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

Xavier Allamigeon, Stéphane Gaubert, Ricardo David Katz, Mateusz Skomra

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

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

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

Ricardo D. Katz
CIFASIS-CONICET, Bv. 27 de Febrero 210 bis, 2000 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.

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 [23, 57] 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 Gurvich, Karzanov and Khachyan [30] to solve the optimality equation of deterministic mean-payoff games. Zwick and Paterson [57] 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 [31], and albeit being

experimentally fast on typical instances, is generally exponential [25]. We refer to the survey of [10] for more information and additional references.

Entropy games have been introduced by Asarin et al. [11]. 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 [52, 47, 46, 56], risk sensitive control [32, 9], mathematical finance [6], or matrix multiplication games [11]. Asarin et al. showed that entropy games also belong to the class NP ∩ 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 30. 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 [18], 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 [30]. 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 Theorem 61. These results concern the extended model of entropy games introduced in [1], taking into account weights. Then,

entropy games in the sense of [11] (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 [57]. 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 [18, 17]. The authors of [18] generalize the "pumping" algorithm, developed for deterministic games by Gurvich, Karzanov, and Khachiyan [30], to the case of stochastic games. The resulting algorithm is also pseudopolynomial if the number of random positions is fixed, see Remark 47 for a detailed comparison. The algorithm of Ibsen-Jensen and Miltersen [33] 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 [28] and more recently by Auger, Badin de Montjoye and Strozecki [12]. The value iteration algorithm for concurrent mean-payoff games, under an ergodicity condition, has been studied by Chatterjee and Ibsen-Jensen [21]. Theorem 18 there gives a $O(|\log \epsilon|/\epsilon)$ bound for the number of iterations needed to get an ϵ-approximation of the mean payoff. When specialized to this case, Theorem 16 below improves this bound by a factor of $|\log \epsilon|$. Another possible general approach to compute the mean payoff would be based on *relative value iteration* combined with Krasnoselskii-Mann damping [13], see [27]. Under some ergodicity condition, this method allows one to get an ϵ-approximation of the mean payoff as well as an approximate optimality certificate in $O(1/\epsilon^2)$ iterations, to be compared with the $O(1/\epsilon)$ iterations of the present algorithm, see Remark 17.

We build on the operator approach for zero-sum games, see [15, 42, 45]. Our study of entropy games is inspired by the works of Asarin et al. [11] 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 [16], 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 [43, 26, 2, 5].

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 [8], by investigating suitable notions of condition numbers [7]. An initial version of some of the present results (concerning turn based stochastic games) appeared in the PhD thesis of one of the authors [50].

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.

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: $\mathscr{V} = \mathscr{V}_{\mathrm{Min}} \uplus \mathscr{V}_{\mathrm{Max}} \uplus \mathscr{V}_{\mathrm{Nat}}$. There are two players, called *Min* and *Max*, and a half-player, *Nature*. The sets $\mathscr{V}_{\mathrm{Min}}$, $\mathscr{V}_{\mathrm{Max}}$ and $\mathscr{V}_{\mathrm{Nat}}$ represent the sets of states in which Min, Max, and Nature respectively play. The set of edges $\mathscr{E}$ represents the allowed moves. We assume $\mathscr{E} \subset \mathscr{V}_{\mathrm{Min}} \times \mathscr{V}_{\mathrm{Max}} \cup \mathscr{V}_{\mathrm{Max}} \times \mathscr{V}_{\mathrm{Nat}} \cup \mathscr{V}_{\mathrm{Nat}} \times \mathscr{V}_{\mathrm{Min}}$, meaning that Min, Max, and Nature alternate their moves. More precisely, a turn consists of three successive moves: when the current state is $j \in \mathscr{V}_{\mathrm{Min}}$, Min selects and edge (j, i) in $\mathscr{E}$ and the next state is $i \in \mathscr{V}_{\mathrm{Max}}$. Then, Max selects an edge (i, k) in $\mathscr{E}$ and the next state is $k \in \mathscr{V}_{\mathrm{Nat}}$. Next, Nature chooses an edge $(k, j') \in \mathscr{E}$ and the next state is $j' \in \mathscr{V}_{\mathrm{Min}}$. 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 $j \in \mathscr{V}_{\mathrm{Min}}, i \in \mathscr{V}_{\mathrm{Max}}$, and $k \in \mathscr{V}_{\mathrm{Nat}}$, the sets $\{i' : (j, i') \in \mathscr{E}\}$, $\{k' : (i, k') \in \mathscr{E}\}$ and $\{j' : (k, j') \in \mathscr{E}\}$ are non-empty.*

Furthermore, every state $k \in \mathscr{V}_{\mathrm{Nat}}$ controlled by Nature is equipped with a probability distribution on its outgoing edges, i.e., we are given a vector $(P_{kj})_{j \in \mathscr{V}_{\mathrm{Min}}}$ with rational entries such that $P_{kj} \geqslant 0$ for all j 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, an integer A_{ij} is associated with each edge (j, i) in $\mathscr{E} \cap (\mathscr{V}_{\mathrm{Min}} \times \mathscr{V}_{\mathrm{Max}})$, and an integer B_{ik} is associated with each edge (i, k) in $\mathscr{E} \cap (\mathscr{V}_{\mathrm{Max}} \times \mathscr{V}_{\mathrm{Nat}})$. These integers encode the payoffs of the game in the following way: if the current state of the game is $j \in \mathscr{V}_{\mathrm{Min}}$ and Min selects the edge (j, i), then Min pays to Max the amount $-A_{ij}$. Similarly, if the current state of the game is $i \in \mathscr{V}_{\mathrm{Max}}$ and Max selects the 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 $j_0 \in \mathscr{V}_{\mathrm{Min}}$, is defined by

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

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 $j \in \mathscr{V}_{\mathrm{Min}}$ 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* $v^N := (v_j^N)_{j \in \mathscr{V}_{\mathrm{Min}}}$ keeps track of the values of all initial states. A classical dynamic programming argument, see e.g. [41, Th. IV.3.2], shows that

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

where $\mathbf{0}$ denotes the vector that has all entries equal to 0 and the *Shapley operator F* is the map from $\mathbb{R}^{\mathscr{V}_{\mathrm{Min}}}$ to $\mathbb{R}^{\mathscr{V}_{\mathrm{Min}}}$ defined by

$$F_j(x) := \min_{(j,i) \in \mathscr{E}} \left(-A_{ij} + \max_{(i,k) \in \mathscr{E}} \left(B_{ik} + \sum_{(k,l) \in \mathscr{E}} P_{kl} x_l \right) \right), \text{ for all } j \in \mathscr{V}_{\mathrm{Min}}\,. \tag{1}$$

Assumption 1 guarantees that F is well defined.

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

$$g_{j_0}(\sigma, \tau) := \liminf_{N \to +\infty} \mathbb{E}_{\sigma\tau} \left(\frac{1}{N} \sum_{p=0}^{N-1} (-A_{i_p j_p} + B_{i_p k_p}) \right). \tag{2}$$

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 [38] 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 $j \in \mathscr{V}_{\mathrm{Min}}$ and pair of non-necessarily positional strategies (σ, τ) of players Min and Max. A result of Mertens and Neyman [40] 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.,

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

where $F^N = F \circ \cdots \circ F$ denotes the Nth iterate of F.

▶ **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 [18], 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 [45, 42]. 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.

We set $[n] := \{1, \ldots, n\}$. We shall use the sup-norm $\|x\|_\infty := \max_{i \in [n]} |x_i|$, and also the *Hilbert's seminorm* [26], which is defined by $\|x\|_{\mathrm{H}} := \mathbf{t}(x) - \mathbf{b}(x)$, where $\mathbf{t}(x) := \max_{i \in [n]} x_i$

(read "top") and $\mathbf{b}(x) \coloneqq \min_{i \in [n]} x_i$ (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 $x = \mathbf{0}$, 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 [42], whereas the generalization to o-minimal structures (see [53] for background), which is needed in the application to entropy games, is established in [16].

▶ **Theorem 3** ([42] and [16]). *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 (meaning that its domain can be covered by finitely many polyhedra such that F restricted to any of them is 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 structure, 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** ([36]). *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 $F^N(z + \beta w) = z + (\beta + N)w$, and so, by nonexpansiveness of F, for all $x \in \mathbb{R}^n$, $F^N(x) = N\chi(F) + O(1)$ as $N \to \infty$. This expansion is more precise than Theorem 3, which only states that $F^N(x) = N\chi(F) + o(N)$.

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 $(F^N(x)/N)_N$ 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:

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

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

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

It follows from Fekete's subadditive lemma that the two limits $\lim_{N\to\infty} \mathbf{t}(F^N(\mathbf{0})/N)$ and $\lim_{N\to\infty} \mathbf{b}(F^N(\mathbf{0})/N)$, which may be thought of as upper and lower regularizations of the escape rate, always exist, see [26]. In the examples of interest to us, the escape rate $\chi(F)$ does exist, it represents the mean-payoff vector, and then $\lim_{N\to\infty} \mathbf{t}(F^N(\mathbf{0})/N) = \mathbf{t}(\chi(F)) = \max_j \chi_j(F)$ is the maximum of the mean payoff among all the initial states. Similarly, $\lim_{N\to\infty} \mathbf{b}(F^N(\mathbf{0})/N) = \mathbf{b}(\chi(F))$ 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 $F^N(z) \leqslant N\mu + z$ for all $N \in \mathbb{N}$, and, by nonexpansiveness of F, $\lim_{N\to\infty} \mathbf{t}(F^N(\mathbf{0})/N) = \lim_{N\to\infty} \mathbf{t}(F^N(z)/N) \leqslant \mu$. Similarly, if $F(z) \geqslant \mu + z$, we deduce that $\lim_{N\to\infty} \mathbf{b}(F^N(\mathbf{0})/N) \geqslant \mu$. The following result of [26], which can also be obtained as a corollary of a minimax result of Nussbaum [43], see [2], shows that these bounds are optimal.

▶ **Theorem 5** ([26, 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, $\lim_{N\to\infty} \mathbf{t}(F^N(x)/N) = \overline{\mathrm{cw}}(F)$ and $\lim_{N\to\infty} \mathbf{b}(F^N(x)/N) = \underline{\mathrm{cw}}(F)$ 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 [4]. 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 $\lambda \in \bar{\mathbb{R}}$. 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 [20] 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 [11]. 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 $\mathscr{V}$ 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 in $\mathscr{V}_{\mathrm{Min}}$, $\mathscr{V}_{\mathrm{Max}}$ and $\mathscr{V}_{\mathrm{Nat}}$ 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 [11], 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 as the product of the multiplicities 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 $V^N \in \mathbb{R}^{\mathscr{V}_D}_{>0}$ of this game does exist, and that it satisfies the recurrence

$$V^0 = \mathbf{1}, \qquad V^N = T(V^{N-1}) \,, \tag{9}$$

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

$$T_d(x) := \min_{(d,t) \in \mathscr{E}} \max_{(t,p) \in \mathscr{E}} \sum_{(p,l) \in \mathscr{E}} m_{pl} x_l, \text{ for all } d \in \mathscr{V}_D \,. \tag{10}$$

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 [35]. 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

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

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 $F \colon \mathbb{R}^{\mathscr{V}_D} \to \mathbb{R}^{\mathscr{V}_D}$,

$$F := \log \circ T \circ \exp \tag{11}$$

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 (10) are nonnegative, the operator F is order-preserving. Moreover, using the morphism property of the maps $\log$ 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 [55], 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 [40].

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 $\left(\tau(\sigma(k)), l \right) \in \mathscr{E}$ 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 [14] for background.

▶ **Proposition 8** ([47], [56, Th. 5.1]). *The value of the subgame with initial state d, induced by a pair of strategies σ, τ, coincides with*

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

```
1: procedure VALUEITERATION(F)
2:     ▷ F a Shapley operator from ℝⁿ to ℝⁿ
3:     u := 0 ∈ ℝⁿ
4:     repeat  u := F(u)    ▷ At iteration ℓ, u = Fℓ(0) is the value vector of the game in finite horizon ℓ
5:     until t(u) ⩽ 0 or b(u) ⩾ 0
6:     if t(u) ⩽ 0 then return "cw̄(F) ⩽ 0"     ▷ Player Min wins for all initial states
7:     else  return "cw(F) ⩾ 0"    ▷ Player Max wins for all initial states
8:     end
9: end
```

■ **Figure 1** Basic value iteration algorithm.

▶ Remark 9. In the original model of Asarin et al. [11], 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 a state $t \in P$ such that $(d, a, t) \in \Delta$. Then, Tribune selects an action $b \in \Sigma$, and People moves to a 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 [11] to a succession of two turns in the game of [1]. Another difference is that the payment, in [11], corresponds to $\max_{d \in D} \lim \sup_{N \to \infty} (R_d^N)^{1/N}$, and this is equivalent to letting Tribune choose the initial state before playing the game. Then, the value of the game in [11] coincides with the maximum of the values of the initial states, $\max_d V_d^\infty$, see [1, Prop. 11]. Finally in [11], 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 Figure 1.

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

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

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

$$S_\lambda(F) := \{v \in \mathbb{R}^n : \lambda + v \leqslant F(v)\} \ \text{ and } \ S^\lambda(F) := \{v \in \mathbb{R}^n : \lambda + v \geqslant F(v)\} .$$

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

$$\inf \left\{ \frac{\|v\|_{\mathrm{H}}}{\lambda} : \lambda > 0 \, , \ v \in S_\lambda(F) \cup S^{-\lambda}(F) \right\} . \tag{12}$$

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

```
 1: procedure FPVALUEITERATION(F̃)
 2:     u := 0 ∈ ℝⁿ, ℓ := 0 ∈ ℕ, ε ∈ ℝ_{>0}
 3:     repeat u := F̃(u); ℓ := ℓ + 1    ▷ We suppose that the operator F is evaluated in approximate
        arithmetics, so that F̃(u) is at most at distance ε in the sup-norm from its true value F(u).
 4:     until ℓε + t(u) ⩽ 0 or −ℓε + b(u) ⩾ 0
 5:     if ℓε + t(u) ⩽ 0 then return "c̅w̅(F) ⩽ 0"    ▷ Player Min wins for all initial states
 6:     end
 7:     if −ℓε + b(u) ⩾ 0 then return "cw(F) ⩾ 0"    ▷ Player Max wins for all initial states
 8:     end
 9: end
```

■ **Figure 2** Value iteration in finite precision arithmetics.

To prove Theorem 10 and some other results, we need the following lemma.

▶ **Lemma 11** ([26, Theorem 8]). *Let F be an order-preserving and additively homogeneous self-map of $\mathbb{R}^n$. Then, $\mathbf{b}(F^\ell(\mathbf{0})) \leqslant \ell\underline{\mathrm{cw}}(F) \leqslant \ell\overline{\mathrm{cw}}(F) \leqslant \mathbf{t}(F^\ell(\mathbf{0}))$ for any $\ell \in \mathbb{N}$.*

Proof of Theorem 10. Since $v + \mathbf{b}(w - v) \leqslant w \leqslant v + \mathbf{t}(w - v)$ for any $v, w \in \mathbb{R}^n$, and F^ℓ is order-preserving and additively homogeneous for any $\ell \in \mathbb{N}$, note that

$$F^\ell(v) - \mathbf{t}(v) = F^\ell(v) + \mathbf{b}(-v) \leqslant F^\ell(\mathbf{0}) \leqslant F^\ell(v) + \mathbf{t}(-v) = F^\ell(v) - \mathbf{b}(v) \ ,$$

for all $v \in \mathbb{R}^n$ and $\ell \in \mathbb{N}$.

To prove the theorem, in the first place suppose that $\underline{\mathrm{cw}}(F) > 0$. Then, by the definition (6) of $\underline{\mathrm{cw}}(F)$ we know that there exist $\lambda > 0$ and $v \in \mathbb{R}^n$ such that $v \in S_\lambda(F)$. Moreover, observe that in this case for any $\lambda > 0$ and $v \in S_\lambda(F) \cup S^{-\lambda}(F)$ we necessarily have $v \in S_\lambda(F)$, because otherwise we would have $v \in S^{-\lambda}(F)$, which implies that $\overline{\mathrm{cw}}(F) < 0$ and so that $\underline{\mathrm{cw}}(F) < 0$, contradicting our assumption. Therefore, for any $\lambda > 0$ and $v \in \mathbb{R}^n$ such that $v \in S_\lambda(F) \cup S^{-\lambda}(F)$, we have $\lambda + v \leqslant F(v)$ and so

$$\ell\lambda - \|v\|_{\mathrm{H}} = \ell\lambda + \mathbf{b}(v) - \mathbf{t}(v) = \mathbf{b}(\ell\lambda + v - \mathbf{t}(v)) \leqslant \mathbf{b}(F^\ell(v) - \mathbf{t}(v)) \leqslant \mathbf{b}(F^\ell(\mathbf{0}))$$

for all $\ell \in \mathbb{N}$. Thus, we conclude that $0 \leqslant \mathbf{b}(F^\ell(\mathbf{0}))$ if $\ell \in \mathbb{N}$ is greater than (12). This shows that Procedure VALUEITERATION terminates in a number of iterations N_{vi} bounded by (12) when $\underline{\mathrm{cw}}(F) > 0$. The proof in the case $\overline{\mathrm{cw}}(F) < 0$ is analogous.

Finally, observe that the correctness of Procedure VALUEITERATION readily follows from Lemma 11. ◀

A special case of Theorem 10 in which the existence of a bias vector is assumed appeared (without proof) in [7].

▶ Remark 12. The infimum in (12) 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 $S^{-\lambda}(F) \neq \emptyset$ 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 (12) is equal to $\frac{\log 2}{\alpha}$ but it is not attained.

3.2 Value Iteration in Finite Precision Arithmetics

The algorithm in Figure 1 can be adapted to work in finite precision arithmetic. Consider the variant of the main body of this algorithm, given in Figure 2. Now 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 what follows, 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:

$$\|\tilde{F}(x) - F(x)\|_\infty \leqslant \epsilon \text{ for all } x \in \mathbb{R}^n. \tag{13}$$

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

▶ **Theorem 13.** *Procedure FPVALUEITERATION (Figure 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*

$$\inf \left\{ \frac{\|v\|_{\mathrm{H}}}{\lambda - 2\epsilon} : \lambda > 2\epsilon \,,\, v \in S_\lambda(F) \cup S^{-\lambda}(F) \right\} \,. \tag{14}$$

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

The proof relies on the next lemma.

