)$$



(a) All roots **(b) Roots around 1**

Figure 1: Real and complex roots of all polynomials of degree at most 5 with integer coefficients ranging from -4 to 4. The horizontal axis is the real axis and the vertical axis is the imaginary axis. Roots of quadratic, cubic, quartic, and quintic polynomials are in grey, cyan, red, and black, respectively. The big hole in the middle is centered at 0, and the second biggest holes are at ±1. (source: [3])

We consider the Taylor expansion of P around 1. We have

$$P(x) = P(1) + \sum_{j=1}^{N} \frac{P^j(1)}{j!}(x-1)^j .$$

We know that $P(1) = 0$, $P(\alpha) = 0$, and $P^j(1) = 0$ for all $j < k$. Without loss of generality, we assume

$$|\alpha - 1| \leq \frac{1}{N+1} ,$$

otherwise the result follows immediately. By algebraic manipulation and triangle inequality, we get

$$\begin{aligned}
\frac{|P^k(1)|}{k!} &\leq \sum_{j=k+1}^{N} \frac{|P^j(1)|}{j!}|\alpha-1|^{j-k} \\
&\leq \sum_{j=k+1}^{N} \frac{W(N+1)^{j+1}}{j!}|\alpha-1|^{j-k} && (\text{Eqn. (1)}) \\
&\leq W(N+1)^{k+2}|\alpha-1| \sum_{j=k+1}^{N} \frac{1}{j!} && \left(|\alpha-1| \leq \frac{1}{N+1}\right) \\
&< \frac{2W(N+1)^{k+2}}{(k+1)!}|\alpha-1| . && \left(\sum_{j=k+1}^{N} \frac{1}{j!} < \frac{2}{(k+1)!}\right)
\end{aligned}$$

We know that $P^k(1) \neq 0$, is integer, and divisible by $k!$. Therefore,

$$\frac{|P^k(1)|}{k!} \geq 1 .$$

Hence,

$$\frac{(k+1)!}{2W(N+1)^{k+2}} < |\alpha - 1| ,$$

which completes the proof. □

To prove an upper bound on the order of 1 as a root (Lemma 4.6), we first present Chebyshev polynomials and their basic property, and then, we present a lemma on the existence of a specific class of polynomials (Lemma 4.5), and finally, we prove Lemma 4.6.

Chebyshev Polynomials. We denote by T_t the Chebyshev polynomial of degree t defined recursively as follows

(1) $T_0(x) = 1$,
(2) $T_1(x) = x$,
(3) $T_{t+1}(x) = 2xT_t(x) - T_{t-1}(x)$.

Lemma 4.4 (Folklore). *For a positive integer t, the Chebyshev polynomial T_t satisfies that, for all $\theta \in \mathbb{R}$, we have $T_t(\cos\theta) = \cos t\theta$.*

Proof. We present the proof for completeness. The proof proceeds by induction.

Base case $t = 0, 1$. We have $T_0(\cos\theta) = 1$ and $T_1(\cos\theta) = \cos\theta$, which completes the case.

Induction case $t > 1$. We have

$$\begin{aligned}
T_{t+1}(\cos\theta) &= 2\cos\theta T_t(\cos\theta) - T_{t-1}(\cos\theta) \\
&= 2\cos\theta\cos t\theta - \cos(t-1)\theta \\
&= \cos(t+1)\theta ,
\end{aligned}$$

where in the first equality we use the definition of T_{t+1}, in the second equality we use induction, and in the third equality we use $2\cos x \cos y = \cos(x+y) + \cos(x-y)$, which concludes the induction case and yields the result. □

Lemma 4.5. *For every positive integers N and W, there exists a polynomial F of degree k where $k \geq \left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 4$ such that*

$$F(0) > W \sum_{i=1}^{N} |F(i)| .$$

Proof. We define

$$\mu := \left\lfloor \frac{4}{7} W^{1/4}\sqrt{N} \right\rfloor + 1, \quad g(x) := \frac{1}{2}T_0(x) + \sum_{t=1}^{\mu} T_t(x) ,$$

where T_t denotes the Chebyshev polynomial of degree t. Note that $g(1) = \mu + \frac{1}{2}$, and for $0 < x \leq \pi$, we have

$$\begin{aligned}
g(\cos x) &= \frac{1}{2} + \sum_{t=1}^{\mu} \cos tx \\
&= \Re\left(\sum_{t=0}^{\mu} e^{itx}\right) - \frac{1}{2} \\
&= \Re\left(\frac{e^{i(\mu+1)x} - 1}{e^{ix} - 1}\right) - \frac{1}{2} \\
&= \Re\left(\frac{e^{\frac{i(\mu+1)x}{2}}}{e^{\frac{ix}{2}}} \frac{e^{\frac{i(\mu+1)x}{2}} - e^{-\frac{i(\mu+1)x}{2}}}{e^{\frac{ix}{2}} - e^{-\frac{ix}{2}}}\right) - \frac{1}{2} \\
&= \frac{2\Re\left(e^{\frac{i\mu x}{2}}\right) \sin\frac{(\mu+1)x}{2} - \sin\frac{x}{2}}{2\sin\frac{x}{2}} \\
&= \frac{2\cos\frac{\mu x}{2} \sin\frac{(\mu+1)x}{2} - \sin\frac{x}{2}}{2\sin\frac{x}{2}} \\
&= \frac{\sin\left(\left(\mu + \frac{1}{2}\right)x\right)}{2\sin\frac{x}{2}} \\
&= \frac{\sin\left(\left(\mu + \frac{1}{2}\right)x\right)}{\sqrt{2(1-\cos x)}} \,,
\end{aligned}$$

where in the first equality we use the definition of $g(\cos x)$, in the second equality we use $\cos x = \Re\left(e^{ix}\right)$, in the third equality we use geometric sum, in the fourth equality we use algebraic manipulation, in the fifth equality we use $\sin x = \frac{e^{ix} - e^{-ix}}{2i}$, in the sixth equality we use $\cos x = \Re\left(e^{ix}\right)$, in the seventh equality we use $2\cos x \sin y = \sin(x+y) + \sin(x-y)$, and in the eighth equality we use $\sin^2\frac{x}{2} = \frac{1-\cos x}{2}$. Therefore, for all $x \in (-1, 1]$, we have

$$|g(x)| \leq \frac{1}{\sqrt{2(1-x)}} \,.$$

We define

$$F(x) \coloneqq \left(g\left(1 - \frac{2x}{N}\right)\right)^4, \quad k \coloneqq 4\mu \leq \left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 4 \,.$$

