



GENERIC UNIQUENESS OF THE BIAS VECTOR OF FINITE ZERO-SUM STOCHASTIC GAMES WITH PERFECT INFORMATION

MARIANNE AKIAN, STÉPHANE GAUBERT, AND ANTOINE HOCHART

ABSTRACT. Mean-payoff zero-sum stochastic games can be studied by means of a nonlinear spectral problem. When the state space is finite, the latter consists in finding an eigenpair (u, λ) solution of $T(u) = \lambda e + u$, where $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the Shapley (or dynamic programming) operator, λ is a scalar, e is the unit vector, and $u \in \mathbb{R}^n$. The scalar λ yields the mean payoff per time unit, and the vector u , called *bias*, allows one to determine optimal stationary strategies in the mean-payoff game. The existence of the eigenpair (u, λ) is generally related to ergodicity conditions. A basic issue is to understand for which classes of games the bias vector is unique (up to an additive constant). In this paper, we consider perfect-information zero-sum stochastic games with finite state and action spaces, thinking of the transition payments as variable parameters, transition probabilities being fixed. We show that the bias vector, thought of as a function of the transition payments, is generically unique (up to an additive constant). The proof uses techniques of nonlinear Perron-Frobenius theory. As an application of our results, we obtain an explicit perturbation scheme allowing one to solve degenerate instances of stochastic games by policy iteration.

1. INTRODUCTION

1.1. The ergodic equation for stochastic games. Repeated zero-sum games describe long-term interactions between two agents, called players, with opposite interests. In this paper, we consider *perfect-information zero-sum stochastic games*, in which the players choose repeatedly and alternatively an action, being informed of all the events that have previously occurred (state of nature and chosen actions). These choices determine at each stage of the game a payment, as well as the next state by a stochastic process. Given a finite horizon k and an initial state i , one player intends to minimize the sum of the payments of the k first stages, while the other player intends to maximize it. This gives rise to the value of the k -stage game, denoted by v_i^k .

A major topic in the theory of zero-sum stochastic games is the asymptotic behavior of the mean values per time unit (v_i^k/k) as the horizon k tends to infinity. The limit, when it exists, is referred to as the *mean payoff*. This question was first addressed in the case of a finite state space by Everett [Eve57], Kohlberg [Koh74], and

Date: July 19, 2017.

2010 Mathematics Subject Classification. 47J10; 91A20,93E20.

Key words and phrases. Zero-sum games, ergodic control, nonexpansive mappings, fixed point sets, policy iteration.

This work was performed when A. Hochart was with Inria and CMAP, Ecole polytechnique, CNRS, Université Paris-Saclay. The authors are also partially supported by the PGMO program of EDF and FMJH and by the ANR (MALTHY Project, number ANR-13-INSE-0003).

Bewley and Kohlberg [BK76]. See also Rosenberg and Sorin [RS01], Sorin [Sor04], Renault [Ren11], Bolte, Gaubert and Vigeral [BGV15] and Ziliotto [Zil16] for more recent developments. We refer the reader to [NS03] for more background on stochastic games.

A way to study the asymptotic behavior of the values $(v_i^k)_k$ consists in exploiting the recursive structure of the game. This structure is encompassed in the *dynamic programming* or *Shapley operator* of the game. In this paper, since the state space is assumed to be finite, say $\{1, \dots, n\}$, the latter is a map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Then, a basic tool to study the asymptotic properties of the sequence $(v_i^k)_k$ is the following nonlinear spectral problem, called the *ergodic equation*:

$$(1.1) \quad T(u) = \lambda e + u \quad ,$$

where e is the unit vector of \mathbb{R}^n . Indeed, if there exist a vector $u \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$ solution of (1.1), then, not only the sequence $(v_i^k/k)_k$ converges as the horizon k tends to infinity, but also the limit is independent of the initial state i , and equal to λ . This scalar, which is unique, is called the *ergodic constant* or the *(additive) eigenvalue* of T , and the vector u , called *bias vector* or *(additive) eigenvector*, gives optimal stationary strategies in the mean-payoff game.

A first question is to understand when the ergodic equation has a solution. In [AGH15], we considered the Shapley operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ of a game with finite state space and bounded transition payment function, and gave necessary and sufficient conditions under which the ergodic equation is solvable for all the operators $g + T$ with $g \in \mathbb{R}^n$, or equivalently for all perturbations g of the payments that only depend on the state. Moreover, assuming the compactness of the action sets of players and some continuity of the transition functions, these conditions can be characterized in terms of reachability in directed hypergraphs.