▶ **Lemma 14.** *Denote $u^\ell \coloneqq F^\ell(\mathbf{0})$ and $\tilde{u}^\ell \coloneqq \tilde{F}^\ell(\mathbf{0})$ for $\ell \in \mathbb{N}$. Then, we have $\|\tilde{u}^\ell - u^\ell\|_\infty \leqslant \ell\epsilon$, $|\mathbf{t}(\tilde{u}^\ell) - \mathbf{t}(u^\ell)| \leqslant \ell\epsilon$, and $|\mathbf{b}(\tilde{u}^\ell) - \mathbf{b}(u^\ell)| \leqslant \ell\epsilon$ for any $\ell \in \mathbb{N}$.*

Proof. We prove the first claim by induction on ℓ. We have $\|\tilde{u}^1 - u^1\|_\infty \leqslant \epsilon$ by (13). Furthermore, since F is nonexpansive we get

$$\begin{aligned}
\|\tilde{u}^\ell - u^\ell\|_\infty &= \|\tilde{F}(\tilde{u}^{\ell-1}) - F(u^{\ell-1})\|_\infty \\
&\leqslant \|\tilde{F}(\tilde{u}^{\ell-1}) - F(\tilde{u}^{\ell-1})\|_\infty + \|F(\tilde{u}^{\ell-1}) - F(u^{\ell-1})\|_\infty \\
&\leqslant \epsilon + \|\tilde{u}^{\ell-1} - u^{\ell-1}\|_\infty \leqslant \ell\epsilon \,.
\end{aligned}$$

To prove the other claims, fix $\ell \in \mathbb{N}$ and let $k \in [n]$ be such that $\mathbf{t}(u^\ell) = u_k^\ell$. Then, we have $\mathbf{t}(u^\ell) = u_k^\ell \leqslant \tilde{u}_k^\ell + \ell\epsilon \leqslant \mathbf{t}(\tilde{u}^\ell) + \ell\epsilon$. Similarly, if $k' \in [n]$ is such that $\mathbf{t}(\tilde{u}^\ell) = \tilde{u}_{k'}^\ell$, then $\mathbf{t}(\tilde{u}^\ell) = \tilde{u}_{k'}^\ell \leqslant u_{k'}^\ell + \ell\epsilon \leqslant \mathbf{t}(u^\ell) + \ell\epsilon$. Thus, we get $|\mathbf{t}(\tilde{u}^\ell) - \mathbf{t}(u^\ell)| \leqslant \ell\epsilon$. The proof of the fact that $|\mathbf{b}(\tilde{u}^\ell) - \mathbf{b}(u^\ell)| \leqslant \ell\epsilon$ is analogous. ◀

Proof of Theorem 13. To prove that procedure FPVALUEITERATION returns the correct answer, suppose that it stops at the ℓth iteration and, in the first place, that the condition $-\ell\epsilon + \mathbf{b}(u) = -\ell\epsilon + \mathbf{b}(\tilde{F}^\ell(\mathbf{0})) \geqslant 0$ is satisfied. Then, we have $\mathbf{b}(F^\ell(\mathbf{0})) \geqslant 0$ by Lemma 14, and so we conclude that $\underline{\mathrm{cw}}(F) \geqslant 0$ by Lemma 11. Therefore, player Max wins for all initial states if the condition $-\ell\epsilon + \mathbf{b}(\tilde{F}^\ell(\mathbf{0})) \geqslant 0$ is satisfied. If this condition is not satisfied, we necessarily have $\ell\epsilon + \mathbf{t}(\tilde{F}^\ell(\mathbf{0})) \leqslant 0$ due to the stopping condition. Then, we can prove that $\overline{\mathrm{cw}}(F) \leqslant 0$ using symmetric arguments to the ones applied in the case in which the condition $-\ell\epsilon + \mathbf{b}(\tilde{F}^\ell(\mathbf{0})) \geqslant 0$ is satisfied, and so player Min wins for all initial states.

In order to bound the number of iterations required by FPVALUEITERATION, suppose that $\underline{\mathrm{cw}}(F) > 2\epsilon$. Then, by definition there exist $\lambda > 2\epsilon$ and $v \in S_\lambda(F)$. Moreover, if λ and v are such that $\lambda > 2\epsilon$ and $v \in S_\lambda(F) \cup S^{-\lambda}(F)$, then we have $v \in S_\lambda(F)$, as $v \in S^{-\lambda}(F)$ would imply that $\overline{\mathrm{cw}}(F) < -2\epsilon$. Therefore $\lambda + v \leqslant F(v)$ and, by Lemma 14,

$$\ell\lambda - \|v\|_{\mathrm{H}} = \mathbf{b}(\ell\lambda + v - \mathbf{t}(v)) \leqslant \mathbf{b}(F^\ell(v) - \mathbf{t}(v)) \leqslant \mathbf{b}(F^\ell(\mathbf{0})) \leqslant \mathbf{b}(\tilde{F}^\ell(\mathbf{0})) + \ell\varepsilon \,.$$

Hence, if $\ell \geqslant \frac{\|v\|_{\mathrm{H}}}{\lambda - 2\epsilon}$, we have $-\ell\epsilon + \mathbf{b}(\tilde{F}^\ell(\mathbf{0})) \geqslant 0$. Therefore, FPVALUEITERATION terminates in a number of iterations bounded by (14). The proof in the case where $\overline{\mathrm{cw}}(F) < -2\epsilon$ is analogous. ◀

```
 1: procedure APPROXIMATECONSTANTMEANPAYOFF(F)
 2:    u, x, y := 0 ∈ ℝⁿ, ℓ := 0 ∈ ℕ, δ ∈ ℝ>0    ▷ The number δ is the desired precision of approximation.
 3:    repeat u := F̃(u); ℓ := ℓ + 1    ▷ The operator F is evaluated in approximate arithmetic, so that
        F̃(u) is at most at distance ε := δ/8 in the sup-norm from its true value F(u).
 4:    until t(u) − b(u) ⩽ (3/4)δℓ
 5: κ := b(u)/ℓ; λ := t(u)/ℓ
 6: u := 0
 7:    for i = 1, 2, …, ℓ − 1 do u := F̃(u); x := max{x, −iκ + u}; y := min{y, −iλ + u}
 8:    done
 9: return "[cw(F), cw(F)] is included in the interval [κ − δ/8, λ + δ/8], which is of width at most δ.
        Furthermore, we have κ − δ/8 + x ⩽ F(x) and λ + δ/8 + y ⩾ F(y)."    ▷ All initial states have a value
        in [κ − δ/8, λ + δ/8].
10: end
```

■ **Figure 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.

When the value of a mean-payoff game is independent of the initial state, Procedure AP-PROXIMATECONSTANTMEANPAYOFF of Figure 3 returns sub and super-eigenvectors x and y satisfying $\kappa - \delta/8 + x \leqslant F(x)$ and $\lambda + \delta/8 + y \geqslant F(y)$, for some $\kappa, \lambda \in \mathbb{R}$ such that the interval $[\kappa - \delta/8, \lambda + \delta/8]$ is of width at most some desired precision δ. As discussed above, by Theorem 5 this 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 [26, Proof of Lemma 2].

▶ **Lemma 15** (cf. [26, Lemma 2]). *Suppose that* $\mathbf{b}(\tilde{F}^\ell(\mathbf{0})) \geqslant \lambda\ell$ *for some* $\lambda \in \mathbb{R}$ *and* $\ell \in \mathbb{N}$. *If we define*

$$\hat{u} := \mathbf{0} \vee \left(-\lambda + \tilde{F}(\mathbf{0})\right) \vee \cdots \vee \left(-(\ell-1)\lambda + \tilde{F}^{\ell-1}(\mathbf{0})\right),$$

then $\lambda - \epsilon + \hat{u} \leqslant F(\hat{u})$. *Analogously, if* $\mathbf{t}(\tilde{F}^\ell(\mathbf{0})) \leqslant \lambda\ell$ *and we define*

$$\bar{u} := \mathbf{0} \wedge \left(-\lambda + \tilde{F}(\mathbf{0})\right) \wedge \cdots \wedge \left(-(\ell-1)\lambda + \tilde{F}^{\ell-1}(\mathbf{0})\right),$$

then $\lambda + \epsilon + \bar{u} \geqslant F(\bar{u})$.

Proof. Since F is order-preserving, we have $F(x \vee y) \geqslant F(x) \vee F(y)$ for every $x, y \in \mathbb{R}^n$. Therefore, we get

$$
\begin{aligned}
F(\hat{u}) &\geqslant F(\mathbf{0}) \vee \left(-\lambda + F\left(\tilde{F}(\mathbf{0})\right)\right) \vee \cdots \vee \left(-(\ell-1)\lambda + F\left(\tilde{F}^{\ell-1}(\mathbf{0})\right)\right) \\
&\geqslant -\epsilon + \tilde{F}(\mathbf{0}) \vee \left(-\lambda - \epsilon + \tilde{F}^2(\mathbf{0})\right) \vee \cdots \vee \left(-(\ell-1)\lambda - \epsilon + \tilde{F}^\ell(\mathbf{0})\right) \\
&\geqslant -\epsilon + \tilde{F}(\mathbf{0}) \vee \left(-\lambda - \epsilon + \tilde{F}^2(\mathbf{0})\right) \vee \cdots \vee \left(-(\ell-2)\lambda - \epsilon + \tilde{F}^{\ell-1}(\mathbf{0})\right) \vee (\lambda - \epsilon) \\
&= \lambda - \epsilon + \hat{u}.
\end{aligned}
$$

The proof of the fact that $\lambda + \epsilon + \bar{u} \geqslant F(\bar{u})$ if $\mathbf{t}(\tilde{F}^\ell(\mathbf{0})) \leqslant \lambda\ell$ is analogous. ◀

▶ **Theorem 16.** *Suppose that* $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$, *and let* ρ *denote this common value. Then, Procedure* APPROXIMATECONSTANTMEANPAYOFF *(Figure 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.*

Proof. Assume $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$. Let $\epsilon \coloneqq \delta/8$ as in Figure 3, $\rho \coloneqq \overline{\mathrm{cw}}(F)$ and $R \coloneqq \max\{\|v\|_{\mathrm{H}}, \|w\|_{\mathrm{H}}\}$, where $v, w \in \mathbb{R}^n$ are any two vectors that satisfy $\rho - \epsilon + v \leqslant F(v)$ and $\rho + \epsilon + w \geqslant F(w)$ (note that such vectors exist due to the fact that $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F) = \rho$ and $\epsilon > 0$). Since $v \leqslant \mathbf{t}(v) + \mathbf{0}$, for every $\ell \in \mathbb{N}$ we get $\ell\rho - \ell\epsilon + v \leqslant F^\ell(v) \leqslant \mathbf{t}(v) + F^\ell(\mathbf{0})$. Thus, $F^\ell(\mathbf{0}) \geqslant \ell\rho - \ell\epsilon + v - \mathbf{t}(v) \geqslant \ell\rho - \ell\epsilon - \|v\|_{\mathrm{H}}$. Analogously, we get $F^\ell(\mathbf{0}) \leqslant \ell\rho + \ell\epsilon + \|w\|_{\mathrm{H}}$. Therefore, if we denote $\tilde{u}^\ell \coloneqq \tilde{F}^\ell(\mathbf{0})$, Lemma 14 shows that $\mathbf{t}(\tilde{u}^\ell) \leqslant \ell\rho + \ell\epsilon + R + \ell\epsilon = \ell\rho + \ell\delta/4 + R$, and analogously that $\mathbf{b}(\tilde{u}^\ell) \geqslant \ell\rho - \ell\delta/4 - R$. Hence $\mathbf{t}(\tilde{u}^\ell) - \mathbf{b}(\tilde{u}^\ell) \leqslant \ell\delta/2 + 2R$. In particular, for every $\ell \geqslant 8R/\delta$ we have $\mathbf{t}(\tilde{u}^\ell) - \mathbf{b}(\tilde{u}^\ell) \leqslant (3/4)\delta\ell$ and so the stopping condition of the first loop is achieved within the first $\lceil 8R/\delta \rceil$ iterations. Moreover, combining Lemmas 11 and 14 we get

$$\mathbf{b}(\tilde{u}^\ell) - \ell\epsilon \leqslant \mathbf{b}\big(F^\ell(\mathbf{0})\big) \leqslant \ell\rho \leqslant \mathbf{t}\big(F^\ell(\mathbf{0})\big) \leqslant \mathbf{t}(\tilde{u}^\ell) + \ell\epsilon$$

for every $\ell \in \mathbb{N}$. Therefore, ρ belongs to the interval $[\mathbf{b}(\tilde{u}^\ell)/\ell - \delta/8, \mathbf{t}(\tilde{u}^\ell)/\ell + \delta/8]$ and this interval has length at most δ whenever $\mathbf{t}(\tilde{u}^\ell) - \mathbf{b}(\tilde{u}^\ell) \leqslant (3/4)\delta\ell$. Furthermore, Lemma 15 implies that $\kappa - \delta/8 + x = \kappa - \epsilon + x \leqslant F(x)$ and $\lambda + \delta/8 + y \geqslant F(y)$. ◄

▶ **Remark 17.** An alternative approach to compute the ergodic constant ρ of the game with Shapley operator F is to apply relative value iteration with Krasnoselskii-Mann damping, see [27], which consists in computing the sequences $y^k = F(x^{k-1}) - \mathbf{t}(F(x^{k-1}))\mathbf{1}$, and $x^k = \alpha y^k + (1-\alpha)x^{k-1}$, for any $0 < \alpha < 1$. We have $\mathbf{b}(F(x^k) - x^k) \leqslant \rho \leqslant \mathbf{t}(F(x^k) - x^k)$, and it is shown there, as a consequence of a general result on Krasnoselskii-Mann iteration in Banach spaces [13], that as soon as F admits a bias vector, $\|x^k - F(x^k)\|_H \leqslant O(1/\sqrt{k})$, see [27, Coro. 13]. In this way, we obtain an approximation of ρ with a precision ϵ in $O(1/\epsilon^2)$ iterations. In contrast, Theorem 16 provides an approximation of the same quality in only $O(1/\epsilon)$ iterations. However, note that relative value iteration with Krasnoselskii-Mann damping produces a vector $x \coloneqq x^k$ such that $\rho + x - O(1/\sqrt{k}) \leqslant F(x) \leqslant \rho + x + O(1/\sqrt{k})$ when it stops after k iterations, whereas Procedure APPROXIMATECONSTANTMEANPAYOFF returns *two* vectors x and y such that $\rho + x - O(1/k) \leqslant F(x)$ and $F(y) \leqslant \rho + y + O(1/k)$. In particular, this procedure returns a *pair (x, y)* of approximate optimality certificates. Hence, by "relaxing" the constraint that "$x = y$" in the optimality certificates, we passed from an iteration complexity of $O(1/\epsilon^2)$ to $O(1/\epsilon)$ to get an ϵ-approximation. Note also that Procedure APPROXIMATECONSTANTMEANPAYOFF differs from Krasnoselskii-Mann damping in the fact that it replaces a linear averaging of x^{k-1} and y^k by a non-linear "averaging" operation, taking a supremum or infimum of a normalized orbit, as in Lemma 15.

3.3   Deciding Whether the Value Is Independent of the Initial State

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. 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 $F^{\mathscr{S}} \colon \bar{\mathbb{R}}^{\mathscr{S}} \to \bar{\mathbb{R}}^{\mathscr{S}}$ as $F^{\mathscr{S}} \coloneqq \mathbf{p}^{\mathscr{S}} \circ F \circ \mathbf{i}^{\mathscr{S}}$, where $\mathbf{p}^{\mathscr{S}} \colon \bar{\mathbb{R}}^n \to \bar{\mathbb{R}}^{\mathscr{S}}$ is the projection on the coordinates in $\mathscr{S}$ which is defined as usual by $\mathbf{p}_j^{\mathscr{S}}(x) = x_j$ for $j \in \mathscr{S}$, and $\mathbf{i}^{\mathscr{S}} \colon \bar{\mathbb{R}}^{\mathscr{S}} \to \bar{\mathbb{R}}^n$ is defined by $\mathbf{i}_j^{\mathscr{S}}(x) = x_j$ if $j \in \mathscr{S}$ and $\mathbf{i}_j^{\mathscr{S}}(x) = -\infty$ otherwise. The next two lemmas provide elementary properties of the operators $F^{\mathscr{S}}$.

▶ **Lemma 18.** *If $\mathscr{S} \subset [n]$ is any set, then the operator $F^{\mathscr{S}}$ is continuous, order-preserving, and additively homogeneous.*

Proof. By definition, $F^{\mathscr{S}}$ is a composition of continuous, order-preserving, and additively homogeneous maps. ◄

▶ **Lemma 19.** *If $\mathscr{S}_1$ and $\mathscr{S}_2$ are two nonempty subsets of $[n]$ such that $\mathscr{S}_1 \subset \mathscr{S}_2$, then $\left(F^{\mathscr{S}_1}\right)_j^\ell(\mathbf{0}) \leqslant \left(F^{\mathscr{S}_2}\right)_j^\ell(\mathbf{0})$ for all $\ell \in \mathbb{N}$ and $j \in \mathscr{S}_1$. In particular, we have $\overline{\mathrm{cw}}(F^{\mathscr{S}_1}) \leqslant \overline{\mathrm{cw}}(F^{\mathscr{S}_2})$.*

Proof. Observe that if $x \in \bar{\mathbb{R}}^{\mathscr{S}_1}$ and $y \in \bar{\mathbb{R}}^{\mathscr{S}_2}$ are such that $x_j \leqslant y_j$ for all $j \in \mathscr{S}_1$, then $F_j^{\mathscr{S}_1}(x) \leqslant F_j^{\mathscr{S}_2}(y)$ for all $j \in \mathscr{S}_1$. Indeed, the facts that $x_j \leqslant y_j$ for all $j \in \mathscr{S}_1$ and that $\mathscr{S}_1 \subset \mathscr{S}_2$ imply $\mathbf{i}^{\mathscr{S}_1}(x) \leqslant \mathbf{i}^{\mathscr{S}_2}(y)$, and so for any $j \in \mathscr{S}_1$ we have $F_j^{\mathscr{S}_1}(x) = F_j(\mathbf{i}^{\mathscr{S}_1}(x)) \leqslant F_j(\mathbf{i}^{\mathscr{S}_2}(y)) = F_j^{\mathscr{S}_2}(y)$ because F is order-preserving. Hence, by setting $(x, y) := (\mathbf{0}, \mathbf{0})$ we get the first claim for $\ell = 1$. Then, this claim follows by induction setting $x := (F^{\mathscr{S}_1})^{\ell-1}(\mathbf{0})$ and $y := (F^{\mathscr{S}_2})^{\ell-1}(\mathbf{0})$.