Then, F is a polynomial of degree k. We show that

$$F(0) \geq W \sum_{i=1}^{N} |F(i)| \,.$$

Indeed, we have

$$\begin{aligned}
W \sum_{i=1}^{N} |F(i)| &\leq W \sum_{i=1}^{N} \left(\frac{4i}{N}\right)^{-2} && \left(|g(x)| \leq \frac{1}{\sqrt{2(1-x)}}\right) \\
&= \frac{WN^2}{16} \sum_{i=1}^{N} \frac{1}{i^2} && \text{(rearrange)} \\
&< \frac{\pi^2}{96} WN^2 && \left(\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}\right) \\
&\leq \mu^4 && \left(\mu = \left\lfloor \frac{4}{7} W^{1/4}\sqrt{N} \right\rfloor + 1\right) \\
&< F(0) \,, && \left(F(0) = \left(\mu + \frac{1}{2}\right)^4\right)
\end{aligned}$$

which concludes the proof. □

Lemma 4.6. *Consider a polynomial P of degree at most N with integer coefficients bounded by W. If P has a root of order k at 1, then*

$$k \leq \left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 4 \,.$$

Proof. Let $P(x) = \sum_{i=0}^{N} a_i x^i$. Consider wlog that $a_0 \neq 0$. Indeed, divide P by the monomial x until the first coefficient of the resulting polynomial is not 0. Note that this operator does not change the order of root at 1. We claim that for all polynomials F of degree $(k-1)$, we have

$$\sum_{i=0}^{N} a_i F(i) = 0. \tag{2}$$

Indeed, let $F(x) = \sum_{j=0}^{k-1} b_j x^j$. Then,

$$\sum_{i=0}^{N} a_i F(i) = \sum_{i=0}^{N} a_i \sum_{j=0}^{k-1} b_j i^j = \sum_{j=0}^{k-1} b_j \sum_{i=0}^{N} a_i i^j \,.$$

Therefore, it is enough to show that if $j < k$, then $\sum_{i=0}^{N} a_i i^j = 0$. Consider the operator $\left(x\frac{d\cdot}{dx}\right)$. Observe that

$$x\frac{dP}{dx} = \sum_{i=0}^{N} i a_i x^i \,.$$

Therefore, by applying this operator j times on P, we get a polynomial $Q \coloneqq \sum_{i=0}^{N} i^j a_i x^i$. Note that Q has a root at 1, i.e.,

$$Q(1) = \sum_{i=0}^{N} i^j a_i = 0 \,.$$

Hence, for all polynomials F of degree $(k-1)$, we have

$$\sum_{i=0}^{N} a_i F(i) = 0 \,.$$

For the sake of contradiction, assume P has a root at 1 of order at least

$$\left(\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 5\right) \,.$$

On the one hand, Lemma 4.5 shows that there exists a polynomial F of degree $(k-1)$ such that

$$F(0) > \sum_{i=1}^{N} W|F(i)|\,.$$

On the other hand, we know that $a_0 \neq 0$. By Eqn. (2),

$$\sum_{i=0}^{N} a_i F(i) = 0\,.$$

Therefore, by rearranging and triangle inequality, we get

$$F(0) \leq |a_0||F(0)| \leq \sum_{i=1}^{N} |a_i F(i)| \leq W \sum_{i=1}^{N} |F(i)|\,, \quad (|a_i| \leq W)$$

which contradicts with the property of polynomial F. Therefore, the order of root at 1 of P is at most

$$\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 4\,,$$

which concludes the proof. □

Proof of Theorem 4.2-(1). By Lemma 4.6, we know that P has a root at 1 of order at most

$$\left(\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N} \right\rfloor + 4 \right)\,.$$

By Lemma 4.3, for all roots $\alpha \neq 1$ of P, we have

$$\begin{aligned}|1-\alpha| &> \frac{\left(\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N}\right\rfloor + 5\right)!}{2W(N+1)^{\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N}\right\rfloor + 6}} \qquad \text{(Lemma 4.3)}\\ &> \frac{\left\lfloor \frac{16}{7} W^{1/4}\sqrt{N}\right\rfloor !}{2W(N+1)^{\frac{16}{7} W^{1/4}\sqrt{N} + 6}}\,,\end{aligned}$$

which concludes the proof. □

4.2 Upper bounds

In this section, we show two upper bounds on the roots of polynomials with integer coefficients (Theorem 4.2-(2,3)). Given a polynomial with a root of order k at 1, we construct a new polynomial with a root close to 1 depending on k (Lemma 4.7). For constructive upper bound, we present an explicit polynomial, and for nonconstructive upper bound, we use the existence of a polynomial with a root at 1 of higher order than the previous case.

