Solving Irreducible Stochastic Mean-Payoff Games and Entropy Games by Relative Krasnoselskii–Mann Iteration

Marianne Akian
INRIA and CMAP, École polytechnique, IP Paris, CNRS, France

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

Ulysse Naepels
École polytechnique, IP Paris, France

Basile Terver
École polytechnique, IP Paris, France

Abstract

We analyse an algorithm solving stochastic mean-payoff games, combining the ideas of relative value iteration and of Krasnoselskii–Mann damping. We derive parameterized complexity bounds for several classes of games satisfying irreducibility conditions. We show in particular that an ϵ-approximation of the value of an irreducible concurrent stochastic game can be computed in a number of iterations in $O(|\log \epsilon|)$ where the constant in the $O(\cdot)$ is explicit, depending on the smallest non-zero transition probabilities. This should be compared with a bound in $O(\epsilon^{-1}|\log(\epsilon)|)$ obtained by Chatterjee and Ibsen-Jensen (ICALP 2014) for the same class of games, and to a $O(\epsilon^{-1})$ bound by Allamigeon, Gaubert, Katz and Skomra (ICALP 2022) for turn-based games. We also establish parameterized complexity bounds for entropy games, a class of matrix multiplication games introduced by Asarin, Cervelle, Degorre, Dima, Horn and Kozyakin. We derive these results by methods of variational analysis, establishing contraction properties of the relative Krasnoselskii–Mann iteration with respect to Hilbert's semi-norm.

2012 ACM Subject Classification Theory of computation → Algorithmic game theory

Keywords and phrases Stochastic mean-payoff games, concurrent games, entropy games, relative value iteration, Krasnoselskii–Mann fixed point algorithm, Hilbert projective metric

Digital Object Identifier 10.4230/LIPIcs.MFCS.2023.10

Related Version *Extended Version*: https://arxiv.org/pdf/2305.02458.pdf

Acknowledgements We thank the reviewers for helpful comments.

1 Introduction

1.1 Motivation and context

Stochastic mean-payoff games are a fundamental class of zero-sum games, appearing in various guises. In *turn-based* games, two players play sequentially, alternating moves, or choices of an action, being aware of the previous decision of the other player. Turn-based games with mean-payoff and finite state and action spaces are among the unsettled problems in complexity theory: they belong to the complexity class NP ∩ coNP [14, 40] but are not known to be polynomial-time solvable. We refer the reader to the survey [7] for more information on the different classes of turn-based games. In contrast, in *concurrent games*, at each stage, the two players choose simultaneously one action, being unaware of the choice of the other player at the same stage. Turn-based games are equivalent to a subclass of concurrent games (in which in each state, one of the two players is a dummy). The existence

48th International Symposium on Mathematical Foundations of Computer Science (MFCS 2023).
Editors: Jérôme Leroux, Sylvain Lombardy, and David Peleg; Article No. 10; pp. 10:1–10:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of the value for concurrent stochastic mean-payoff games is a celebrated result of Mertens and Neyman [28]. This builds on earlier results by Bewley and Kohlberg, connecting mean-payoff concurrent games with discounted concurrent games, by making the discount factor tend to 1, see [11]. Concurrent games are hard to solve exactly: the value is an algebraic number whose degree may be exponential in the number of states [21]. Moreover, concurrent reachability games are square-root sum hard [16].

Another class consists of *entropy games*, introduced by Asarin, Cervelle, Degorre, Dima, Horn and Kozyakin as an interesting category of "matrix multiplication games" [8]. Entropy games capture a variety of applications, arising in risk sensitive control [23, 6], portfolio optimization [3], growth maximization and population dynamics [36, 33, 32, 39]. Asarin et al. showed that entropy games belong to the class NP $\cap$ coNP, showing an analogy with turn based games. In [1], Akian, Gaubert, Grand-Clément and Guillaud showed that entropy games are actually special cases of stochastic mean-payoff games, in which action spaces are infinite sets (simplices), and payments are given by Kullback-Leibler divergences.

A remarkable subclass of stochastic mean-payoff games arises when imposing *ergodicity* or *irreducibility* conditions. Such conditions entail that the value of the game is independent of the initial state. The simplest condition of this type requires that every pair of policies (Markovian stationary strategies) of the two players induces an irreducible Markov chain. Then, the solution of the game reduces to solving a nonlinear eigenproblem of the form $T(u) = \lambda e + u$, in which $u \in \mathbb{R}^n$ is a non-linear eigenvector, λ is a non-linear eigenvalue, which provides the value of the mean-payoff game, e is the unit vector of $\mathbb{R}^n$, and T is a self-map of $\mathbb{R}^n$, the dynamic programming operator of the game, which we shall refer to as the "Shapley" operator. In fact, Shapley originally introduced a variant of this operator, adapted to the discounted case [34]. The undiscounted mean-payoff case was subsequently considered by Gillette [20]. We refer the reader to [29, 31] for background on Shapley operators and on the "operator approach" to games, and to [2] for a discussion of the non-linear eigenproblem.

In the one-player case, White [38] introduced *relative value iteration*, which consist in fixed point iterations up to additive constants $\lambda_k \in \mathbb{R}$, i.e. $x_{k+1} = T(x_k) - \lambda_k e$. This solves the non-linear eigenproblem $T(u) = \lambda e + u$ under a primitivity assumption. However, this assumption appears to be too restrictive in the light of the classical Krasnoselkii–Mann algorithm [25, 27], which allows one to find a fixed point of a nonexpansive self-map T of a finite dimensional normed space, by constructing the "damped" sequence $x_{k+1} = (1-\theta)T(x_k) + \theta x_k$, where $0 < \theta < 1$. Indeed, it was proposed in [19] to apply this algorithm to the non-linear eigenproblem $T(u) = \lambda e + u$, thought of as a fixed point problem in the quotient vector space $\mathbb{R}^n/\mathbb{R}e$. We will refer to this algorithm as the *relative Krasnoselskii–Mann* value iteration. An error bound in $O(1/\sqrt{k})$ was derived in [19] for this algorithm, as a consequence of a general theorem of Baillon and Bruck [10], and the existence of an asymptotic geometric convergence rate was established in a special case. This left open the question of obtaining stronger iteration complexity bounds, in a "white box model", for specific classes of stochastic mean-payoff games.

1.2 Contribution

We apply the relative Krasnoselskii–Mann value iteration algorithm to deduce complexity bounds for several classes of stochastic games. We consider in particular *unichain* concurrent stochastic mean-payoff games, in which every pair of policies of the two players induces a unichain transition matrix (i.e., a stochastic matrix with a unique final class). We define $p_{\min}$ to be the smallest non-zero off-diagonal transition probability in the model. Corollary 20 shows that the relative Krasnoselskii–Mann iteration yields an ϵ-approximation of the value of the game, after $C|\log \epsilon|$ iterations. The factor C is exponential in the bit-size of the input, it has an essential term of the form $k\theta^{-k}$, in which $k \leqslant n$ is a certain "unichain index", which is equal to 1 if all the transition probabilities are positive, $\theta = p_{\min}/(1+p_{\min})$, and n denotes the number of states. Then, we consider the special case of unichain turn-based games, with rational transition probabilities whose denominator divides M. Theorem 23 shows that optimal policies can be obtained after a number of iterations of order M^k. The main tool is Theorem 19, which shows that a suitable iterate of the Shapley operator of a unichain concurrent game is a contraction in Hilbert's seminorm. This theorem is proved using techniques of variational analysis, in particular we use a classical result of Mills [30], characterizing the directional derivative of the value of a matrix game, and properties of nonsmooth semidifferentiable maps.