The second claim follows from the first claim and Theorem 5. ◄

▶ **Definition 20.** *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 [8, 4], 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 [34], in which a dominion is required in addition to consist only of initial states that are winning for this player.

▶ **Lemma 21.** *A set $\mathscr{D} \subset [n]$ is a dominion if and only if $F^{\mathscr{D}}(\mathbf{0}) \in \mathbb{R}^{\mathscr{D}}$.*

Proof. If $\mathscr{D} \subset [n]$ is a dominion, then $F^{\mathscr{D}}(\mathbf{0}) \in \mathbb{R}^{\mathscr{D}}$. Conversely, given any $x \in \mathbb{R}^{\mathscr{D}}$, we have $F^{\mathscr{D}}(x) \geqslant \mathbf{b}(x) + F^{\mathscr{D}}(\mathbf{0})$ because $x \geqslant \mathbf{b}(x) + \mathbf{0}$ and $F^{\mathscr{D}}$ is order-preserving and additively homogeneous. Thus, if $F^{\mathscr{D}}(\mathbf{0}) \in \mathbb{R}^{\mathscr{D}}$, we conclude that $F^{\mathscr{D}}(x) \in \mathbb{R}^{\mathscr{D}}$ for all $x \in \mathbb{R}^{\mathscr{D}}$. ◄

The procedures 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 $\chi^{\mathscr{D}} := \lim_{\ell \to \infty} \frac{(F^{\mathscr{D}})^\ell(\mathbf{0})}{\ell} \in \mathbb{R}^{\mathscr{D}}$ exists for every dominion $\mathscr{D} \subset [n]$. Furthermore, we assume that the set $\mathscr{D}_{\max} := \{j \in [n]\colon \chi_j^{[n]} = \overline{\mathrm{cw}}(F)\}$ 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 22.** 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 23.** 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 $\chi^{\mathscr{D}} = \lim_{\ell \to \infty} \frac{(F^{\mathscr{D}})^\ell(\mathbf{0})}{\ell}$ exists will allow us to apply value iteration to the Shapley operator of the reduced game, $F^{\mathscr{D}}$.

▶ **Lemma 24.** *If $\mathscr{D}_1$ and $\mathscr{D}_2$ are two dominions such that $\mathscr{D}_1 \subset \mathscr{D}_2$, then $\chi_j^{\mathscr{D}_1} \leqslant \chi_j^{\mathscr{D}_2}$ for all $j \in \mathscr{D}_1$.*

Proof. Since $\chi^{\mathscr{D}_1} = \lim_{\ell \to \infty} \frac{(F^{\mathscr{D}_1})^\ell(\mathbf{0})}{\ell}$ and $\chi^{\mathscr{D}_2} = \lim_{\ell \to \infty} \frac{(F^{\mathscr{D}_2})^\ell(\mathbf{0})}{\ell}$, the lemma readily follows from Lemma 19. ◄

```
1: procedure DECIDECONSTANTVALUE(F, δ, R)
2:     u := 0 ∈ ℝⁿ, ℓ := 0 ∈ ℕ.
3:     F̃ := any map such that F̃(u) is at most at distance ϵ := δ/8 in the sup-norm from F(u).
4:     repeat u := F̃(u); ℓ := ℓ + 1
5:     until t(u) − b(u) ⩽ (3/4)δℓ or ℓ = 1 + ⌈8R/δ⌉
6:     if ℓ = 1 + ⌈8R/δ⌉ then
7:         𝒮 := {i: uᵢ = b(u)}
8:         return 𝒮     ▷ The value of the game depends on the initial state. We have χᵢ < cw̄(F) for all
           i ∈ 𝒮.
9:     else
10:        return ∅     ▷ The value of the game is independent of the initial state.
11:    end
12: end
```

■ **Figure 4** Algorithm that decides if the value is constant.

▶ **Lemma 25.** *Suppose that $\mathscr{D} \subset [n]$ is a dominion that contains the set of states of maximal value, i.e., $\{j \in [n]: \chi_j = \overline{\mathrm{cw}}(F)\} \subset \mathscr{D}$. Then, $\overline{\mathrm{cw}}(F) = \overline{\mathrm{cw}}(F^{\mathscr{D}})$ and $\{j \in [n]: \chi_j = \overline{\mathrm{cw}}(F)\} = \{j \in \mathscr{D}: \chi_j^{\mathscr{D}} = \overline{\mathrm{cw}}(F^{\mathscr{D}})\}$.*

Proof. We have $\chi_j^{\mathscr{D}} \leqslant \chi_j$ for all $j \in \mathscr{D}$ by Lemma 24. Let $\mathscr{D}_{\max} := \{j \in [n]: \chi_j = \overline{\mathrm{cw}}(F)\} \subset \mathscr{D}$. Then, $\mathscr{D}_{\max}$ is a dominion of the operator F by Assumption 2. Moreover, since $\mathscr{D}_{\max} \subset \mathscr{D}$, we have $\chi_j^{\mathscr{D}_{\max}} \leqslant \chi_j^{\mathscr{D}} \leqslant \chi_j = \overline{\mathrm{cw}}(F)$ for all $j \in \mathscr{D}_{\max}$ by Lemma 24. As $\chi_j^{\mathscr{D}_{\max}} = \overline{\mathrm{cw}}(F)$ for all $j \in \mathscr{D}_{\max}$ by Assumption 2, it follows that $\chi_j^{\mathscr{D}} = \overline{\mathrm{cw}}(F)$ for all $j \in \mathscr{D}_{\max}$. Furthermore, if $j \notin \mathscr{D}_{\max}$, then $\chi_j^{\mathscr{D}} \leqslant \chi_j < \overline{\mathrm{cw}}(F)$, which finishes the proof. ◀

From now on, we denote $\chi := \chi^{[n]}$ and $\mathscr{D}_{\max} := \{j \in [n]: \chi_j = \overline{\mathrm{cw}}(F)\}$. 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 Assumption 2.

▶ **Theorem 26.** *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 (Figure 4) is correct.*

The proof relies on the following lemma.

▶ **Lemma 27.** *Let $\tilde{F}: \mathbb{R}^n \to \mathbb{R}^n$ be such that $\|\tilde{F}(x) - F(x)\|_\infty \leqslant \epsilon$ for all $x \in \mathbb{R}^n$. Suppose that $\mathscr{S} \subset [n]$, $v \in \mathbb{R}^{\mathscr{S}}$ and $\gamma > 0$ are such that $\overline{\mathrm{cw}}(F) - \gamma + v \leqslant F^{\mathscr{S}}(v)$. Then, for all $\ell \in \mathbb{N}$ and $j \in \mathscr{S}$, we have $\tilde{F}_j^\ell(\mathbf{0}) \geqslant \ell(\overline{\mathrm{cw}}(F) - \gamma - \epsilon) - \|v\|_{\mathrm{H}}$.*

Proof. In the first place, note that $\overline{\mathrm{cw}}(F) - \gamma + \mathbf{i}^{\mathscr{S}}(v) \leqslant F(\mathbf{i}^{\mathscr{S}}(v))$. Indeed, if $j \in \mathscr{S}$ we have $\overline{\mathrm{cw}}(F) - \gamma + \mathbf{i}_j^{\mathscr{S}}(v) \leqslant F_j(\mathbf{i}^{\mathscr{S}}(v))$ because $\mathbf{i}_j^{\mathscr{S}}(v) = v_j$ and $F_j(\mathbf{i}^{\mathscr{S}}(v)) = F_j^{\mathscr{S}}(v)$, and if $j \notin \mathscr{S}$ we also have $\overline{\mathrm{cw}}(F) - \gamma + \mathbf{i}_j^{\mathscr{S}}(v) \leqslant F_j(\mathbf{i}^{\mathscr{S}}(v))$ because $\mathbf{i}_j^{\mathscr{S}}(v) = -\infty$.

Now, since $\mathbf{0} \geqslant -\mathbf{t}(v) + \mathbf{i}^{\mathscr{S}}(v)$ (note that $\mathbf{t}(v) \neq -\infty$ because $v \in \mathbb{R}^{\mathscr{S}}$), we get $F^\ell(\mathbf{0}) \geqslant -\mathbf{t}(v) + F^\ell(\mathbf{i}^{\mathscr{S}}(v)) \geqslant -\mathbf{t}(v) + \ell(\overline{\mathrm{cw}}(F) - \gamma) + \mathbf{i}^{\mathscr{S}}(v)$. Hence, for every $j \in \mathscr{S}$ we have $F_j^\ell(\mathbf{0}) \geqslant -\mathbf{t}(v) + \ell(\overline{\mathrm{cw}}(F) - \gamma) + \mathbf{i}_j^{\mathscr{S}}(v) = -\mathbf{t}(v) + \ell(\overline{\mathrm{cw}}(F) - \gamma) + v_j \geqslant \ell(\overline{\mathrm{cw}}(F) - \gamma) - \|v\|_{\mathrm{H}}$. Therefore, by Lemma 14, we conclude that $\tilde{F}_j^\ell(\mathbf{0}) \geqslant \ell(\overline{\mathrm{cw}}(F) - \gamma - \epsilon) - \|v\|_{\mathrm{H}}$. ◀

Proof of Theorem 26. In the first place, note that vectors $v, w \in \mathbb{R}^{\mathscr{D}_{\max}}$ satisfying the condition of the theorem exist owing to Assumption 2.

```
 1: procedure EXTEND(𝒟, 𝒮)
 2:    ▷ 𝒮 ⊂ 𝒟 is a nonempty set
 3:    while True do
 4:        𝒮' := ∅
 5:        define x ∈ ℝ̄^𝒟 as x_j := −∞ for j ∈ 𝒮 and x_j = 0 otherwise
 6:        for j ∈ 𝒟 do
 7:            if F_j^𝒟(x) = −∞  then
 8:                𝒮' := 𝒮' ∪ {j}
 9:            end
10:        done
11:        if 𝒮 ∪ 𝒮' ≠ 𝒮 then
12:            𝒮 := 𝒮 ∪ 𝒮'
13:        else
14:            return 𝒮
15:        end
16:    done
17: end
```

■ **Figure 5** Procedure that extends the set laying outside some dominion.

Denote $\tilde{u}^\ell := \tilde{F}^\ell(\mathbf{0})$ for $\ell \in \mathbb{N}$. As in the proof of Theorem 16, we note that Lemmas 11 and 14 imply that we have the inequality

$$\mathbf{b}(\tilde{u}^\ell) - \ell(\delta/8) \leqslant \ell\underline{\mathrm{cw}}(F) \leqslant \ell\overline{\mathrm{cw}}(F) \leqslant \mathbf{t}(\tilde{u}^\ell) + \ell(\delta/8). \tag{15}$$

Hence, if the condition $\mathbf{t}(\tilde{u}^\ell) - \mathbf{b}(\tilde{u}^\ell) \leqslant (3/4)\delta\ell$ is satisfied for some ℓ, then $\overline{\mathrm{cw}}(F) - \underline{\mathrm{cw}}(F) \leqslant \delta$ and so $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$ by our assumption on δ. Conversely, if $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$, then $\mathscr{D}_{\max} = [n]$ and Theorem 16 shows that the condition $\mathbf{t}(\tilde{u}^\ell) - \mathbf{b}(\tilde{u}^\ell) \leqslant (3/4)\delta\ell$ is satisfied for some $\ell \leqslant \lceil 8R/\delta \rceil$. Hence, the algorithm correctly decides if the value of the game is constant.

To finish the proof, suppose the value of the game is not constant, so $\overline{\mathrm{cw}}(F) - \underline{\mathrm{cw}}(F) > \delta$ by our assumption on δ, and let $\ell := \lceil 8R/\delta \rceil + 1$. Then, by Lemma 27 for any $j \in \mathscr{D}_{\max}$ we have $\tilde{u}_j^\ell \geqslant \ell(\overline{\mathrm{cw}}(F) - \delta/8 - \epsilon) - \|v\|_{\mathrm{H}} \geqslant \ell(\overline{\mathrm{cw}}(F) - \delta/8 - \epsilon) - R > \ell\overline{\mathrm{cw}}(F) - 3\ell(\delta/8)$ since $R < \ell(\delta/8)$ and $\epsilon = \delta/8$ in Procedure DECIDECONSTANTVALUE. On the other hand, (15) implies $\mathbf{b}(\tilde{u}^\ell) - \ell(\delta/8) \leqslant \ell\underline{\mathrm{cw}}(F) < \ell(\overline{\mathrm{cw}}(F) - \delta)$. Hence, we have $\mathbf{b}(\tilde{u}^\ell) < \ell\overline{\mathrm{cw}}(F) - 7\ell(\delta/8) < \tilde{u}_j^\ell$. We conclude that, if we take any $k \in [n]$ such that $\tilde{u}_k = \mathbf{b}(\tilde{u}^\ell)$, then $k \notin \mathscr{D}_{\max}$.   ◀

3.4 Finding the States of Maximal Value

We now refine Theorem 26, showing in Theorem 30 below that we can in fact extract all the initial states with maximal value. To this end, we introduce and analyze two procedures, EXTEND and TOPCLASS, which are used to obtain Theorem 30.

▶ **Lemma 28.** *Procedure* EXTEND *(Figure 5) has the following properties. Let $\mathscr{D} \subset [n]$ be a dominion and $\mathscr{S} \subset \mathscr{D}$ be a non-empty set. If $\mathscr{D}'$ is a dominion such that $\mathscr{D}' \subset \mathscr{D}$ and $\mathscr{S} \cap \mathscr{D}' = \emptyset$, then* EXTEND$(\mathscr{D}, \mathscr{S}) \cap \mathscr{D}' = \emptyset$. *Besides, $\mathscr{D} \setminus$* EXTEND$(\mathscr{D}, \mathscr{S})$ *is a dominion.*

Proof. At each execution of the while loop, the procedure either stops or strictly increases the cardinality of $\mathscr{S} \subset \mathscr{D}$. Therefore, the procedure stops after at most $|\mathscr{D}| - 1$ executions of the while loop.

To prove the first property of the procedure, let $\mathscr{D}' \subset \mathscr{D}$ be a dominion such that $\mathscr{S} \cap \mathscr{D}' = \emptyset$. Since $\mathscr{S} \cap \mathscr{D}' = \emptyset$ and $\mathscr{D}' \subset \mathscr{D}$, we have $\mathbf{i}^{\mathscr{D}'}(\mathbf{0}) \leqslant \mathbf{i}^{\mathscr{D}}(x)$, where x is the vector defined in the while loop of the procedure (i.e., $x := \mathbf{i}^{\mathscr{D}\setminus\mathscr{S}}(\mathbf{0})$). Hence, $F(\mathbf{i}^{\mathscr{D}'}(\mathbf{0})) \leqslant F(\mathbf{i}^{\mathscr{D}}(x))$ because F is order-preserving. Since $\mathscr{D}'$ is a dominion, we get $-\infty < F_j^{\mathscr{D}'}(\mathbf{0}) = F_j(\mathbf{i}^{\mathscr{D}'}(\mathbf{0})) \leqslant F_j(\mathbf{i}^{\mathscr{D}}(x)) = F_j^{\mathscr{D}}(x)$ for all $j \in \mathscr{D}'$. Therefore, $\mathscr{S}' \cap \mathscr{D}' = \emptyset$ and so $(\mathscr{S} \cup \mathscr{S}') \cap \mathscr{D}' = \emptyset$. In

```
 1: procedure TopClass(F, δ, R)
 2:     𝒟 := [n]
 3:     while True do
 4:         𝒮 := DecideConstantValue(F^𝒟, δ, R)
 5:         if 𝒮 = ∅ then
 6:             return 𝒟      ▷ 𝒟 is the set of states that have the maximal value
 7:         end
 8:         𝒟 := 𝒟 \ Extend(𝒟, 𝒮)
 9:     done
10: end
```

■ **Figure 6** Procedure that finds the set of states with maximal value.

other words, the set $\mathscr{S} \cup \mathscr{S}'$ obtained by a single execution of the while loop in the procedure is disjoint from $\mathscr{D}'$. Therefore, we have $\text{Extend}(\mathscr{D}, \mathscr{S}) \cap \mathscr{D}' = \emptyset$.