Lemma 4.7. *Consider a polynomial F with $\{-1, 0, +1\}$ coefficients such that $F(x) = (x-1)^k f(x)$, where k is an integer greater or equal to 9 and f is a polynomial with integer coefficients such that $f(1) \neq 0$. For every positive integer $d \geq (\mathfrak{d}(F)+2)^{3/2}$, there exists a polynomial P of degree $(d(\mathfrak{d}(F)+2)-1)$ with $\{-2,-1,0,+1,+2\}$ coefficients and a root $\alpha < 1$ such that*

$$|1-\alpha| \leq 2 \cdot d^{-(k+2)}\,.$$

Proof. We define polynomial $\widehat{F}(x) := (x-1)F(x)$. We claim that the desired polynomial is given by

$$H(x) := \widehat{F}\left(x^d\right)\left(\frac{1-x^d}{1-x}\right) + F(x)\,.$$

We define

$$N := d(\mathfrak{d}(F)+1) + d - 1\,.$$

Note that H is a polynomial of degree N with $\{-2,-1,0,+1,+2\}$ coefficients and has a root of order k at 1. We show that H has a root $\alpha < 1$ such that

$$|1-\alpha| \leq \frac{2}{d^{k+2}}\,.$$

We take Taylor expansion of H around 1 and get

$$H(x) = (x-1)^k \left[f(1) + (x-1)f^1(1) + (x-1)d^{k+2}f(1) + E(x) \right]\,,$$

where $E(x)$ is the residue polynomial. It is enough to show that H changes its sign in the interval

$$\left(1 - \frac{2}{d^{k+2}}, 1\right)\,,$$

which implies the existence of the desired root. The three key terms are $E(x)$, $f(1)$, and $f^1(1)$. We bound these three terms to deduce the existence of the desired root. Note that

$$E(x) = \sum_{j=k+2}^{N} (x-1)^{j-k} \frac{H^j(1)}{j!}$$

Therefore, for $|x-1| \leq \frac{1}{N+1}$, we have

$$\begin{aligned}|E(x)| &\leq \sum_{j=k+2}^{N} |x-1|^{j-k} \frac{|H^j(1)|}{j!} && \text{(triangle inequality)}\\ &\leq \sum_{j=k+2}^{N} |x-1|^{j-k} \frac{2(N+1)^{j+1}}{j!} && \left(|H^j(1)| \leq 2(N+1)^{j+1}\right)\\ &\leq 2|x-1|^2 (N+1)^{k+3} \sum_{j=k+2}^{N} \frac{1}{j!} && \left(|x-1| \leq \frac{1}{N+1}\right)\\ &\leq \frac{4(N+1)^{k+3}}{(k+2)!}|x-1|^2\,. && \left(\sum_{j=k+2}^{N} \frac{1}{j!} < \frac{2}{(k+2)!}\right)\end{aligned}$$

We denote the sign function by $\text{sign}(x)$ defined by

$$\text{sign}(x) = \begin{cases} -1 & \text{if } x < 0,\\ 0 & \text{if } x = 0,\\ 1 & \text{if } x > 0.\end{cases}$$

Without loss of generality, we consider $f(1) \geq 1$ (recall that f is a polynomial with integer coefficients). We define

$$\beta := 1 - \frac{2}{d^{k+2}}\,.$$

We claim that

$$\text{sign}\left(H(\beta)\right) = -\lim_{x \to 1^-} \text{sign}(H(x))\,.$$

It is enough to show that

$$\begin{aligned}&\lim_{x\to 1^-} \text{sign}\left(f(1) + (x-1)f^1(1) + (x-1)d^{k+2}f(1) + E(x)\right)\\ &\qquad = -\text{sign}\left(f(1) + (\beta-1)f^1(1) + (\beta-1)d^{k+2}f(1) + E(\beta)\right)\,.\end{aligned}$$

Indeed, for the LHS, the term $f(1)$ is dominating. Hence, we get

$$\lim_{x\to 1^-} \text{sign}\left(f(1) + (x-1)f^1(1) + (x-1)d^{k+2}f(1) + E(x)\right) = 1\,. \tag{3}$$

For the RHS, we have

$$\begin{aligned}
&f(1) + (\beta - 1)f^1(1) + (\beta - 1)d^{k+2}f(1) + E(\beta) \\
&\quad = \frac{-2f^1(1)}{d^{k+2}} - f(1) + E(\beta) \\
&\quad \leq \frac{-2f^1(1)}{d^{k+2}} - f(1) + \frac{4(N+1)^{k+3}}{(k+2)!}(1-\beta)^2 \\
&\quad = \frac{-2f^1(1)}{d^{k+2}} - f(1) + \frac{16(N+1)^{k+3}}{(k+2)!\, d^{2k+4}} \\
&\quad \leq \frac{-2f^1(1)}{d^{k+2}} - 1 + \frac{16(N+1)^{k+3}}{(k+2)!\, d^{2k+4}} \\
&\quad \leq \frac{2(\mathfrak{d}(F))^{k+2}}{d^{k+2}(k+1)!} - 1 + \frac{16(N+1)^{k+3}}{(k+2)!\, d^{2k+4}} \\
&\quad \leq \frac{2}{(k+1)!} - 1 + \frac{16(N+1)^{k+3}}{(k+2)!\, d^{2k+4}} \\
&\quad \leq \frac{2}{(k+1)!} - 1 + \frac{16}{(k+2)!} \\
&\quad < 0\,, \qquad (4)
\end{aligned}$$

where in the first equality we use the definition of β, in the first inequality we use

$$E(\beta) \leq \frac{4(N+1)^{k+3}}{(k+2)!}(\beta - 1)^2\,,$$

in the second equality we again use the definition of β, in the second inequality $f(1) \geq 1$, in the third inequality we use

$$|f^1(1)| = \frac{|F^{k+1}(1)|}{(k+1)!} \leq \frac{(\mathfrak{d}(F)^{k+2})}{(k+1)!}\,,$$

in the fourth inequality we use $\mathfrak{d}(F) \leq d$, in the fifth equality we use $(N+1)^{3/5} \leq d$ and $9 \leq k$, and in the sixth inequality we again use $9 \leq k$. Therefore, we have

$$\operatorname{sign}\left(f(1) + (\beta - 1)f^1(1) + (\beta - 1)d^{k+2}f(1) + E(\beta)\right) = -1\,.$$