Finally, we introduce a variant of the relative Krasnoselkii–Mann algorithm, adapted to entropy games. Theorem 26 shows that an irreducible entropy game can be solved exactly in a time of order $(1 + \mathcal{A}/\underline{m})^k$ where $k \leqslant n$ is a certain "irreducibility index", $\underline{m} \geqslant 1$ is the smallest multiplicity of an off-diagonal transition, and $\mathcal{A}$ is a measure of the *ambiguity* of the game. In particular, we have $W \leqslant \mathcal{A} \leqslant n^{1-1/n}W$ where W is the maximal multiplicity of a transition. The proof exploits the Birkhoff-Hopf theorem, which states that a positive matrix is a contraction in Hilbert's projective metric.

The proofs of the present results can be found in the extended version of this article [4].

1.3 Related work

The algorithmic approach of stochastic mean-payoff games games satisfying irreducibility conditions goes back to the work of Hoffman and Karp [22], applying policy iteration. Chatterjee and Ibsen-Jensen [13] studied concurrent stochastic mean-payoff games, under appropriate conditions of ergodicity. They showed in particular that the problem of approximation of the value is in FNP, and that this approximation problem, restricted to turn-based ergodic games, is at least as hard as the decision problem for simple stochastic games. They also showed that value iteration provides and ϵ-approximation of the value of a concurrent stochastic game statisfying an irreducibility condition in $O(\tau\epsilon^{-1}|\log \epsilon|)$ iterations, where τ denotes a bound of the passage time between any two states under an arbitrary strategy, see Theorem 18, *ibid.* A recent "universal bound" on value iteration by Allamigeon, Gaubert, Katz and Skomra [5, Th. 13] entails an improvement of this bound to $O(\tau\epsilon^{-1})$. Corollary 20 further improves this bound to get $C|\log \epsilon|$. However, the later result requires an unichain assumption, whereas the assumption of [5, Th. 13] is milder.

The question of computing the value of a concurrent discounted stochastic game has been studied by Hansen, Koucký, Lauritzen, Miltersen and Tsigaridas in [21], who showed, using semi-algebraic geometry techniques, that an ϵ-approximation of the value of a general concurrent game can be obtained in polynomial time if the number n of states is fixed. The exponent of the polynomial is of order $O(n)^{n^2}$ and it was remarked in [21] that "getting a better dependence on n is a very interesting open problem". Boros, Gurvich, Elbassioni and Makino considered the notion of ϵ-ergodicity of a concurrent mean-payoff game, requiring that the mean-payoff of two initial states differ by at most ϵ. They provided a potential-reduction algorithm allowing one to decide ϵ-ergodicity, and to get an ϵ-approximation of the value, with a dependence in ϵ of order $\epsilon^{-O(2^{2^n} n \max(|A|,|B|))}$, see [12]. Attia and Oliu-Barton developed in [9] a bisection algorithm, with a complexity bound polynomial in $|\log \epsilon|$ and in $|A|^n$ and $|B|^n$ where A, B are the action spaces. In contrast to these three works, our approach only applies to the subclass of *unichain* concurrent games, but its complexity has a better dependence in the number of states; in particular, the exponents in our bound is at most n, and the execution time grows only polynomially with the numbers of actions $|A|$ and $|B|$. Moreover, our approach applies more generally to infinite (compact) action spaces (we only need an oracle evaluating the value of a possibly infinite matrix game up to a given accuracy).

The analysis of relative value iteration, using contraction techniques, goes back to the work of Federguen, Schweitzer and Tijms [17], dealing with the one-player and finite action spaces case, under a primitivity condition. The novelty here is the analysis of the concurrent two-player case, as well as the analysis of the effect of the Krasnoselskii–Mann damping, allowing one to replace earlier primitivity conditions by a milder unichain condition. Moreover, even in the one-player case, our formula for the contraction rate given in Theorem 19 improves the one of [17].

Our results of Section 9 dealing with entropy games are inspired by the series of works [8, 1, 5]. The subclass of "Despot-free" entropy games can be solved in polynomial time [1], and it is an open question whether general entropy games can be solved in polynomial time. The approach of [5] entails that one can get an ϵ-approximation of the value of an entropy game in $O(\epsilon^{-1})$ iterations, where the factor in the $O(\cdot)$ is exponential in the parameters of the game. This bound is refined here to $O(|\log \epsilon|)$, in which the factor in the $O(\cdot)$ depends on a measure of "ambiguity" – but our approach requires an irreducibility assumption.

2 Preliminary results on Shapley operators

Let n be an integer. A map $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *order-preserving* when: $\forall x, y \in \mathbb{R}^n, x \leqslant y \implies T(x) \leqslant T(y)$, where $\leqslant$ denotes the standard partial order of $\mathbb{R}^n$. It is *additively homogeneous* when: $\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, T(x + \lambda e) = T(x) + \lambda e$ where e is the vector of $\mathbb{R}^n$ having 1 in each coordinate.

▶ **Definition 1.** *A map $T : \mathbb{R}^n \to \mathbb{R}^n$ is an (abstract)* Shapley operator *if it is order-preserving and additively homogeneous.*

We will justify the terminology "Shapley operator" in the next section, where we give concrete examples, arising as dynamic programming operators of different classes of zero-sum repeated games. We set $[n] := \{1, \dots n\}$. For any $x \in \mathbb{R}^n$, we denote $\mathbf{t}(x) := \max_{i \in [n]} x_i$ and $\mathbf{b}(x) := \min_{i \in [n]} x_i$ (read "top" and "bottom"). We define the *Hilbert's seminorm* of x by: $\|x\|_{\mathrm{H}} = \mathbf{t}(x) - \mathbf{b}(x)$. Since $\|x\|_{\mathrm{H}} = 0$ iff $x \in \mathbb{R}e$, we get that $\| \cdot \|_{\mathrm{H}}$ is actually a norm on the quotient vector space $\mathbb{R}^n/\mathbb{R}e$. We also notice that $\|x\|_\infty = \inf\{\lambda \in \mathbb{R}_+ \mid -\lambda e \leqslant x \leqslant \lambda e\}$ and $\|x\|_H = \inf\{\beta - \alpha \in \mathbb{R}_+ \mid \alpha, \beta \in \mathbb{R}, \alpha e \leqslant x \leqslant \beta e\}$. It is easy to show, thanks to these expressions, that a Shapley operator T is non-expansive (i.e., 1-Lipschitz) for $\| \cdot \|_{\mathrm{H}}$ and for $\| \cdot \|_\infty$. Then, it induces a self-map $\overline{T}$ on the quotient vector space $\mathbb{R}^n/\mathbb{R}e$, sending the equivalence class $x + \mathbb{R}e$ to $T(x) + \mathbb{R}e$, and which is non-expansive.

▶ **Definition 2.** *We define the* escape rate $\chi(T)$ *of a Shapley operator T as $\lim_{k\to\infty} k^{-1}T^k(v)$, where v is an arbitrary vector in $\mathbb{R}^n$. The lower and upper escape rates are defined respectively by $\underline{\chi}(T) = \lim_{k\to\infty} k^{-1}\mathbf{b}(T^k(v))$ and $\overline{\chi}(T) = \lim_{k\to\infty} k^{-1}\mathbf{t}(T^k(v))$.*

Since T is nonexpansive in the sup-norm, the existence and the values of these limits are independent of the choice of $v \in \mathbb{R}^n$. In general, the escape rate $\chi(T) = \lim_{k\to\infty} k^{-1}T^k(v)$ may not exist, but a subadditive argument shows that the lower and upper escape rates always exist, see e.g. [18]. A fundamental tool to establish the existence of the escape rate is to consider the following ergodic equation.