A second question concerns the structure of the set of bias vectors. For one-player problems, i.e., for discrete optimal control, the ergodic equation (1.1), also known as the *average case optimality equation*, has been much studied, either in the deterministic or in the stochastic case (Markov decision problems). Then, the representation of bias vectors and their relation with optimal strategies is well understood.

Indeed, in the deterministic case, the analysis of the ergodic equation relies on max-plus spectral theory, which goes back to the work of Romanovsky [Rom67], Gondran and Minoux [GM77] and Cuninghame-Green [CG79]. Kontorov and Yakovenko [KY92] deal specially with infinite horizon optimization and mean-payoff problems. We refer the reader to the monographies [BCOQ92, KM97, But10] or surveys [Bap98, ABG13] for more background on max-plus spectral theory. One of the main result of this theory shows that the set of bias vectors has the structure of a max-plus (tropical) cone, i.e., that it is invariant by max-plus linear combinations, and it has a unique minimal generating family consisting of certain “extreme” generators, which can be identified by looking at the support of the maximizing measures in the linear programming formulation of the optimal control problem, or at the “recurrence points” of infinite optimal trajectories. A geometric approach to some of these results, in terms of polyhedral fans, has been recently given by Sturmfels and Tran [ST13].

The eigenproblem (1.1) has been studied in the more general infinite-dimensional state space case, see Kolokoltsov and Maslov [KM97], Mallet-Paret and Nussbaum [MPN02], and Akian, Gaubert and Walsh [AGW09] for an approach in terms

of horoboundaries. It has also been studied in the setting of weak KAM theory, for which we refer the reader to Fathi [Fat08]. In the weak KAM setting, the bias vector becomes the solution of an ergodic Hamilton-Jacobi PDE; Figalli and Rifford showed in [FR13] that this solution is unique for generic perturbations of a Hamiltonian by a potential function. In the stochastic case, the structure of the set of bias vectors is still known when the state space is finite, see Akian and Gaubert [AG03].

In the two-player case, the structure of the set of bias vectors is less well known, although the description of this set remains a fundamental issue. In particular, the uniqueness of the bias vector up to an additive constant is an important matter for algorithmic purposes. Indeed, the nonuniqueness of the bias typically leads to numerical instabilities or degeneracies. In particular, the standard Hoffman and Karp policy iteration algorithm [HK66] may fail to converge in situations in which the bias vector is not unique. The standard approach to handle such degeneracies is to approximate the ergodic problem by the discounted problem. This was proposed by Puri [Pur95] in the case of deterministic games and this was further analyzed by Paterson and Zwick [ZP96], see also the discussion by Friedmann in [Fri11]. A different approach was proposed by Cochet-Terrasson and Gaubert [CTG06], Akian, Cochet-Terrasson, Detournay and Gaubert [ACTDG12], and by Bourque and Raghavan [BR14], allowing one to circumvent such degeneracies at the price of an increased complexity of the algorithm (handling the nonuniqueness of the bias). Hence, it is of interest to understand when such technicalities can be avoided.

1.2. Main results. We address the question of the uniqueness of the bias vector for stochastic games with perfect information and finite state and action spaces, restricting our attention to games for which the ergodic equation (1.1) is solvable *for all state-dependent perturbations of the transition payments*. Our main result, Theorem 3.2, shows that the bias vector is generically unique up to an additive constant. More precisely, we show that the set of perturbation vectors for which the bias vector is not unique belongs to a polyhedral complex the cells of which have codimension one at least. A first ingredient in the proof relies on nonlinear Perron-Frobenius theory [AG03]. A second ingredient is a general result, showing that the set of fixed points of a nonexpansive self-map of \mathbb{R}^n is a retract of \mathbb{R}^n , see Theorem 3.10. This allows us to infer the uniqueness of the bias vector of a Shapley operator from the uniqueness of the bias vector of the reduced Shapley operators obtained by fixing the strategy of one player.

