Universal Complexity Bounds Based on Value Iteration and Application to Entropy Games

Xavier Allamigeon
INRIA, Palaiseau, France
CMAP, École polytechnique, IP Paris, CNRS, Palaiseau, France

Stéphane Gaubert
INRIA, Palaiseau, France
CMAP, École polytechnique, IP Paris, CNRS, Palaiseau, France

Ricardo D. Katz
CIFASIS-CONICET, Rosario, Argentina

Mateusz Skomra
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

Abstract

We develop value iteration-based algorithms to solve in a unified manner different classes of combinatorial zero-sum games with mean-payoff type rewards. These algorithms rely on an oracle, evaluating the dynamic programming operator up to a given precision. We show that the number of calls to the oracle needed to determine exact optimal (positional) strategies is, up to a factor polynomial in the dimension, of order R/sep, where the "separation" sep is defined as the minimal difference between distinct values arising from strategies, and R is a metric estimate, involving the norm of approximate sub and super-eigenvectors of the dynamic programming operator. We illustrate this method by two applications. The first one is a new proof, leading to improved complexity estimates, of a theorem of Boros, Elbassioni, Gurvich and Makino, showing that turn-based mean payoff games with a fixed number of random positions can be solved in pseudo-polynomial time. The second one concerns entropy games, a model introduced by Asarin, Cervelle, Degorre, Dima, Horn and Kozyakin. The *rank* of an entropy game is defined as the maximal rank among all the ambiguity matrices determined by strategies of the two players. We show that entropy games with a fixed rank, in their original formulation, can be solved in polynomial time, and that an extension of entropy games incorporating weights can be solved in pseudo-polynomial time under the same fixed rank condition.

2012 ACM Subject Classification Theory of computation → Algorithmic game theory

Keywords and phrases Mean-payoff games, entropy games, value iteration, Perron root, separation bounds, parameterized complexity

Digital Object Identifier 10.4230/LIPIcs.ICALP.2022.110

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version *Full Version*: https://arxiv.org/abs/2206.09044

Acknowledgements We thank the reviewers for detailed and helpful comments.

1 Introduction

1.1 Motivation

Deterministic and turn-based stochastic Mean Payoff games are fundamental classes of games with an unsettled complexity. They belong to the complexity class NP ∩ coNP [21, 47] but they are not known to be polynomial-time solvable. Various algorithms have been developed and analyzed. The pumping algorithm is a pseudo-polynomial iterative scheme introduced by

49th International Colloquium on Automata, Languages, and Programming (ICALP 2022).
Editors: Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff;
Article No. 110; pp. 110:1–110:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Gurvich, Karzanov and Khachyan [25] to solve the optimality equation of deterministic mean payoff games. Zwick and Paterson [47] derived peudo-polynomial bounds for the same games by analyzing value iteration. Friedmann showed that policy iteration, originally introduced by Hoffman and Karp in the setting of zero-sum games [26], and albeit being experimentally fast on typical instances, is generally exponential [22]. We refer to the survey of [9] for more information and additional references.

Entropy games have been introduced by Asarin et al. [10]. They are combinatorial games, in which one player, called Tribune, wants to maximize a topological entropy, whereas its opponent, called Despot, wishes to minimize it. This topological entropy quantifies the freedom of a half-player, called People. Although the formalization of entropy game is recent, specific classes or variants of entropy games appeared earlier in several fields, including the control of branching processes, population dynamics and growth maximization [42, 38, 37, 46], risk sensitive control [27, 8], mathematical finance [5], or matrix multiplication games [10]. Asarin et al. showed that entropy games also belong to the class NP $\cap$ coNP. Akian et al. showed in [1] that entropy games reduce to ordinary stochastic mean payoff games with infinite action spaces (actions consist of probability measures and the payments are given by relative entropies), and deduced that the subclass of entropy games in which Despot has a fixed number of significant positions (positions with a non-trivial choice) can be solved in polynomial time. The complexity of entropy games without restrictions on the number of (significant) Despot positions is an open problem.

1.2 Main Results

We develop value iteration-based algorithms to solve in a unified manner different classes of combinatorial zero-sum games with mean-payoff type rewards. These algorithms rely on an oracle, evaluating approximately the dynamic programming operator of the game. Our main results include universal estimates, providing explicit bounds for the error of approximation of the value, as a function of two characteristic quantities, of a metric nature. The first one is the *separation* sep, defined as the minimal difference between distinct values induced by (positional) strategies. The second one, R, is defined in terms the norm of approximate sub and super-optimality certificates. These certificates are vectors, defined as sub or super-solutions of non-linear eigenproblems. For games such that the mean payoff is independent of the initial state, we show that (exact) optimal strategies can be found in a number of calls to the oracle bounded by the ratio R/sep, up to a factor polynomial in the number of states, see Theorems 10 and 18. We also obtain a similar complexity bound for games in which the mean payoff does depend on the initial state, under additional assumptions.

We provide two applications of this method.

The first application is a new proof of an essential part of the theorem of Boros, Elbassioni, Gurvich and Makino [16], showing that turn-based stochastic mean payoff games with a fixed number of random positions can be solved in pseudo-polynomial time. The original proof relies on a deep analysis of a generalization to the stochastic case of the "pumping algorithm" of [25]. Our analysis of value iteration leads to improved complexity estimates. Indeed, we bound the characteristic numbers R and sep in a tight way, by exploiting bit-complexity estimates for the solutions of Fokker–Planck and Poisson-type equations of discrete Markov chains.

The second application concerns entropy games. Let us recall that in such a game, the value of a pair of (positional) strategies of the two players is given by the Perron root of a certain principal submatrix of a nonnegative matrix, which we call the *ambiguity matrix*, as it measures the number of nondeterministic choices of People. We show that entropy games with a fixed rank, and in particular, entropy games with a fixed number of People's states, can be solved in pseudo-polynomial time; see Corollary 32. These results concern the extended model of entropy games introduced in [1], taking into account weights. Then, entropy games in the sense of [10] (implying a unary encoding of weights) that have a fixed rank can be solved in polynomial time. These results rely on separation bounds for algebraic numbers arising as the eigenvalues of integer matrices with a fixed rank.

1.3 Related Work

The idea of applying value iteration to analyze the complexity of deterministic mean-payoff games goes back to the classical work of Zwick and Paterson [47]. In some sense, the present approach extends this idea to more general classes of games. When specialized to stochastic mean payoff games with perfect information, our bounds should be compared with the ones of Boros, Elbassioni, Gurvich, and Makino [16, 15]. The authors of [16] generalize the "pumping" algorithm, developed for deterministic games by Gurvich, Karzanov, and Khachiyan [25], to the case of stochastic games. The resulting algorithm is also pseudopolynomial if the number of random positions is fixed, see Remark 26 for a detailed comparison. The algorithm of Ibsen-Jensen and Miltersen [28] yields a stronger bound in the case of simple stochastic games, still assuming that the number of random positions is fixed. A different approach, based on an analysis of strategy iteration, was developed by Gimbert and Horn [24] and more recently by Auger, Badin de Montjoye and Strozecki [11]. The value iteration algorithm for concurrent mean payoff games, under an ergodicity condition, has been studied by Chatterjee and Ibsen-Jensen [19]. Theorem 18 there bounds the number of iterations needed to get an ϵ-approximation of the mean-payoff. When specialized to this case, Theorem 13 below improves this bound by a factor of $|\log \epsilon|$.

We build on the operator approach for zero-sum games, see [13, 34, 36]. Our study of entropy games is inspired by the works of Asarin et al. [10] and Akian et al. [1]. We rely on the existence of optimal positional strategies for entropy games, established in [1] by an o-minimal geometry approach [14] and also, on results of non-linear Perron–Frobenius theory, especially the Collatz–Wielandt variational formulation of the escape rate of an order preserving and additively homogeneous mapping [35, 23, 2, 4].

The present work, providing complexity bounds based on value iteration, grew out from an effort to understand the surprising speed of value iteration on random stochastic games examples arising from tropical geometry [7], by investigating suitable notions of condition numbers [6]. An initial version of some of the present results (concerning turn based stochastic games) appeared in the PhD thesis of one of the authors [40].

1.4 Organization of the Paper

In Section 2 we recall the definitions and basic properties of turn-based stochastic mean payoff games and entropy games, and also key notions in the "operator approach" of zero-sum games, including the Collatz–Wielandt optimality certificates.

The universal complexity bounds based on value iteration are presented in Section 3. First, we deal with games whose value is independent of the initial state, and then, we extend these results to determine the set of initial states with a maximal value.

The applications to turn-based stochastic mean payoff games and to entropy games are provided in Section 4 and Section 5. The detailed proofs can be found in the extended version of the present paper.

2 Preliminaries on Dynamic Programming Operators and Games

2.1 Introducing Shapley Operators: The Example of Stochastic Turn-Based Zero-Sum Games

Shapley operators are the two-player version of the Bellman operators (a.k.a. dynamic programming or one-day operators) which are classically used to study Markov decision processes. In this section, we introduce the simplest example of Shapley operator, arising from stochastic turn-based zero-sum games.

A *stochastic turn-based zero-sum game* is a game played on a digraph $(\mathscr{V}, \mathscr{E})$ in which the set of vertices $\mathscr{V}$ has a non-trivial partition $\mathscr{V} = \mathscr{V}_{\text{Min}} \uplus \mathscr{V}_{\text{Max}} \uplus \mathscr{V}_{\text{Nat}}$. There are two players, called *Min* and *Max*, and a half-player, *Nature*. The sets $\mathscr{V}_{\text{Min}}$, $\mathscr{V}_{\text{Max}}$ and $\mathscr{V}_{\text{Nat}}$ represent the sets of states at which Min, Max, and Nature respectively play. The set of edges $\mathscr{E}$ represents the allowed moves. We assume $\mathscr{E} \subset \mathscr{V}_{\text{Min}} \times \mathscr{V}_{\text{Max}} \cup \mathscr{V}_{\text{Max}} \times \mathscr{V}_{\text{Nat}} \cup \mathscr{V}_{\text{Nat}} \times \mathscr{V}_{\text{Min}}$, meaning that Min, Max, and Nature alternate their moves. More precisely, a turn consists of three successive moves: when the current state is $j \in \mathscr{V}_{\text{Min}}$, Min selects and edge (j, i) in $\mathscr{E}$ and the next state is $i \in \mathscr{V}_{\text{Max}}$. Then, Max selects an edge (i, k) in $\mathscr{E}$ and the next state is $k \in \mathscr{V}_{\text{Nat}}$. Next, Nature chooses an edge $(k, j') \in \mathscr{E}$ and the next state is $j' \in \mathscr{V}_{\text{Min}}$. This process can be repeated, alternating moves of Min, Max, and Nature.