▶ **Definition 3.** *We say that the* ergodic equation *has a solution when there exists $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^n$ such that : $T(u) = \lambda e + u$.*

▶ **Observation 4.** *If the above ergodic equation is solvable, then $\chi(T) = \lambda e$. More generally, if $\alpha e + v \leqslant T(v) \leqslant \beta e + v$ for some $v \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha \leqslant \underline{\chi}(T) \leqslant \overline{\chi}(T) \leqslant \beta$.*

Proof. By an immediate induction, and as T is order-preserving and additively homogeneous we have : $k\alpha e + v \leqslant T^k(v) \leqslant k\beta e + v$. Then, $k\alpha + \mathbf{b}(v) \leqslant \mathbf{b}(T^k(v)) \leqslant \mathbf{t}(T^k(v)) \leqslant k\beta + \mathbf{t}(v)$. Dividing by k and letting k tend to infinity, we obtain the second statement. ◀

We are inspired by the following observation from fixed point theory.

▶ **Observation 5.** *Suppose now that T^q is γ-contraction in Hilbert's seminorm $\|\cdot\|_H$, for some $q \geqslant 1$ and $0 < \gamma < 1$. Then, the ergodic equation is solvable.*

Shapley operators include (finite dimensional) Markov operators, which are of the form $T(x) = Mx$, where M is a $n \times n$ stochastic matrix (meaning that M has nonnegative entries and row sums one). In this case, an exact formula is known for the contraction rate. In fact, one can consider the operator norm of M, thought of as a linear map acting on the quotient vector space $\mathbb{R}^n/\mathbb{R}e$, $\|M\|_{\mathrm{H}} = \sup\limits_{u \notin \mathbb{R}e} \frac{\|Mu\|_{\mathrm{H}}}{\|u\|_{\mathrm{H}}}$.

▶ **Theorem 6** (Corollary of [15]). $\|M\|_{\mathrm{H}} = \delta(M) \coloneqq 1 - \min_{1 \leqslant i < j \leqslant n} \left\{ \sum_{k \in [n]} \min(M_{ik}, M_{jk}) \right\}$.

The term $\delta(M)$ is known as *Dobrushin ergodicity coefficient.*

3 Two classes of zero-sum two-player repeated games

We next recall the definition and basic properties of two classes of zero-sum two-player games with finite state spaces. More details can be found in [29] for stochastic games and in [8, 1] for entropy games.

3.1 Concurrent repeated zero-sum stochastic two-player games

We assume that the state space is equal to $[n] = \{1, \dots, n\}$. We call the two players "Min" and "Max". The game is specified by the following data. For every state $i \in [n]$, we are given two non-empty compact sets $A(i)$ and $B(i)$, representing the admissible actions of players Min and Max, respectively. For every $i \in [n]$ and every choice of actions $(a, b) \in A(i) \times B(i)$, we are given a real number r_i^{ab}, representing an instantaneous payment, and a stochastic vector $P_i^{a,b} = (P_{ij}^{ab})_{j \in [n]}$, meaning that $P_{ij}^{ab} \geqslant 0$ and that $\sum_{j \in [n]} P_{ij}^{ab} = 1$. We assume that the functions $(a, b) \mapsto r_i^{ab}$ and $(a, b) \mapsto P_i^{a,b}$ are continuous.

The concurrent game is played in successive stages, starting from a known initial state i_0 at stage 0. We denote by a_k and b_k the actions selected by Players Min and Max at stage k, respectively, and by i_k the state at this stage. The history until stage k consists of the sequence $H_k = ((i_\ell, a_\ell, b_\ell)_{0 \leqslant \ell < k}, i_k)$. A randomized strategy of Player Min (resp. Max) is a collection of measurable functions assigning to every history H_k a probability measure α_k (resp. β_k) on the compact set $A(i_k)$ (resp. $B(i_k)$). At stage k, being informed of the history H_k up to this stage, Player Min draws a random action a_k according to the probability measure α_k, and similarly, Player Max draws a random action b_k according to the probability measure β_k. Then, Player Min makes to Player Max an instantaneous payment of $r_{i_k}^{a_k, b_k}$, and the next state i_{k+1} is drawn randomly according to the probability measure $(P_{i_k,j}^{a_k b_k})_{j \in [n]}$ on the state space $[n]$, i.e., the conditional probability that $i_{k+1} = j$, given the history H_k and actions a_k, b_k, is given by $P_{i_k,j}^{a_k b_k}$. We shall say that a strategy is *pure* or *deterministic* if the action of the player is chosen as a deterministic function of the history. We denote by σ_k (resp. τ_k) the strategy of Player Min (resp. Max) at stage k, and denote by σ and τ

the sequences $(\sigma_k)_{k\geqslant 0}$ and $(\tau_k)_{k\geqslant 0}$. In this way, to any initial state $i_0 \in [n]$ and any pair of strategies (σ, τ) of the two players is associated the infinite random sequence $(i_k, a_k, b_k)_{k\geqslant 0}$. We denote by $\mathbb{E}_{i_0}^{\sigma,\tau}$ the expectation operator with respect to this process.

We shall consider special classes of strategies. We denote by Δ_X the set of probability measures on a compact set X. A *randomized policy* of Player Min is a map $\alpha : [n] \to \cup_{i\in[n]} \Delta_{A(i)}, i \mapsto \alpha_i$. For each $i \in [n]$, $d\alpha_i(a)$ determines the probability an action $a \in A_i$ is chosen, according to this policy. Thus, the set of randomized policies of Min, $\Pi_{\mathrm{R}}^{\mathrm{Min}}$, can be identified to $\prod_{i\in[n]} \Delta_{A(i)}$. A policy of Min is said to be *pure* if for all $i \in [n]$, α_i is a Dirac measure, so that $d\alpha_i = \delta_{a_i}$ for some $a_i \in A_i$. Such a policy prescribes to play the deterministic action a_i when in state i. Therefore, a pure policy is uniquely specified by the map $i \mapsto a_i$. This allows us to identify the set of pure policies, denoted by $\Pi_{\mathrm{P}}^{\mathrm{Min}}$, to the product $\prod_{i\in[n]} A_i$. A randomized *Markovian* strategy σ of Player Min is a strategy such that the decision prescribed by σ_k depends only on the current state i_k. In other words, it is obtained by selecting, at each time step, a randomized policy of Player Min, and playing the action according to this policy. A Markovian strategy is *pure* if only pure policies are selected. It is *stationary* if the same policy is applied at every time step k. In this way, a pure (resp. randomized) Markovian stationary strategy can be identified to a pure (resp. randomized) policy. We shall use the same notation and terminology for Player Max, mutatis mutandis. In particular, we denote by $\Pi_{\mathrm{R}}^{\mathrm{Max}}$ and $\Pi_{\mathrm{P}}^{\mathrm{Max}}$ the sets of randomized and pure policies of Player Max. We shall also denote by $\Pi_{\mathrm{P}} = \Pi_{\mathrm{P}}^{\mathrm{Min}} \times \Pi_{\mathrm{P}}^{\mathrm{Max}}$ and $\Pi_{\mathrm{R}} = \Pi_{\mathrm{R}}^{\mathrm{Min}} \times \Pi_{\mathrm{R}}^{\mathrm{Max}}$ the spaces of pairs of policies.