We then present an algorithmic application of our results. Hoffman and Karp introduced a policy iteration algorithm to solve mean-payoff zero-sum stochastic games with perfect information and finite state and action spaces [HK66]. They showed that policy iteration does terminate if every pair of strategies of the two players yields an irreducible Markov chain. If this irreducibility assumption is not satisfied, policy iteration may cycle. However, the irreducibility assumption is not satisfied by many classes of games – in particular, it is essentially never satisfied for deterministic games. The cycling of the Hoffman-Karp algorithm is due to the nonuniqueness of the bias vector. Hence, we deduce from our results that the Hoffman-Karp policy iteration algorithm does converge if the payment is generic. Moreover, we provide a family of effective perturbations of the payment for which the bias vector is unique (up to an additive constant). This leads to an explicit perturbation scheme, allowing one to solve nongeneric instances by policy iteration,

avoiding the classical irreducibility condition or the use of vanishing discount based perturbation schemes.

The paper is organized as follows. After some preliminaries on stochastic games in Section 2, we establish the generic uniqueness of the bias vector in Section 3. The application to policy iteration is presented in Section 4.

We finally point out that some of the present results have been announced in the conference article [AGH14].

2. PRELIMINARIES ON ZERO-SUM STOCHASTIC GAMES

2.1. Games with perfect information. In this paper we consider *finite (zero-sum) stochastic games with perfect information*, where the second player, called player MAX, always makes a move after being informed of the action chosen by the first player, called player MIN. Such a game is characterized by the following:

- a finite state space, that we denote by $S = \{1, \dots, n\}$;
- finite action spaces, denoted by A_i for player MIN when the current state is $i \in S$, and by $B_{i,a}$ for player MAX in state i and after action $a \in A_i$ has been chosen by player MIN;
- a transition payment $r_i^{ab} \in \mathbb{R}$, paid by player MIN to player MAX, when the current state is $i \in S$ and the last actions selected by the players are $a \in A_i$ and $b \in B_{i,a}$;
- a transition probability $P_i^{ab} \in \Delta(S)$ (where $\Delta(S)$ denotes the set of probability measures over S) which gives the law according to which the next state is determined, when the current state is $i \in S$ and the last actions selected by the players are $a \in A_i$ and $b \in B_{i,a}$.

Denoting by $K_A := \bigcup_{i \in S} \{i\} \times A_i$ the action set of player MIN and by $K_B := \bigcup_{(i,a) \in K_A} \{(i,a)\} \times B_{i,a}$ the action set of player MAX, a finite stochastic game with perfect information is thus defined by a 5-tuple $\Gamma := (S, K_A, K_B, r, P)$ where the three first sets are finite.

Such a game is played in stages, starting from a given initial state i_0 , as follows: at step ℓ , if the current state is i_ℓ , player MIN chooses an action $a_\ell \in A_{i_\ell}$, and player MAX subsequently chooses an action $b_\ell \in B_{i_\ell, a_\ell}$. Then, player MIN pays $r_{i_\ell}^{a_\ell b_\ell}$ to player MAX and the next state is chosen according to the probability law $P_{i_\ell}^{a_\ell b_\ell}$. The information is perfect, meaning that at each step, the players have a perfect knowledge of all the previously chosen actions, as well as the state previously visited.

Denote by $H_k^{\text{MIN}} := (K_B)^k \times S$ the set of histories of length k of player MIN and by $H^{\text{MIN}} := \bigcup_{k \in \mathbb{N}} H_k^{\text{MIN}}$ the set of all finite histories of player MIN. Likewise, define $H_k^{\text{MAX}} := (K_B)^k \times K_A$ and $H^{\text{MAX}} := \bigcup_{k \in \mathbb{N}} H_k^{\text{MAX}}$ to be, respectively, the set of histories of length k and the set of all finite histories of player MAX. Let $H_\infty := (K_B)^\mathbb{N}$ be the set of infinite histories. A strategy of player MIN is a map

$$\sigma : H^{\text{MIN}} \rightarrow \bigcup_{i \in S} \Delta(A_i)$$

(where $\Delta(X)$ denote the set of probability measures over any set X) such that, for every finite history $h_k = (i_0, a_0, b_0, \dots, i_k) \in H_k^{\text{MIN}}$, we have $\sigma(\cdot \mid h_k) \in \Delta(A_{i_k})$.

Likewise, a strategy of player MAX is a map

$$\tau : H^{\text{MAX}} \rightarrow \bigcup_{(i,a) \in K_A} \Delta(B_{i,a})$$

such that, for every finite history $h'_k = (i_0, a_0, b_0, \dots, i_k, a_k) \in H_k^{\text{MAX}}$, we have $\tau(\cdot \mid h'_k) \in \Delta(B_{i_k, a_k})$.