By combining Eqns. (3) and (4), we get

$$\lim_{x \to 1^-} \operatorname{sign}(H(x)) = -\operatorname{sign}\left(H(\beta)\right)\,.$$

Hence, by continuity of H, there exists a root α of polynomial H such that

$$|\alpha - 1| \leq \frac{2}{d^{k+2}}\,,$$

which concludes the proof. □

Proof of Theorem 4.2-(2). Fix

$$M := \left\lfloor N^{2/5} \right\rfloor - 3, \quad d := \left\lceil (M+2)^{3/2} \right\rceil.$$

We define the polynomial

$$P := x^{M - 2^{\lfloor \log M \rfloor} + 1} \prod_{i=0}^{\lfloor \log M \rfloor - 1} \left(x^{2^i} - 1\right).$$

Note that P is a polynomial of degree M with $\{-1, 0, +1\}$ coefficients and has a root of order $\lfloor \log M \rfloor$ at 1. By Lemma 4.7, there exists a polynomial Q of degree $d(M+2) - 1$ with $\{-2, -1, 0, +1, +2\}$ coefficients and a root $\alpha < 1$ such that

$$\begin{aligned}
1 - \alpha &\leq \frac{2}{d^{\lfloor \log M \rfloor + 2}} \leq \frac{2}{d^{\lfloor \log M \rfloor}} && \text{(Lemma 4.7)} \\
&\leq \frac{2}{(M+2)^{\frac{3}{2}\lfloor \log M \rfloor}} && \left(d = \left\lceil (M+2)^{3/2} \right\rceil\right) \\
&\leq \frac{2}{\left(\left\lfloor N^{2/5} \right\rfloor - 1\right)^{\frac{3}{2}\left\lfloor \log\left(\lfloor N^{2/5} \rfloor - 3\right)\right\rfloor}}\,, && \left(M = \left\lfloor N^{2/5} \right\rfloor - 3\right)
\end{aligned}$$

which yields the result. □

For the nonconstructive upper bound, we recall a result of [5] on the existence of polynomials with high order roots at 1.

Lemma 4.8 ([5, Theorem 2.7]). *For a positive integer $N \geq 9$, there exists a polynomial P of degree N with $\{-1, 0, +1\}$ coefficients such that P has a root at 1 of order*

$$\left(\left\lfloor \sqrt{\frac{N}{\log N}} \right\rfloor - 2\right).$$

Proof. We proceed with the proof by pigeonhole principle. Let A_N be the set of all polynomials of degree at most N with $\{0, 1\}$ coefficients. Given a positive integer k, for all polynomials $Q \in A_N$, we define the mapping $Q \mapsto \left(Q(1), Q^1(1), \ldots, Q^{k-1}(1)\right)$, where $Q^j(1)$ is the j-th derivative of Q at 1. On one hand, the number of different outputs of the mapping is at most

$$\begin{aligned}
&\prod_{j=0}^{k-1} (N+1)^{j+1} && \text{(Eqn. (1))} \\
&= (N+1)^{k(k+1)/2}\,.
\end{aligned}$$

On the other hand, the number of polynomials in A_N, i.e., $|A_N|$ is 2^{N+1}. Therefore, if

$$(N+1)^{k(k+1)/2} < 2^{N+1}\,,$$

then there exist two polynomials $Q_1, Q_2 \in A_N$ such that for all $0 \leq j < k$, we have $Q_1^j(1) = Q_2^j(1)$, which implies that the polynomial $Q_1 - Q_2$ has a root at 1 of order at least k. Observe that the coefficients of the polynomial $Q_1 - Q_2$ belong to the set $\{-1, 0, +1\}$. By setting $k = \left(\left\lfloor \sqrt{\frac{N}{\log N}} \right\rfloor - 2\right)$, we obtain $(N+1)^{k(k+1)/2} < 2^{N+1)}$, which completes the proof. □

Proof of Theorem 4.2-(3). Fix

$$M := \left\lfloor N^{2/5} \right\rfloor - 3, \quad d := \left\lceil (M+2)^{3/2} \right\rceil$$

By Lemma 4.8, there exists a polynomial P of degree M with $\{-1, 0, +1\}$ coefficients and a root at 1 of order

$$\left(\left\lfloor \sqrt{\frac{M}{\log M}} \right\rfloor - 2\right).$$

By Lemma 4.7, there exists a polynomial Q of degree $d(M+2)-1$ with $\{-2,-1,0,+1,+2\}$ coefficients and a root $\alpha<1$ such that

$$\begin{aligned}
1-\alpha &\le \frac{2}{d^{\left\lfloor\sqrt{\frac{M}{\log M}}\right\rfloor}} && \text{(Lemma 4.7)}\\
&\le \frac{2}{(M+2)^{\frac{3}{2}\left\lfloor\sqrt{\frac{M}{\log M}}\right\rfloor}} && \left(d=\left\lceil (M+2)^{3/2}\right\rceil\right)\\
&\le \frac{2}{\left(\lfloor N^{2/5}\rfloor-1\right)^{\frac{3}{2}\left\lfloor\sqrt{\frac{\lfloor N^{2/5}\rfloor-3}{\log(N^{2/5})}}\right\rfloor}}, && \left(M=\left\lfloor N^{2/5}\right\rfloor-3\right)
\end{aligned}$$

which concludes the proof. □

5 IMPROVED ANALYSIS OF STRATEGY ITERATION ALGORITHM FOR DISCVAL

In this section, we first define the basic notions related to the strategy iteration algorithm (Section 5.1), then present the procedure SI (Section 5.2), and finally, we analyze the time complexity of the algorithm (Section 5.3).