Given an initial state i_0 and a pair of strategies (σ, τ) of the two players, the expected payment received by Player Max in horizon N is defined by

$$J_{i_0}^N(\sigma, \tau) := \mathbb{E}_{i_0}^{\sigma,\tau} \left[\sum_{k=0}^{N-1} r_{i_k}^{a_k, b_k} \right] \ .$$

We shall denote by $J^N(\sigma, \tau)$ the vector of $\mathbb{R}^n$ with the above i_0 entry, for each $i_0 \in [n]$. The finite horizon game has a value $v^N \in \mathbb{R}^n$ and has a pair of optimal (randomized) strategies (σ^*, τ^*), meaning that

$$J^N(\sigma^*, \tau) \leqslant v^N = J^N(\sigma^*, \tau^*) \leqslant J^N(\sigma, \tau^*) \ , \tag{1}$$

for all pairs (σ, τ) of strategies, see [29]. Moreover, one can choose the pair of optimal strategies (σ^*, τ^*) to be Markovian, that is $(\sigma_k^*, \tau_k^*) \in \Pi_{\mathrm{R}}$ for all $k \leqslant N$ (but it generally depends on k and N). These optimal strategies can be obtained by using the dynamic programming equation of the game, as follows.

For any $i, j \in [n]$, $\alpha_i \in \Delta_{A(i)}$ and $\beta_i \in \Delta_{B(i)}$, let us denote

$$r_i^{\alpha_i,\beta_i} = \int_{A(i)\times B(i)} r_i^{a,b} d\alpha_i(a) d\beta_i(b) \quad \text{and} \quad P_{i,j}^{\alpha_i,\beta_i} = \int_{A(i)\times B(i)} P_{i,j}^{a,b} d\alpha_i(a) d\beta_i(b) \ . \tag{2}$$

This extends the functions $(a, b) \mapsto r_i^{a,b}$ and $(a, b) \mapsto P_{i,j}^{a,b}$ from $A(i) \times B(i)$ to $\Delta_{A(i)} \times \Delta_{B(i)}$. We then define the Shapley operator T of the concurrent game as the map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T_i(v) = \min_{\alpha_i\in\Delta_{A(i)}} \max_{\beta_i\in\Delta_{B(i)}} \Big(r_i^{\alpha_i,\beta_i} + \sum_{j\in[n]} P_{ij}^{\alpha_i,\beta_i} v_j \Big), \quad \text{for } i \in [n], \ v \in \mathbb{R}^n \ . \tag{3}$$

Note that in the above expression the infimum and supremum commute, owing to the compactness of action spaces, and continuity assumptions on the functions $(a, b) \mapsto r_i^{a,b}$ and $(a, b) \mapsto P_{ij}^{ab}$ (this follows from Sion's minimax theorem). Moreover, the operator T satisfies the properties of Definition 1.

Then, the value of the concurrent game in finite horizon is obtained from the recurrence equations: $v^0 = 0, \quad v^N = T(v^{N-1})$. Moreover, optimal strategies of the game when the remaining time is $k < N$ (or at stage $N-k$) are obtained by choosing optimal policies α and β with respect to the vectors v^k, that is such that α_i and β_i are optimal in the expression of $T_i(v^k)$ in (3).

We now describe the *mean-payoff game*, which is obtained by considering the Cesaro limit of the payoff as the horizon N tends to infinity. More precisely, we set:

$$\chi^+_{i_0}(\sigma,\tau) \coloneqq \limsup_{N\to\infty} N^{-1} J^N_{i_0}(\sigma,\tau) \quad \chi^-_{i_0}(\sigma,\tau) \coloneqq \liminf_{N\to\infty} N^{-1} J^N_{i_0}(\sigma,\tau) \enspace .$$

We shall say that the game with mean-payoff has a value $\chi^* \in \mathbb{R}^n$ if for all $\varepsilon > 0$, there exists strategies $\sigma^\varepsilon, \tau^\varepsilon$ of the two players which are ε-optimal, meaning that for every strategies σ and τ, $-\varepsilon e + \chi^+(\sigma^\varepsilon, \tau) \leqslant \chi^* \leqslant \chi^-(\sigma, \tau^\varepsilon) + \varepsilon e$. Mertens and Neyman [28], building on a result of Bewley and Kohlberg [11], showed that when the action spaces $A(i)$ and $B(i)$ are finite, the mean-payoff game has a value (actually, in a stronger *uniform* sense). Moreover, the value coincides with the escape rate of the Shapley operator, i.e., $\chi^* = \lim_k T^k(0)/k$. A counter-example of Vigeral shows that these properties do not carry over to the case of general compact action spaces [37].

One particular case that will interest us is when the ergodic equation is solvable, that is when there exists $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$ such that $T(v) = \lambda e + v$. In that case, $\chi^* = \lambda e$ and there exists optimal randomized strategies for the two players which are both Markovian and stationary. Such a pair of strategies is obtained by choosing a pair (α, β) of policies such that α and β achieve the minimum and the maximum, respectively, in the expression of $T(v)$ in (3). We shall see that the ergodic equation is always solvable under a unichain condition, even in the case of compact action spaces (Theorem 11).

A remarkable subclass of concurrent games consists of *turn-based* games. Then, the actions spaces $A(i)$ and $B(i)$ are required to be *finite*, and for every state $i \in [n]$, we assume that either $A(i)$ or $B(i)$ is a singleton. In other words, there is a bipartition $[n] = I_{\mathrm{Min}} \uplus I_{\mathrm{Max}}$ of the set of states, so that in every state $i \in I_{\mathrm{Min}}$ (resp. I_{Max}), Min (resp. Max) is the only player who has to take a decision. Then, the Shapley operator of the game reduces to $T_i(x) = \min_{a\in A(i)} \max_{b\in B(i)} \left(r_i^{a,b} + \sum_j P^{a,b}_{i,j} x_j\right)$, for $i \in [n]$, where again the min and max commute, because in every $i \in [n]$, either the min or the max is taken over a set reduced to a singleton. When the ergodic equation $T(v) = \lambda e + v$ of a turn-based game is solvable, one obtains *pure* optimal policies in the mean-payoff game, by selecting actions that achieve the minimum and the maximum in each coordinate $[T(v)]_i$ with $i \in [n]$. More generally, the existence of pure optimal policies for turn-based mean-payoff stochastic games was shown by Liggett and Lippman [26]. An illustrative example is given in Appendix A.