We make the following assumption.

▶ **Assumption 1.** *Each player has at least one available action in each state in which he has to play, i.e., for all $j \in \mathscr{V}_{\text{Min}}, i \in \mathscr{V}_{\text{Max}}$, and $k \in \mathscr{V}_{\text{Nat}}$, the sets $\{i' \colon (j, i') \in \mathscr{E}\}$, $\{k' \colon (i, k') \in \mathscr{E}\}$ and $\{j' \colon (k, j') \in \mathscr{E}\}$ are non-empty.*

Furthermore, every state $k \in \mathscr{V}_{\text{Nat}}$ controlled by Nature is equipped with a probability distribution on its outgoing edges, i.e., we are given a vector $(P_{kj})_{j \in \mathscr{V}_{\text{Min}}}$ with rational entries such that $P_{kj} \geqslant 0$ for all i and $\sum_{(k,j) \in \mathscr{E}} P_{kj} = 1$. We suppose that Nature makes its decisions according to this probability distribution, i.e., it chooses an edge (k, j) with probability P_{kj}. Moreover, we are given two integer matrices $A, B \in \mathbb{Z}^{\mathscr{V}_{\text{Max}} \times \mathscr{V}_{\text{Min}}}$. These matrices encode the payoffs of the game in the following way: if the current state of the game is $j \in \mathscr{V}_{\text{Min}}$ and Min selects an edge (j, i), then Player Min pays to Max the amount $-A_{ij}$. Similarly, if the current state of the game is $i \in \mathscr{V}_{\text{Max}}$ and Max selects an edge (i, k), then Max receives from Min the payment B_{ik}.

We first consider the *game in horizon* N, in which each of the two players Min and Max makes N moves, starting from a known initial state, which by convention we require to be controlled by Min. In this setting, a *history* of the game consists of the sequence of states visited up to a given stage. A *strategy* of a player is a function which assigns to a history of the game a decision of this player. A pair of strategies (σ, τ) of players Min and Max induces a probability measure on the set of finite sequences of states. Then, the expected reward of Max, starting from the initial position j_0, is defined by

$$R_{j_0}(\sigma, \tau) := \mathbb{E}_{\sigma\tau}\Big(\sum_{p=0}^{N-1} (-A_{i_p j_p} + B_{i_p k_p})\Big) ,$$

in which the expectation $\mathbb{E}_{\sigma,\tau}$ refers to the probability measure induced by (σ, τ), and $j_0, i_0, k_0, j_1, i_1, k_1, \ldots$ is the random sequence of states visited when applying this pair of strategies. The objective of Max is to maximize this reward, while Min wants to minimize it. The game in horizon N starting from state j is known to have a *value* v_j^N and optimal strategies σ^* and τ^*, meaning that

$$R_j(\sigma^*, \tau) \leqslant v_j^N := R_j(\sigma^*, \tau^*) \leqslant R_j(\sigma, \tau^*) ,$$

for all strategies σ of Min and τ of Max. The *value vector* $v^N := (v_j^N)_{j\in\mathscr{V}_{\rm Min}}$ keeps track of the values of all initial states. A classical dynamic programming argument, see e.g. [33, Th. IV.3.2], shows that

$$v^0 = 0, \qquad v^N = F(v^{N-1}),$$

where the *Shapley operator* F is the map from $\mathbb{R}^{\mathscr{V}_{\rm Min}}$ to $\mathbb{R}^{\mathscr{V}_{\rm Min}}$ defined by

$$F_j(x) := \min_{(j,i)\in\mathscr{E}} \Bigl(-A_{ij} + \max_{(i,k)\in\mathscr{E}} \bigl(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}x_k\bigr)\Bigr), \text{ for all } j \in \mathscr{V}_{\rm Min}\,. \tag{1}$$

Assumption 1 guarantees that F is well defined. One can also consider the *mean-payoff* stochastic game, in which the payment $g_{j_0}(\sigma,\tau)$ received by Player Max becomes the limiting average of the sum of instantaneous payments, i.e.,

$$g_{j_0}(\sigma,\tau) := \liminf_{N\to+\infty} \mathbb{E}_{\sigma\tau}\Bigl(\frac{1}{N}\sum_{p=0}^{N-1}(-A_{i_pj_p} + B_{i_pk_p})\Bigr)\,. \tag{2}$$

We say that a strategy is *positional* if the decision of the player depends only of the current state. A result of Liggett and Lippman [31] entails that a mean payoff game has a value χ_j and that there exists a pair of optimal positional strategies (σ^*,τ^*), meaning that

$$g_j(\sigma^*,\tau) \leqslant \chi_j := g_j(\sigma^*,\tau^*) \leqslant g_j(\sigma,\tau^*)\,,$$

for every initial state $j \in \mathscr{V}_{\rm Min}$ and pair of non-necessarily positional strategies (σ,τ) of players Min and Max. A result of Mertens and Neyman [32] entails in particular that the value of the mean-payoff game coincides with the limit of the normalized value of the games in horizon N, i.e.,

$$\chi = \lim_{N\to\infty}\frac{v^N}{N} = \lim_{N\to\infty}\frac{F^N(0)}{N}\quad,$$

where $F^N = F\circ\cdots\circ F$ denotes the Nth iterate of F and 0 the vector that has all entries equal to 0.

▶ Remark 1. In our model, players Min, Max, and Nature play successively, so that a turn decomposes in three stages, resulting in a Shapley operator of the form (1). Alternative models, like the one of [16], in which a turn consists of a single move, reduce to our model by adding linearly many dummy states, and rescaling the mean payoff by a factor 3.

2.2 The Operator Approach to Zero-Sum Games

We shall develop a general approach, which applies to various classes of zero-sum games with a mean-payoff type payment. To do so, it is convenient to introduce an abstract version of Shapley operators, following the "operator approach" of stochastic games [36, 34]. This will allow us to apply notions from nonlinear Perron–Frobenius theory, especially sub and super eigenvectors, and Collatz-Wielandt numbers, which play a key role in our analysis.

Recall that the sup-norm is defined by $\|x\|_\infty := \max_{i\in[n]}|x_i|$. We also use the *Hilbert's seminorm* [23], which is defined by $\|x\|_{\rm H} := \mathbf{t}(x) - \mathbf{b}(x)$, where $\mathbf{t}(x) := \max_{i\in[n]} x_i$ (read "top") and $\mathbf{b}(x) := \min_{i\in[n]} x_i$ (read "bottom"). We endow $\mathbb{R}$ with the standard order $\leqslant$, which is extended to vectors entrywise.

A self-map F of $\mathbb{R}^n$ is said to be *order-preserving* when

$$x \leqslant y \implies F(x) \leqslant F(y) \text{ for all } x,y\in\mathbb{R}^n\,, \tag{3}$$

and *additively homogeneous* when

$$F(\lambda + x) = \lambda + F(x) \text{ for all } \lambda \in \mathbb{R} \text{ and } x \in \mathbb{R}^n \ , \tag{4}$$

where, for any $z \in \mathbb{R}^n$, $\lambda + z$ stands for the vector with entries $\lambda + z_i$.

▶ **Definition 2.** *A self-map F of $\mathbb{R}^n$ is an* (abstract) Shapley operator *if it is order-preserving and additively homogeneous.*

A basic example is provided by the Shapley operator of a turn-based stochastic mean-payoff game (1). Here, the additive homogeneity axiom captures the absence of discount. We shall see in the next section a different example, arising from entropy games.

We point out that any order-preserving and additively homogeneous self-map F of $\mathbb{R}^n$ is nonexpansive in the sup-norm, meaning that

$$\|F(x) - F(y)\|_\infty \leqslant \|x - y\|_\infty \text{ for all } x, y \in \mathbb{R}^n \ .$$

Using the nonexpansiveness property, we get that the existence and the value of the limit $\lim_{N\to\infty}(F^N(x)/N)$ are independent of the choice of $x \in \mathbb{R}^N$. We call this limit the *escape rate* of F, and denote it by $\chi(F)$. When F is the Shapley operator of a turn-based stochastic mean-payoff game, fixing $x = 0$, we see that $F^N(x)$ coincides with the value vector in horizon N, and so $\chi_j(F)$ yields the mean-payoff when the initial state is j, consistently with our notation χ_j in Section 2.1.

The escape rate is known to exist under some "rigidity" assumptions. The case of semialgebraic maps is treated in [34], whereas the generalization to o-minimal structures (see [43] for background), which is needed in the application to entropy games, is established in [14].

▶ **Theorem 3** ([34] and [14])**.** *Suppose that the function $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive in any norm and that it is semialgebraic, or, more generally, defined in an o-minimal structure. Then, the escape rate $\chi(F)$ does exist.*

This applies in particular to Shapley operators of turn-based mean-payoff games, since in this case the operator F, given by (1), is piecewise affine, and a fortiori semialgebraic. In the case of entropy games, we shall see in the next section that the relevant Shapley operator is defined by a finite expression involving the maps log, exp, as well as the arithmetic operations, and so that it is definable in a richer stucture, which is still o-minimal. We emphasize that no knowledge of o-minimal techniques is needed to follow the present paper, it suffices to admit that the escape rate does exist for all the classes of maps considered here, and this follows from Theorem 3.

When the map F is piecewise-affine, a result finer than Theorem 3 holds:

▶ **Theorem 4** ([30])**.** *A piecewise affine self-map F of $\mathbb{R}^n$ that is nonexpansive in any norm admits an* invariant half-line*, meaning that there exist $z, w \in \mathbb{R}^n$ such that*

$$F(z + \beta w) = z + (\beta + 1)w$$

for any $\beta \in \mathbb{R}$ large enough. In particular, the escape rate $\chi(F)$ exists, and is given by the vector w.