5.1 Basic Notions

Best-response to a strategy. Given a discounted-payoff game G with discount factor λ and player-1 strategy σ, the *best-response* to σ for player 2 is a strategy τ such that for all strategies $\tau'\in\Gamma^P$ and all vertices v, we have

$$\mathrm{Disc}_\lambda\left(G_v^{\sigma,\tau}\right)\le \mathrm{Disc}_\lambda\left(G_v^{\sigma,\tau'}\right).$$

The existence of the best-response strategies follows from Theorem 2.1.

Bellman strategy extractor. Given a discounted-payoff game G with discount factor λ and a function $f\colon V\to\mathbb{R}$, Bellman strategy extractor is defined as follows:

$$\mathcal{B}_{G,\lambda}(f)(v)=\begin{cases}\arg\max_{u\in E(v)} r(v,u)+\lambda f(u), & \text{if } v\in V_1\\ \arg\min_{u\in E(v)} r(v,u)+\lambda f(u), & \text{if } v\in V_2\end{cases}$$

We assume that the ties are resolved independently of discount factor, e.g., given a fixed indexing of vertices, choosing the vertex with the least index.

Polynomials for a strategy profile. Given a discounted-payoff game, a vertex v, and a strategy profile (σ,τ), we define a pair of polynomials (P,Q) with integer coefficients such that $\frac{P(\lambda)}{Q(\lambda)}$ is the discounted value of play $G_v^{\sigma,\tau}$ when the discount factor is λ. Given strategies σ and τ, the lasso-shaped play $G_v^{\sigma,\tau}$ consists in a simple path $\mathcal{P}:=\langle v_0,\dots,v_{p-1}\rangle$ and a cycle $C:=\langle v_p,\dots,v_{p+c-1}\rangle$ repeated forever. Then,

$$\begin{aligned}
\mathrm{Disc}_\lambda(G_v^{\sigma,\tau}) &= \sum_{i=0}^{\infty}\lambda^i r(v_i,v_{i+1})\\
&= \sum_{i=0}^{p-1}\lambda^i r(v_i,v_{i+1})+\sum_{i=p}^{\infty}\lambda^i r(v_i,v_{i+1})\\
&= \sum_{i=0}^{p-1}\lambda^i r(v_i,v_{i+1})+\frac{\sum_{i=0}^{c-1}\lambda^{p+i} r(v_{p+i},v_{p+i+1})}{1-\lambda^c}\\
&= \frac{(1-\lambda^c)\sum_{i=0}^{p-1}\lambda^i r(v_i,v_{i+1})+\sum_{i=0}^{c-1}\lambda^{p+i} r(v_{p+i},v_{p+i+1})}{1-\lambda^c}\\
&=: \frac{P(\lambda)}{Q(\lambda)},
\end{aligned}$$

where in the first equality we use the definition of Disc_λ, in the second equality we partition the play into the simple path and the cycle, in the the third equality we use geometric sum, and in the the fourth we use algebraic manipulation. Notice that the coefficients of P and Q are integers and bounded by $3W$. The degrees of P and Q are at most n.

Example 5.1. Figure 2 illustrates a turn-based game G with three vertices: two player-1 vertices a,b and one player-2 vertex c. The directed edges among them represent possible actions with associated weights: 1 from a to b, 3 from b to c, and -2 from c to b. The initial vertex is a. Since each vertex has one possible action, there exist a positional strategy σ for player 1 and a positional strategy τ for player 2. Given the strategy profile (σ,τ), the lasso-shaped play $G_a^{\sigma,\tau}=\langle a,b,c,b,c,\cdots\rangle$ consists in the simple $\mathcal{P}=\langle a\rangle$ and the cycle $C=\langle b,c\rangle$ repeated forever. For all λ, we have

$$\mathrm{Disc}_\lambda(G_a^{\sigma,\tau})=1+\frac{3\lambda-2\lambda^2}{1-\lambda^2}=\frac{1+3\lambda-3\lambda^2}{1-\lambda^2}=\frac{P(\lambda)}{Q(\lambda)}.$$

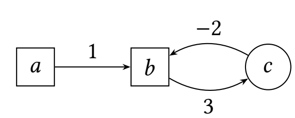


Figure 2: A turn-based game.

5.2 Strategy Iteration Algorithm

Strategy iteration for turn-based games computes the optimal strategy for player 1. It starts with an arbitrary strategy σ^0. Then, at each iteration, it locally improves player-1 strategy. Starting with strategy σ^k at iteration k, it computes the best-response to σ^k for player 2, called τ^k. This computation can be done in time $O(mn^2\log m)$ by linear programming [1]. Then, it improves the strategy σ^k by greedy local improvements using Bellman strategy extraction $\mathcal{B}_{G,\lambda}$. This procedure is guaranteed to reach a fixed point, which is an optimal strategy for player 1. The formal description is shown in Algorithm 1.

Algorithm 1 Strategy Iteration

Input: Game G, discount factor λ, strategy σ^0 for player 1
Output: Optimal strategy σ^* for player 1

```
procedure SI(G, λ, σ^0)
    k ← 0
    repeat
        τ^k ← the best response to σ^k
        σ^{k+1} ← 𝓑_{G,λ} (Disc_λ (G^{σ^k, τ^k}))
        k ← k + 1
    until σ^k = σ^{k−1}
    return σ^k
```

5.3 Improved Time Complexity Analysis

In this section, we analyze the time complexity of SI algorithm (Theorem 3.2). We show that, given a turn-based game G, there exists a discount factor λ_0 such that, for all discount factors $\lambda \in [\lambda_0, 1)$, if we start SI from the same initial strategy for both λ_0 and λ, SI generates the same sequence of strategies (Corollary 5.4). This result, alongside the time complexity of SI in general case (Theorem 3.1), yields an improved complexity of SI. To prove Corollary 5.4, we show that if a strategy profile outperforms another when the discount factor is λ_0, then it also outperforms when the discount factor is $\lambda \in [\lambda_0, 1)$ (Lemma 5.3). Example 5.2 illustrates that the performance of two strategy profiles can be compared by their associated rationals.