3.2 Entropy games

Entropy games were introduced in [8]. We use here the slightly more general model of [1, 5], to which we refer for background. An *entropy game* is a turn-based game played on a (finite) digraph $(\mathcal{V}, \mathcal{E})$, with two players, called "Despot" and "Tribune", and an additional non-deterministic player, called "People". We assume the set of vertices $\mathcal{V}$ has a non-trivial partition: $\mathcal{V} = \mathcal{V}_D \uplus \mathcal{V}_T \uplus \mathcal{V}_P$. Players Despot, Tribune, and People control the states in $\mathcal{V}_D$, $\mathcal{V}_T$ and $\mathcal{V}_P$ respectively, and they alternate their moves, i.e., $\mathcal{E} \subset (\mathcal{V}_D \times \mathcal{V}_T) \cup (\mathcal{V}_T \times \mathcal{V}_P) \cup (\mathcal{V}_P \times \mathcal{V}_D)$. We suppose that every edge $(p, d) \in \mathcal{E}$ with $p \in \mathcal{V}_P$ and $d \in \mathcal{V}_D$ is equipped with a *multiplicity* m_{pd} which is a (positive) natural number. For simplicity of exposition, we shall define here the value of an entropy game using only pure policies. More precisely, a (pure) policy σ of Despot is a map which assigns to every node $d \in \mathcal{V}_D$ a node t such that $(d,t) \in \mathcal{E}$. Similarly, a policy τ of Tribune is a map which assigns to every node $t \in \mathcal{V}_T$ a node $p \in \mathcal{V}_P$. We denote by n the cardinality of $\mathcal{V}_D$. Such a pair of policies determine a $n \times n$ matrix $M^{\sigma,\tau}$, such that $M^{\sigma,\tau}_{d,d'} = m_{\tau(\sigma(d)),d'}$. Given an initial state $\bar{d} \in \mathcal{V}_D$, we measure the "freedom" of Player People by the limit $R(\sigma,\tau) := \lim_{k\to\infty}[((M^{\sigma,\tau})^k e)_{\bar{d}}]^{1/k}$. A pair of (pure) policies determine a subgraph $\mathcal{G}^{\sigma,\tau}$, obtained by keeping only the successor prescribed by σ for every node of $\mathcal{V}_D$, and similarly for τ and $\mathcal{V}_T$. Then, the "freedom" of player People is precisely the geometric growth rate of the number of paths of length $3k$ starting from node $\bar{d}$, counted with multiplicities, as $k \to \infty$. In general, the graph $\mathcal{G}^{\sigma,\tau}$ may have several strongly connected components, and it is observed in [5] that $R(\sigma,\tau)$ coincides with the maximal spectral radii of the diagonal blocks of the matrix $M^{\sigma,\tau}$ corresponding to the strongly connected components to which the initial state $\bar{d}$ has access in $\mathcal{G}^{\sigma,\tau}$. In an entropy game, Despot wishes to minimize the freedom of People, whereas Tribune (a reference to the magistrate of Roman republic) wishes to maximize it. It is shown in [1] that the entropy game has a value in the space of pure policies, meaning that there exists pure policies σ^*, τ^*, such that $R(\sigma^*,\tau) \leqslant R(\sigma^*,\tau^*) \leqslant R(\sigma,\tau^*)$ for all pure policies σ, τ. (Actually, more general, history dependent, strategies are considered in [1], and it is shown there that pure policies are optimal).

The dynamic programming operator of an entropy game is the self-map F of $\mathbb{R}^n_{>0}$ given by $F_d(x) = \min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}} \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}} \sum_{d'\in\mathcal{V}_D,(p,d')\in\mathcal{E}} m_{p,d'}x_{d'}$, for $d \in \mathcal{V}_D$. Then, the operator $T := \log \circ F \circ \exp$ is a Shapley operator. It is shown in [1] that the value of the entropy game with initial state $\bar{d}$ is given by the limit $\lim_{k\to\infty}[(F^k(e))_{\bar{d}}]^{1/k}$.

4 The unichain property

Recall that to every $n \times n$ nonnegative matrix M is associated a digraph with set of nodes $[n]$, such that there is an arc from i to j if $M_{ij} > 0$. The matrix is *irreducible* if this digraph is strongly connected. It is *unichain* if this digraph has a unique final strongly connected component (a strongly components is *final* if any path starting from this component stays in this component). The property of unichainedness is sometimes referred to as *ergodicity* since a stochastic matrix is unichain iff it has only one invariant measure, or equivalently, if the only *harmonic vectors* (i.e. the solutions v of $Mv = v$) are the constant vectors, see the discussion in Theorem 1.1 of [2], and the references therein.

Given a pair $(\sigma,\tau) \in \Pi_{\mathrm{P}}$ of pure policies, we define the stochastic matrix: $P^{\sigma,\tau} = (P^{\sigma(i),\tau(i)}_{i,j})_{i,j\in[n]}$.

▶ **Definition 7.** *We say that a game is* unichain *(resp.* irreducible*) if for all pairs of pure policies σ, τ, the matrix $P^{\sigma,\tau}$ is unichain (resp. irreducible).*

▶ **Definition 8.** *We say that a subset S of the states is* closed *under the action of a matrix $P^{\sigma,\tau}$ if, starting from a state $s \in S$ and playing according to the policies σ and τ, the next state is still in S.*

▶ Remark 9. If S is a set closed under the action of an unichain matrix $P^{\sigma,\tau}$, then S contains the final class of this matrix.

▶ Remark 10. The final class does not have to be the same for all pairs σ, τ of policies in our definition of unichain games.

The following theorem addresses the issue of the existence of a solution to the ergodic equation in the case of a unichain game.

▶ **Theorem 11.** *Let T be the Shapley operator of a unichain concurrent stochastic game. Then, there exists a vector $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that $T(v) = \lambda e + v$. Moreover, there exists a pair of optimal (randomized) Markovian stationary strategies, obtained by selecting actions that achieve the minimum and maximum in the expression of $[T(v)]_i$, for each state $i \in [n]$.*

5 Relative value iteration

Relative value iteration was introduced in [38] to solve one player stochastic mean-payoff games (i.e., average cost Markov decision processes). The "vanilla" value iteration algorithm consists in computing the sequence $x_{k+1} = T(x_k)$, starting from $x_0 = 0$. Then, x_k yields the value vector of the game in horizon k, an so, we expect x_k to go to infinity as $k \to \infty$. The idea of *relative* value is to renormalize the sequence by additive constants. We state in Algorithm 1 a general version of relative value iteration, allowing for *approximate* dynamic programming oracles. This will allow us to obtain complexity results in the Turing model of computation, by computing a rational approximation of the value of the Shapley operator $T(x)$ at a given rational vector x up to a given accuracy.

Algorithm 1 Relative value iteration in approximate arithmetics.

```
input: A final requested numerical precision $\epsilon > 0$ and a parameter $0 < \eta \leqslant \epsilon/3$. An
    oracle $\tilde{T}$ which provides an $\eta$-approximation in the sup-norm of a Shapley operator $T$.
$x := 0 \in \mathbb{R}^n$
repeat
    $x := \tilde{T}(x) - \mathbf{t}(\tilde{T}(x))e$
until $\|x - \tilde{T}(x)\|_{\mathrm{H}} \leqslant \epsilon/3$
$\alpha := \mathbf{b}(\tilde{T}(x) - x); \beta := \mathbf{t}(\tilde{T}(x) - x)$
return $x, \alpha, \beta$        ▷ The lower and upper escape rates of $T$ are included in the interval
    $[\alpha - \epsilon/3, \beta + \epsilon/3]$, which is of width at most $\epsilon$
```

▶ **Theorem 12.** *Suppose that T is a Shapley operator. Then,*

1. *When it terminates, Algorithm 1 returns a valid interval of width at most ϵ containing the lower and upper escape rates of T.*
2. *If there is an integer q and a scalar $0 < \gamma < 1$ such that T^q is a γ-contraction in Hilbert's seminorm, and if η is chosen small enough, in such a way that $\eta(12 + 24q/(1-\gamma)) \leqslant \epsilon$, then Algorithm 1 terminates in at most $q(\log \|T(0)\|_H + \log 6 + |\log \epsilon|)/|\log \gamma|$ iterations.*

The proof exploits the nonexpansiveness of the operator T in Hilbert's seminorm.