This entails that $F^k(z + \beta w) = z + (\beta + k)w$, and so, by nonexpansiveness of F, for all $x \in \mathbb{R}^n$, $F^k(x) = k\chi(F) + O(1)$ as $k \to \infty$. This expansion is more precise than Theorem 3, which only states that $F^k(x) = k\chi(F) + o(k)$.

For a general order-preserving and additively homogeneous self-map of $\mathbb{R}^n$, and in particular, for the Shapley operators of the entropy games considered below, an invariant half-line may not exist. However, we can still recover information about the sequences $(F^k(x)/k)_k$ through non-linear spectral theory methods. Assuming that F is an order-preserving and additively homogeneous self-map of $\mathbb{R}^n$, the *upper Collatz–Wielandt number* of F is defined by:

$$\overline{\mathrm{cw}}(F) := \inf\{\mu \in \mathbb{R} \colon \exists z \in \mathbb{R}^n, F(z) \leqslant \mu + z\}\,, \tag{5}$$

and the *lower Collatz–Wielandt number* of F by:

$$\underline{\mathrm{cw}}(F) := \sup\{\mu \in \mathbb{R} \colon \exists z \in \mathbb{R}^n, F(z) \geqslant \mu + z\}\,. \tag{6}$$

It follows from Fekete's subadditive lemma that the two limits $\lim_{k\to\infty} \mathbf{t}(F^k(0)/k)$ and $\lim_{k\to\infty} \mathbf{b}(F^k(0)/k)$, which may be thought of as upper and lower regularizations of the escape rate, always exist, see [23]. In the examples of interest to us, the escape rate $\chi(F)$ does exist, it represents the mean-payoff vector, and then $\lim_{k\to\infty} \mathbf{t}(F^k(0)/k) = \mathbf{t}(\chi(F)) = \max_j \chi_j(F)$ is the maximum of the mean payoff among all the initial states. Similarly, $\lim_{k\to\infty} \mathbf{b}(F^k(0)/k) = \mathbf{b}(\chi(F))$ is the minimum of these mean payoffs.

The interest of the vectors z arising in the definition of Collatz-Wielandt numbers is to provide *approximate optimality certificates*, allowing us to bound mean payoffs from above and from below. Indeed, if $F(z) \leqslant \mu + z$, using the order-preserving property and additively homogeneity of F, we get that $F^k(z) \leqslant k\mu + z$ for all $k \in \mathbb{N}$, and, by nonexpansiveness of F, $\lim_{k\to\infty} \mathbf{t}(F^k(0)/k) = \lim_{k\to\infty} \mathbf{t}(F^k(z)/k) \leqslant \mu$. Similarly, if $F(z) \geqslant \mu + z$, we deduce that $\lim_{k\to\infty} \mathbf{b}(F^k(0)/k) \geqslant \mu$. The following result of [23], which can also be obtained as a corollary of a minimax result of Nussbaum [35], see [2], shows that these bounds are optimal.

▶ **Theorem 5** ([23, Prop. 2.1], [2, Lemma 2.8 and Rk. 2.10])**.** *Let F be an order-preserving and additively homogeneous self-map of $\mathbb{R}^n$. Then, $\lim_{k\to\infty} \mathbf{t}(F^k(x)/k) = \overline{\mathrm{cw}}(F)$ and $\lim_{k\to\infty} \mathbf{b}(F^k(x)/k) = \underline{\mathrm{cw}}(F)$ for any $x \in \mathbb{R}^n$.*

Thus, when F is the Shapley operator of a game, the quantities $\overline{\mathrm{cw}}(F)$ and $\underline{\mathrm{cw}}(F)$ respectively correspond to the greatest and smallest mean payoff among all the initial states.

A simpler situation arises when there is a vector $v \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$ such that

$$F(v) = \lambda + v\,. \tag{7}$$

The scalar λ, which is unique, is known as the *ergodic constant*, and (7) is referred to as the *ergodic equation*. Then, $\underline{\mathrm{cw}}(F) = \overline{\mathrm{cw}}(F) = \lambda$. The vector v is known as a *bias* or *potential*. It will be convenient to have a specific notation for the ergodic constant λ when the ergodic equation is solvable, then, we set $\mathrm{erg}(F) := \lambda$. The existence of a solution (λ, v) of (7) is guaranteed by certain "ergodicity" assumptions [3]. When the Shapley operator F is piecewise affine, it follows form Kohlberg's theorem (Theorem 4) that the ergodic equation (7) is solvable if and only if the mean payoff is independent of the initial state.

Denote $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\}$. Properties (3) and (4) also make sense for self-maps of $\bar{\mathbb{R}}^n$, by requiring them to hold for all $x, y \in \bar{\mathbb{R}}^n$ and $\lambda \in \mathbb{R}$. Any order-preserving and additively homogeneous self-map F of $\mathbb{R}^n$ admits a unique continuous extension $\bar{F}$ to $\bar{\mathbb{R}}^n$, obtained by setting, for $x \in \bar{\mathbb{R}}^n$,

$$\bar{F}(x) := \inf\{F(y) \colon y \in \mathbb{R}^n, y \geqslant x\}\ \ . \tag{8}$$

Moreover, $\bar{F}$ is still order-preserving and additively homogeneous, see [18] for details. Hence, in the sequel, we assume that any order-preserving and additively homogeneous self-map F of $\mathbb{R}^n$ is canonically extended to $\bar{\mathbb{R}}^n$, and we will not distinguish between F and $\bar{F}$.

2.3 Entropy Games

Entropy games were introduced in [10]. We follow the presentation of [1] since it extends the original model, see Remark 9 for a comparison.

Similarly to stochastic turn-based zero-sum games, an *entropy game* is played on a digraph $(\mathscr{V}, \mathscr{E})$ in which the set of vertices $(\mathscr{V}, \mathscr{E})$ has a non-trivial partition $\mathscr{V} = \mathscr{V}_{\mathrm{Min}} \uplus \mathscr{V}_{\mathrm{Max}} \uplus \mathscr{V}_{\mathrm{Nat}}$. As in the case of stochastic turn-based games, players Min, Max, and Nature control the states $\mathscr{V}_{\mathrm{Min}}$, $\mathscr{V}_{\mathrm{Max}}$, $\mathscr{V}_{\mathrm{Nat}}$ respectively and they alternate their moves, i.e., $\mathscr{E} \subset \mathscr{V}_{\mathrm{Min}} \times \mathscr{V}_{\mathrm{Max}} \cup \mathscr{V}_{\mathrm{Max}} \times \mathscr{V}_{\mathrm{Nat}} \cup \mathscr{V}_{\mathrm{Nat}} \times \mathscr{V}_{\mathrm{Min}}$. We also suppose that the underlying graph satisfies Assumption 1. In the context of entropy games, player Min is called *Despot*, player Max is called *Tribune*, and Nature is called *People*. For this reason, we denote $\mathscr{V}_D := \mathscr{V}_{\mathrm{Min}}$, $\mathscr{V}_T := \mathscr{V}_{\mathrm{Min}}$, and $\mathscr{V}_P := \mathscr{V}_{\mathrm{Nat}}$. The name "Tribune" coined in [10], refers to the magistrates interceding on behalf of the plebeians in ancient Rome.

The first difference between stochastic turn-based games and entropy games lies in the behavior of Nature: while in stochastic games Nature makes its decisions according to some fixed probability distribution, in entropy games People is a *nondeterministic* player, i.e., nothing is assumed about the behavior of People. The second difference lies in the definition of the payoffs received by Tribune. We suppose that every edge $(p, d) \in \mathscr{E}$ with $p \in \mathscr{V}_P$ and $d \in \mathscr{V}_D$ is equipped with a *multiplicity* m_{pd} which is a (positive) natural number. The *weight* of a path is defined to be the product of the weights of the arcs arising on this path. For instance, the path $(d_0, t_0, p_0, d_1, t_1, p_1, d_2, t_2)$ where $d_i \in \mathscr{V}_D$, $t_i \in \mathscr{V}_T$ and $p_i \in \mathscr{V}_P$, makes 2 and 1/3 turn, and its weight is $m_{p_0 d_1} m_{p_1 d_2}$. A *game in horizon* N is then defined as follows: if (σ, τ) is a pair of strategies of Despot and Tribune, then we denote by $R_d^N(\sigma, \tau)$ the sum of the weights of paths with initial state d that make N turns and that are consistent with the choice of (σ, τ). Tribune wants to maximize this quantity, while Despot wants to minimize it. As for stochastic turn-based games, a dynamic programming argument given in [1] shows that the value $V^N \in \mathbb{R}_{>0}^{\mathscr{V}_D}$ of this game does exist, and that it satisfies the recurrence

$$V^0 = \mathbf{1}, \qquad V^N = T(V^{N-1}),$$

where $\mathbf{1}$ is the vector whose entries are identically one and the operator $T \colon \mathbb{R}_{>0}^{\mathscr{V}_D} \to \mathbb{R}_{>0}^{\mathscr{V}_D}$ is defined by

$$T_d(x) := \min_{(d,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl} x_l, \text{ for all } d \in \mathscr{V}_D . \tag{9}$$

To define a game that lasts for an infinite number of turns, we consider the limit

$$V_d^\infty(\sigma, \tau) := \limsup_{N\to+\infty} (R_d^N(\sigma, \tau))^{1/N} ,$$

which may be thought of as a measure of the freedom of People. The logarithm of this limit is known as a *topological entropy* in symbolic dynamics. The following result shows that the value of the entropy game V_d^∞ does exist and that it coincides with the limit of the renormalized value $(V_d^N)^{1/N} = [T^N(\mathbf{1})]_d^{1/N}$ of the finite horizon entropy game, so that the situation is similar to the case of stochastic turn-based games, albeit the renormalization now involves a Nth geometric mean owing to the multiplicative nature of the payment.

▶ **Theorem 6** ([1]). *The entropy game with initial state d has a value V_d^∞. Moreover, there are (positional) strategies σ^* and τ^* of Despot and Tribune, such that, for all $d \in \mathscr{V}_D$,*

$$V_d^\infty(\sigma^*, \tau) \leqslant V_d^\infty = V_d^\infty(\sigma^*, \tau^*) \leqslant V_d^\infty(\sigma, \tau^*),$$

for all strategies σ and τ of the two players. In addition, the value vector $V^\infty := (V_d^\infty)_{d\in\mathscr{V}_D}$ coincides with the vector

$$\lim_{N\to\infty} \big(T^N(\mathbf{1})\big)^{1/N} \in \mathbb{R}_{>0}^{\mathscr{V}_D} \enspace ,$$

in which the operation $\cdot^{1/N}$ is understood entrywise.