Example 5.2. Figure 3 illustrates a turn-based game G with nine vertices, where all vertices are player-1 vertices. The directed edges among them represent possible actions with associated weights. The initial vertex is a. There are two positional strategies for player 1: σ_1 (going left from a) and σ_2 (going right from a). Since player 2 does not have any vertices, there is one positional strategy τ. For all discount factors λ, we have

$$\mathrm{Disc}_\lambda\left(G_a^{\sigma_1,\tau}\right) = \frac{\lambda + 2\lambda^4}{1-\lambda^4}\,, \quad \mathrm{Disc}_\lambda\left(G_a^{\sigma_2,\tau}\right) = \frac{2\lambda^2 + \lambda^3}{1-\lambda^4}\,.$$

The difference between the discounted payoff of σ_1 and σ_2 is

$$\begin{aligned}\mathrm{Disc}_\lambda\left(G_a^{\sigma_1,\tau}\right) - \mathrm{Disc}_\lambda\left(G_a^{\sigma_2,\tau}\right) &= \frac{\lambda - 2\lambda^2 - \lambda^3 + 2\lambda^4}{1-\lambda^4} \\ &= \frac{\lambda(1+\lambda)(1-\lambda)(1-2\lambda)}{1-\lambda^4}\,.\end{aligned}$$

As λ varies in $(0,1)$, the performance of two strategies is compared as follows.

- If $\lambda \in (0, 1/2)$, then σ_1 outperforms σ_2.
- If $\lambda = 1/2$, then the performance of σ_1 and σ_2 are the same.
- If $\lambda \in (1/2, 1)$, then σ_2 outperforms σ_1.

Lemma 5.3. *Consider a discounted-payoff game G, a vertex v, and two strategy profiles (σ_1, τ_1) and (σ_2, τ_2). Fix discount factor*

$$\lambda_0 \coloneqq 1 - \left(24W(2n+1)^{7W^{1/4}\sqrt{n}+6}\right)^{-1}\,.$$

If $\mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma_1,\tau_1}\right) \geq \mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma_2,\tau_2}\right)$, then for all $\lambda \in [\lambda_0, 1)$, we have

$$\mathrm{Disc}_\lambda\left(G_v^{\sigma_1,\tau_1}\right) \geq \mathrm{Disc}_\lambda\left(G_v^{\sigma_2,\tau_2}\right)\,.$$

Proof. Let (P_1, Q_1) (resp. (P_2, Q_2)) be the polynomials corresponding to a strategy profile (σ_1, τ_1) (resp. (σ_2, τ_2)), and a vertex v. We know that $\mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma_1,\tau_1}\right) \geq \mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma_2,\tau_2}\right)$. Therefore, we have

$$\frac{P_1(\lambda_0)}{Q_1(\lambda_0)} \geq \frac{P_2(\lambda_0)}{Q_2(\lambda_0)}\,,$$

or equivalently,

$$(P_1Q_2 - P_2Q_1)(\lambda_0) \geq 0\,.$$

We define

$$F \coloneqq P_1Q_2 - P_2Q_1\,.$$

Note P_1, Q_1, P_2, and Q_2 are polynomials of degree n with integer coefficients bounded by $3W$, and Q_1 and Q_2 are of the form $1-\lambda^{c_1}$ and $1-\lambda^{c_2}$ where c_1 and c_2 are the length of the cycles in the lasso-shaped plays given the strategy profiles. Therefore, polynomial F is of degree at most $2n$, and its coefficients are integer and are bounded by $12W$. If F is identically 0, then for all $\lambda \in [0,1)$, we have $\mathrm{Disc}_\lambda(G_v^{\sigma_1,\tau_1}) = \mathrm{Disc}(G_v^{\sigma_2,\tau_2})$. Otherwise, by Theorem 4.2-(1), for all roots $\lambda < 1$ of F, we have

$$\begin{aligned}1-\lambda &\geq \frac{\left\lfloor \frac{16}{7}(12W)^{1/4}\sqrt{2n}\right\rfloor!}{24W(2n+1)^{\frac{16}{7}(12W)^{1/4}\sqrt{2n}+6}} && \text{(Theorem 4.2-(1))} \\ &> \frac{1}{24W(2n+1)^{\frac{16}{7}(12W)^{1/4}\sqrt{2n}+6}} \\ &> \frac{1}{24W(2n+1)^{7W^{1/4}\sqrt{n}+6}}\,, && \left(\frac{16}{7}12^{1/4}\sqrt{2} < 7\right)\end{aligned}$$

which implies that F does not have any roots in the interval $[\lambda_0, 1)$. Therefore, by continuity of F, for all discount factors $\lambda \in [\lambda_0, 1)$, we have

$$(P_1Q_2 - P_2Q_1)(\lambda) = F(\lambda) \geq 0\,,$$

which implies that

$$\left(\frac{P_1}{Q_1}\right)(\lambda) \geq \left(\frac{P_2}{Q_2}\right)(\lambda)\,.$$

Hence, we have

$$\mathrm{Disc}_\lambda\left(G_v^{\sigma_1,\tau_1}\right) \geq \mathrm{Disc}_\lambda\left(G_v^{\sigma_2,\tau_2}\right)\,,$$

which concludes the proof. □

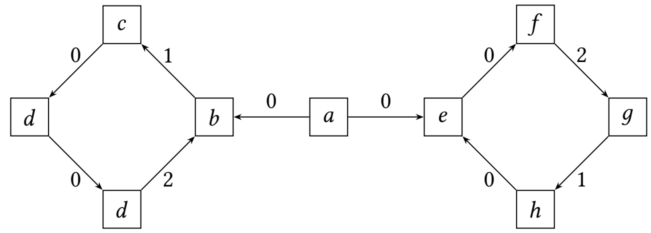


Figure 3: A turn-based game with two cycles.