6 Krasnoselskii–Mann damping

We shall see that for turn-based or concurrent games, it is useful to replace the original Shapley operator by a Krasnoselskii–Mann damped version of this operator. This will allow the relative-value iteration algorithm to converge under milder conditions.

▶ **Definition 13.** *If T is a Shapley operator, and $0 < \theta < 1$, we define $T_\theta = \theta I + (1-\theta)T$ where I is the identity operator.*

We will call *Krasnoselskii–Mann operator* the T_θ operator. It is easy to show that it is also a Shapley operator. The following observation relates the ergodic constant of a damped Shapley operator with the ergodic constant of the original Shapley operator.

▶ **Lemma 14.** *Let T be a Shapley operator, $u \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then, $T(u) = \lambda e + u$ if and only if $T_\theta(u) = (1-\theta)\lambda e + u$. In particular, $\chi(T) = (1-\theta)^{-1}\chi(T_\theta)$ holds as soon as the ergodic equation $T(u) = \lambda e + u$ is solvable.*

Proof. The equivalence is straightforward, and $\chi(T) = (1-\theta)^{-1}\chi(T_\theta)$ follows from Obs. 4. ◀

We consider the iteration $x_{k+1} = T_\theta(x_k) - \mathsf{t}(T_\theta(x_k))e$, obtained by applying relative value iteration (as in Algorithm 1) to the Krasnoselskii–Mann operator T_θ, with an arbitrary initial condition $x_0 \in \mathbb{R}^n$. Ishikawa showed that the ordinary Krasnoselskii–Mann iteration applied to a nonexpansive self-map of a finite dimensional normed space does converge, as soon as a fixed point exists [24]. This entails the following result.

▶ **Theorem 15** (Compare with [19])**.** *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a Shapley operator, and $0 < \theta < 1$. Then, the sequence x_k obtained by applying relative value iteration to the Krasnoselskii–Mann operator T_θ converges if and only if $\exists u \in \mathbb{R}^n, \lambda \in \mathbb{R}$, $T(u) = \lambda e + u$.*

A multiplicative variant of this result was proved in Theorem 11 of [19].

7 Contraction properties of unchain games under pure policies

We define the following parameter, representing the minimal value of a non-zero off-diagonal transition probability, $p_{\min} = \min_{i,j\in[n], i\neq j,\, (a,b)\in A\times B} \{P^{a,b}_{i,j} : P^{a,b}_{i,j} > 0\}$, and set $\theta := p_{\min}/(1+p_{\min})$. For every pair of policies σ, τ of the two players, we set $Q^{\sigma,\tau} := \theta I + (1-\theta)P^{\sigma,\tau}$. For any sequence of pairs of pure policies $\sigma_1, \tau_1, \dots, \sigma_k, \tau_k$, we define, for all $i \in [n]$, $S_i(\sigma_1, \tau_1, \dots, \sigma_k, \tau_k) := \{j \mid [Q^{\sigma_1,\tau_1} \dots Q^{\sigma_k,\tau_k}]_{ij} > 0\}$.

▶ **Lemma 16.** *Suppose a concurrent game is unichain. Then, there is an integer $k \leqslant n$ such that for all $i_1, i_2 \in [n]$, and for all sequences of pairs of pure policies $\sigma_1, \tau_1, \dots, \sigma_k, \tau_k$, $S_{i_1}(\sigma_1, \tau_1, \dots, \sigma_k, \tau_k) \cap S_{i_2}(\sigma_1, \tau_1, \dots, \sigma_k, \tau_k) \neq \varnothing$.*

We call the *unichain index* of the game, and denote by k_{uni} the smallest integer k satisfying the property of Lemma 16. Similarly, we call *irreducibility index* of an irreducible game, and denote by k_{irr}, the smallest integer k such that for every sequence of pure policies $\sigma_1, \tau_1, \cdots, \sigma_k, \tau_k$, the matrix $Q^{\sigma_1,\tau_1} \dots Q^{\sigma_k,\tau_k}$ is positive. We have $1 \leqslant k_{\mathrm{uni}} \leqslant k_{\mathrm{irr}}$.

The following result will allow us to obtain a geometric contraction rate. The proofs of this theorem and of the next proposition show in particular that $k_{\mathrm{uni}} \leqslant n$ if the game is unichain and $k_{\mathrm{irr}} \leqslant n$ if the game is irreducible.

▶ **Theorem 17.** *Let us suppose that a concurrent game with n states is unichain, with unichain index $k = k_{\mathrm{uni}}$. Then, for all sequences $\sigma_1, \tau_1, \cdots, \sigma_k, \tau_k$ of pairs of pure policies of the two players, $\|Q^{\sigma_1,\tau_1} \dots Q^{\sigma_k,\tau_k}\|_{\mathrm{H}} \leqslant 1 - \theta^k$.*

The following proposition improves the bound on the contraction rate provided by Theorem 17, in the special case of irreducible games.

▶ **Proposition 18.** *Let us suppose that a concurrent game with n states is irreducible, and let $k = k_{\mathrm{irr}}$ be the irreducibility index of the game. Then, for all sequences $\sigma_1, \tau_1, \cdots, \sigma_k, \tau_k$ of pairs of pure policies of the two players, $\|Q^{\sigma_1,\tau_1} \dots Q^{\sigma_k,\tau_k}\|_{\mathrm{H}} \leqslant 1 - n\theta^k$.*

8 Solving concurrent and turn-based games by relative Krasnoselskii–Mann iteration

We first establish a general bound for concurrent unichain games. Recall that $\theta = p_{\min}/(1 + p_{\min})$.

▶ **Theorem 19.** *Let T_θ be the Krasnoselskii–Mann operator of a concurrent and unichain game. Then, $T_\theta^{k_{\text{uni}}}$ is a contraction in Hilbert's seminorm, with rate bounded by $1 - \theta^{k_{\text{uni}}}$. Moreover, if the game is irreducible, $T_\theta^{k_{\text{irr}}}$ is a contraction in Hilbert's seminorm, with rate bounded by $1 - n\theta^{k_{\text{irr}}}$.*

Combining this result with Theorem 12, we obtain the following result, in which we denote by $\|r\|_\infty := \max_{i,a,b} |r_i^{ab}|$ the sup-norm of the payment function.

▶ **Corollary 20.** *Let T be the Shapley operator of a concurrent unichain game, and $\epsilon \in (0, 1)$. Algorithm 1, applied to the Krasnoselskii–Mann operator T_θ, with the precision η prescribed in Theorem 12, provides an ϵ-approximation of the value of the game in at most $(|\log(\epsilon)| + \log 24 + \log \|r\|_\infty)k_{\text{uni}}\theta^{-k_{\text{uni}}}$ iterations.*

We now consider the special case of turn-based games. Then, the value is a rational number, and there are optimal pure policies. We now apply our approach to compute exactly the value and to find optimal pure policies.

▶ **Assumption 21.** *We now assume that the probabilities $P_{i,j}^{a,b}$ are rational numbers with a common denominator denoted by M. We also assume that the payments $r_i^{a,b}$ are* integers.