Entropy games can be cast in the general operator setting of Section 2.2, by introducing the conjugate operator $F\colon \mathbb{R}^{\mathscr{V}_D} \to \mathbb{R}^{\mathscr{V}_D}$,

$$F := \log \circ T \circ \exp \tag{10}$$

in which $\exp\colon \mathbb{R}^{\mathscr{V}_D} \mapsto \mathbb{R}_{>0}^{\mathscr{V}_D}$ is the map which applies the exponential entrywise, and $\log := \exp^{-1}$. Since the maps log and exp are order preserving, and since the weights m_{pl} appearing in the expression of $T(x)$ in (9) are nonnegative, the operator F is order preserving. Moreover, using the morphism property of the map log and exp with respect to multiplication and addition, we see that F is also additively homogeneous, hence, it is an abstract Shapley operator in the sense of Definition 2. Moreover, it is definable in the real exponential field, which was shown to be an o-minimal structure by Wilkie [45], and this is precisely how Theorem 6 is derived in [1] from Theorem 3. Actually, entropy games are studied in [1] in a more general setting, allowing history dependent strategies and showing that positional strategies are optimal. It is also shown there that the game has a uniform value in the sense of Mertens and Neyman [32].

When the (positional) strategies σ, τ are fixed, the value can be characterized by a classical result of Perron–Frobenius theory.

▶ **Definition 7.** *Given a pair of strategies (σ, τ) of Despot and Tribune, we define the* ambiguity matrix *$M^{\sigma,\tau} \in \mathbb{R}_{\geqslant 0}^{\mathscr{V}_D\times\mathscr{V}_D}$, with entries $(M^{\sigma,\tau})_{k,l} = m_{\tau(\sigma(k)),l}$ if $\big(\tau(\sigma(k)), l\big) \in \mathscr{E}$ and $(M^{\sigma,\tau})_{k,l} = 0$ otherwise, i.e., this is the weighted transition matrix of the subgraph $\mathscr{G}^{\sigma,\tau}$ obtained by keeping only the arcs $\mathscr{V}_D \to \mathscr{V}_T$ and $\mathscr{V}_T \to \mathscr{V}_P$ determined by the two strategies.*

The digraph $\mathscr{G}^{\sigma,\tau}$ can generally be decomposed in strongly connected components $\mathscr{C}_1, \dots, \mathscr{C}_s$, and each of these components, $\mathscr{C}_i$, determines a principal submatrix of $M^{\sigma,\tau}$, denoted by $M^{\sigma,\tau}[\mathscr{C}_i]$, obtained by keeping only the rows and columns in $\mathscr{C}_i \cap \mathscr{V}_D$. We denote by $\rho(\cdot)$ the spectral radius of a matrix, which is also known as the *Perron root* when the matrix is nonnegative and irreducible, see [12] for background.

▶ **Proposition 8** ([38], [46, Th. 5.1]). *The value of the subgame with initial state d, induced by a pair of strategies σ, τ, coincides with*

$$\max\{\rho(M^{\sigma,\tau}[\mathscr{C}_i])\colon \textit{there is a dipath } d \to \mathscr{C}_i \textit{ in } \mathscr{G}^{\sigma,\tau}\} \enspace .$$

▶ Remark 9. In the original model of Asarin et al. [10], an entropy game is specified by finite sets of states of Depot and Tribune, D and T, respectively, by a finite alphabet Σ representing actions, and by a transition relation $\Delta \subset T \times \Sigma \times D \cup D \times \Sigma \times T$. A turn consists of four successive moves by Despot, People, Tribune, and People: in state $d \in D$, Despot selects an action $a \in \Sigma$, then, People moves to one state $t \in P$ such that $(d, a, t) \in \Delta$. Then, Tribune selects an action $b \in \Sigma$, and People moves to one state $d' \in D$ such that $(t, b, d') \in \Delta$. This reduces to the model of [1] by introducing dummy states, identifying a turn in the game of [10] to a succession of two turns in the game of [1]. Another difference is that the payment, in [10], corresponds to $\max_{d\in D} \limsup_{N\to\infty} (R_d^N)^{1/N}$, and this is equivalent

Algorithm 1 Basic value iteration algorithm.

```
procedure ValueIteration($F$)
    ▷ $F$ a Shapley operator from $\mathbb{R}^n$ to $\mathbb{R}^n$
    $u := 0 \in \mathbb{R}^n$
    repeat $u := F(u)$    ▷ At iteration $\ell$, $u = F^\ell(0)$ is the value vector of the game in finite horizon $\ell$
    until $\mathbf{t}(u) \leqslant 0$ or $\mathbf{b}(u) \geqslant 0$
    if $\mathbf{t}(u) \leqslant 0$ then return "$\overline{\mathrm{cw}}(F) \leqslant 0$"    ▷ Player Min wins for all initial states
    else return "$\underline{\mathrm{cw}}(F) \geqslant 0$"    ▷ Player Max wins for all initial states
    end
end
```

to letting Tribune choose the initial state before playing the game. Then, the value of the game in [10] coincides with the maximum of the values of the initial states, $\max_d V_d^\infty$, see [1, Prop. 11]. Finally in [10], the arcs have multiplicity one, whereas we allow integer multiplicies (coded in binary), as in [1].

3 Bounding the Complexity of Value Iteration

In this section, F is an (abstract) Shapley operator, i.e., an order-preserving and additively homogeneous self-map of $\mathbb{R}^n$.

3.1 A Universal Complexity Bound for Value Iteration

The most straightforward idea to solve a mean-payoff game is probably value iteration: we infer whether or not the mean-payoff game is winning by solving the finite horizon game, for a large enough horizon. This is formalized in Algorithm 1.

When the non-linear eigenproblem $F(w) = \mathrm{erg}(F) + w$ is solvable, we shall use the following metric estimate, which represents the minimal Hilbert's seminorm of a bias vector

$$R(F) := \inf \{ \|w\|_{\mathrm{H}} \colon w \in \mathbb{R}^n ,\ F(w) = \mathrm{erg}(F) + w \} \ .$$

In general, however, this non-linear eigenproblem may not be solvable. Then, we consider, for $\lambda \in \mathbb{R}$,

$$S_\lambda(F) = \{ v \in \mathbb{R}^n \colon \lambda + v \leqslant F(v) \}, \qquad S^\lambda(F) = \{ v \in \mathbb{R}^n \colon \lambda + v \geqslant F(v) \} \ .$$

▶ **Theorem 10.** *Procedure ValueIteration (Algorithm 1) is correct as soon as $\underline{\mathrm{cw}}(F) > 0$ or $\overline{\mathrm{cw}}(F) < 0$, and it terminates in a number of iterations N_{vi} bounded by*

$$\inf \left\{ \frac{\|v\|_{\mathrm{H}}}{\lambda} \colon \lambda > 0 \, , \ v \in S_\lambda(F) \cup S^{-\lambda}(F) \right\} \ . \tag{11}$$

In particular, if F has a bias vector and $\mathrm{erg}(F) \neq 0$, we have $N_{vi} \leqslant \frac{R(F)}{|\mathrm{erg}(F)|}$.

We prove this theorem by using the Collatz-Wielandt variational characterization of the limits $\lim_{k\to\infty} \mathbf{t}(F^k(x)/k)$ and $\lim_{k\to\infty} \mathbf{b}(F^k(x)/k)$, see Theorem 5. A special case of Theorem 10 in which the existence of a bias vector is assumed appeared in [6] (without proof).

▶ Remark 11. The infimum in (11) is generally not attained. Consider for instance $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $F(x) = (\log(\exp(x_1) + \exp(x_2)), x_2) - \alpha$, where $\alpha > 0$. Then, since $F_2(x) = x_2 - \alpha < x_2$, we have $S_\lambda(F) = \emptyset$ for $\lambda > 0$. Besides, since $x - \lambda \geqslant F(x)$ if and only if $x_1 - \lambda \geqslant \log(\exp(x_1) + \exp(x_2)) - \alpha$ and $x_2 - \lambda \geqslant x_2 - \alpha$, it follows that $S^{-\lambda}(F) \neq \emptyset$

Algorithm 2 Value iteration in finite precision arithmetics.

```
procedure FPValueIteration($\tilde{F}$)
    $u := 0 \in \mathbb{R}^n$, $\ell := 0 \in \mathbb{N}$, $\epsilon \in \mathbb{R}_{>0}$
    repeat $u := \tilde{F}(u)$; $\ell := \ell + 1$    ▷ We suppose that the operator $F$ is evaluated in approximate arithmetics, so that $\tilde{F}(u)$ is at most at distance $\epsilon$ in the sup-norm from its true value $F(u)$.
    until $\ell\epsilon + \mathbf{t}(u) \leqslant 0$ or $-\ell\epsilon + \mathbf{b}(u) \geqslant 0$
    if $\ell\epsilon + \mathbf{t}(u) \leqslant 0$ then return "$\overline{\mathrm{cw}}(F) \leqslant 0$"    ▷ Player Min wins for all initial states
    end
    if $-\ell\epsilon + \mathbf{b}(u) \geqslant 0$ then return "$\underline{\mathrm{cw}}(F) \geqslant 0$"    ▷ Player Max wins for all initial states
    end
end
```

Algorithm 3 Approximating the value of a mean-payoff game when it is independent of the initial state, and computing approximate optimality certificates, working in finite precision arithmetic.

```
procedure ApproximateConstantMeanPayoff($F$)
    $u, x, y := 0 \in \mathbb{R}^n$, $\ell := 0 \in \mathbb{N}$, $\delta \in \mathbb{R}_{>0}$    ▷ The number $\delta$ is the desired precision of approximation.
    repeat $u := \tilde{F}(u)$; $\ell := \ell + 1$    ▷ The operator $F$ is evaluated in approximate arithmetic, so that $\tilde{F}(u)$ is at most at distance $\epsilon := \delta/8$ in the sup-norm from its true value $F(u)$.
    until $\mathbf{t}(u) - \mathbf{b}(u) \leqslant (3/4)\delta\ell$
$\kappa := \mathbf{b}(u)/\ell$; $\lambda := \mathbf{t}(u)/\ell$
$u := 0$
    for $i = 1, 2, \dots, \ell - 1$ do $u := \tilde{F}(u)$; $x := \max\{x, -i\kappa + u\}$; $y := \min\{y, -i\lambda + u\}$
    done
return "$[\underline{\mathrm{cw}}(F), \overline{\mathrm{cw}}(F)]$ is included in the interval $[\kappa - \delta/8, \lambda + \delta/8]$, which is of width at most $\delta$. Furthermore, we have $\kappa - \delta/8 + x \leqslant F(x)$ and $\lambda + \delta/8 + y \geqslant F(y)$."    ▷ All initial states have a value in $[\kappa - \delta/8, \lambda + \delta/8]$.
end
```

if and only if $\lambda < \alpha$. Now let $v \in S^{-\lambda}(F)$ for some $\lambda < \alpha$. Without loss of generality, we may assume $\mathbf{b}(v) = 0$. Then, we have $v_1 - \lambda \geqslant \log(\exp(v_1) + \exp(v_2)) - \alpha \geqslant \log 2 - \alpha$ and so $\frac{\|v\|_{\mathrm{H}}}{\lambda} \geqslant 1 + \frac{\log 2 - \alpha}{\lambda}$. We conclude that the infimum in (11) is equal to $\frac{\log 2}{\alpha}$ but it is not attained.