Remark 5.1. If there is equality, by symmetry, one concludes that the functions of the respective profiles are equal in the interval $[\lambda_0, 1)$.

The above result shows that for any vertex v, the ordering of strategy profiles is the same for both discount factors λ_0 and λ. Lemma 5.3 yields that, if SI starts from the same initial strategy for both λ_0 and λ, then SI outputs the same optimal strategies.

Corollary 5.4. *Fix discount factor*

$$\lambda_0 := 1 - \left(24W(2n+1)^{7W^{1/4}\sqrt{n}+6}\right)^{-1} .$$

Consider a discounted-payoff game G with discount factor $\lambda \in [\lambda_0, 1)$, and player-1 strategy σ^0. Let σ^k and $\overline{\sigma}^k$ be the strategy for player 1 after k iterations of $\mathrm{SI}(G, \lambda_0, \sigma^0)$ and $\mathrm{SI}(G, \lambda, \sigma^0)$, respectively. Then, for all $k \geq 0$, we have $\sigma^k = \overline{\sigma}^k$. In particular,

$$\mathrm{SI}(G, \lambda_0, \sigma^0) = \mathrm{SI}(G, \lambda, \sigma^0) .$$

Proof. We proceed with the proof by induction on the number of iterations.

Base case $k = 0$. Since the initial strategy for both procedure calls is the same, then we have $\sigma^0 = \overline{\sigma}^0$, which completes the case.

Induction case $k > 0$. We claim that $\tau^k = \overline{\tau}^k$, where τ^k and $\overline{\tau}^k$ are the best responses to σ^k and $\overline{\sigma}^k$ for player 2 with respect to λ and λ_0, respectively. Indeed, since τ^k is the best response to σ^k for player 2, then, for all vertices v, we have

$$\mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma^k,\tau^k}\right) \leq \mathrm{Disc}_{\lambda_0}\left(G_v^{\sigma^k,\overline{\tau}^k}\right) .$$

by Lemma 5.3, we have

$$\mathrm{Disc}_{\lambda}\left(G_v^{\sigma^k,\tau^k}\right) \leq \mathrm{Disc}_{\lambda}\left(G_v^{\sigma^k,\overline{\tau}^k}\right)$$

The inequality also holds in the other direction since $\overline{\tau}^k$ is the best response to $\overline{\sigma}^k$. Since ties are broken independently of the discount factor, we have that $\tau^k = \overline{\tau}^k$. Similarly, we can show that $\sigma^{k+1} = \overline{\sigma}^{k+1}$, which concludes the induction case and yields the result. □

Remark 5.2.

Proof of Theorem 3.2. Fix

$$\lambda_0 := 1 - \left(24W(2n+1)^{7W^{1/4}\sqrt{n}+6}\right)^{-1} .$$

We proceed with a proof by cases on the size of λ.

- Case $\lambda \leq \lambda_0$. By Theorem 3.1, the procedure SI terminates after $n^{O\left(W^{1/4}\sqrt{n}\right)}$ iterations, which completes the case.
- Case $\lambda > \lambda_0$. By Corollary 5.4, the procedure SI for discount factors λ and λ_0 terminates after the same number of iterations. Therefore, it terminates after $n^{O\left(W^{1/4}\sqrt{n}\right)}$ iterations, which concludes the case and yields the result.

□

6 DISCUSSION AND CONCLUSION

We discuss the novelty of our work and future directions. The novelty of our work is not a new algorithm, but a new improved analysis of a classical and simple algorithm (SI algorithm) that has been widely studied. Moreover, most results in the literature focus on establishing lower bounds for the SI algorithm [19, 21, 22]; in contrast we focus on a better upper bound. The novel aspects of our analysis are as follows: (a) establishing the connection of analysis of SI with a problem about lower bounds on roots of a class of polynomials, which is the key insight; and (b) establishing lower and upper bounds on roots of the required class of polynomials. While our analysis only requires the lower bounds on roots, the significance of the upper bounds on roots is to show that our technique does not yield a polynomial-time bound without further non-trivial insights. Our result shows that for the DiscVal-Wun problem we have a deterministic sub-exponential time algorithm. Discounted-sum games lie in between mean-payoff games and stochastic games with reachability objectives, i.e., linear-time reduction exists from mean-payoff to discounted-sum games and from discounted-sum games to stochastic games, but no reductions are known for the converse direction. Whether discounted-sum games are more similar to mean-payoff games or stochastic games is an intriguing question. When the weights are expressed in unary, mean-payoff games admit polynomial-time algorithm, whereas stochastic games with rewards 0 and 1, and all probabilities are half are as hard as general stochastic games. Thus deterministic sub-exponential time bound for unary stochastic games is a major open question. Our result shows the difference of unary discounted-sum games compared to unary stochastic games by establishing deterministic sub-exponential time upper bound. Whether there is a polynomial-time algorithm for unary discounted-sum games or there is a deterministic sub-exponential time algorithm for unary stochastic games are interesting questions for future work.

Acknowledgements. This research was partially supported by the ERC CoG 863818 (ForM-SMArt) grant.