To prove the second property, let $\mathscr{D}' := \mathscr{D} \setminus \text{Extend}(\mathscr{D}, \mathscr{S})$, and let $x \in \bar{\mathbb{R}}^n$ be the vector defined as $x_j = -\infty$ for $j \in \text{Extend}(\mathscr{D}, \mathscr{S})$ and $x_j = 0$ otherwise (i.e., let $x := \mathbf{i}^{\mathscr{D}'}(\mathbf{0})$). The stopping criterion of the procedure implies that $\{j \in \mathscr{D} : F_j^{\mathscr{D}}(\mathbf{p}^{\mathscr{D}}(x)) = -\infty\} \subset \text{Extend}(\mathscr{D}, \mathscr{S})$. Since $x = \mathbf{i}^{\mathscr{D}}(\mathbf{p}^{\mathscr{D}}(x))$, we have $F_j^{\mathscr{D}}(\mathbf{p}^{\mathscr{D}}(x)) = F_j(x)$ for every $j \in \mathscr{D}$. Therefore, $\{j \in \mathscr{D} : F_j(x) = -\infty\} \subset \text{Extend}(\mathscr{D}, \mathscr{S})$ and so $F_j(x) > -\infty$ for all $j \in \mathscr{D}'$. In particular, we have $F_j^{\mathscr{D}'}(\mathbf{0}) = F_j(\mathbf{i}^{\mathscr{D}'}(\mathbf{0})) = F_j(x) > -\infty$ for all $j \in \mathscr{D}'$. Hence, $F^{\mathscr{D}'}(\mathbf{0}) \in \mathbb{R}^{\mathscr{D}'}$ and $\mathscr{D}'$ is a dominion by Lemma 21. ◀

Let

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

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

▶ **Theorem 29.** *Let $\delta > 0$ be such that $\delta < \text{sep}(F)$, $\mathscr{D}_{\max}$ 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 $\overline{\text{cw}}(F) - \delta/8 + v \leqslant F^{\mathscr{D}_{\max}}(v)$ and $\overline{\text{cw}}(F) + \delta/8 + w \geqslant F^{\mathscr{D}_{\max}}(w)$. Then, Procedure TopClass(F, δ, R) (Figure 6) halts after at most n iterations of the while loop, and correctly computes the set of initial states with maximal value.*

Proof. Fix any dominion $\mathscr{D} \subset [n]$ such that $\mathscr{D}_{\max} \subset \mathscr{D}$ and consider two cases. If $\mathscr{D} = \mathscr{D}_{\max}$, then we have $\overline{\text{cw}}(F) = \overline{\text{cw}}(F^{\mathscr{D}}) = \underline{\text{cw}}(F^{\mathscr{D}})$ by Assumption 2. By applying Theorem 26 to $F^{\mathscr{D}}$, Procedure DecideConstantValue$(F^{\mathscr{D}}, \delta, R)$ outputs $\emptyset$. If $\mathscr{D} \neq \mathscr{D}_{\max}$, then we have $\overline{\text{cw}}(F) = \overline{\text{cw}}(F^{\mathscr{D}}) \neq \underline{\text{cw}}(F^{\mathscr{D}})$ by Lemma 25. Thus, $\overline{\text{cw}}(F^{\mathscr{D}}) - \underline{\text{cw}}(F^{\mathscr{D}}) \geqslant \text{sep}(F) > \delta$ by the definition of $\text{sep}(F)$. Hence, by applying Theorem 26 to $F^{\mathscr{D}}$, Procedure DecideConstantValue$(F^{\mathscr{D}}, \delta, R)$ outputs some nonempty set $\mathscr{S} \subset \mathscr{D}$ such that $\chi_i^{\mathscr{D}} < \overline{\text{cw}}(F)$ for all $i \in \mathscr{S}$. In particular, $\mathscr{S} \cap \mathscr{D}_{\max} = \emptyset$ by Lemma 25. Therefore, by Lemma 28, the set $\mathscr{D} \setminus \text{Extend}(\mathscr{D}, \mathscr{S})$ is a dominion that contains $\mathscr{D}_{\max}$ and is strictly smaller than $\mathscr{D}$. By induction, if $\mathscr{D}_\ell$ denotes the set $\mathscr{D}$ at the ℓth iteration of the while loop in TopClass(F, δ, R), then we have $[n] = \mathscr{D}_1 \supsetneq \mathscr{D}_2 \supsetneq \cdots \supset \mathscr{D}_{\max}$ and all of the sets $\mathscr{D}_\ell$ are dominions. Furthermore, Procedure TopClass(F, δ, R) stops only when it finds a set $\mathscr{D}_p$ such that $\mathscr{D}_p = \mathscr{D}_{\max}$ and outputs $\mathscr{D}_p$ as the set of states with maximal value. ◀

To state the final result of this section, we will suppose that we have access to an oracle that approximates (the canonical extension (8) of) 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 $y_j = -\infty \iff F_j(x) = -\infty$ and $|F_j(x) - y_j| \leqslant \epsilon$ for all j such that $F_j(x) \neq -\infty$.

▶ **Theorem 30.** *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 an oracle that approximates F to precision $\epsilon := \delta/8$.*

Proof. Denote the oracle by $\tilde{F} \colon \bar{\mathbb{R}}^n \to \bar{\mathbb{R}}^n$ and let $\mathscr{D}_1, \dots, \mathscr{D}_p$ be as in the proof of Theorem 29. If $\mathscr{D} := \mathscr{D}_\ell$ for some $\ell \in [p]$, then $|F_j^{\mathscr{D}}(x) - \tilde{F}_j(\mathbf{i}^{\mathscr{D}}(x))| \leqslant \delta/8$ for every $x \in \mathbb{R}^{\mathscr{D}}$ and $j \in \mathscr{D}$. Indeed, since $\mathscr{D}$ is a dominion, we have $F_j^{\mathscr{D}}(x) = F_j(\mathbf{i}^{\mathscr{D}}(x)) > -\infty$ for all $j \in \mathscr{D}$ and therefore $|F_j^{\mathscr{D}}(x) - \tilde{F}_j(\mathbf{i}^{\mathscr{D}}(x))| \leqslant \delta/8$ by the definition of $\tilde{F}$. Hence, the map that approximates $F^{\mathscr{D}} \colon \mathbb{R}^{\mathscr{D}} \to \mathbb{R}^{\mathscr{D}}$ as required by Procedure DECIDECONSTANTVALUE$(F^{\mathscr{D}}, \delta, R)$ is obtained by calling $\tilde{F}$. Furthermore, Procedure DECIDECONSTANTVALUE$(F^{\mathscr{D}}, \delta, R)$ finishes after making at most $1 + \lceil 8R/\delta \rceil$ calls to the oracle. Moreover, for every $x \in \bar{\mathbb{R}}^{\mathscr{D}}$ and $j \in \mathscr{D}$ we have $F_j^{\mathscr{D}}(x) = -\infty \iff \tilde{F}_j(\mathbf{i}^{\mathscr{D}}(x)) = -\infty$ by the definition of $\tilde{F}$. Therefore, we can replace every call to $F^{\mathscr{D}}(x)$ in EXTEND$(\mathscr{D}, \mathscr{S})$ by a call to $\tilde{F}(\mathbf{i}^{\mathscr{D}}(x))$. We also note that EXTEND$(\mathscr{D}, \mathscr{S})$ makes at most $n-1$ calls to the oracle. Therefore, a single iteration of the while loop in TOPCLASS(F, δ, R) can be done by making at most $n + \lceil 8R/\delta \rceil$ calls to the oracle. The claim follows from the fact that TOPCLASS finishes after at most n iterations of the while loop. ◀

4 Application to Stochastic Mean-Payoff Games

In this section, we apply our results to stochastic mean-payoff games. We follow the notation concerning stochastic mean-payoff games as introduced in Section 2.1. In particular, we suppose that F is of the form given in (1). Before starting, we note that the continuous extension of F to $\bar{\mathbb{R}}^{\mathscr{V}_{\mathrm{Min}}}$ is still given by the same formula (1), with the convention that $0 \cdot (-\infty) = 0$. Indeed, under this convention, the operations min, max, and $x \to \alpha \cdot x$ for $\alpha \geqslant 0$ are continuous in $\bar{\mathbb{R}}$.

4.1 Dominions of Stochastic Mean-Payoff Games

We first describe the dominions in the special case of stochastic mean-payoff games, and deduce that this class of games satisfies Assumption 2.

Intuitively speaking, $\mathscr{D} \subset \mathscr{V}_{\mathrm{Min}}$ is a dominion if player Max can force the game to stay in $\mathscr{D}$ provided that the initial state of the game belongs to $\mathscr{D}$. To make this more precise, we use the following notation. If $\mathscr{S} \subset \mathscr{V}_{\mathrm{Min}}$ is any subset, then we denote by $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{S}} \subset \mathscr{V}_{\mathrm{Nat}}$ the set of vertices controlled by Nature whose all outgoing edges go to $\mathscr{S}$, i.e., $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{S}} := \{k \in \mathscr{V}_{\mathrm{Nat}} \colon \sum_{l \in \mathscr{S}} P_{kl} = 1\}$. Moreover, we denote by $\mathscr{V}_{\mathrm{Max}}^{\mathscr{S}} \subset \mathscr{V}_{\mathrm{Max}}$ the set of vertices controlled by Max which have at least one outgoing edge that goes to $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{S}}$, i.e., $\mathscr{V}_{\mathrm{Max}}^{\mathscr{S}} := \{i \in \mathscr{V}_{\mathrm{Max}} \colon \exists k \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{S}}, (i, k) \in \mathscr{E}\}$. Furthermore, we denote by $(\mathscr{V}^{\mathscr{S}}, \mathscr{E}^{\mathscr{S}})$ the subgraph of $(\mathscr{V}, \mathscr{E})$ induced by $\mathscr{S} \uplus \mathscr{V}_{\mathrm{Max}}^{\mathscr{S}} \uplus \mathscr{V}_{\mathrm{Nat}}^{\mathscr{S}}$.

▶ **Lemma 31.** *A set $\mathscr{D} \subset \mathscr{V}_{\mathrm{Min}}$ is a dominion of a stochastic mean-payoff game if and only if all edges leaving $\mathscr{D}$ go to $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$, i.e., $\{i \in \mathscr{V}_{\mathrm{Max}} \colon \exists j \in \mathscr{D}, (j, i) \in \mathscr{E}\} \subset \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$. In particular, if $\mathscr{D}$ is a dominion, then the sets $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}}$ and $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$ are nonempty. Furthermore, if $\mathscr{D}$ is a dominion, then we have the equality*

$$F_j^{\mathscr{D}}(x) = \min_{(j,i) \in \mathscr{E}^{\mathscr{S}}} \Big(-A_{ij} + \max_{(i,k) \in \mathscr{E}^{\mathscr{S}}} \big(B_{ik} + \sum_{(k,l) \in \mathscr{E}^{\mathscr{S}}} P_{kl} x_l \big) \Big) \ \textit{for all } j \in \mathscr{D} \,.$$

Proof. Let $\mathscr{D} \subset \mathscr{V}_{\mathrm{Min}}$ be any set. In the first place, note that for every $x \in \bar{\mathbb{R}}^{\mathscr{D}}$, by the definition of $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}}$, for every $k \in \mathscr{V}_{\mathrm{Nat}}$ we have

$$\sum_{(k,l) \in \mathscr{E}} P_{kl} \mathbf{i}^{\mathscr{D}}(x)_l = \begin{cases} \sum_{(k,l) \in \mathscr{E}^{\mathscr{D}}} P_{kl} x_l & \text{if } k \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}}, \\ -\infty & \text{otherwise.} \end{cases}$$

Thus, by the definition of $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$, it follows that

$$\max_{(i,k) \in \mathscr{E}} \Big(B_{ik} + \sum_{(k,l) \in \mathscr{E}} P_{kl} \mathbf{i}^{\mathscr{D}}(x)_l \Big) = \begin{cases} \max_{(i,k) \in \mathscr{E}^{\mathscr{D}}} \big(B_{ik} + \sum_{(k,l) \in \mathscr{E}^{\mathscr{D}}} P_{kl} x_l \big) & \text{if } i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}, \\ -\infty & \text{otherwise,} \end{cases}$$

for all $i \in \mathscr{V}_{\mathrm{Max}}$. We conclude that

$$F_j\big(\mathbf{i}^{\mathscr{D}}(x)\big) = \tag{16}$$
$$\begin{cases} \displaystyle \min_{(j,i) \in \mathscr{E}^{\mathscr{D}}} \Big(-A_{ij} + \max_{(i,k) \in \mathscr{E}^{\mathscr{D}}} \big(B_{ik} + \sum_{(k,l) \in \mathscr{E}^{\mathscr{D}}} P_{kl} x_l \big) \Big) & \text{if } i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}} \text{ for all } (j,i) \in \mathscr{E}, \\ -\infty & \text{otherwise,} \end{cases}$$

for all $j \in \mathscr{V}_{\mathrm{Min}}$.

If we assume that $\mathscr{D}$ is a dominion, we necessarily have that $i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$ for every $j \in \mathscr{D}$ and $i \in \mathscr{V}_{\mathrm{Max}}$ such that $(j,i) \in \mathscr{E}$, because otherwise (16) would imply that $F_j^{\mathscr{D}}(\mathbf{0}) = -\infty$ for some $j \in \mathscr{D}$, contradicting Lemma 21. Then, when $\mathscr{D}$ is a dominion, (16) also shows that

$$F_j^{\mathscr{D}}(x) = \min_{(j,i) \in \mathscr{E}^{\mathscr{D}}} \Big(-A_{ij} + \max_{(i,k) \in \mathscr{E}^{\mathscr{D}}} \big(B_{ik} + \sum_{(k,l) \in \mathscr{E}^{\mathscr{D}}} P_{kl} x_l \big) \Big)$$

for every $x \in \bar{\mathbb{R}}^{\mathscr{D}}$ and $j \in \mathscr{D}$.

Assume now that $\mathscr{D}$ has the property that for every $j \in \mathscr{D}$ and $i \in \mathscr{V}_{\mathrm{Max}}$ such that $(j,i) \in \mathscr{E}$ we have $i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$. Then, by (16) we have $F_j^{\mathscr{D}}(\mathbf{0}) = \min_{(j,i) \in \mathscr{E}^{\mathscr{D}}} \big(-A_{ij} + \max_{(i,k) \in \mathscr{E}^{\mathscr{D}}} (B_{ik}) \big)$ for every $j \in \mathscr{D}$. Besides, since every vertex in $\mathscr{V}_{\mathrm{Min}}$ has at least one outgoing edge by Assumption 1, for every $j \in \mathscr{D}$ there exists $i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$ such that $(j,i) \in \mathscr{E}^{\mathscr{D}}$. It follows that $F_j^{\mathscr{D}}(\mathbf{0}) > -\infty$ for every $j \in \mathscr{D}$ because $\max_{k \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}}}(B_{ik}) > -\infty$ for every $i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$ by the definition of $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$. Thus, by Lemma 21 we conclude that $\mathscr{D}$ is a dominion. ◀

Lemma 31 implies that if $\mathscr{D}$ is a dominion, then $F^{\mathscr{D}}$ is the Shapley operator of a stochastic mean-payoff game on the subgraph $(\mathscr{V}^{\mathscr{D}}, \mathscr{E}^{\mathscr{D}})$. We refer to this game as the *subgame induced by $\mathscr{D}$*. We note that this subgame satisfies Assumption 1.

▶ **Lemma 32.** *If $\mathscr{D} \subset \mathscr{V}_{\mathrm{Min}}$ is a dominion, then the subgame induced by $\mathscr{D}$ satisfies Assumption 1.*

Proof. Lemma 31 implies that every vertex $j \in \mathscr{D}$ has an outgoing edge to a vertex $i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$. The vertices in $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}}$ and $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}}$ have at least one outgoing edge in $\mathscr{E}^{\mathscr{D}}$ by definition. ◀

The following lemma characterizes the recession operators of Shapley operators of stochastic mean-payoff games. Recall that if F is a self-map of $\mathbb{R}^n$, the *recession operator* of F is defined by

$$\hat{F}(x) := \lim_{\beta \to \infty} \beta^{-1} F(\beta x), \qquad x \in \mathbb{R}^n .$$

It is immediate that when G and H are Lipschitz continuous self-maps of $\mathbb{R}^n$ that admit recession functions, so does $G \circ H$, and $\widehat{G \circ H} = \hat{G} \circ \hat{H}$. Moreover, if F is a self-map of $\mathbb{R}^n$ that is order-preserving and additively homogeneous, so is $\hat{F}$. All the Shapley operators of the concrete games considered in this paper admit recession functions. In particular:

▶ **Lemma 33.** *For every $j \in \mathscr{V}_{\mathrm{Min}}$ we have*

$$\hat{F}_j(x) = \min_{(j,i)\in\mathscr{E}} \max_{(i,k)\in\mathscr{E}} \sum_{(k,l)\in\mathscr{E}} P_{kl}x_l \,. \qquad \blacktriangleleft$$

▶ **Lemma 34.** *The Shapley operators of stochastic mean-payoff games satisfy Assumption 2.*

Proof. The fact that F satisfies the first part of Assumption 2 follows from Remark 22. Alternatively, to prove this part it is enough to note that F has an escape rate by Theorem 4 and, by Lemma 31, the same is true for the operator $F^{\mathscr{D}}$ when $\mathscr{D}$ is a dominion.

In order to prove the second part of Assumption 2, let τ be an optimal strategy of Max in the stochastic mean-payoff game described by F. Consider the Shapley operator

$$F_j^\tau(x) := \min_{(j,i)\in\mathscr{E}} \Big(-A_{ij} + B_{i\tau(i)} + \sum_{(\tau(i),l)\in\mathscr{E}} P_{\tau(i)l}x_l \Big),$$

of the game in which Max plays according to τ. Since τ is optimal, we have $\chi = \lim_{\ell\to\infty} \frac{(F^\tau)^\ell(\mathbf{0})}{\ell}$. Let $\hat{F}^\tau$ be the recession operator of F^τ. By Lemma 33 we have

$$\hat{F}_j^\tau(x) = \min_{(j,i)\in\mathscr{E}} \sum_{(\tau(i),l)\in\mathscr{E}} P_{\tau(i)l}x_l$$

Furthermore, by a standard observation, $\hat{F}^\tau(\chi) = \chi$, see e.g. [45] or [3, Prop. 3.1]. Note that $\sum_{(k,l)\in\mathscr{E}} P_{kl}\chi_l = \overline{\mathrm{cw}}(F)$ for every $k \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$ and $\sum_{(k,l)\in\mathscr{E}} P_{kl}\chi_l < \overline{\mathrm{cw}}(F)$ for all $k \in \mathscr{V}_{\mathrm{Nat}} \setminus \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$. Therefore, the equality $\hat{F}^\tau(\chi) = \chi$ implies that for every $j \in \mathscr{D}_{\max}$ and every $i \in \mathscr{V}_{\mathrm{Max}}$ such that $(j,i) \in \mathscr{E}$ we have $\tau(i) \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$. In particular, every such i belongs to $\mathscr{V}_{\mathrm{Max}}^{\mathscr{D}_{\max}}$ and $\mathscr{D}_{\max}$ is a dominion by Lemma 31.