3.2 Value Iteration in Finite Precision Arithmetics

Algorithm 1 can be adapted to work in finite precision arithmetic. Consider the variant given in Algorithm 2. We assume that each evaluation of the Shapley operator F is performed with an error of at most $\epsilon > 0$ in the sup-norm. In this section, we denote by $\tilde{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ the operator which approximates F, as in Procedure FPValueIteration, so it satisfies

$$\|\tilde{F}(x) - F(x)\|_\infty \leqslant \epsilon \text{ for all } x \in \mathbb{R}^n. \tag{12}$$

The following result is established by exploiting nonexpansiveness properties of Shapley operators.

▶ **Theorem 12.** *Procedure FPValueIteration (Algorithm 2) is correct as soon as $\underline{\mathrm{cw}}(F) > 2\epsilon$ or $\overline{\mathrm{cw}}(F) < -2\epsilon$, and it terminates in a number of iterations N_{vi}^{ϵ} bounded by*

$$\inf\left\{\frac{\|v\|_{\mathrm{H}}}{\lambda - 2\epsilon} : \lambda > 2\epsilon\,,\ v \in S_\lambda(F) \cup S^{-\lambda}(F)\right\}\ . \tag{13}$$

In particular, if F has a bias vector and $|\operatorname{erg}(F)| > 2\epsilon$, we have $N_{vi}^{\epsilon} \leqslant \frac{R(F)}{|\operatorname{erg}(F)| - 2\epsilon}$.

Procedure ApproximateConstantMeanPayoff returns sub and super-eigenvectors x and y, satisfying $\kappa - \delta/8 + x \leqslant F(x)$ and $\lambda + \delta/8 + y \geqslant F(y)$, which, by Theorem 5, entails that $[\underline{\mathrm{cw}}(F), \overline{\mathrm{cw}}(F)]$ is included in the interval $[\kappa - \delta/8, \lambda + \delta/8]$. The construction of these sub and sup-eigenvectors, by taking infima and suprema of normalized orbits of F, is inspired by [23, Proof of Lemma 2].

▶ **Theorem 13.** *Suppose that $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$, and let ρ denote this common value. Then, Procedure* ApproximateConstantMeanPayoff *(Algorithm 3) halts and is correct for any given desired precision of approximation $\delta \in \mathbb{R}_{>0}$. Furthermore, if $R := \max\{\|v\|_{\mathrm{H}}, \|w\|_{\mathrm{H}}\}$, where $v, w \in \mathbb{R}^n$ are any two vectors that satisfy $\rho - \delta/8 + v \leqslant F(v)$ and $\rho + \delta/8 + w \geqslant F(w)$, then this procedure stops after at most $\lceil 8R/\delta \rceil$ iterations of the first loop.*

3.3 Finding the States of Maximal Value

In this section, we will show how the value iteration algorithm can be adapted to decide whether or not a given game has constant value, and to find the set of states that have the maximal value. Our analysis is based on an abstract notion of dominion. As previously, we suppose that $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is an order-preserving and additively homogeneous operator. Recall that thanks to (8), F is canonically extended to define a self-map of $\bar{\mathbb{R}}^n$. Furthermore, given a nonempty set $\mathscr{S} \subset [n]$, we define the operator $F^{\mathscr{S}} : \bar{\mathbb{R}}^{\mathscr{S}} \to \bar{\mathbb{R}}^{\mathscr{S}}$ as $F^{\mathscr{S}} := \mathbf{p}^{\mathscr{S}} \circ F \circ \mathbf{i}^{\mathscr{S}}$, where $\mathbf{p}^{\mathscr{S}} : \bar{\mathbb{R}}^n \to \bar{\mathbb{R}}^{\mathscr{S}}$ is the projection on the coordinates in $\mathscr{S}$ which is defined as usual by $\mathbf{p}^{\mathscr{S}}_j(x) = x_j$ for $j \in \mathscr{S}$, and $\mathbf{i}^{\mathscr{S}} : \bar{\mathbb{R}}^{\mathscr{S}} \to \bar{\mathbb{R}}^n$ is defined by $\mathbf{i}^{\mathscr{S}}_j(x) = x_j$ if $j \in \mathscr{S}$ and $\mathbf{i}^{\mathscr{S}}_j(x) = -\infty$ otherwise.

▶ **Definition 14.** *A* dominion (of Player Max) *is a nonempty set $\mathscr{D} \subset [n]$ such that $F^{\mathscr{D}}$ preserves $\mathbb{R}^{\mathscr{D}}$, i.e., such that $F^{\mathscr{D}}(x) \in \mathbb{R}^{\mathscr{D}}$ for all $x \in \mathbb{R}^{\mathscr{D}}$.*

As discussed in [7, 3], for stochastic mean-payoff games (with finite action spaces), a dominion of a player can be interpreted as a set of states such that the player can force the game to stay in this set if the initial state belongs to it. This terminology differs from the one of [29], in which a dominion is required in addition to consist only of initial states that are winning for this player. The algorithms that we discuss in this section require an additional assumption on the structure of the Shapley operator F.

▶ **Assumption 2.** *We assume that the limit $\chi^{\mathscr{D}} := \lim_{\ell\to\infty} \frac{(F^{\mathscr{D}})^{\ell}(0)}{\ell} \in \mathbb{R}^{\mathscr{D}}$ exists for every dominion $\mathscr{D} \subset [n]$. Furthermore, we assume that the set $\mathscr{D}_{\max} := \{j \in [n] \colon \chi^{[n]}_j = \overline{\mathrm{cw}}(F)\}$ is a dominion and that it satisfies $\underline{\mathrm{cw}}(F^{\mathscr{D}_{\max}}) = \overline{\mathrm{cw}}(F^{\mathscr{D}_{\max}}) = \overline{\mathrm{cw}}(F)$.*

▶ Remark 15. We note that the first part of Assumption 2 holds automatically when the Shapley operator $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is definable in an o-minimal structure. Indeed, in this case the relation (8) implies that $F^{\mathscr{D}}$ is definable in the same structure for every dominion $\mathscr{D}$, so $\chi^{\mathscr{D}}$ exists by Theorem 3. We will see that the second part of the assumption applies to the games considered in this paper.

▶ Remark 16. Assumption 2 will allow us to make an induction on the number states, by a reduction to a simpler game with a reduced state space $\mathscr{D}$. In particular, the assumption that the limit $\chi^{\mathscr{D}} = \lim_{\ell\to\infty} \frac{(F^{\mathscr{D}})^{\ell}(0)}{\ell}$ exists will allow us to apply value iteration to the Shapley operator of the reduced game, $F^{\mathscr{D}}$.

From now on, we denote $\chi := \chi^{[n]}$ and $\mathscr{D}_{\max} := \{j \in [n] \colon \chi_j = \overline{\mathrm{cw}}(F)\}$. The following theorem applies to Shapley operators for which an a priori separation bound is known: if $\overline{\mathrm{cw}}(F) > \underline{\mathrm{cw}}(F)$, it requires an apriori bound $\delta > 0$ such that $\overline{\mathrm{cw}}(F) - \underline{\mathrm{cw}}(F) > \delta$. We note that the existence of the approximate sub and super-eigenvectors v and w used in this theorem follows from Theorem 5 and from Assumption 2.

Algorithm 4 Deciding if the value is constant.

```
procedure DecideConstantValue($F, \delta, R$)
    $u := 0 \in \mathbb{R}^n$, $\ell := 0 \in \mathbb{N}$.
    $\tilde{F} :=$ any map such that $\tilde{F}(u)$ is at most at distance $\epsilon := \delta/8$ in the sup-norm from $F(u)$.
    repeat $u := \tilde{F}(u)$; $\ell := \ell + 1$
    until $\mathbf{t}(u) - \mathbf{b}(u) \leqslant (3/4)\delta\ell$ or $\ell = 1 + \lceil 8R/\delta \rceil$
    if $\ell = 1 + \lceil 8R/\delta \rceil$ then
        $\mathscr{S} := \{i \colon u_i = \mathbf{b}(u)\}$
        return $\mathscr{S}$    ▷ The value of the game depends on the initial state. We have $\chi_i < \overline{\mathrm{cw}}(F)$ for all $i \in \mathscr{S}$.
    else
        return $\varnothing$    ▷ The value of the game is independent of the initial state.
    end
end
```

▶ **Theorem 17.** *Suppose that F is such that either $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$ or $\overline{\mathrm{cw}}(F) - \underline{\mathrm{cw}}(F) > \delta$ for some $\delta > 0$. Let $\mathscr{D}_{\max}$ be the set of states of maximal value and $R := \max\{\|v\|_{\mathrm{H}}, \|w\|_{\mathrm{H}}\}$, where $v, w \in \mathbb{R}^{\mathscr{D}_{\max}}$ are any two vectors that satisfy $\overline{\mathrm{cw}}(F) - \delta/8 + v \leqslant F^{\mathscr{D}_{\max}}(v)$ and $\overline{\mathrm{cw}}(F) + \delta/8 + w \geqslant F^{\mathscr{D}_{\max}}(w)$. Then, Procedure* DecideConstantValue *(Algorithm 4) is correct.*

Let

$$\mathrm{sep}(F) := \inf_{\mathscr{D}} \left(\overline{\mathrm{cw}}(F^{\mathscr{D}}) - \underline{\mathrm{cw}}(F^{\mathscr{D}})\right)$$

where the infimum is taken over all the dominions $\mathscr{D}$ of F which contain all the states of maximal value and satisfy $\overline{\mathrm{cw}}(F^{\mathscr{D}}) - \underline{\mathrm{cw}}(F^{\mathscr{D}}) > 0$.