A strategy σ (resp. τ) of player MIN (resp. MAX) is pure and Markovian if it depends only on the current stage and state (resp. the current stage, state and action of player MIN) and is deterministic, that is its values are Dirac probabilities. Such a strategy is stationary if it does not depend on stage. We denote by \mathcal{S}_p the finite set of (deterministic) *policies* of player MIN, i.e., the set of maps $\sigma : S \rightarrow \bigcup_{i \in S} A_i$ such that $\sigma(i) \in A_i$ for every state $i \in S$. Likewise, we denote by \mathcal{T}_p the set of (deterministic) policies of player MAX, i.e., the set of maps $\tau : K_A \rightarrow \bigcup_{i \in S, a \in A_i} B_{i,a}$ such that $\tau(i, a) \in B_{i,a}$ for every state $i \in S$ and action $a \in A_i$. Then, a pure Markovian strategy of player MIN (resp. MAX) can be identified to a sequence of policies: for every history of length k of player MIN, $h_k = (i_0, a_0, b_0, \dots, i_k) \in H_k^{\text{MIN}}$, $\sigma(\cdot \mid h_k)$ is the Dirac measure at $\sigma_k(i_k)$. Likewise, for every history of length k of player MAX, $h'_k = (i_0, a_0, b_0, \dots, i_k, a_k) \in H_k^{\text{MAX}}$, $\tau(\cdot \mid h'_k)$ is the Dirac measure at $\tau_k(i_k, a_k)$. Moreover, a pure Markovian stationary strategy can be identified to a single policy.

Given an initial state i and strategies σ and τ of the players, the triple (i, σ, τ) defines a probability law on H_∞ the expectation of which is denoted by $\mathbb{E}_{i, \sigma, \tau}$. The *payoff* of the k -stage game is the following additive function of the transition payments:

$$J_i^k(\sigma, \tau) := \mathbb{E}_{i, \sigma, \tau} \left[\sum_{\ell=0}^{k-1} r_{i_\ell}^{a_\ell b_\ell} \right].$$

Player MIN intends to choose a strategy minimizing the payoff J_i^k , whereas player MAX intends to maximize the same payoff. The value of the k -stage game starting at state i is then defined by

$$v_i^k := \inf_{\sigma} \sup_{\tau} J_i^k(\sigma, \tau) = \sup_{\tau} \inf_{\sigma} J_i^k(\sigma, \tau),$$

when the equality holds, where the infimum and the supremum are taken over the set of all strategies of players MIN and MAX, respectively. Since the game has finite state and action spaces, the value exists, with the infima and suprema realized by pure Markovian strategies of both players [Sha53].

In this paper, we are interested in the asymptotic behavior of the sequence of mean values per time unit $(v^k/k)_{k \geq 1}$. When the latter ratio converges, the limit will be called the *mean-payoff vector*. A stronger condition is the existence of a *uniform value* v^U , meaning that

$$(2.1) \quad \inf_{\sigma} \limsup_{k \rightarrow \infty} \sup_{\tau} \frac{1}{k} J_i^k(\sigma, \tau) \leq v_i^U \leq \sup_{\tau} \liminf_{k \rightarrow \infty} \inf_{\sigma} \frac{1}{k} J_i^k(\sigma, \tau),$$

where infima and suprema are taken over all strategies (see for instance [Sor02, Ren12]). This implies that $\lim_k v^k/k = v^U$. Note that the first (resp. second) inequality means that player MIN (resp. MAX) uniformly guarantees v_i^U .

Moreover, following [LR17], *optimal uniform strategies* are defined as strategies σ^* and τ^* for players MIN and MAX respectively such that

$$(2.2) \quad \limsup_{k \rightarrow \infty} \sup_{\tau} \frac{1}{k} J_i^k(\sigma^*, \tau) = v_i^U = \liminf_{k \rightarrow \infty} \inf_{\sigma} \frac{1}{k} J_i^k(\sigma, \tau^*) .$$

Note that Condition (2.2) is stronger than (2.1). When v^U exists, it also coincides with the value of the zero-sum game in which one considers the following *limiting average payoffs*:

$$(2.3) \quad \begin{aligned} J_i^+(\sigma, \tau) &:= \limsup_{k \rightarrow \infty} \frac{1}{k} J_i^k(\sigma, \tau) , \\ J_i^-(\sigma, \tau) &:= \liminf_{k \rightarrow \infty} \frac{1}{k} J_i^k(\sigma, \tau) . \end{aligned}$$

Moreover, optimal uniform strategies are also optimal for these games. Mertens and Neyman [MN81] proved that for stochastic games with finite state and action spaces and simultaneous moves of the players the uniform value exists. It follows that the uniform value also exists for finite state and action spaces games with perfect information – the latter can be reduced to degenerate instances of games with simultaneous moves in which in each state, only one of the players has a choice of action. We shall also recall in Theorem 2.1 how for this class of games, the existence of the uniform value and of uniform optimal strategies, follows from a result of Kohlberg.