To complete the proof, recall that for all $j \in \mathscr{D}_{\max}$ we have $\chi_j^{\mathscr{D}_{\max}} \leqslant \overline{\mathrm{cw}}(F)$ by Lemma 24. To prove that $\overline{\mathrm{cw}}(F) \leqslant \chi_j^{\mathscr{D}_{\max}}$ for $j \in \mathscr{D}_{\max}$, let $\sigma^{\mathscr{D}_{\max}} : \mathscr{D}_{\max} \to \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}_{\max}}$ be an optimal strategy of Min in the subgame induced by $\mathscr{D}_{\max}$ and let $\sigma : \mathscr{V}_{\mathrm{Min}} \to \mathscr{V}_{\mathrm{Max}}$ be any strategy of Min that agrees with $\sigma^{\mathscr{D}_{\max}}$ on $\mathscr{D}_{\max}$. Likewise, let $\tau^{\mathscr{D}_{\max}} : \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}_{\max}} \to \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$ be any strategy of Max in the subgame induced by $\mathscr{D}_{\max}$ that agrees with τ on the set $\{i \in \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}_{\max}} : \exists j \in \mathscr{D}_{\max}, (j,i) \in \mathscr{E}\}$. Note that there is at least one such strategy, because we have shown above that $\tau(i) \in \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$ for any i that belongs to this set. Furthermore, if the initial state j of the game belongs to $\mathscr{D}_{\max}$ and the players use the strategies (σ, τ), then the game never leaves the set of states $\mathscr{D}_{\max} \uplus \mathscr{V}_{\mathrm{Max}}^{\mathscr{D}_{\max}} \uplus \mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$ by the definition of $\mathscr{V}_{\mathrm{Nat}}^{\mathscr{D}_{\max}}$. Even more, (σ, τ) and $(\sigma^{\mathscr{D}_{\max}}, \tau^{\mathscr{D}_{\max}})$ generate the same probability measures on the possible trajectories starting at j. In particular, we have $g_j(\sigma, \tau) = g_j(\sigma^{\mathscr{D}_{\max}}, \tau^{\mathscr{D}_{\max}})$ for all $j \in \mathscr{D}_{\max}$. Hence, the optimality of $\sigma^{\mathscr{D}_{\max}}$ and τ implies $\chi_j^{\mathscr{D}_{\max}} \geqslant g_j(\sigma^{\mathscr{D}_{\max}}, \tau^{\mathscr{D}_{\max}}) = g_j(\sigma, \tau) \geqslant \chi_j = \overline{\mathrm{cw}}(F)$. ◀

4.2 Bit-Complexity Estimates for 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 payoffs A_{ij} and B_{ik} are integers. This is not more special than assuming that they are rational numbers (we may always rescale rational payments so that they become integers). We set

$$W := \max \left\{ |A_{ij} - B_{ik}| : i \in \mathscr{V}_{\mathrm{Max}}, j \in \mathscr{V}_{\mathrm{Min}}, k \in \mathscr{V}_{\mathrm{Nat}} \right\} . \qquad (17)$$

We also assume that the probabilities P_{kj} are rational, and that they have a common denominator $M \in \mathbb{N}_{>0}$, $P_{kj} = n_{kj}/M$, where $n_{kj} \in \{0, \dots, M\}$ 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 $s(P)$ (or simply s, when P is clear from the context) the number of significant random states. The following separation bound improves an estimate in [18].

▶ **Theorem 35.** *We have* $\mathrm{sep}(F) > 1/(nM^{\min\{s, n-1\}})^2$.

We will deduce this theorem from an optimal bit-complexity result for Markov chains, established in [51].

▶ **Lemma 36** ([51]). *Suppose that a Markov chain with transition matrix Q and n states is irreducible, and that the transition probabilities are rational numbers whose denominators divide the integer M. Let $s := s(Q)$, and let $\pi \in (0, 1]^{n \times n}$ be the invariant measure of the chain. Then, the least common denominator of the rational numbers π_j, $j \in [n]$, is not greater than $nM^{\min\{s, n-1\}}$.*

▶ **Lemma 37.** *Both $\overline{\mathrm{cw}}(F)$ and $\underline{\mathrm{cw}}(F)$ are rational numbers whose denominators are not greater than $nM^{\min\{s, n-1\}}$.*

Proof. Let σ and τ be optimal strategies of players Min and Max respectively. Consider the Shapley operator

$$F_j^{\sigma, \tau}(x) := -A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))} + \sum_{(\tau(\sigma(j)), l) \in \mathscr{E}} P_{\tau(\sigma(j))l} x_l$$

of the game in which players play according to (σ, τ). Note that $\overline{\mathrm{cw}}(F) = \overline{\mathrm{cw}}(F^{\sigma, \tau})$ and $\underline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F^{\sigma, \tau})$ because the strategies σ, τ are optimal. Furthermore, let the vector $r \in \mathbb{R}^n$ and the row-stochastic matrix $Q \in \mathbb{R}^{n \times n}$ defined as $r_j := -A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))}$ for all $j \in \mathscr{V}_{\mathrm{Min}}$ and

$$Q_{jl} = \begin{cases} P_{\tau(\sigma(j))l} & \text{if } (\tau(\sigma(j)), l) \in \mathscr{E}, \\ 0 & \text{otherwise} \end{cases}$$

for all $j, l \in \mathscr{V}_{\mathrm{Min}}$. In this way, we get $F^{\sigma, \tau}(x) = r + Qx$ for all $x \in \bar{\mathbb{R}}^n$. In other words, $F^{\sigma, \tau}$ is an operator describing a Markov chain with rewards, in which the Markov chain has transition probabilities given by Q and rewards given by r. The ergodic theorem of finite Markov chains [44, Appendix A.4] implies that this Markov chain has two (possibly identical) recurrent classes U_1, U_2 such that $\overline{\mathrm{cw}}(F^{\sigma, \tau}) = \sum_{j \in U_1} r_j \pi_j^{(1)}$, $\underline{\mathrm{cw}}(F^{\sigma, \tau}) = \sum_{j \in U_2} r_j \pi_j^{(2)}$, where $\pi^{(1)}, \pi^{(2)}$ are the invariant measures of the restrictions of the Markov chain to these classes. Hence, the claim follows from Lemma 36. ◀

Proof of Theorem 35. By Lemma 37, both $\overline{\mathrm{cw}}(F)$ and $\underline{\mathrm{cw}}(F)$ are rational numbers with denominators not greater than $nM^{\min\{s, n-1\}}$. Therefore, they are either equal or satisfy $\overline{\mathrm{cw}}(F) - \underline{\mathrm{cw}}(F) > 1/(nM^{\min\{s, n-1\}})^2$. By Lemma 31, the same is true for any operator $F^{\mathscr{D}}$ where $\mathscr{D} \subset [n]$ is a dominion. ◀

We now provide a bit-complexity estimate for the bias vector of the Shapley operator of a stochastic mean-payoff game.

▶ **Theorem 38.** *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*

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

Before detailing the proof, we sketch the main ideas. 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 then the main difficulty is to find a "short" bias. The one which will be constructed in the proof of this theorem 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 $\mathbb{E}_{\sigma\tau} \sum_{p=0}^{\infty} (1-\alpha)^p (-A_{i_p j_p} + B_{i_p k_p})$, where $0 < \alpha < 1$ and $1 - \alpha$ is the discount factor. The discounted game with initial state j has a value, $x_j(\alpha)$, and the value vector, $x(\alpha) = (x_j(\alpha)) \in \mathbb{R}^n$, 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. Kohlberg proved that $x(\alpha)$ admits a Laurent series expanstion with a pole of order at most 1. In fact, the result of Kohlberg applies more generally to piecewise-affine maps that are nonexpansive in any norm (not just to Shapley operators of stochastic mean-payoff games).

▶ **Theorem 39** ([36]). *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a piecewise-affine function that is nonexpansive in any norm. Then, for every $j \in [n]$ there exists a Laurent series $x_j \in \mathbb{R}((\alpha))$ of the form*

$$x_j(\alpha) = c_{j,-1}\alpha^{-1} + c_{j,0} + c_{j,1}\alpha + c_{j,2}\alpha^2 + \dots$$

such that $x_j(\alpha)$ converges for all small $\alpha > 0$ and the point $x(\alpha) = \big(x_1(\alpha), \dots, x_n(\alpha)\big) \in \mathbb{R}((\alpha))^n$ satisfies $f\big((1-\alpha)x(\alpha)\big) = x(\alpha)$ for all small $\alpha > 0$. Furthermore, if we denote $\eta = (c_{1,-1}, \dots, c_{n,-1}) \in \mathbb{R}^n$ and $u = (c_{1,0}, \dots, c_{n,0}) \in \mathbb{R}^n$, then the equality $F\big(t\eta + u\big) = (t+1)\eta + u$ is satisfied for all sufficiently large $t > 0$.

In fact, the vector η collecting the coefficients of α^{-1} in the expansions of $x_j(\alpha)$ coincides with the escape rate vector of F, and the vector u collecting the coefficients of the term of order 1 yields a bias vector, in a special case.

▶ **Proposition 40.** *Under the assumptions of Theorem 39, we have $\chi(F) = \eta$. Moreover, if F is the Shapley operator of a stochastic mean-payoff game, and if $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$, then $F(u) = \eta + u$, i.e., u is a bias vector.*

Proof. For all sufficiently large t, and for all $N \in \mathbb{N}$, we have $F^N(t\eta + u) = (t + N)\eta + u$, from which we deduce that $\chi(F) = \lim_{N \to \infty} (F^N(t\eta + u))/N = \eta$.

Moreover, we have $F(t\eta + u) = \eta + t\eta + u$. When $\overline{\mathrm{cw}}(F) = \underline{\mathrm{cw}}(F)$ and F is the Shapley operator of a stochastic mean-payoff game, using the fact that F commutes with the addition of a constant vector, we deduce that $F(u) = \eta + u$. ◀

We shall see that the bias vector u satisfies a Poisson-type equation $\chi(F) + u = r + Qu$, and that this special bias vector has the remarkable property of having a zero expectation with respect to all invariant measures of Q. Hence, the following lemma will allow us to bound the bit-complexity of this bias vector.

▶ **Lemma 41** ([51]). *Take a (possibly reducible) Markov chain with n states and transition probabilities given by the row-stochastic matrix $Q \in \mathbb{R}^{n \times n}$. Suppose that the transition probabilities are rational numbers, and let s and M be as above. Let $r \in \mathbb{Z}^n$ be an integer vector and suppose that $(\eta, u) \in \mathbb{R}^{2n}$ is a solution of the system*

$$\begin{cases} Q\eta &= \eta \\ Qu &= -r + \eta + u \,. \end{cases} \tag{18}$$

Furthermore, suppose that u is orthogonal to all the invariant measures of the Markov chain, i.e., $u^\mathsf{T}\pi = 0$ whenever $\pi^\mathsf{T}Q = \pi$. Then, $\|u\|_\infty \leqslant 4\|r\|_\infty n M^{\min\{s,n-1\}}$.

Proof of Theorem 38. Since F (understood as a function from $\mathbb{R}^n$ to $\mathbb{R}^n$) is piecewise-linear and nonexpansive in the sup-norm, Theorem 39 shows that there exists two vectors $\eta, u \in \mathbb{R}^n$ such that the equality $F(t\eta + u) = (t+1)\eta + u$ holds for all sufficiently large $t > 0$. If we take $v = t_0\eta + u$ for one such t_0, then we get $F(v) = \eta + v$, which (by the fact that $\underline{\mathrm{cw}}(F) = \overline{\mathrm{cw}}(F)$) implies that $\eta = \chi = \overline{\mathrm{cw}}(F)\mathbf{1}$. In particular, we have $F\big(t_0\overline{\mathrm{cw}}(F) + u\big) = (t_0 + 1)\overline{\mathrm{cw}}(F) + u$, which gives $F(u) = \overline{\mathrm{cw}}(F) + u$. Furthermore, we can suppose that the vector u comes from a vector $x(\alpha) \in \mathbb{R}((\alpha))^n$ as described in Theorem 39. In particular, for all small $\alpha > 0$ we have

$$F\big((1-\alpha)x(\alpha)\big) = x(\alpha) \tag{19}$$

and, for every $j \in [n]$, the series $x_j(\alpha)$ satisfies $x_j(\alpha) = \overline{\mathrm{cw}}(F)\alpha^{-1} + u_j + o(1)$. Define $y(\alpha) := (1-\alpha)x(\alpha) \in \mathbb{R}((\alpha))^n$ and note that this series satisfies $y(\alpha) = \overline{\mathrm{cw}}(F)\alpha^{-1} + u - \overline{\mathrm{cw}}(F) + o(1)$. For every $i \in \mathscr{V}_{\mathrm{Max}}$ consider the expression

$$\max_{(i,k)\in\mathscr{E}} \Big(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}y(\alpha)_l\Big). \tag{20}$$

Observe that for all sufficiently small α, the minima in (20) are achieved by the same indices. In other words, for every $i \in \mathscr{V}_{\mathrm{Max}}$ there exists $\tau(i) \in \mathscr{V}_{\mathrm{Nat}}$ such that the equality $\max_{(i,k)\in\mathscr{E}}\big(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}y(\alpha)_l\big) = B_{i\tau(i)} + \sum_{(\tau(i),l)\in\mathscr{E}} P_{\tau(i)l}y_l(\alpha)$ holds for all small α. In particular, for every $j \in \mathscr{V}_{\mathrm{Min}}$ we have the equality

$$F_j\big(y(\alpha)\big) = \min_{(j,i)\in\mathscr{E}}\Big(-A_{ij} + B_{i\tau(i)} + \sum_{(\tau(i),l)\in\mathscr{E}} P_{\tau(i)l}y_l(\alpha)\Big).$$

As before, the maxima on the right-hand side of this equality are achieved by the same indices if α is small. Hence, for every $j \in \mathscr{V}_{\mathrm{Min}}$ there exists $\sigma(j) \in \mathscr{V}_{\mathrm{Max}}$ such that the equality

$$\begin{aligned}
x_j(\alpha) = F_j\big(y(\alpha)\big) &= -A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))} + \sum_{(\tau(\sigma(j)),l)\in\mathscr{E}} P_{\tau(\sigma(j))l}y_l(\alpha) \\
&= \overline{\mathrm{cw}}(F)\alpha^{-1} - A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))} - \overline{\mathrm{cw}}(F) + \sum_{(\tau(\sigma(j)),l)\in\mathscr{E}} P_{\tau(\sigma(j))l}u_l + o(1)
\end{aligned} \tag{21}$$

holds for all small $\alpha > 0$. Therefore, we get

$$u_j = -A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))} - \overline{\mathrm{cw}}(F) + \sum_{(\tau(\sigma(j)),l)\in\mathscr{E}} P_{\tau(\sigma(j))l}u_l$$

for all $j \in \mathscr{V}_{\mathrm{Min}}$. Let $Q \in \mathbb{R}^{n\times n}$ be the row-stochastic matrix defined as $Q_{jl} := P_{\tau(\sigma(j))l}$ if $(\tau(\sigma(j)),l) \in \mathscr{E}$ and $Q_{jl} := 0$ otherwise, and let $r \in \mathbb{R}^n$ be the vector given by $r_j := -A_{\sigma(j)j} + B_{\sigma(j)\tau(\sigma(j))}$. If $\pi \in \mathbb{R}^n$ satisfies $\pi^\mathsf{T}Q = \pi$, then (21) gives $\pi^\mathsf{T}x(\alpha) = \pi^\mathsf{T}r + \pi^\mathsf{T}y(\alpha)$ for all small α and therefore $\pi^\mathsf{T}r = \alpha\pi^\mathsf{T}x(\alpha) = \overline{\mathrm{cw}}(F) + \alpha\pi^\mathsf{T}u + o(\alpha)$ for all small α, which gives $\pi^\mathsf{T}r = \overline{\mathrm{cw}}(F)$ and $\pi^\mathsf{T}u = 0$. By defining $\eta = \overline{\mathrm{cw}}(F)\mathbf{1}$, the pair (η, u) is a solution of the linear system given in Theorem 38 for the pair (r, Q). Hence, we have $\|u\|_\infty \leqslant 4nWM^{\min\{s,n-1\}}$. The claim follows from the fact that $\|u\|_{\mathrm{H}} \leqslant 2\|u\|_\infty$. ◀

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

▶ **Corollary 42.** *Let F be a Shapley operator as above, supposing that F has a bias vector and that $\mathrm{erg}(F)$ is nonzero. Then, procedure* VALUEITERATION *stops after*

$$N_{vi} \leqslant 8n^2 W M^{2\min\{s,n-1\}} \tag{22}$$

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

Proof. By Theorem 38 we have $R(F) \leqslant 8nW M^{\min\{s,n-1\}}$. Moreover, by Lemma 37, $\mathrm{erg}(F)$ is a rational number with denominator at most $nM^{\min\{s,n-1\}}$. In particular, $|\mathrm{erg}(F)| \geqslant (nM^{\min\{s,n-1\}})^{-1}$. Hence, the claim follows from Theorem 10. ◀

▶ **Remark 43.** When specialized to deterministic mean-payoff games, i.e., when $s = 0$, Corollary 42 yields $N_{\mathrm{vi}} = O(n^2 W)$ which is precisely the bound that follows from the analysis of value iteration by Zwick and Paterson [57].

▶ **Corollary 44.** *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*

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

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 strategies can be obtained from the approximate optimality certificates generated by the procedure.

Before proving this corollary, let us explain a key idea, how the optimal strategies can be 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 $\lambda + \delta/8 + y \geqslant F(y)$. By selecting, for each state $j \in \mathscr{V}_{\mathrm{Min}}$, a minimizing state $i \in \mathscr{V}_{\mathrm{Max}}$ in the expression

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

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 will deduce these strategies guarantee a value of $\mathrm{erg}(F)$ to each of the players, and so, they are optimal.