To state the final result of this section, we will suppose that we have an access to an oracle that approximates F to a given precision $\epsilon > 0$. More precisely, given a point $x \in \bar{\mathbb{R}}^n$, the oracle is supposed to output a point $y \in \bar{\mathbb{R}}^n$ that satisfies $y_j = -\infty$ for all $j \in [n]$ such that $F_j(x) = -\infty$ and $|F_j(x) - y_j| \leqslant \epsilon$ for all j such that $F_j(x) \neq -\infty$. We have

▶ **Theorem 18.** *Let $\delta > 0$ be such that $\delta < \mathrm{sep}(F)$, $\mathscr{D}_{\max}$ be the set of states of maximal value and $R := \max\{\|v\|_{\mathrm{H}}, \|w\|_{\mathrm{H}}\}$, where $v, w \in \mathbb{R}^{\mathscr{D}_{\max}}$ are any two vectors that satisfy $\overline{\mathrm{cw}}(F) - \delta/8 + v \leqslant F^{\mathscr{D}_{\max}}(v)$ and $\overline{\mathrm{cw}}(F) + \delta/8 + w \geqslant F^{\mathscr{D}_{\max}}(w)$. Then, the set of initial states of maximal value can be found by making at most $n^2 + n\lceil 8R/\delta \rceil$ calls to oracle that approximates F to precision $\epsilon := \delta/8$.*

4 Application to Stochastic Mean-Payoff Games

In this section, we apply our results to stochastic mean-payoff games. We start by bounding the separation sep and the metric estimate $R(F)$, when F is the Shapley operator of a stochastic turn-based zero-sum game as in (1). We recall that the entries of A and B are integers. This is not more special than assuming that the entries of A and B are rational numbers (we may always rescale rational payments so that they become integers). We set

$$W := \max\left\{|A_{ij} - B_{ik}| \colon i \in \mathscr{V}_{\mathrm{Max}},\, j \in \mathscr{V}_{\mathrm{Min}},\, k \in \mathscr{V}_{\mathrm{Nat}}\right\} . \tag{14}$$

We also assume that the probabilities P_{kj} are rational, and that they have a common denominator $M \in \mathbb{N}_{>0}$, $P_{kj} = Q_{kj}/M$, where $Q_{kj} \in [M]$ for all $k \in \mathscr{V}_{\mathrm{Nat}}$ and $j \in \mathscr{V}_{\mathrm{Min}}$. We say that a state $k \in \mathscr{V}_{\mathrm{Nat}}$ is a *significant random state* if there are at least two indices $j, j' \in \mathscr{V}_{\mathrm{Min}}$ such that $P_{kj} > 0$ and $P_{kj'} > 0$. We denote by s the number of significant random states and by $n := |\mathscr{V}_{\mathrm{Min}}|$ the number of states controlled by Min. The following estimates follow from optimal bit-complexity results for Markov chains, established in [41]. These improve an estimate in [16].

▶ **Lemma 19.** *We have* $\operatorname{sep}(F) > 1/(nM^{\min\{s,n-1\}})^2$.

▶ **Lemma 20.** *Suppose that* $\underline{\mathrm{cw}}(F) = \overline{\mathrm{cw}}(F)$. *Then, there exists a vector* $u \in \mathbb{R}^{\mathscr{V}_{\mathrm{Min}}}$ *such that* $F(u) = \overline{\mathrm{cw}}(F) + u$ *and*

$$R(F) \leqslant \|u\|_{\mathrm{H}} \leqslant 8nWM^{\min\{s,n-1\}} \enspace .$$

The existence of the bias vector follows from Kohlberg's theorem (Theorem 4). The bias is generally not unique (even up to an additive constant) and the main difficulty then is to find a "short" bias. The one which is constructed in the proof of this lemma relies on the notion of Blackwell optimality. This notion requires to consider the *discounted* version of the game, in which the payment (2) is replaced by $\mathbb{E}_{\sigma\tau}\sum_{p=0}^{\infty}(1-\alpha)^p(-A_{i_pj_p}+B_{i_pk_p})$, where $0<\alpha<1$ and $1-\alpha$ is the discount factor. The discounted game with initial state i has a value, $x_i(\alpha)$, and the value vector, $x(\alpha) = (x_i(\alpha)) \in \mathbb{R}^n$ is the unique solution of the fixed point problem $x(\alpha) = F((1-\alpha)x(\alpha))$. Then, a strategy of a player is *Blackwell optimal* if it is optimal in all the discounted games with a discount factor sufficiently close to 1. It can be obtained by selecting minimizing or maximizing actions when evaluating the expression $F((1-\alpha)x(\alpha))$, for $\alpha > 0$ close enough to 0. Moreover, it follows from Kohlberg's proof that $x(\alpha)$ admits an expansion $x(\alpha) = \chi(F)/\alpha + u + o(\alpha)$ when $\alpha \to 0^+$, where u is a bias vector. If Q is the stochastic matrix determined by a pair of Blackwell optimal strategies of the two players, and if r is the associated one-stage payment vector, then, the bias vector u satisfies a Poisson-type equation $\chi(F) + u = r + Qu$. Moreover, this special bias vector has the remarkable property of having a zero expectation with respect to all invariant measures of Q, and together with the bit-complexity estimate of [41, Th. 1.5] for the solution of Poisson-type equations, this leads to the proof of Lemma 20.

Thanks to these estimates, we arrive at the following corollaries.

▶ **Corollary 21.** *Let F be a Shapley operator as above, supposing that F has a bias vector and that* $\operatorname{erg}(F)$ *is nonzero. Then, procedure* ValueIteration *stops after*

$$N_{vi} \leqslant 8n^2WM^{2\min\{s,n-1\}} \tag{15}$$

iterations and correctly decides which of the two players is winning.

▶ Remark 22. When specialized to deterministic mean-payoff games, i.e., when $s = 0$, Corollary 21 yields $N_{\mathrm{vi}} = O(n^2W)$ which is precisely the bound that follows from the analysis of value iteration by Zwick and Paterson [47].

▶ **Corollary 23.** *Suppose that F has a bias vector and let* $\mu := nM^{\min\{s,n-1\}}$. *Then, Procedure* ApproximateConstantMeanPayoff, *applied to F and to* $\delta := \mu^{-2}$, *terminates in at most*

$$128n^3WM^{3\min\{s,n-1\}}$$

calls to the oracle. Moreover the interval returned by this procedure contains a unique rational number of denominator at most μ, which coincides with the value, and optimal policies can be obtained from the approximate optimality certificates generated by the procedure.

Let us explain how the optimal strategies are obtained from the output of Procedure ApproximateConstantMeanPayoff. This procedure returns sub and super-eigenvectors x and y that satisfy $\kappa - \delta/8 + x \leqslant F(x)$ and $\lambda + \delta/8 + y \geqslant F(y)$ where $\lambda = \operatorname{erg}(F)$. By selecting, for each state $j \in \mathscr{V}_{\mathrm{Min}}$, a minimizing action in the expression

$$F_j(y) = \min_{(j,i)\in\mathscr{E}}\Bigl(-A_{ij} + \max_{(i,k)\in\mathscr{E}}\bigl(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}y_k\bigr)\Bigr),$$

one gets a positional strategy which guarantees to Min a value at most $\lambda + \delta/8$. A similar method is used to construct a positional strategy of Max. Then, since $\delta = \mu^{-2}$ is smaller than the separation bound between values of different strategies, we deduce these policies guarantee a value of λ to each of the players, and so, they are optimal.

▶ **Corollary 24.** *Let $\mu := nM^{\min\{s,n-1\}}$. Then, we can find the set of states with maximal value of a stochastic mean-payoff game by performing at most $65n^4WM^{3\min\{s,n-1\}}$ calls to the oracle approximating F with precision $\delta := 1/\mu^2$.*

▶ Remark 25. If we are given the matrices A, B, P explicitly, then the operator F can be evaluated exactly in $O(E)$ complexity, where E is the number of edges of the graph representing a stochastic mean-payoff game. In particular, there is no need to construct an approximation oracle in order to apply the results of this section. Nevertheless, even in this case it may be beneficial to use an approximation oracle. Indeed, if we evaluate F exactly, then each value iteration $u := F(u)$ increases the number of bits needed to encode u. As a result, value iteration would require exponential memory. In order to avoid this problem, one can replace F with an approximation oracle $\tilde{F}$ obtained as follows. Let $\mu := nM^{\min\{s,n-1\}}$ and $\epsilon := 1/(8\mu^2)$. Given $x \in \bar{\mathbb{R}}^{\mathscr{V}_{\mathrm{Min}}}$, we first compute $y := F(x)$ exactly and then round the finite coordinates of y in such a way that the rounded vector $\tilde{y}$ satisfies $|y_j - \tilde{y}_j| \leqslant \epsilon$ whenever $y_j \neq -\infty$ and $\tilde{y}_j$ is a rational number with denominator at most $8\mu^2$. One can check that if we use $\tilde{F}$ obtained in this way as an approximation oracle, then all algorithms presented in this section require $O\big(nE\log(nMW)\big)$ memory, which is polynomial in the size of the input.

▶ Remark 26. Since a single call to the oracle approximating F can be done in $O(E)$ arithmetic operations, by combining Corollary 24 with Corollary 23 we see that the set of states with maximal value, and a pair of optimal strategies within this set can be found in $O(n^4EWM^{3\min\{s,n-1\}})$ complexity. This should be compared with the algorithm BWR-FindTop from [16] which achieves the same aim using a pumping algorithm instead of value iteration. If we combine the estimate from [41] with the complexity bound presented in [16] for the pumping algorithm, then we get that BWR-FindTop has $O(V^6EWs2^sM^{4s} + V^3EW\log W)$ complexity, where V is the number of vertices of the graph. In particular, our result gives a better complexity bound. Furthermore, the authors of [16] show that, given an oracle access to BWR-FindTop and to another oracle that solves deterministic mean-payoff games, one can completely solve stochastic mean-payoff games with pseudopolynomial number of calls to these oracles, provided that s is fixed. Hence, we can speed-up this algorithm by replacing the oracle BWR-FindTop with our algorithms.