In the computer science literature [ZP96, AM09], mean-payoff games are defined in a slightly different manner, following Ehrenfeucht and Mycielski [EM79], as non-zero sum games in which player MIN wishes to minimize $J_i^+(\sigma, \tau)$ whereas player MAX wishes to maximize $J_i^-(\sigma, \tau)$. Liggett and Lippman [LL69] showed that such games admit optimal policies σ^*, τ^* and a value v^* , meaning that

$$J_i^+(\sigma^*, \tau) \leq v^* = J_i^+(\sigma^*, \tau^*) = J_i^-(\sigma^*, \tau^*) \leq J_i^-(\sigma, \tau^*)$$

for all pair of strategies σ, τ of players MIN and MAX. The latter property is implied by the existence of the uniform value and the existence of pure Markovian stationary uniform optimal strategies (also called uniform optimal policies), so that $v^* = v^U$.

Therefore, in the sequel, we will use the term *mean-payoff games* with a general meaning, understanding that the different approaches that we just discussed lead to the same notion of value. In particular, the notion of value and optimal policies can always be understood in the strongest sense (uniform value and uniform optimal policies).

2.2. The operator approach. The study of the value vector $v^k = (v_i^k)_{i \in S}$ involves the *dynamic programming operator*, or *Shapley operator* of the game. The latter is a map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose i th coordinate is given by

$$(2.4) \quad T_i(x) = \min_{a \in A_i} \max_{b \in B_{i,a}} (r_i^{ab} + P_i^{ab}x) , \quad x \in \mathbb{R}^n .$$

Note that an element $P \in \Delta(S)$ is seen as a row vector $P = (P_j)_{j \in S}$ of \mathbb{R}^n , so that Px means $\sum_{j \in S} P_j x_j$. Also note that, given a vector $g \in \mathbb{R}^n$, the operator $g + T$ appears as the Shapley operator of the game $(S, K_A, K_B, \tilde{r}, P)$ where the transition payment function satisfies $\tilde{r}_i^{ab} = g_i + r_i^{ab}$. The latter game is almost identical to the initial game (S, K_A, K_B, r, P) , except that the transition payments are perturbed

with quantities that only depend on the state, hence the designation of g as an *additive (state-dependent) perturbation vector*.

The Shapley operator allows one to determine recursively the value vector of the k -stage game:

$$(2.5) \quad v^k = T(v^{k-1}) \quad , \quad v^0 = 0 \quad .$$

Also, T is *monotone* and *additively homogeneous*, meaning that it satisfies the following two properties, respectively:

$$\begin{aligned} (\text{monotonicity}) \quad & x \leq y \implies T(x) \leq T(y) \quad , \quad x, y \in \mathbb{R}^n \quad , \\ (\text{additive homogeneity}) \quad & T(x + \lambda e) = T(x) + \lambda e \quad , \quad x \in \mathbb{R}^n, \lambda \in \mathbb{R} \quad , \end{aligned}$$

where \mathbb{R}^n is endowed with its usual partial order and e is the unit vector of \mathbb{R}^n . A first consequence of the additive homogeneity of T is that, for any bias vector u and any scalar $\alpha \in \mathbb{R}$, the vector $u + \alpha e$ is also a bias: we say that u is defined up to an *additive constant*. More generally, the importance of the above axioms in stochastic control and game theory has been known for a long time [CT80]. In particular, they imply that T is *sup-norm nonexpansive*:

$$\|T(x) - T(y)\|_\infty \leq \|x - y\|_\infty \quad , \quad x, y \in \mathbb{R}^n \quad .$$

According to the dynamic programming principle (2.5), the mean-payoff vector defined above is given by

$$\chi(T) := \lim_{k \rightarrow \infty} \frac{T^k(0)}{k} \quad ,$$

where $T^k := T \circ \dots \circ T$ denotes the k th iterate of T . Observe that, since T is sup-norm nonexpansive, 0 could be replaced by any vector $x \in \mathbb{R}^n$ in the above limit.