▶ **Lemma 22** (Coro. of [35]). *Let P be a $n \times n$ unichain matrix whose entries are rational numbers with a common denominator M. Then, the entries of the unique invariant measure of P are rational numbers of denominator at most nM^{n-1}.*

When Algorithm 1 halts, returning a vector $x \in \mathbb{R}^n$, we select two pure policies σ^* and τ^* that reach the minimum and maximum in the expression of $T(x)$, meaning that, for $i \in [n]$, we have:

$$T_i(x) = \max_{b \in B(i)} \Big(r_i^{\sigma^*(i),b} + \sum_{j \in [n]} P_{ij}^{\sigma^*(i),b} x_j \Big) = \min_{a \in A(i)} \Big(r_i^{a,\tau^*(i)} + \sum_{j \in [n]} P_{ij}^{a,\tau^*(i)} x_j \Big) \ . \tag{4}$$

▶ **Theorem 23.** *Consider a unichain turn-based stochastic game satisfying Assumption 21. Let us choose $\epsilon = (1 - \theta)(n^2 M^{2(n-1)})^{-1}$, so that Algorithm 1 applied to T_θ runs in at most*

$$(2 \log n + 2(n - 1) \log M + \log 24 + \log \|r\|_\infty)\theta^{-k_{\text{uni}}} k_{\text{uni}} \tag{5}$$

iterations. Let x^* be the vector returned by the algorithm. Let us select pure policies σ^* and τ^* reaching respectively the minimum and maximum in the expression of $T(x^*)$, as in (4). Then, these policies are optimal.

9 Multiplicative Krasnoselskii–Mann Damping applied to Entropy Games

In the case of entropy games, the ergodic eigenproblem, for the operator F defined in Section 3.2, consists in finding $u \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that $\exp(\lambda)\exp(u) = F(u)$. Equivalently, $\lambda e + u = T(u)$ where $T = \log \circ F \circ \exp$. If this equation is solvable, then $\exp(\lambda)$ is the value of the entropy game, for all initial states $d \in \mathcal{V}_D$. To solve this equation, we fix a positive number $\vartheta > 0$, and consider the following "multiplicative" variant of the Krasnoselskii–Mann operator:

$$[T_{\mathrm{m},\vartheta}(v)]_d = \log \min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}} \ \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}} \Big(\vartheta \exp(v_d) + \sum_{d'\in\mathcal{V}_D,(p,d')\in\mathcal{E}} m_{p,d'} \exp(v_{d'})\Big) \ ,$$

recalling that $m_{p,d'}$ denotes the multiplicity of the arc (p,d'). Unlike in the additive case, we do not perform a "convex combination" of the identity map and of the Shapley operator, but we only add the "diagonal term" $\vartheta \exp(v_d)$, where ϑ can still interpreted as a "damping intensity", albeit in a multiplicative sense. If $T(u) = \lambda e + u$, then, one readily checks that $T_{\mathrm{m},\vartheta}(u) = \mu e + u$, where $\mu = \log(\vartheta + \exp(\lambda))$, and vice versa, so the non-linear eigenproblems for T and $T_{\mathrm{m},\vartheta}$ are equivalent. As in the additive case, the damping intensity must be tuned to optimize the complexity bounds. We shall say that the multiplicity $m_{p,d'}$ is *off-diagonal* if there is no path $d' \to t \to p \to d'$ in the graph of the game. Equivalently, for any choices of policies σ, τ of the two players, the entry $m_{p,d'}$ does not appear on the diagonal of the matrix $M^{\sigma,\tau}$, defined in Section 3.2. Then, we denote by $\underline{m}$ the minimum of off-diagonal multiplicities, observe that $\underline{m}$ is precisely the minimum of all off-diagonal entries of the matrices $M^{\sigma,\tau}$ associated to all pairs of policies. We set $\vartheta := \underline{m}$.

We shall say that an entropy game is *irreducible* if for every pair of policies σ, τ, the matrix $M^{\sigma,\tau}$ is irreducible. The *irreducibility index* k_{irr} of an irreducible entropy game is the smallest integer k such that for all policies $\sigma_1, \tau_1, \dots, \sigma_k, \tau_k$, the matrix $M^{\sigma_1\tau_1} \dots M^{\sigma_k\tau_k}$ has positive entries. Arguing as in the case of stochastic concurrent games, we get that $k_{\mathrm{irr}} \leqslant n$ as soon as the game is irreducible. We define the *l-ambiguity* of the entropy game $\mathcal{A}_l := \max_{d,d'\in\mathcal{V}_D} \max_{\sigma_1,\tau_1,\dots,\sigma_l,\tau_l} (M^{\sigma_1\tau_1} \dots M^{\sigma_l\tau_l})_{d,d'}$. Observe that $(M^{\sigma_1\tau_1} \dots M^{\sigma_l\tau_l})_{d,d'}$ is the number of paths from d to d' counted with multiplicities, in the finite horizon game induced by the policies $\sigma_1, \tau_1, \dots, \sigma_l, \tau_l$ (this motivates the term "l-ambiguity"). If the game is irreducible, we define the *ambiguity* of the game $\mathcal{A} := \max_{1\leqslant l\leqslant k_{\mathrm{irr}}} \mathcal{A}_l^{1/l}$. We set $W := \max_{(p,d)\in\mathcal{E}\cap(\mathcal{V}_P\times\mathcal{V}_D)} m_{p,d}$, and observe that $W \leqslant \mathcal{A} \leqslant n^{1-1/k_{\mathrm{irr}}} W$.

▶ **Theorem 24.** *Let $T_{m,\vartheta}$ be the multiplicative Krasnoselskii–Mann operator of an irreducible entropy game. Then, $T_{m,\vartheta}^{k_{\mathrm{irr}}}$ is a contraction in Hilbert's seminorm, with contraction rate bounded by $\frac{\bar{\mathcal{M}}-1}{\bar{\mathcal{M}}+1}$, where $\bar{\mathcal{M}} := (1 + \mathcal{A}/\underline{m})^{k_{\mathrm{irr}}}$.*

We recall the following separation bound.

▶ **Theorem 25** (Coro. of [5])**.** *Suppose two pairs of (pure) policies yield distinct values in an entropy game with n Despot's states. Then, these values differ at least by ν_n^{-1} where*

$$\nu_n := 2^n (n+1)^{8n} n^{2n^2+n+1} e^{4n^2} \max(1, W/2)^{4n^2} \ .$$

Then, using Theorem 12, we deduce:

▶ **Theorem 26.** *Consider an irreducible entropy game, with irreducibility index k_{irr}. Let us choose $\epsilon = (1 + (\underline{m} + W)\nu_n)^{-1}$, so that Algorithm 1 applied to $T_{m,\vartheta}$ runs in at most $(\log(1 + (\underline{m} + W)\nu_n) + \log 6) k_{\mathrm{irr}} \bar{\mathcal{M}}/2$ iterations. Moreover, let x^* be the vector returned by the algorithm. Let us select pure policies σ^* and τ^* reaching respectively the minimum and maximum in the expression of $T_{m,\vartheta}(x^*)$. Then, these policies are optimal.*

10 Concluding Remarks

We have established parameterized complexity bounds for relative value iteration applied to several classes of stochastic games satisfying irreducibility conditions. These bounds rely on contraction properties in Hilbert's seminorm. It would be interesting to see whether these contraction properties can also be exploited to derive complexity bounds for policy iteration, instead of value iteration.

References

1 M. Akian, S. Gaubert, J. Grand-Clément, and J. Guillaud. The operator approach to entropy games. *Theory of Computing Systems*, 63:1089–1130, 2019.