Proof of Corollary 44. By Theorem 38, the value $R := 8nW M^{\min\{s,n-1\}}$ satisfies the conditions of Theorem 16. Therefore, APPROXIMATECONSTANTMEANPAYOFF stops after at most $\lceil 8R/\delta \rceil = 64n^3 W M^{3\min\{s,n-1\}}$ iterations of its first loop. In particular, it makes no more than $128n^3 W M^{3\min\{s,n-1\}}$ calls to the oracle during its entire execution. Furthermore, it outputs an interval $[a,b]$ that contains $\mathrm{erg}(F)$ and is of width at most δ. Since $\mathrm{erg}(F)$ is a rational number of denominator at most μ by Lemma 37 and $\delta := \mu^{-2}$, $\mathrm{erg}(F)$ is the unique rational number in $[a,b]$ of denominator at most μ. In particular, the exact value of $\mathrm{erg}(F)$ can be found by the rational search technique, see, e.g., [37, 24]. Furthermore, the procedure outputs two vectors $x, y \in \mathbb{R}^n$ such that $a + x \leqslant F(x)$ and $b + y \geqslant F(y)$. A pair of optimal strategies (σ, τ) is then constructed as follows. For every $i \in \mathscr{V}_{\mathrm{Max}}$ let $\tau(i) \in \mathscr{V}_{\mathrm{Nat}}$ be a vertex such that

$$\max_{(i,k)\in\mathscr{E}} \Big(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}x_l \Big) = B_{i\tau(i)} + \sum_{(\tau(i),l)\in\mathscr{E}} P_{\tau(i)l}x_l \,.$$

Analogously, for every $j \in \mathscr{V}_{\text{Min}}$, let $\sigma(j) \in \mathscr{V}_{\text{Max}}$ be a vertex such that

$$\min_{(j,i)\in\mathscr{E}} \left(-A_{ij} + \max_{(i,k)\in\mathscr{E}} \left(B_{ik} + \sum_{(k,l)\in\mathscr{E}} P_{kl}y_l\right)\right) = -A_{\sigma(j)j} + \max_{(i,k)\in\mathscr{E}} \left(B_{\sigma(j)k} + \sum_{(k,l)\in\mathscr{E}} P_{kl}y_l\right).$$

We claim that σ is optimal for Min and τ is optimal for Max. Indeed, let F^σ be the Shapley operator of the game in which Min uses σ. Then, the definition of σ gives the inequality $F^\sigma(y) = F(y) \leqslant b + y$. Hence, $\overline{\text{cw}}(F^\sigma) \leqslant b$. Even more, since Lemma 37 applies to F^σ, we get $\overline{\text{cw}}(F^\sigma) \leqslant \text{erg}(F)$. If τ' is any strategy of Max, and we denote by $F^{\sigma,\tau'}$ the operator obtained by fixing both strategies, then we have $F^{\sigma,\tau'}(z) \leqslant F^\sigma(z)$ for any $z \in \mathbb{R}^n$ and hence $g_j(\sigma, \tau') \leqslant \overline{\text{cw}}(F^{\sigma,\tau'}) \leqslant \overline{\text{cw}}(F^\sigma) \leqslant \text{erg}(F)$ for every $j \in \mathscr{V}_{\text{Min}}$. Analogously, if τ is a strategy of Max defined above and σ' is any strategy of Min, then $\text{erg}(F) \leqslant g_j(\sigma', \tau)$ for all $j \in \mathscr{V}_{\text{Min}}$. Hence, the strategies (σ, τ) are optimal. ◀

▶ **Corollary 45.** *The set of initial states with maximal value of a stochastic mean-payoff game can be found by performing at most $65n^4WM^{3\min\{s,n-1\}}$ calls to an oracle approximating its Shapley operator F with precision $\epsilon := 1/(8\mu^2)$, where $\mu := nM^{\min\{s,n-1\}}$.*

Proof. Let $\delta := 1/\mu^2$. By Theorem 35 we have $\delta < \text{sep}(F)$. Furthermore, if we denote by $\mathscr{D}_{\text{max}} \subset \mathscr{V}_{\text{Min}}$ the set of initial states with maximal value, then by applying Theorem 38 to $F^{\mathscr{D}_{\text{max}}}$ we see that the number $R := 8W\mu$ satisfies the conditions of Theorem 29, so that the procedure TOPCLASS applied to the Shapley operator F, with $\delta := 1/\mu^2$ and $R := 8W\mu$, is correct, and finds the set of initial states with maximal value. Furthermore, by Theorem 30, TOPCLASS can be executed by performing at most $n^2 + n\lceil 8R/\delta\rceil = n^2 + 64n^4WM^{3\min\{s,n-1\}}$ calls to the oracle that approximates F. ◀

▶ **Remark 46.** If we are given explicitly the graph of the game together with the probabilities P_{kj} and the payoffs A_{ij} and B_{ik}, then the operator F can be evaluated exactly in $O(E)$ complexity, where E is the number of edges of the graph representing the 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 $x \in \bar{\mathbb{R}}^{\mathscr{V}_{\text{Min}}}$, 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 such a procedure as an approximation oracle, then all the algorithms presented in this section require $O(nE\log(nMW))$ memory, which is polynomial in the size of the input.

▶ **Remark 47.** Since a single call to the oracle approximating F can be done in $O(E)$ arithmetic operations, by combining Corollary 45 with Corollary 44 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 [18] which achieves the same aim using a pumping algorithm instead of value iteration. If we combine the estimate from [51] with the complexity bound presented in [18] 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 [18] 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 (10) and (11). As in the last section, we denote $n := |\mathscr{V}_D|$ and we put $W := \max_{(p,k)\in\mathscr{E}} m_{pk}$.

5.1 Dominions of Entropy Games

We first describe the dominions of entropy games, and verify that this class of games satisfies Assumption 2. Given a set $\mathscr{S} \subset \mathscr{V}_D$, we denote by $\mathscr{V}_P^{\mathscr{S}} \subset \mathscr{V}_P$ the set of states of People that have at least one outgoing edge that goes to $\mathscr{S}$, i.e., $\mathscr{V}_P^{\mathscr{S}} := \{p \in \mathscr{V}_P : \exists l \in \mathscr{S}, (p,l) \in \mathscr{E}\}$. In the same way, we denote by $\mathscr{V}_T^{\mathscr{S}} \subset \mathscr{V}_T$ the set of states of Tribune that have at least one outgoing edge that goes to $\mathscr{V}_P^{\mathscr{S}}$, i.e., $\mathscr{V}_T^{\mathscr{S}} := \{t \in \mathscr{V}_T : \exists p \in \mathscr{V}_P^{\mathscr{S}}, (t,p) \in \mathscr{E}\}$. The following lemma characterizes the dominions of entropy games.

▶ **Lemma 48.** *A set $\mathscr{D} \subset \mathscr{V}_D$ is a dominion of the operator F if and only if every outgoing edge of every state of $\mathscr{D}$ goes to $\mathscr{V}_T^{\mathscr{D}}$, i.e., if for every pair $(k,t) \in \mathscr{D} \times \mathscr{V}_T$ we have $(k,t) \in \mathscr{E} \implies t \in \mathscr{V}_T^{\mathscr{D}}$. Furthermore, if $\mathscr{D}$ is a dominion, then $F^{\mathscr{D}} = \log \circ T^{\mathscr{D}} \circ \exp$, where for all $x \in \mathbb{R}_{>0}^{\mathscr{D}}$ and all $k \in \mathscr{D}$ we define*

$$T_k^{\mathscr{D}}(x) := \min_{\substack{(k,t)\in\mathscr{E} \\ t\in\mathscr{V}_T^{\mathscr{D}}}} \max_{\substack{(t,p)\in\mathscr{E} \\ p\in\mathscr{V}_P^{\mathscr{D}}}} \sum_{\substack{(p,l)\in\mathscr{E} \\ l\in\mathscr{D}}} m_{pl}x_l\,.$$

Proof. The proof is similar to the proof of Lemma 31. By Lemma 21, $\mathscr{D}$ is a dominion if and only if for every $k \in \mathscr{D}$ and every $t \in \mathscr{V}_T$ such that $(k,t) \in \mathscr{E}$ we have

$$\max_{(t,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl}\mathbf{1}_{l\in\mathscr{D}} > 0\,. \tag{23}$$

By definition, we have $\sum_{(p,l)\in\mathscr{E}} m_{pl}\mathbf{1}_{l\in\mathscr{D}} > 0$ if and only if $p \in \mathscr{V}_P^{\mathscr{D}}$, and so the relation (23) holds if and only if $t \in \mathscr{V}_T^{\mathscr{D}}$. The second part of the claim follows from the definition of $F^{\mathscr{D}}$. ◀

In particular, as in the case of stochastic mean-payoff games, $F^{\mathscr{D}}$ is the Shapley operator of a smaller entropy game that takes place on the state space $\mathscr{D} \uplus \mathscr{V}_T^{\mathscr{D}} \uplus \mathscr{V}_P^{\mathscr{D}}$.

▶ **Lemma 49.** *If $\mathscr{D}$ is a dominion, then the subgame induced by $\mathscr{D}$ satisfies Assumption 1.*

Proof. This is an immediate consequence of Lemma 48. ◀

▶ **Lemma 50.** *The recession operator of F is given by*

$$\hat{F}_k(x) = \min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \max_{(p,l)\in\mathscr{E}} x_l, \qquad \forall k \in \mathscr{V}_D\ .$$

Proof. For every $s > 0$ and every $x \in \mathbb{R}^{\mathscr{V}_D}$ we have

$$s^{-1}F(sx) = s^{-1}\log\Big(\min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl}\exp(x_l)^s\Big)$$

$$= \min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} s^{-1}\log\Big(\sum_{(p,l)\in\mathscr{E}} m_{pl}\exp(x_l)^s\Big)\,.$$

To finish the proof, note that the inequality

$$\max_{(p,l)\in\mathscr{E}} \exp(x_l)^s \leqslant \sum_{(p,l)\in\mathscr{E}} m_{pl} \exp(x_l)^s \leqslant nW \max_{(p,l)\in\mathscr{E}} \exp(x_l)^s$$

implies that $s^{-1}\log\left(\sum_{(p,l)\in\mathscr{E}} m_{pl}\exp(x_l)^s\right) \to \max_{(p,l)\in\mathscr{E}} x_l$.   ◀

▶ **Lemma 51.** *The Shapley operator F of an entropy game satisfies Assumption 2. Further-more, if τ is an optimal strategy of Tribune, then for any $k \in \mathscr{D}_{\max}$ and any $t \in \mathscr{V}_T$ such that $(k,t) \in \mathscr{E}$ we have $\tau(t) \in \mathscr{V}_P^{\mathscr{D}_{\max}}$. Even more, if σ is an optimal strategy of Despot and $k \notin \mathscr{D}_{\max}$, then $\sigma(k) \notin \mathscr{V}_T^{\mathscr{D}_{\max}}$.*

Proof. The operator F satisfies the first part of Assumption 2 by Remark 22. To prove the remaining claims, we proceed as in the proof of Lemma 34. Let χ be the escape rate of F, and let $\mathscr{D}_{\max}$ be the set of states of maximal value. Furthermore, let $\tau\colon \mathscr{V}_T \to \mathscr{V}_P$ be an optimal strategy of Tribune. We consider a reduced operator

$$T_k^\tau(x) := \min_{(k,t)\in\mathscr{E}} \sum_{(\tau(t),l)\in\mathscr{E}} m_{\tau(t)l}x_l$$

obtained from a game in which Tribune plays according to τ. Let $F^\tau = \log\circ T^\tau \circ \exp$. By the optimality of τ we get $\chi(F^\tau) = \chi$. Since $\hat{F}^\tau(\chi) = \chi$, Lemma 50 implies that

$$\mathbf{t}(\chi) = \min_{(k,t)\in\mathscr{E}} \max_{(\tau(t),l)\in\mathscr{E}} \chi_l$$

for all $k \in \mathscr{D}_{\max}$. Note that $\mathbf{t}(\chi) = \max_{(p,l)\in\mathscr{E}} \chi_l$ if and only if $p \in \mathscr{V}_P^{\mathscr{D}_{\max}}$. Hence, for every $k \in \mathscr{D}_{\max}$ we have

$$(k,t) \in \mathscr{E} \implies \tau(t) \in \mathscr{V}_P^{\mathscr{D}_{\max}}.$$

This proves the second claim. In particular, if $k \in \mathscr{D}_{\max}$ and $(k,t) \in \mathscr{E}$, then t belongs to $\mathscr{V}_T^{\mathscr{D}_{\max}}$ and so $\mathscr{D}_{\max}$ is a dominion by Lemma 48. To prove the third claim, let $\sigma\colon \mathscr{V}_D \to \mathscr{V}_T$ be an optimal strategy of Despot. As above, we consider the reduced operator

$$T_k^\sigma(x) := \max_{(\sigma(k),p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl}x_l$$

and we let $F^\sigma = \log\circ T^\sigma \circ \exp$. Since $\hat{F}^\sigma(\chi) = \chi$, Lemma 50 implies that

$$\mathbf{t}(\chi) = \max_{(\sigma(k),p)\in\mathscr{E}} \max_{(p,l)\in\mathscr{E}} \chi_l.$$

Hence, we get $k \in \mathscr{D}_{\max}$ whenever $\sigma(k) \in \mathscr{V}_T^{\mathscr{D}_{\max}}$. This proves the third claim.

It remains to prove that $\underline{\mathrm{cw}}(F^{\mathscr{D}_{\max}}) = \mathbf{t}(\chi)$. To do so, let $\tilde{\sigma}\colon \mathscr{D}_{\max} \to \mathscr{V}_T^{\mathscr{D}_{\max}}$ be an optimal strategy of Despot in the subgame induced by $\mathscr{D}_{\max}$. We extend $\tilde{\sigma}$ to a strategy $\tilde{\sigma}\colon \mathscr{V}_D \to \mathscr{V}_T$ by setting $\tilde{\sigma}(k) := \sigma(k)$ for all $k \notin \mathscr{D}_{\max}$. Let $T'(x) := M^{\tilde{\sigma},\tau}x$, where $M^{\tilde{\sigma},\tau} \in \mathbb{R}^{n\times n}$ is the associated ambiguity matrix. In this way, $T'(x)$ is the operator of a game in which both strategies are fixed. In particular, if we take $F' := \log\circ T' \circ \exp$, then $\chi(F') \geqslant \chi$ by the optimality of τ. Furthermore, if $k \notin \mathscr{D}_{\max}$, then the third claim implies that the dipaths in $\mathscr{G}^{\tilde{\sigma},\tau}$ that start in k are the same as in the graph $\mathscr{G}^{\sigma,\tau}$. Therefore, by Proposition 8, $\chi(F')_k = \chi_k$ for all $k \notin \mathscr{D}_{\max}$ and so

$$k \notin \mathscr{D}_{\max} \implies \chi(F')_k < \mathbf{t}(\chi).\tag{24}$$

Moreover, if $k \in \mathscr{D}_{\max}$, then $\chi(F')_k \geqslant \mathbf{t}(\chi)$ and Proposition 8 shows that there exists a strongly connected component $\mathscr{C}$ of $\mathscr{G}^{\tilde{\sigma},\tau}$ such that $\rho(M^{\tilde{\sigma},\tau}[\mathscr{C}]) \geqslant \mathbf{t}(\chi)$ and such that there exists a dipath from k to $\mathscr{C}$ in $\mathscr{G}^{\tilde{\sigma},\tau}$. By (24), this dipath does not go through any vertex in $\mathscr{V}_D \setminus \mathscr{D}_{\max}$. Therefore, this dipath only goes through vertices in

$$\mathscr{S} := \mathscr{D}_{\max} \uplus \{\tilde{\sigma}(k) \colon k \in \mathscr{D}_{\max}\} \uplus \{\tau(\tilde{\sigma}(k)) \colon k \in \mathscr{D}_{\max}\} \subset \mathscr{D}_{\max} \uplus \mathscr{V}_T^{\mathscr{D}_{\max}} \uplus \mathscr{V}_P^{\mathscr{D}_{\max}}$$

and the component $\mathscr{C}$ is included in $\mathscr{S}$. Consider the subgame induced by $\mathscr{D}_{\max}$ and suppose that in this game Tribune uses a strategy $\tilde{\tau} \colon \mathscr{V}_T^{\mathscr{D}_{\max}} \to \mathscr{V}_P^{\mathscr{D}_{\max}}$ that agrees with τ on the set $\{t \in \mathscr{V}_T^{\mathscr{D}_{\max}} \colon \exists k \in \mathscr{D}_{\max}, (k,t) \in \mathscr{E}\}$. Note that there exists at least one such strategy by the second claim. Furthermore, since $\mathscr{C}$ is included in $\mathscr{S}$, Proposition 8 implies that the value of state k in the game obtained by fixing $(\tilde{\sigma}, \tilde{\tau})$ is not smaller than its value in the original game. By the optimality of $\tilde{\sigma}$ and since $k \in \mathscr{D}_{\max}$ was arbitrary, we have $\underline{\mathrm{cw}}(F^{\mathscr{D}_{\max}}) \geqslant \mathbf{t}(\chi)$. The other inequality follows from Lemma 24.

◀

5.2   Bit-Complexity Bounds for Entropy Games

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 will be established by combining a separation bound of Rump [48] 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 $\{0,1\}$ [22].

▶ **Theorem 52.** *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*

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

Proof. Step 1. First, we note that if C is a $k \times k$ matrix with entries in $[0, W]$, we have

$$|\det C| \leqslant (k+1)\sqrt{k}^k (W/2)^k . \tag{25}$$

Indeed, this follows by writing $C = B + WJ/2$, where J is the matrix whose entries are all equal to one and $B = C - WJ/2$, so that $-W/2 \leqslant B_{ij} \leqslant W/2$, then by expanding $\det C$ by multilinearity as a function of columns, i.e., $\det C = \det B + \sum_{i=1}^{k} (\pm 1) \det(\hat{B}_i, (W/2)\mathbf{1})$ in which $\hat{B}_i$ denotes the matrix obtained by deleting column i of B, noting that the other terms of this multilinear expansion are zero owing to repeated appearances of columns proportional to $\mathbf{1}$. By Hadamard inequality, $|\det B| \leqslant k^{k/2} (W/2)^k$, and similarly for each $|\det(\hat{B}_i, (W/2)\mathbf{1})|$, which gives (25). (We note that when all the entries of C belong to $\{0,1\}$, the bound can be refined to $|\det C| \leqslant \sqrt{k+1}^{k+1}/2^k$, see [22].)