5 Solving Entropy Games With Bounded Rank

Recall that the dynamic programming operator T of an entropy game, as well as its conjugate F, which we call the Shapley operator of an entropy game, were defined in (9),(10). As in the last section, we denote $n := |\mathscr{V}_D|$ and we put $W := \max_{(p,k)\in\mathscr{E}} m_{pk}$.

We define the *rank* of the entropy game to be the maximum of the ranks of the ambiguity matrices, see Definition 7. The following result is established by combining a separation bound of Rump [39] for algebraic numbers, with bounds on determinants of nonnegative matrices with entries in an interval, building on the study of Hadamard's maximal determinant problem for matrices with entries in $\{0,1\}$ [20].

▶ **Theorem 27.** *Suppose two pairs of strategies yield distinct values in an entropy game of rank r, with n Despot's states. Then, these values differ at least by $\nu_{n,r}^{-1}$ where*

$$\nu_{n,r} := 2^r(r+1)^{8r}r^{-2r^2+r+1}(ne)^{4r^2}\big(1 \vee \frac{W}{2}\big)^{4r^2} .$$

We show that the value of an entropy game is always in the interval $[1, nW]$. Then, a separation bound for values of different pairs of strategies entails a separation bound for their logarithm, differing only by a nW factor:

▶ **Corollary 28.** *Suppose two pairs of strategies yield distinct values in an entropy game of rank r, with n Despot's states. Then, the logarithms of these values differ at least by $\hat{\nu}_{n,r}^{-1}$ where*

$$\hat{\nu}_{n,r} := nW\nu_{n,r} \enspace .$$

▶ **Proposition 29.** *Let $0 < \delta < 1$. Then, there exist vectors $w, z \in \mathbb{R}_{>0}^{\mathscr{V}_D}$ such that $e^{-\delta}\mathbf{b}(V^\infty)w \leqslant T(w)$, $e^{\delta}\mathbf{t}(V^\infty)z \geqslant T(z)$, and $\max\{\|\log w\|_{\mathrm{H}}, \|\log z\|_{\mathrm{H}}\} \leqslant 1200(n^3 \log W + n^2 \log \delta^{-1})$.*

This proposition is established by observing that for a given value of δ, w and z are defined by semi-linear constraints, and by using bitlength estimates on the generators and vertices of polyhedra defined by inequalities.

By applying Theorem 18 to the Shapley operator $F = \log \circ T \circ \exp$, and by using the bounds Corollary 28 and Proposition 29, we get:

▶ **Theorem 30.** *In an entropy game of rank at most r, we can find the set of initial states with maximal value by performing $O(nR_{n,r}\hat{\nu}_{n,r})$ calls to an oracle approximating F with precision $\delta/8$ where $\delta = (\hat{\nu}_{n,r})^{-1}$.*

The following decomposition property for entropy games extends a classical property of deterministic mean payoff games. Once the set of Despot's states with maximal value is known, it allows one to determine the value of the other states by reduction to an entropy game induced by the other states of Despot. To state this property formally, we denote by $\mathscr{D}_{\max} \subset \mathscr{V}_D$ the states of Despot of maximal value, and we put ${\mathscr{V}_P}^{\mathscr{D}_{\max}} := \{p \in \mathscr{V}_P : \exists k \in \mathscr{D}_{\max}, (p, k) \in \mathscr{E}\}$ and ${\mathscr{V}_T}^{\mathscr{D}_{\max}} := \{t \in \mathscr{V}_T : \exists p \in {\mathscr{V}_P}^{\mathscr{D}_{\max}}, (t, p) \in \mathscr{E}\}$.

▶ **Lemma 31** (Decomposition property). *Let $\mathscr{S}_1 := \mathscr{D}_{\max} \uplus {\mathscr{V}_T}^{\mathscr{D}_{\max}} \uplus {\mathscr{V}_P}^{\mathscr{D}_{\max}}$ and $\mathscr{S}_2 := \mathscr{V} \setminus \mathscr{S}_1$. Furthermore, suppose that $\mathscr{S}_2$ is nonempty. Consider the induced digraphs $\mathscr{G}[\mathscr{S}_1]$ and $\mathscr{G}[\mathscr{S}_2]$ of the original graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$. Then, the entropy games arising by restricting the graph to $\mathscr{G}[\mathscr{S}_1]$ and $\mathscr{G}[\mathscr{S}_2]$ satisfy Assumption 1. Furthermore, if (σ_1, τ_1) are optimal strategies of Despot an Tribune in the induced entropy game on $\mathscr{G}[\mathscr{S}_1]$ and (σ_2, τ_2) are optimal strategies of Despot and Tribune in the induced entropy game on $\mathscr{G}[\mathscr{S}_2]$, then the joint strategies*

$$\forall k \in \mathscr{V}_D,\ \sigma(k) = \begin{cases} \sigma_1(k) & \textit{if } k \in \mathscr{D}_{\max}, \\ \sigma_2(k) & \textit{otherwise,} \end{cases} \quad \forall t \in \mathscr{V}_T,\ \tau(t) = \begin{cases} \tau_1(t) & \textit{if } t \in {\mathscr{V}_T}^{\mathscr{D}_{\max}}, \\ \tau_2(t) & \textit{otherwise.} \end{cases} \tag{16}$$

are optimal in the original game.

We also note that the Shapley operator of an entropy game can be approximated in polynomial time, this follows by using a result of Borwein and Borwein [17] on the approximation of the log and exp maps, together with a scaling argument, see [1, Lemma 27]. Then, by combining Theorem 30 and Lemma 31, we get:

▶ **Corollary 32.** *A pair of optimal policies of an entropy game of rank r can be found in $O(n^2 R_{n,r}\hat{\nu}_{n,r})$ calls to an oracle that return F with a precision of $1/(16\hat{\nu}_{n,r})$. Then, entropy games in the original model of Asarin et al. [10] and with a fixed rank are polynomial-time solvable, whereas entropy games with weights, in the model of Akian et al. [1], and with a fixed rank, are pseudo-polynomial time solvable.*

▶ **Corollary 33.** *Entropy games with weights and with a fixed number of People's positions are pseudo-polynomial time solvable.*

We say that a state of a game is *significant* if there are several options in this state, in particular, a state p of People is significant if there are at least two distinct arcs (p, k) and (p, l) in $\mathscr{E}$. We may ask whether the statement of Corollary 33 carries over to entropy games with a fixed number of significant People's states. The following result shows that this can not be derived from the universal value iteration bounds, since value iteration needs $\Omega(W^{n-1})$ iterations to recognize the optimal strategy.

▶ **Theorem 34.** *There is a family of $G_n(W)$ of Despot-free entropy games, and a constant $C > 0$, with the following properties:*

1. *$G_n(W)$ has arc weights $\leqslant W$, only one significant Tribune's position, with two actions, and $2n + 1$ People's positions among which there are only 4 significant positions;*
2. *The action of Tribune that is optimal in the mean-payoff entropy game is never played, if Tribune plays optimally in the entropy game of finite horizon k, for all $k \leqslant CW^{n-1}$.*

Let us explain how the game $G_n(W)$ is constructed. We start by estimating the positive root of a special polynomial p_n.

▶ **Proposition 35.** *Consider the polynomial $p_n(x) = x^n - W(x^{n-1} + \cdots + 1)$, where $W > 0$. Then, p_n has a unique positive root, $x_n(W)$, which satisfies*

$$x_n(W) = W + 1 - 1/W^{n-1} + o(1/W^{n-1}) \ , \qquad \text{as } W \to \infty \ .$$

This is established by applying the Newton-Puiseux algorithm [44] to the equation $p_n(x) = 0$ parameterized by W. Then, we define the companion matrix of p_n, $A_n = A_n(W)$, together with the sequence $(z(k))_{k \in \mathbb{N}}$,

$$A_n(W) := \left(\begin{array}{ccc|c} W & \dots & W & W \\ \hline & I_{n-1} & & 0 \end{array}\right) \ , \ z(k) = \max(\mathbf{1}_n^\top A_n^k \mathbf{1}_n, \alpha \mathbf{1}_{n-1}^\top A_{n-1}^k \mathbf{1}_{n-1}) \ , \tag{17}$$

where for all $k \geqslant 1$, $\mathbf{1}_k$ is the vector of dimension k with unit entries, I_k the identity matrix of dimension k, and $\alpha > 1$. We note that A_n is an irreducible nonnegative matrix, so, the unique positive root of p_n is actually the Perron root of A_n. Using Proposition 35, as well as explicit bounds for the left Perron eigenvector of A_n, we can show that the maximum in (17) is achieved by the rightmost term for $k \leqslant k^*$ and by the leftmost term for $k > k^*$ where $k^* = (\log 2)W^{n-1} + o(W^{n-1})$. Moreover, $z(k)$ can be interpreted as the value in horizon $k + 1$ of an entropy game satisfying the conditions of the theorem. Indeed, the leftmost term $\mathbf{1}_n^\top A_n^k \mathbf{1}_n$ can be interpreted as the value in horizon $k + 1$ of a Despot-free and Tribune-free entropy game, with $n + 1$ People's states, among which there are only two significant states: one encoding the first row of A_n, and another one encoding the row vector $\mathbf{1}_n^\top$. The term $\alpha \mathbf{1}_{n-1}^\top A_{n-1}^k \mathbf{1}_{n-1}$ also admits an interpretation as the value of a similar game. We finally construct the entropy game $G_n(W)$ by allowing Tribune to choose whichever of these elementary games is played. This can be implemented by taking the disjoint union of the graphs of the two elementary games, and adding one significant state of Tribune, with only two actions. One action gives rise to the left term of the maximum in (17), whereas the other action gives rise to the right term, so that the value of the corresponding entropy game in horizon k is precisely $z(k - 1)$. Since $\lambda_n > \lambda_{n-1}$, in the mean-payoff entropy game, the optimal action for Tribune is to choose the term with the highest geometric growth, i.e., to play "left", which guarantees a geometric growth of λ_n. However, for $k \leqslant k^* + 1$, the optimal action of Tribune in the game of horizon k is always to play "right". This shows Theorem 34.