2 M. Akian, S. Gaubert, and A. Hochart. Ergodicity conditions for zero-sum games. *Discrete Contin. Dyn. Syst.*, 35(9):3901–3931, 2015.

3 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.

4 Marianne Akian, Stéphane Gaubert, Ulysse Naepels, and Basile Terver. Solving irreducible stochastic mean-payoff games and entropy games by relative Krasnoselskii–Mann iteration, 2023. Extended version of the present article, arXiv:2305.02458.

5 X. Allamigeon, S. Gaubert, R. D. Katz, and M. Skomra. Universal Complexity Bounds Based on Value Iteration and Application to Entropy Games. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 110:1–110:20, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

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

7 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.

8 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.

9 L. Attia and M. Oliu-Barton. A formula for the value of a stochastic game. *PNAS*, 52(116):26435–26443, 2019.

10 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.

11 T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Math. Oper. Res.*, 1(3):197–208, 1976.

12 E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A potential reduction algorithm for two-person zero-sum mean payoff stochastic games. *Dynamic Games and Applications*, 8(1):22–41, July 2018.

13 K. Chatterjee and R. Ibsen-Jensen. The complexity of ergodic mean-payoff games. Extended version of a paper published in the proceedings of ICALP, 2014. `arXiv:1404.5734`.

14 A. Condon. The complexity of stochastic games. *Inform. and Comput.*, 96(2):203–224, 1992.

15 R. L. Dobrushin. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, January 1956.

16 K. Etessami and M. Yannakakis. Recursive concurrent stochastic games. *Logical Methods in Computer Science*, 4(4), November 2008.

17 A Federgruen, P.J Schweitzer, and H.C Tijms. Contraction mappings underlying undiscounted Markov decision problems. *Journal of Mathematical Analysis and Applications*, 65(3):711–730, 1978.

18 S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Trans. of AMS*, 356(12):4931–4950, 2004.

19 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.

20 D. Gillette. *Stochastic games with zero stop probabilities*, volume III, chapter 9, pages 179–188. Princeton University Press, 1958.

21 K. Arnsfelt Hansen, M. Koucky, N. Lauritzen, P. Bro Miltersen, and E. P. Tsigaridas. Exact algorithms for solving stochastic games. In *STOC 2011*, 2011.

22 A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Manag. Sci.*, 12(5):359–370, 1966.

23 R. A. Howard and J. E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, 1972.

24 S. Ishikawa. Fixed points and iteration of a nonexpansive mapping in a Banach space. *Proceedings of the American Mathematical Society*, 59(1):65–71, 1976.

25 M. A. Krasnosel'skiĭ. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10:123–127, 1955.

26 T. M. Liggett and S. A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Rev.*, 11:604–607, 1969.

27 W. R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.

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

29 J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated games*, volume 55 of *Econom. Soc. Monogr.* Cambridge University Press, Cambridge, 2015.

30 H.D. Mills. Marginal values of matrix games and linear programs. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38 of *Annals of Mathematics Studies*, pages 183–194. Princeton University Press, 1956.

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

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

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

34 L. S. Shapley. Stochastic games. *Proc. Natl. Acad. Sci. USA*, 39(10):1095–1100, 1953.

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

36 K. Sladký. *On dynamic programming recursions for multiplicative Markov decision chains*, pages 216–226. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.

37 G. Vigeral. A zero-sum stochastic game with compact action sets and no asymptotic value. *Dynamic Games and Applications*, 3(2):172–186, January 2013.

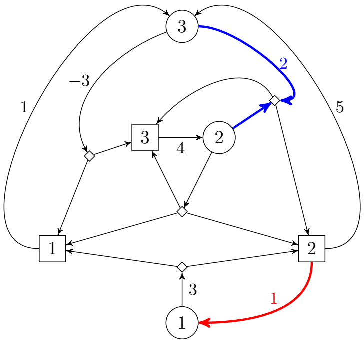
38 D.J White. Dynamic programming, Markov chains, and the method of successive approximations. *Journal of Mathematical Analysis and Applications*, 6(3):373–376, 1963.

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

40 U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theoret. Comput. Sci.*, 158(1–2):343–359, 1996.

A Example of turn-based stochastic mean-payoff game

A turn-based stochastic mean-payoff game is represented below. Min states are represented by squares; Max states are represented by circles; Nature states are represented by small diamonds. The payments made by Min to Max are shown on the arcs. For every Nature state, the next state is chosen with the uniform distribution among the successors. The associated Shapley operator is the map $T : \mathbb{R}^3 \to \mathbb{R}^3$ shown at right.



$$T_1(x)=1+\max(2+\tfrac{x_2+x_3}{2},-3+\tfrac{x_1+x_3}{2})$$

$$T_2(x)=\min\left(5+\max(2+\tfrac{x_2+x_3}{2},-3+\tfrac{x_1+x_3}{2}),1+3+\tfrac{x_1+x_2}{2}\right)$$

$$T_3(x)=4+\max(\tfrac{x_2+x_3}{2},\tfrac{x_1+x_2+x_3}{3})$$

The unichain index defined in Section 7 is $k_{\text{uni}} = 1$. Indeed, for all pairs of policies (σ_1, τ_1), we have $S_1(\sigma_1, \tau_1) \supset \{1, 3\}$, $S_3(\sigma_1, \tau_1) \supset \{2, 3\}$, and $S_2(\sigma_1, \tau_1) \supset \{2, 3\}$ if σ_1 sends Min state 2 to Max state 3, and $S_2(\sigma_1, \tau_1) = \{1, 2\}$ if σ_1 sends Min state 2 to Max state 1. In all cases, we have $S_i(\sigma_1, \tau_1) \cap S_j(\sigma_1, \tau_1) \neq \varnothing$ for $i \neq j$. We have $p_{\min} = 1/3$, and $\theta = p_{\min}/(1 + p_{\min}) = 1/4$. It follows from Theorem 19 that the damped Shapley operator T_θ is a contraction of rate $3/4$. We know from Theorem 11 that the ergodic eigenproblem is solvable. By applying Algorithm 1, we find $T(u) = \lambda e + u$ with $(-1, -0.5, 0)$ and $\lambda = 3.75$. An approximation of u of precision $< 10^{-8}$ in the sup norm is reached after only 15 iterations, to be compared with the precision of order $(3/4)^{15} \simeq 10^{-2}$ given by the theoretical upper bound, for the same number of iterations. Thus, the convergence may be faster in practice than the one shown in Corollary 20. We deduce from $T(u) = \lambda e + u$ that the value of the mean-payoff game is 3.75 regardless of the initial state. Optimal policies σ and τ of both players are obtained by selecting the actions that achieve the minimum or the maximum in the expression of $T(u)$. The non-trivial actions of these optimal policies are as follows: from Min state 2 (square at bottom right), go to Max state 3 (circle at the top level), from Max state 2 (circle at the middle level), and also from Max state 3, got to the top right state (diamond) of Nature. These policies are shown on the figure above (red: policy of Min; blue: policy of Max). The stochastic matrix $P^{\sigma,\tau}$ and payment vector $r^{\sigma,\tau}$ associated to these policies are given by

$$r^{\sigma,\tau} = \begin{pmatrix} 3 \\ 4 \\ 4 \end{pmatrix}, \qquad P^{\sigma,\tau} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

The unique invariant measure of the matrix $P^{\sigma,\tau}$ is $\pi = (1/4, 1/2, 1/4)$, and we have $\pi r^{\sigma,\tau} = 15/4 = 3.75$, consistently with the value of the mean-payoff already found.