Step 2. Let A be the $n \times n$ ambiguity matrix associated with a pair of strategies. Since A has rank at most r, it has at most r non-zero eigenvalues, and so, the $n - r$ first coefficients of the characteristic polynomial $Q^A := \det(XI - A) = \sum_{k=0}^{n} Q_k^A X^k$ are zero. The coefficient Q_{n-k}^A of X^{n-k} in this polynomial is given by the sum of minors of order k of A. Hence, using the inequality established in Step 1, together with the inequality $\binom{n}{r} \leqslant (ne/r)^r$, we get

$$S_{n,r} := \sum_{k=0}^{n} |Q_{n-k}^A| = \sum_{k=0}^{r} |Q_{n-k}^A| \leqslant \sum_{k=0}^{r} \binom{n}{k} (k+1)\sqrt{k}^k \left(\frac{W}{2}\right)^k \leqslant (r+1)^2 \left(\frac{ne}{r}\right)^r \sqrt{r}^r \left(1 \vee \frac{W}{2}\right)^r .$$

Step 3. Now, if A, B are two $n \times n$ ambiguity matrices, consider the product $Q^A Q^B$ of the characteristic polynomials of A and B. Let S denote the sum of absolute values of the coefficients of $Q^A Q^B$. Then,

$$S \leqslant S_{n,r}^2 \leqslant a_{n,r} := (r+1)^4 \Big(\frac{ne}{r}\Big)^{2r} r^r \big(1 \vee \frac{W}{2}\big)^{2r} \ .$$

Now, we use a theorem of Rump, [48, Th. 3], showing that for a polynomial of degree d with integer coefficients whose sum of absolute values is bounded by S, the distance between any two distinct real roots is at least

$$2\sqrt{2}\big(d^{\frac{d}{2}+1}(S+1)^d\big)^{-1} \ . \tag{26}$$

We apply this result to the polynomial $Q^A Q^B$, which is of degree $d = 2r$, showing that the roots of this polynomial are separated at least by the inverse of the following quantity

$$
\begin{aligned}
2^{-3/2}(2r)^{r+1}\Big(a_{n,r}+1\Big)^{2r} &\leqslant 2^{-3/2}(2r)^{r+1}a_{n,r}^{2r} \exp\big(2r\log(1+a_{n,r}^{-1})\big) \\
&\leqslant 2^{-3/2}(2r)^{r+1}a_{n,r}^{2r}\exp(2ra_{n,r}^{-1}) \leqslant (2r)^{r+1}a_{n,r}^{2r} \\
&= 2^r r^{r+1}(r+1)^{8r}\Big(\frac{ne}{r}\Big)^{4r^2} r^{2r^2}\big(1\vee\frac{W}{2}\big)^{4r^2} \ ,
\end{aligned}
$$

in which the second inequality follows from the concavity of the logarithm, and the third one follows from the fact that for all $n \geqslant r \geqslant 1$,

$$2ra_{n,r}^{-1} \leqslant 2r\big((r+1)^4\big(\frac{ne}{r}\big)^{2r} r^r\big)^{-1} \leqslant \frac{2}{(r+1)^3}\frac{r}{r+1}\frac{\exp(-2r)}{r^r}\Big(\frac{r}{n}\Big)^n \leqslant 2\exp(-2) \ ,$$

together with $2^{-3/2}\exp(2\exp(-2)) \leqslant 1$.

Since the value of an entropy game coincides with a non-zero eigenvalue of the ambiguity matrix associated with a pair of strategies, the result is established. ◀

▶ **Remark 53.** When $W > 2$, the separation bound of $\nu_{n,r}^{-1}$ may be written as $C_{n,r}^{-1}W^{-4r^2}$. The exponent of W may be improved at the price of increasing the combinatorial factor $C_{n,r}$, by using a theorem of Mahler [39], showing that for a polynomial of degree d, with integer coefficients bounded by M, and without distinct roots, the inverse of the distance between two distinct roots is bounded by

$$\vartheta_{d,M} = 3^{-1/2}(d+1)^{(2d+1)/2}M^{d-1} \ .$$

As explained in [48], one can deduce from this a separation root for polynomials P with possibly multiple roots, after replacing P by $\tilde{P} := P/\gcd(P, P')$. Using Malher's bound instead of (26), we would arrive at a bound $(C'_{n,r})^{-1}W^{-4r^2+2r}$ instead of $C_{n,r}^{-1}W^{-4r^2}$.

Theorem 52 readily entails a separation bound for values of strategies. Using the observation fact that the value of an entropy game is always in the interval $[1, nW]$, we next deduce a separation bound for the logarithms of these values, differing only by a nW factor:

▶ **Corollary 54.** *The value of an entropy games lies in the interval $[1, nW]$. Moreover, 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*

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

Proof. We first recall the following property of monotonicity of the spectral radius: if U, V are nonnegative matrices, then,

$$U \leqslant V \implies \rho(U) \leqslant \rho(V) \tag{27}$$

see [14, Chap. 2, 1.5].

Let λ, μ denote the distinct values determined by two pairs of strategies, and suppose, without loss of generality, that $\lambda > \mu$. By Proposition 8, the numbers λ and μ coincide with the Perron roots of some irreducible principal submatrices of ambiguity matrices. In particular, suppose that $\lambda = \rho(U[\mathscr{C}])$ in which U is an ambiguity matrix, and $\mathscr{C}$ the subset of indices of rows and columns used to extract the matrix. Since all the players alway have at least one allowed move, we deduce there are distinct elements $i_1, \ldots, i_s$ of $\mathscr{C}$ such that $U_{i_1 i_2} \ldots U_{i_s i_1} \neq 0$, and since the entries of the ambiguity matrix are integer, $U_{i_1 i_2}, \ldots, U_{i_s i_1} \geqslant 1$. We deduce that $U[\mathscr{C}] \geqslant V$ where V is a $\{0, 1\}$ matrix whose only non-zero entries are in the positions $(i_1, i_2), \ldots, (i_s, i_1)$, and by (27), $\rho(U[\mathscr{C}]) \geqslant \rho(V) = 1$. Moroever, since all the multiplicities are bounded by W, we have that $U[\mathscr{C}] \leqslant WJ$, where J is the $\mathscr{C} \times \mathscr{C}$ matrix whose entries are all equal to 1, and so, $\rho(U[\mathscr{C}]) \leqslant \rho(WJ) = |\mathscr{C}|W \leqslant nW$.

This entails that $1 \leqslant \lambda, \mu \leqslant nW$. By the intermediate value theorem, $\log \lambda - \log \mu = (\lambda - \mu)/\xi$ for some $\xi \in (1, nW)$, and so, $\log \lambda - \log \mu \geqslant (\lambda - \mu)/(nW) \geqslant \hat{\nu}_{n,r}^{-1}$. ◄

▶ **Proposition 55.** *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 $\max\{\|\log w\|_H, \|\log z\|_H\} \leqslant 1200(n^3 \log W + n^2 \log \delta^{-1})$.*

This proposition will be 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. To do so, we first state, for convenience, the following consequences of Theorem 5, which follow by applying this theorem to the Shapley operator $F = \log \circ T \circ \exp$,

$$\mathbf{b}(V^\infty) = \sup\{\lambda > 0 \colon \exists w \in \mathbb{R}_{>0}^n, \ \lambda w \leqslant T(w)\} \tag{28}$$

$$\mathbf{t}(V^\infty) = \inf\{\mu > 0 \colon \exists z \in \mathbb{R}_{>0}^n, \ \mu z \geqslant T(z)\} \ . \tag{29}$$

We also recall that the *encoding length* of an integer number r is defined as $\langle r \rangle := \lceil \log_2(|r| + 1) \rceil + 1$. Moreover, if $r = p/q$ is a rational number then its encoding length is defined as $\langle r \rangle := \langle p \rangle + \langle q \rangle$. The encoding length of an affine inequality with rational coefficients $ax \leqslant b$ is defined as $\langle b \rangle + \langle a_1 \rangle + \cdots + \langle a_n \rangle$. In the proof, we use the following straightforward inequalities:

▶ **Lemma 56.** *If $r \geqslant 1$ is a natural number, then $\log_2(r) \leqslant \langle r \rangle \leqslant 3 + \log_2(r)$. If $r_1, \ldots, r_m$ are rational numbers, then $\langle r_1 + \cdots + r_m \rangle \leqslant 2(\langle r_1 \rangle + \cdots + \langle r_m \rangle)$ and $\langle r_1 r_2 \rangle \leqslant \langle r_1 \rangle + \langle r_2 \rangle$.* ◄

Proof of Proposition 55. We have $T(\mathbf{1}) \leqslant nW\mathbf{1}$, so that $T^\ell(\mathbf{1}) \leqslant (nW)^\ell \mathbf{1}$ for all ℓ and therefore $V^\infty \leqslant nW\mathbf{1}$. Likewise, since m_{pl} are natural numbers, we have $T(\mathbf{1}) \geqslant \mathbf{1}$ and so $V^\infty \geqslant \mathbf{1}$. In particular, we have $\mathbf{b}(V^\infty) \geqslant 1$, so the open interval $]e^{-\delta} \mathbf{b}(V^\infty), \mathbf{b}(V^\infty)[$ is of length at least $1 - e^{-\delta} \geqslant \frac{\delta}{1+\delta}$. Thus, it contains a rational number $q \in \mathbb{Q}$ with denominator at most $\frac{2(1+\delta)}{\delta} = 2 + \frac{2}{\delta} \leqslant \frac{4}{\delta}$. Since $\mathbf{b}(V^\infty) \leqslant nW$, the numerator of q is not greater than $\frac{4nW}{\delta}$, and so the encoding length of q satisfies

$$\langle q \rangle \leqslant 10 + \log_2 n + \log_2 W + 2 \log_2 \delta^{-1} \ .$$

Let $w \in \mathbb{R}_{>0}^n$ be any vector that satisfies $qw \leqslant T(w)$. Since $q < \mathbf{b}(V^\infty)$, the existence of such a vector follows from (28). For every t, let p_t be such that $(t, p_t) \in \mathscr{E}$ and $\max_{(t,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl} w_l = \sum_{(p_t,l)\in\mathscr{E}} m_{p_t l} w_l$. Likewise, for every k let t_k be such that $(k, t_k) \in \mathscr{E}$ and

$$\min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl} w_l = \max_{(t_k,p)\in\mathscr{E}} \sum_{(p,l)\in\mathscr{E}} m_{pl} w_l \ .$$

In this way, we have

$$T_k(w) = \sum_{(p_{(t_k)},l)\in\mathscr{E}} m_{p_{(t_k)} l} w_l$$

for all k. Consider now the polyhedron $\mathcal{W} \subset \mathbb{R}^n$ defined by the inequalities

$$\begin{aligned}
\forall (t,p) \in \mathscr{E}, \ \sum_{(p,l)\in\mathscr{E}} m_{pl} x_l &\leqslant \sum_{(p_t,l)\in\mathscr{E}} m_{p_t l} x_l \\
\forall (k,t) \in \mathscr{E}, \ \sum_{(p_t,l)\in\mathscr{E}} m_{p_t l} x_l &\geqslant \sum_{(p_{(t_k)},l)\in\mathscr{E}} m_{p_{(t_k)} l} x_l \\
\forall k, \ qx_k &\leqslant \sum_{(p_{(t_k)},l)\in\mathscr{E}} m_{p_{(t_k)} l} x_l \\
\forall k, \ x_k &\geqslant 0 \, .
\end{aligned} \tag{30}$$

Then, we have $w \in \mathcal{W} \cap \mathbb{R}_{>0}^n$ so $\mathcal{W}$ is nonempty. Even more, any vector $x \in \mathcal{W} \cap \mathbb{R}_{>0}^n$ satisfies $qx \leqslant T(x)$. Let

$$\phi := 1 + 4n(3 + \log_2 W) + 2\langle q \rangle \leqslant 50(n \log_2 W + \log_2 \delta^{-1}) \, .$$

Then, Lemma 56 shows that every inequality that describes $\mathcal{W}$ has encoding length at most ϕ. Hence, by [29, Lemma 6.2.4], there exists two finite sets $X, Y \subset \mathbb{Q}^n$ such that $\mathcal{W} = \text{conv}(X) + \text{cone}(Y)$ and every entry of every vector in $X \cup Y$ has encoding length at most $4n\phi$. Furthermore, we have $X \subset \mathbb{Q}_{\geqslant 0}^n$ because $X \subset \mathcal{W}$ and $Y \subset \mathbb{Q}_{\geqslant 0}^n$ because $x + \lambda y \in \mathcal{W}$ for any $(x, y) \in X \times Y$ and $\lambda \geqslant 0$. By Carathéodory's theorem [49, Section 7.7], there exist $x_0, \ldots x_n \in X$, $y_1, \ldots, y_n \in Y$ such that $w \in \text{conv}(x_0, \ldots, x_n) + \text{cone}(y_1, \ldots, y_n)$. Since $w \in \mathbb{R}_{>0}^n$ and $X, Y \subset \mathbb{Q}_{\geqslant 0}^n$, the point $\tilde{w} \in \mathcal{W}$ defined as

$$\tilde{w} := \frac{1}{n+1} x_0 + \cdots + \frac{1}{n+1} x_{n+1} + y_1 + \cdots + y_n$$

also satisfies $\tilde{w} \in \mathbb{R}_{>0}^n$. Moreover, the inequalities from Lemma 56 show that every entry of $\tilde{w}$ has encoding length at most $60n^2\phi$. Therefore, for all k we can write $\tilde{w}_k = p_k/q_k$, where the numbers p_k, q_k are natural and their encoding length is bounded by $60n^2\phi$. In particular, we have $\|\log_2 \tilde{w}\|_{\text{H}} \leqslant 2\|\log_2 \tilde{w}\|_\infty \leqslant 2\max_k(\log_2 p_k + \log_2 q_k) \leqslant 240n^2\phi \leqslant 1200(n^3 \log_2 W + \log_2 \delta^{-1})$ and $\tilde{w}$ satisfies the claim. The proof of the other part is analogous, using the fact that the open interval $]\mathbf{t}(V^\infty), e^\delta \mathbf{t}(V^\infty)[$ has length at least $e^\delta - 1 \geqslant \delta$, and so it contains a rational number with denominator at most $2/\delta$. ◀

We deduce the following parameterized complexity result.

▶ **Theorem 57.** *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 $\delta = (\hat{\nu}_{n,r})^{-1}$.*

Proof. The claim follows by combining Theorem 30 with the estimates from Corollary 54 and Proposition 55. ◀

▶ **Remark 58.** We note that the oracle used in Theorem 57, approximating the Shapley operator of an entropy game up to a given precision, can be implemented in polynomial time, this follows by using a result of Borwein and Borwein [19] on the approximation of the log and exp maps, together with a scaling argument, see [1, Lemma 27].

As an intermediate step, we solve strategically the special case of entropy games whose value is independent of the initial state. To do so, we suppose that we have access to the graph of the game as well as to an oracle that approximates the operator F.

▶ **Proposition 59.** *In an entropy game of rank at most r, such that the value is independent of the initial state, we can find a pair of optimal strategies for both players by performing $O(nR_{n,r}\hat{\nu}_{n,r})$ calls to an oracle approximating F with precision $\delta/16$ where $\delta = (\hat{\nu}_{n,r})^{-1}$.*

Proof. The proof is similar to the proof of Corollary 44. We use APPROXIMATECONSTANT-MEANPAYOFF to obtain two vectors $x, y \in \mathbb{R}^{\mathscr{V}_D}$ and an interval $[a, b]$ of width at most $\delta/2$ such that $a + x \leqslant F(x)$ and $b + y \geqslant F(x)$. By Theorem 16 combined with the estimates from Corollary 54 and Proposition 55, this requires $O(nR_{n,r}\hat{\nu}_{n,r})$ calls to an oracle approximating F. By Corollary 54, the interval $[a, b]$ contains $\mathrm{erg}(F)$. For every $k \in \mathscr{V}_D$ we have

$$b + y_k \geqslant \min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \log\Big(\sum_{(p,l)\in\mathscr{E}} m_{pl}e^{y_l} \Big).$$

For every $p \in \mathscr{V}_P$ we approximate the expression $\log\Big(\sum_{(p,l)\in\mathscr{E}} e^{y_l}\Big)$ to precision $\delta/8$ using the procedure from [1, Lemma 27]. Let $Q_p \in \mathbb{R}$ denote this approximation and let σ be a strategy of Despot that satisfies

$$\min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} Q_p = \max_{(\sigma(k),p)\in\mathscr{E}} Q_p$$

for all $k \in \mathscr{V}_D$. We have

$$b + y_k \geqslant \min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} \log\Big(\sum_{(p,l)\in\mathscr{E}} m_{pl}e^{y_l} \Big) \geqslant -\delta/8 + \min_{(k,t)\in\mathscr{E}} \max_{(t,p)\in\mathscr{E}} Q_p$$

$$= -\delta/8 + \max_{(\sigma(k),p)\in\mathscr{E}} Q_p \geqslant -\delta/4 + \max_{(\sigma(k),p)\in\mathscr{E}} \log\Big(\sum_{(p,l)\in\mathscr{E}} m_{pl}e^{y_l} \Big).$$

Hence, if we denote by F^σ the Shapley operator obtained by fixing σ, then $b + \delta/4 + y \geqslant F^\sigma(y)$. In particular, $\overline{\mathrm{cw}}(F^\sigma) \leqslant b + \delta/4$. Since the interval $[a, b + \delta/4]$ is of length smaller than δ and contains $\mathrm{erg}(F)$, Corollary 54 implies that $\overline{\mathrm{cw}}(F^\sigma) \leqslant \mathrm{erg}(F)$ and σ is optimal. We can construct an optimal strategy of Tribune in an analogous way. ◀

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.

▶ **Lemma 60** (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

$$\forall k \in \mathscr{V}_D, \ \sigma(k) = \begin{cases} \sigma_1(k) & \text{if } k \in \mathscr{D}_{\max}, \\ \sigma_2(k) & \text{otherwise,} \end{cases} \quad \forall t \in \mathscr{V}_T, \ \tau(t) = \begin{cases} \tau_1(t) & \text{if } t \in \mathscr{V}_T^{\mathscr{D}_{\max}}, \\ \tau_2(t) & \text{otherwise.} \end{cases} \tag{31}$$

are optimal in the original game.

Proof. The game on $\mathscr{G}[\mathscr{S}_1]$ satisfies Assumption 1 by Lemmas 48 and 51. The other game satisfies this assumption by the definition of the set $\mathscr{S}_2$. Indeed, if $p \in \mathscr{V}_P \cap \mathscr{S}_2$, then all the outgoing edges of p go to $\mathscr{V}_D \setminus \mathscr{D}_{\max}$, so they are in $\mathscr{G}[\mathscr{S}_2]$. Likewise, if $t \in \mathscr{V}_T \cap \mathscr{S}_2$, then all the outgoing edges of t go to $\mathscr{V}_P \setminus \mathscr{V}_P^{\mathscr{D}_{\max}}$, so they are in $\mathscr{G}[\mathscr{S}_2]$. If $k \in \mathscr{V}_D \cap \mathscr{S}_2$, then Lemmas 48 and 51 shows that k has an outgoing edge that goes to $\mathscr{V}_T \setminus \mathscr{V}_T^{\mathscr{D}_{\max}}$, so this edge is in $\mathscr{G}[\mathscr{S}_2]$.