6 Concluding Remarks

We developed generic value iteration algorithms, which apply to various classes of zero-sum games with mean payoffs. These algorithms admit universal complexity bounds, in an approximate oracle model – we only need an oracle evaluating approximately the Shapley operator. These bounds involve three fundamental ingredients: the number of states, a separation bound between the values induced by different strategies, and a bound on the norms of Collatz-Wielandt vectors. We showed that entropy games with a fixed rank (and in particular, entropy games with a fixed number of People's states) are pseudo-polynomial time solvable. This should be compared with the result of [1], showing that entropy games with a fixed number of Despot positions are polynomial-time solvable. Since fixing the number of states of Despot or People leads to improved complexity bounds, one may ask whether entropy games with a fixed number of significant Tribune states are polynomial or at least pseudo-polynomial, this is still an open question.

References

1 M. Akian, S. Gaubert, J. Grand-Clément, and J. Guillaud. The operator approach to entropy games. *Theor. Comp. Sys.*, 63(5):1089–1130, July 2019. doi:10.1007/s00224-019-09925-z.

2 M. Akian, S. Gaubert, and A. Guterman. Tropical polyhedra are equivalent to mean payoff games. *Int. J. Algebra Comput.*, 22(1):125001 (43 pages), 2012. doi:10.1142/S0218196711006674.

3 M. Akian, S. Gaubert, and A. Hochart. A game theory approach to the existence and uniqueness of nonlinear Perron-Frobenius eigenvectors. *Discrete & Continuous Dynamical Systems - A*, 40:207–231, 2020. doi:10.3934/dcds.2020009.

4 M. Akian, S. Gaubert, and R. Nussbaum. A Collatz-Wielandt characterization of the spectral radius of order-preserving homogeneous maps on cones. arXiv:1112.5968, 2011.

5 M. Akian, A. Sulem, and M. I. Taksar. Dynamic optimization of long-term growth rate for a portfolio with transaction costs and logarithmic utility. *Mathematical Finance*, 11(2):153–188, April 2001. doi:10.1111/1467-9965.00111.

6 X. Allamigeon, S. Gaubert, R. D. Katz, and M. Skomra. Condition numbers of stochastic mean payoff games and what they say about nonarchimedean semidefinite programming. In *Proceedings of the 23rd International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 160–167, 2018. URL: http://mtns2018.ust.hk/media/files/0213.pdf.

7 X. Allamigeon, S. Gaubert, and M. Skomra. Solving generic nonarchimedean semidefinite programs using stochastic game algorithms. *J. Symbolic Comput.*, 85:25–54, 2018. doi:10.1016/j.jsc.2017.07.002.

8 V. Anantharam and V. S. Borkar. A variational formula for risk-sensitive reward. *SIAM J. Contro Optim.*, 55(2):961–988, 2017. arXiv:1501.00676.

9 D. Andersson and P. B. Miltersen. The complexity of solving stochastic games on graphs. In *Proceedings of the 20th International Symposium on Algorithms and Computation (ISAAC)*, volume 5878 of *Lecture Notes in Comput. Sci.*, pages 112–121. Springer, 2009. doi:10.1007/978-3-642-10631-6_13.

10 E. Asarin, J. Cervelle, A. Degorre, C. Dima, F. Horn, and V. Kozyakin. Entropy games and matrix multiplication games. In *Proceedings of the 33rd International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 47 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 11:1–11:14, Wadern, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.STACS.2016.11.

11 D. Auger, X. Badin de Montjoye, and Y. Strozecki. A generic strategy improvement method for simple stochastic games. In Filippo Bonchi and Simon J. Puglisi, editors, *46th International Symposium on Mathematical Foundations of Computer Science, MFCS 2021, August 23-27,*

2021, Tallinn, Estonia, volume 202 of *LIPIcs*, pages 12:1–12:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.MFCS.2021.12.

12 A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences.* SIAM, 1994.

13 T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Math. Oper. Res.*, 1(3):197–208, 1976. doi:10.1287/moor.1.3.197.

14 J. Bolte, S. Gaubert, and G. Vigeral. Definable zero-sum stochastic games. *Mathematics of Operations Research*, 40(1):171–191, 2014. doi:10.1287/moor.2014.0666.

15 E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A convex programming-based algorithm for mean payoff stochastic games with perfect information. *Optim. Lett.*, 11(8):1499–1512, 2017. doi:10.1007/s11590-017-1140-y.

16 E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A pseudo-polynomial algorithm for mean payoff stochastic games with perfect information and few random positions. *Inform. and Comput.*, 267:74–95, 2019. doi:10.1016/j.ic.2019.03.005.

17 J. M. Borwein and P. B. Borwein. On the complexity of familiar functions and numbers. *SIAM Review*, 30(4):589–601, 1988.

18 A. D. Burbanks, R. D. Nussbaum, and C. T. Sparrow. Extension of order-preserving maps on a cone. *Proc. Roy. Soc. Edinburgh Sect. A*, 133(1):35–59, 2003. doi:10.1017/S0308210500002274.

19 K. Chatterjee and R. Ibsen-Jensen. The complexity of ergodic mean-payoff games. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 8573 of *Lecture Notes in Comput. Sci.*, pages 122–133. Springer, 2014. doi:10.1007/978-3-662-43951-7_11.

20 John H. E. Cohn. On the value of determinants. *Proceedings of the American Mathematical Society*, 14(4):581–588, 1963.

21 A. Condon. The complexity of stochastic games. *Inform. and Comput.*, 96(2):203–224, 1992. doi:10.1016/0890-5401(92)90048-K.

22 O. Friedmann. An exponential lower bound for the latest deterministic strategy iteration algorithms. *Logical Methods in Computer Science*, 7(3:19):1–42, 2011.

23 S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Trans. Amer. Math. Soc.*, 356(12):4931–4950, 2004. doi:10.1090/S0002-9947-04-03470-1.

24 H. Gimbert and F. Horn. Simple stochastic games with few random vertices are easy to solve. In *Proceedings of the 11th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS)*, volume 4962 of *Lecture Notes in Comput. Sci.*, pages 5–19. Springer, 2008. doi:10.1007/978-3-540-78499-9_2.

25 V. A. Gurvich, A. V. Karzanov, and L. G. Khachiyan. Cyclic games and finding minimax mean cycles in digraphs. *Zh. Vychisl. Mat. Mat. Fiz.*, 28(9):1406–1417, 1988. doi:10.1016/0041-5553(88)90012-2.

26 A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Manag. Sci.*, 12(5):359–370, 1966. doi:10.1287/mnsc.12.5.359.

27 R. A. Howard and J. E. Matheson. Risk-sensitive markov decision processes. *Management Science*, 18(7):356–369, 1972. doi:10.1287/mnsc.18.7.356.

28 R. Ibsen-Jensen and P. B. Miltersen. Solving simple stochastic games with few coin toss positions. In *Proceedings of the 20th Annual European Symposium on Algorithms (ESA)*, volume 7501 of *Lecture Notes in Comput. Sci.*, pages 636–647. Springer, 2012. doi:10.1007/978-3-642-33090-2_55.

29 M. Jurdziński, M. Paterson, and U. Zwick. A deterministic subexponential algorithm for solving parity games. *SIAM J. Comput.*, 38(4):1519–1532, 2008. doi:10.1137/070686652.

30 E. Kohlberg. Invariant half-lines of nonexpansive piecewise-linear transformations. *Math. Oper. Res.*, 5(3):366–372, 1980. doi:10.1287/moor.5.3.366.

31 T. M. Liggett and S. A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Rev.*, 11(4):604–607, 1969. doi:10.1137/1011093.

32 J.-F. Mertens and A. Neyman. Stochastic games. *Internat. J. Game Theory*, 10(2):53–66, 1981. doi:10.1007/BF01769259.

33 J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated games*, volume 55 of *Econom. Soc. Monogr.* Cambridge University Press, Cambridge, 2015. doi:10.1017/CBO9781139343275.

34 A. Neyman. Stochastic games and nonexpansive maps. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C*, pages 397–415. Kluwer Academic Publishers, 2003. doi:10.1007/978-94-010-0189-2_26.

35 R. D. Nussbaum. Convexity and log convexity for the spectral radius. *Linear Algebra Appl.*, 73:59–122, 1986. doi:10.1016/0024-3795(86)90233-8.

36 D. Rosenberg and S. Sorin. An operator approach to zero-sum repeated games. *Israel J. Math.*, 121(1):221–246, 2001. doi:10.1007/BF02802505.

37 U. G. Rothblum. Multiplicative markov decision chains. *Mathematics of Operations Research*, 9(1):6–24, 1984.

38 U. G. Rothblum and P. Whittle. Growth optimality for branching markov decision chains. *Mathematics of Operations Research*, 7(4):582–601, 1982.

39 S. M. Rump. Polynomial minimum root separation. *Mathematics of Computation*, 145(33):327–336, 1979.

40 M. Skomra. *Tropical spectrahedra: Application to semidefinite programming and mean payoff games*. PhD thesis, Université Paris-Saclay, 2018. URL: https://pastel.archives-ouvertes.fr/tel-01958741.

41 M. Skomra. Optimal bounds for bit-sizes of stationary distributions in finite Markov chains. arXiv:2109.04976, 2021.

42 K. Sladký. *On dynamic programming recursions for multiplicative Markov decision chains*, pages 216–226. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976. doi:10.1007/BFb0120753.

43 L. van den Dries. *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998. doi:10.1017/CBO9780511525919.

44 R. J. Walker. *Algebraic Curves*. Springer, New York, 1978.

45 A. J. Wilkie. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Amer. Math. Soc.*, 9(4):1051–1094, 1996.

46 W. H. M. Zijm. Asymptotic expansions for dynamic programming recursions with general nonnegative matrices. *J. Optim. Theory Appl.*, 54(1):157–191, 1987. doi:10.1007/BF00940410.

47 U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theoret. Comput. Sci.*, 158(1–2):343–359, 1996. doi:10.1016/0304-3975(95)00188-3.