REFERENCES

[1] Andersson, D. Improved combinatorial algorithms for discounted payoff games. Master's thesis, Uppsala University, 2006.
[2] Andersson, D., and Miltersen, P. B. The complexity of solving stochastic games on graphs. In *International Symposium on Algorithms and Computation* (2009), Springer, pp. 112–121.
[3] Baez, J. The beauty of roots, 2023. https://math.ucr.edu/home/baez/roots/.
[4] Bansal, S., Chatterjee, K., and Vardi, M. Y. On satisficing in quantitative games. *Tools and Algorithms for the Construction and Analysis of Systems 12651*, 20.
[5] Borwein, P., Erdélyi, T., and Kós, G. Littlewood-type problems on [0, 1]. *Proceedings of the London Mathematical Society 79*, 1 (1999), 22–46.
[6] Brim, L., Chaloupka, J., Doyen, L., Gentilini, R., and Raskin, J. Faster algorithms for mean-payoff games. *Formal Methods Syst. Des. 38*, 2 (2011), 97–118.
[7] Calude, C. S., Jain, S., Khoussainov, B., Li, W., and Stephan, F. Deciding parity games in quasi-polynomial time. *SIAM J. Comput. 51*, 2 (2022), 17–152.
[8] Chandra, A. K., Kozen, D. C., and Stockmeyer, L. J. Alternation. *Journal of the ACM (JACM) 28*, 1 (1981), 114–133.
[9] Chatterjee, K., and Fijalkow, N. A reduction from parity games to simple stochastic games. In *GandALF* (2011), pp. 74–86.
[10] Chatterjee, K., Ghosal, A., Henzinger, T. A., Iercan, D., Kirsch, C. M., Pinello, C., and Sangiovanni-Vincentelli, A. Logical reliability of interacting real-time tasks. In *Proceedings of the conference on Design, automation and test in Europe* (2008), pp. 909–914.
[11] Chatterjee, K., and Henzinger, M. Efficient and dynamic algorithms for alternating büchi games and maximal end-component decomposition. *Journal of the ACM (JACM) 61*, 3 (2014), 1–40.
[12] Chatterjee, K., Pavlogiannis, A., and Velner, Y. Quantitative interprocedural analysis. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (2015), pp. 539–551.
[13] Condon, A. On Algorithms for Simple Stochastic Games. *Advances in computational complexity theory* (1990).
[14] Condon, A. The complexity of stochastic games. *Information and Computation 96*, 2 (1992), 203–224.
[15] De Alfaro, L., Henzinger, T. A., and Majumdar, R. Discounting the future in systems theory. In *International Colloquium on Automata, Languages, and Programming* (2003), Springer, pp. 1022–1037.
[16] Dorfman, D., Kaplan, H., and Zwick, U. A faster deterministic exponential time algorithm for energy games and mean payoff games. In *ICALP* (2019).
[17] Ehrenfeucht, A., and Mycielski, J. Positional strategies for mean payoff games. *Int. Journal of Game Theory 8*, 2 (1979), 109–113.
[18] Emerson, E., and Jutla, C. Tree automata, mu-calculus and determinacy. In *FOCS* (1991).

[19] Fearnley, J., and Savani, R. The complexity of the simplex method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing* (2015), pp. 201–208.

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

[21] Friedman, O. A super-polynomial lower bound for the parity game strategy improvement algorithm as we know it. *Logic in Computer Science (LICS). IEEE, Los Alamitos (to appear, 2009)* (2009).

[22] Friedmann, O., Hansen, T. D., and Zwick, U. Subexponential lower bounds for randomized pivoting rules for the simplex algorithm. In *Proceedings of the forty-third annual ACM symposium on Theory of computing* (2011), pp. 283–292.

[23] Gurevich, Y., and Harrington, L. Trees, automata, and games. In *STOC'82* (1982), ACM Press, pp. 60–65.

[24] Gurvich, V. A., Karzanov, A. V., and Khachiyan, L. G. Cyclic games and an algorithm to find minimax cycle means in directed graphs. *USSR Comput. Math. Math. Phys. 28*, 5 (1990), 85–91.

[25] Hansen, T. D., Miltersen, P. B., and Zwick, U. Strategy iteration is strongly polynomial for 2-player turn-based stochastic games with a constant discount factor. *JACM 60*, 1 (2013), 1–16.

[26] Jurdzinski, M. Deciding the winner in parity games is in UP $\cap$ co-UP. *Information Processing Letters 68*, 3 (1998), 119–124.

[27] Jurdzinski, M. Small progress measures for solving parity games. In *STACS* (2000), pp. 290–301.

[28] Jurdzinski, M., Paterson, M., and Zwick, U. A deterministic subexponential algorithm for solving parity games. *SIAM J. Comput. 38*, 4 (2008), 1519–1532.

[29] Kozachinskiy, A. Polyhedral value iteration for discounted games and energy games. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2021), SIAM, pp. 600–616.

[30] Ludwig, W. A subexponential randomized algorithm for the simple stochastic game problem. *Information and Computation 117* (1995), 151–155.

[31] Puterman, M. L. *Markov Decision Processes.* John Wiley and Sons, 1994.

[32] Schewe, S. Solving parity games in big steps. In *FSTTCS 2007* (2007), pp. 449–460.

[33] Shapley, L. S. Stochastic Games. *Proceedings of the National Academy of Sciences 39*, 10 (1953), 1095–1100.

[34] Zielonka, W. Infinite games on finitely coloured graphs with applications to automata on infinite trees. In *Theoretical Computer Science* (1998), vol. 200(1-2), pp. 135–183.

[35] Zwick, U., and Paterson, M. The complexity of mean payoff games on graphs. *Theoretical Computer Science 158*, 1-2 (1996), 343–359.