To prove the second part of the claim, let $\hat{\sigma}, \hat{\tau}$ be any pair of optimal strategies of Despot and Tribune in the original game and let χ be the escape rate of the operator F. Let $\chi^{\sigma,\tau}$ be the escape rate of the operator obtained by fixing (σ, τ). We use analogous notation for other pairs of strategies. By Lemma 51 for every $k \in \mathscr{V}_D \setminus \mathscr{D}_{\max}$ we have $\hat{\sigma}(k) \notin \mathscr{V}_T^{\mathscr{D}_{\max}}$. Hence, Proposition 8 combined with the optimality of $\hat{\sigma}, \tau_2, \sigma_2, \hat{\tau}$ gives the inequality $\chi_k \geqslant \chi_k^{\hat{\sigma},\tau} \geqslant \chi_k^{\sigma,\tau} \geqslant \chi_k^{\sigma,\hat{\tau}} \geqslant \chi_k$. Therefore, $\chi_k^{\sigma,\tau} = \chi_k$ for all such k.

Let $\bar{\sigma}$ be an optimal response to τ, i.e., an optimal strategy of Despot in the game in which Tribune plays according to τ. To prove the optimality of τ it is enough to show that $\chi^{\bar{\sigma},\tau} \geqslant \chi$. Consider the game obtained by fixing $(\bar{\sigma}, \tau)$. Note that if we remove from the graph $\mathscr{G}^{\bar{\sigma},\tau}$ the edges that go from $\mathscr{V}_P^{\mathscr{D}_{\max}}$ to $\mathscr{V}_D \setminus \mathscr{D}_{\max}$, then the value of this game can only decrease.Moreover, by Lemma 51 we have $\bar{\sigma}(k) \in \mathscr{V}_T^{\mathscr{D}_{\max}}$ for all $k \in \mathscr{D}_{\max}$ and $\underline{\text{cw}}(F^{\mathscr{D}_{\max}}) = \overline{\text{cw}}(F^{\mathscr{D}_{\max}}) = \mathbf{t}(\chi)$. Hence, the optimality of τ_1 combined with Proposition 8 give $\chi_k^{\bar{\sigma},\tau} \geqslant \mathbf{t}(\chi) = \chi_k$ for all $k \in \mathscr{D}_{\max}$. Furthermore, we have $\bar{\sigma}(k) \notin \mathscr{V}_T^{\mathscr{D}_{\max}}$ for all $k \in \mathscr{V}_D \setminus \mathscr{D}_{\max}$. Indeed, suppose that $\bar{\sigma}(k) \in \mathscr{V}_T^{\mathscr{D}_{\max}}$. Then, there is a dipath in $\mathscr{G}^{\bar{\sigma},\tau}$ that goes from k to $\mathscr{D}_{\max}$ and so $\chi_k^{\bar{\sigma},\tau} \geqslant \mathbf{t}(\chi)$ by the previous observation. However, $\bar{\sigma}$ is an optimal response to τ and therefore we have $\chi_k^{\bar{\sigma},\tau} \leqslant \chi_k < \mathbf{t}(\chi)$, which gives a contradiction. Thus, $\bar{\sigma}(k) \notin \mathscr{V}_T^{\mathscr{D}_{\max}}$ for all $\mathscr{V}_D \setminus \mathscr{D}_{\max}$. In particular, the optimality of τ_2 gives $\chi_k^{\bar{\sigma},\tau} \geqslant \chi_k^{\sigma,\tau} = \chi_k$.

Analogously, let $\bar{\tau}$ be an optimal response to σ and consider the game obtained by fixing $(\sigma, \bar{\tau})$. The optimality of σ_2 gives $\chi_k^{\sigma,\bar{\tau}} \leqslant \chi_k^{\sigma,\tau} = \chi_k$ for all $k \in \mathscr{V}_D \setminus \mathscr{D}_{\max}$. Fix $k \in \mathscr{D}_{\max}$ and let $\mathscr{C}$ be any strongly connected component of $\mathscr{G}^{\sigma,\bar{\tau}}$ that can be reached form k in this graph. If $\mathscr{C}$ contains a vertex from $\mathscr{V}_D \setminus \mathscr{D}_{\max}$, then the previous observation combined with Proposition 8 gives $\rho(M^{\sigma,\bar{\tau}}[\mathscr{C}]) < \mathbf{t}(\chi)$. If $\mathscr{C}$ does not contain any such vertex, then it is included in $\mathscr{S}_1$. Hence, the optimality of σ_1 combined with Proposition 8 and the equality $\underline{\text{cw}}(F^{\mathscr{D}_{\max}}) = \overline{\text{cw}}(F^{\mathscr{D}_{\max}}) = \mathbf{t}(\chi)$ give $\rho(M^{\sigma,\bar{\tau}}[\mathscr{C}]) \leqslant \mathbf{t}(\chi)$. Hence, $\chi_k^{\sigma,\bar{\tau}} \leqslant \mathbf{t}(\chi) = \chi_k$ for all $k \in \mathscr{D}_{\max}$ and σ is optimal. ◀

Then, by combining Theorem 57 and Lemma 60, we get:

▶ **Theorem 61.** *A pair of optimal strategies 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. [11] 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.*

Sketch of the proof of Theorem 61. We solve entropy games by the following algorithm. We find the states with maximal value using TopClass, find optimal strategies on this dominion using Proposition 59, we split the game in two as in the decomposition property

(Lemma 60) and continue recursively on the smaller game. The bound on the number of calls to the oracle follows from Theorem 57, Lemma 60 and Proposition 59. Since the approximating oracle can be constructed in polynomial time as discussed in Remark 58, this algorithm can be implemented to work in polynomial memory. Furthermore, as in Remark 47, the time complexity of this algorithm is dominated by the number of calls to the oracle multiplied by the time cost of a single call. ◄

▶ **Corollary 62.** *Entropy games with weights and with a fixed number of People's positions are pseudo-polynomial time solvable.*

Proof. The claim follows from Theorem 61 by noting that the rank of an entropy game is bounded by the number People's states. ◄

5.3 A Lower Complexity Bound for Value Iteration for Entropy Games

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 62 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 63.** *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 $k \leqslant CW^{n-1}$.*
To construct this game, we need an estimate of the positive root of a special polynomial p_n.

▶ **Proposition 64.** *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*

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

Proof. Recall that Descartes' rule of sign states that the number of positive roots of a polynomial is bounded by the number of variations of signs of the sequence of its coefficients, and that it is equal modulo 2 to this number. It follows that $p_n(W)$ has a unique positive root, for all $W > 0$. Let us define the Newton polygon Δ as the upper boundary of the convex hull of the points (i, j) such that p_n has a monomial of degree i in x and of degree j in W. Then, the Newton-Puiseux theorem [54] shows that all the roots of p_n have Puiseux series expansions, with a leading term $aW^\alpha + o(W^\alpha)$ where α is the opposite of a slope of the Newton polygon Δ of p_n, and $a \in \mathbb{C} \setminus \{0\}$. Moreover, the horizontal width of an edge determine the number of roots of order W^α, counted with multiplicities. Here, the monomials $-W, \ldots, -Wx^{n-1}$ determine the edge $[(0, 1), (n-1, 1)]$, which is of slope 0, and has horizontal width $n-1$. The monomials $-Wx^{n-1}$ and x^n determine the edge $[(n-1, 1), (n, 0)]$, which is of slope -1 and has horizontal width 1, So there are $n-1$ roots $x_i(W) = a_iW^0 + o(W^0) = a_i + o(1)$ as $W \to \infty$, with $1 \leqslant i \leqslant n-1$, and there is one root $x_n(W) = a_nW^1 + o(W)$. Substituting $x_i(W)$ in $p_n(x_i(W)) = 0$, for $1 \leqslant i \leqslant n-1$, we find that a_i must be a root of $1 + \cdots + x^{n-1}$. Thus, a_i cannot be positive for $i < n$. So, the unique positive root of p_n is $x_n(W) = a_nW + o(W)$. Substituting $x_n(W)$ in $p_n(x_n(W)) = 0$, and cancelling negligible terms, we get $a_n = 1$. Hence,

we can write $x_n(W) = W + y(W)$, where $y(W) = o(W)$. Moreover, using $(x^{n-1} + \cdots + 1) = (x^n - 1)/(x - 1)$, $p_n(x) = 0$ can be rewritten as $x^n(x - W - 1) + W = 0$. Substituting $x_n(W) = W + y(W)$, we end up with $y(W) = 1 - W/x_n(W)^n = 1 - 1/W^{n-1} + o(1/W^n)$, hence, $x_n(W) = 1 + W - 1/W^{n-1} + o(1/W^{n-1})$. ◀

We denote by

$$A_n(W) = \begin{pmatrix} W & \dots & \dots & & W \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

the $n \times n$ companion matrix of the polynomial $p_n(W)$.

▶ **Lemma 65.** *The left Perron eigenvector u^n of the matrix $A_n(W)$ satisfies*

$$\frac{1}{2}u_1^n \leqslant u_i^n \leqslant u_1^n \text{ for all } 2 \leqslant i \leqslant n .$$

Proof. Noting that $p_n(W + 1) > 0 > p_n(W)$, and using the intermediate value theorem, we deduce that the unique positive root $x_n(W)$ of p_n satisfies $W < x_n(W) < W + 1$. The Perron root λ of $A_n(W)$ coincides with $x_n(W)$. We have $\lambda u_1 = Wu_1 + u_2$, $\lambda u_2 = Wu_1 + u_3, \ldots$, $\lambda u_{n-1} = Wu_1 + u_n$, $\lambda u_n = Wu_1$. Since all the entries of u are positive, we deduce that $u_i \geqslant u_1 W/\lambda \geqslant u_1 W/(W + 1) \geqslant u_1/2$, for all $2 \leqslant i \leqslant n$. From $\lambda u_1 = Wu_1 + u_2$, we deduce that $u_2 \leqslant (\lambda - W)u_1 \leqslant u_1$. Then, from $\lambda u_2 = Wu_1 + u_3$, we deduce that $u_3 \leqslant \lambda u_2 - Wu_1 \leqslant \lambda u_1 - Wu_1 \leqslant u_1$. Continuing in this way, we get that $u_i \leqslant u_1$ for all $2 \leqslant i \leqslant n$. ◀

The proof of Theorem 63 also relies on the following lemma. We denote by $\mathbf{1}_n$ the unit column vector of $\mathbb{R}^n$.

▶ **Lemma 66.** *Let $\alpha > 1$. Then, in the expression*

$$z(k) = \max(\mathbf{1}_n^\top A_n^k \mathbf{1}_n, \alpha \mathbf{1}_{n-1}^\top A_{n-1}^k \mathbf{1}_{n-1})$$

the maximum is achieved by the rightmost term, for all k such that

$$k \leqslant k^* := \frac{\log(\alpha(n - 1)/(4n))}{\log(\lambda_n/\lambda_{n-1})} . \tag{32}$$
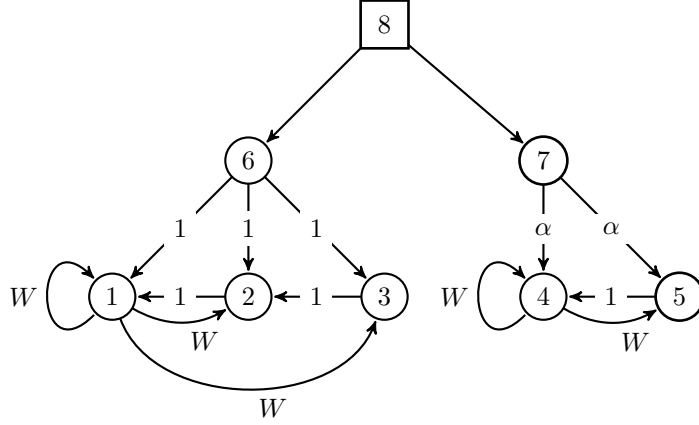
Proof. Let $\lambda_n := x_n(W)$. The Perron eigenvector u^n is defined only up to a positive multiplicative constant, so from Lemma 65, we may assume that $\mathbf{1}_n^\top \leqslant u^n \leqslant 2\mathbf{1}_n^\top$. Hence,

$$\frac{n}{2}\lambda_n^k \leqslant \frac{1}{2}\lambda_n^k u^n \mathbf{1}_n = \frac{1}{2}u^n A_n^k \mathbf{1}_n \leqslant \mathbf{1}_n^\top A_n^k \mathbf{1}_n \leqslant u^n A_n^k \mathbf{1}_n = \lambda_n^k u^n \mathbf{1}_n \leqslant 2n\lambda_n^k .$$

Hence, $\mathbf{1}_n^\top A_n^k \mathbf{1}_n \leqslant \alpha \mathbf{1}_{n-1}^\top A_{n-1}^k \mathbf{1}_{n-1}$ holds as soon as

$$2n\lambda_n^k \leqslant \frac{n - 1}{2}\alpha \lambda_{n-1}^k ,$$

which is the case when (32) holds. ◀

■ **Figure 7** The entropy game constructed in the proof of Theorem 63, here for $n = 3$. The circles represent People's states, whereas the square represents the unique significant Tribune's state. The weights are indicated on the arcs. For $\alpha \gg 1$, when the horizon is small, it is optimal for Tribune to play "bottom right", whereas when the horizon is large, the optimal move of Tribune is "bottom left".

Proof of Theorem 63. We next set $\alpha := 8$, so that $\log(\alpha(n-1)/(4n)) \geqslant \log 2$. Using the asymptotics of $\lambda_n = x_n(W)$ given by Proposition 64, we find that

$$k^* = (\log 2)W^{n-1} + o(W^{n-1}) \tag{33}$$

We now claim that $z(k)$ can be interpreted as the value in horizon $k + 1$ of an entropy game satisfying the conditions of the theorem. Figure 7 illustrates the proof when $n = 3$.

First, we note that the term $\mathbf{1}_n^\top A_n^k \mathbf{1}_n$ can be interpreted a 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 $\mathbf{1}_n^\top$. Recall every turn of an entropy game involves a succession of three stages, with moves made by Despot, Tribune, and People. So, this interpretation requires to insert dummy states of Despot and Tribune for each transition between People states. E.g., on the figure, in which dummy states are not represented, the value of the entropy game in horizon $k + 1$ with initial state 6 is precisely $\mathbf{1}_3^\top A_3^k \mathbf{1}_3$.

The term $\alpha \mathbf{1}_{n-1}^\top A_{n-1}^k \mathbf{1}_{n-1}$ admits a similar interpretation, with n People's states instead of $n + 1$. One significant state encodes the first row of A_{n-1}, whereas the other significant state encodes the row vector $\alpha \mathbf{1}_{n-1}^\top$. E.g., on the figure, the value of the entropy game in horizon $k + 1$ with initial state 7 is precisely $\mathbf{1}_2^\top A_2^k \mathbf{1}_2$.

We complete the construction of the entropy game $G_n(W)$ by adding a significant state of Tribune, with only two options: moving to the state encoding $\mathbf{1}_n^\top$, or moving to the state encoding $\alpha \mathbf{1}_{n-1}^\top$. On the figure, this significant state of Tribune is labeled 8.

Then, using the dynamic programming equation (9), we see that the value of the corresponding entropy game in horizon k, starting from the significant state of Tribune, is precisely $z(k - 1)$. Since $\lambda_n > \lambda_{n-1}$, in the mean-payoff entropy game, the optimal action for Tribune is to move to the state encoding $\mathbf{1}_n^\top$ (move "bottom left" on the figure) which guarantees a geometric growth of λ_n. However, for $k \leqslant k^* + 1$, the optimal action, for the initial move, is to select the term achieving the maximum in the expression of $z(k)$, and so, to move to the state encoding $\alpha \mathbf{1}_{n-1}^\top$ (move "bottom right"). ◀

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. Ergodicity conditions for zero-sum games. *Discrete Contin. Dyn. Syst.*, 35(9):3901–3931, 2015. `doi:10.3934/dcds.2015.35.3901`.

4 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`.

5 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.

6 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`.

7 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`.

8 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`.

9 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.

10 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`.

11 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`.

12 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`.

13 J. B. Baillon and R. E. Bruck. Optimal rates of asymptotic regularity for averaged nonexpansive mappings. In K. K. Tan, editor, *Proceedings of the Second International Conference on Fixed Point Theory and Applications*, pages 27–66. World Scientific Press, 1992.

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

15 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`.

16 J. Bolte, S. Gaubert, and G. Vigeral. Definable zero-sum stochastic games. *Mathematics of Operations Research*, 40(1):171–191, 2014. `arXiv:1301.1967`, `doi:10.1287/moor.2014.0666`.

17 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`.

18 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`.

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

20 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`.

21 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`.

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

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

24 M. Forišek. Approximating rational numbers by fractions. In *Proceedings of the 4th International Conference on Fun with Algorithms (FUN)*, volume 4475 of *Lecture Notes in Comput. Sci.*, pages 156–165. Springer, 2007. `doi:10.1007/978-3-540-72914-3_15`.

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

26 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`.

27 S. Gaubert and N. Stott. A convergent hierarchy of non-linear eigenproblems to compute the joint spectral radius of nonnegative matrices. *Mathematical Control and Related Fields*, 10(3):573–590, 2020. `arXiv:1805.03284`, `doi:10.3934/mcrf.2020011`.

28 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`.

29 M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms Combin.* Springer, Berlin, 1993. `doi:10.1007/978-3-642-78240-4`.

30   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`.

31   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`.

32   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`.

33   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`.

34   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`.

35   M. S. Keane. Ergodic theory and subshifts of finite type. In M. Keane T. Bedford and C. Series, editors, *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*. Oxford University Press, 1991. Chapter 2.

36   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`.

37   S. Kwek and K. Mehlhorn. Optimal search for rationals. *Inform. Process. Lett.*, 86(1):23–26, 2003. `doi:10.1016/S0020-0190(02)00455-6`.

38   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`.

39   K. Mahler. An inequality for the discriminant of a polynomial. *Michigan Math. J.*, 11:257–262, 1964.

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

41   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`.

42   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`.

43   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`.

44   M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Ser. Probab. Stat. Wiley, Hoboken, NJ, 2005.

45   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`.

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

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

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

49   A. Schrijver. *Theory of linear and integer programming*. Wiley-Intersci. Ser. Discrete Math. Optim. Wiley, New York, 1987.

50   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`.

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

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

53 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`.

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

55 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.

56 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`.

57 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`.