Here, since the action spaces are finite, T is a *piecewise affine* map over \mathbb{R}^n . Kohlberg showed in [Koh80] that a piecewise affine self-map T of \mathbb{R}^n that is non-expansive in an arbitrary norm has an *invariant half-line*, meaning that there exist two vectors $u, \nu \in \mathbb{R}^n$ such that

$$(2.6) \quad T(u + \alpha \nu) = u + (\alpha + 1)\nu$$

for every scalar α large enough. Kohlberg's theorem applies to the present setting, since the Shapley operator (2.4) is nonexpansive in the sup-norm. This implies that the limit $\chi(T)$ does exist and coincides with ν . It follows that the vector ν arising in the definition of invariant half-lines is unique.

The existence of uniform optimal policies follows from the existence of an invariant half-line. Moreover, such policies are readily computed from the invariant half-line. Although these properties are surely known to some experts, we could not find a reference for them, so we next state them as Theorem 2.1.

To this end, we define the *reduced Shapley operator* $T^\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$ associated with the policy $\sigma \in \mathcal{S}_p$ of player MIN. Its i th coordinate map is given by

$$(2.7) \quad T_i^\sigma(x) = \max_{b \in B_{i, \sigma(i)}} (r_i^{\sigma(i)b} + P_i^{\sigma(i)b} x) \quad , \quad x \in \mathbb{R}^n \quad .$$

Since the action spaces are finite, we readily have, for all $x \in \mathbb{R}^n$,

$$(2.8) \quad T(x) = \min_{\sigma \in \mathcal{S}_p} T^\sigma(x) \quad ,$$

where by min, we mean that for every $x \in \mathbb{R}^n$, the minimum is attained. Indeed, it suffices to take for $\sigma(i)$ any action $a \in A_i$ achieving the minimum in (2.4).

Similarly, to any policy $\tau \in \mathcal{T}_p$ of player MAX, we associate a *dual reduced Shapley operator* ${}^\tau T$, whose i th coordinate map is given by

$${}^\tau T_i(x) = \min_{a \in A_i} (r_i^{a\tau(i,a)} + P_i^{a\tau(i,a)} x), \quad x \in \mathbb{R}^n .$$

For all $x \in \mathbb{R}^n$, we have

$$(2.9) \quad T(x) = \max_{\tau \in \mathcal{T}_p} {}^\tau T(x) .$$

Theorem 2.1 (Coro. of [Koh80]). *Perfect information stochastic games with finite state and action spaces have a uniform value and uniform optimal policies σ^*, τ^* that are obtained as follows: given an invariant half-line $\alpha \mapsto u + \alpha\nu$ of the Shapley operator T , take for σ^* any policy σ attaining the minimum in (2.8) when x is substituted by $u + \alpha\nu$ for α large enough. Similarly, take for τ^* any policy τ attaining the maximum in (2.9) when x is substituted by the same quantity.*

Proof. Recall that the *germ* of a real function $\mathbb{R} \rightarrow \mathbb{R}$, $\alpha \mapsto f(\alpha)$, at the point $+\infty$ is the equivalence class of f modulo the relation $f \sim g$ if there exists α_0 such that $f(\alpha) = g(\alpha)$ for $\alpha \geq \alpha_0$. We shall use the same notation f for the function and its equivalence class. We shall consider in particular the set \mathbb{A} of germs of affine functions. Observe that \mathbb{A} is invariant by linear combinations, by translation by a constant, and also by the operations \min and \max , because \mathbb{A} is totally ordered (of two affine functions of α , one ultimately dominates the other as $\alpha \rightarrow \infty$). Since the action spaces are finite, if $f \in \mathbb{A}^n$, we may identify $\alpha \mapsto T(f(\alpha))$ to an element of \mathbb{A}^n . In other words, T acts on vectors of germs of affine functions. We now substitute $x = u + \alpha\nu$ in (2.4). Since \mathbb{A} is totally ordered, the minima and maxima in (2.4) are attained by actions independent of α provided that α is large enough. I.e., there exists policies σ^* and τ^* of players MIN and MAX respectively and a constant α_0 such that for all $\alpha \geq \alpha_0$,

$$(2.10) \quad u + (\alpha + 1)\nu = T(u + \alpha\nu) = T^{\sigma^*}(u + \alpha\nu) = {}^{\tau^*}T(u + \alpha\nu) .$$

It will now be convenient to use the operator ${}^\tau T^\sigma$ from \mathbb{R}^n to \mathbb{R}^n , such that:

$${}^\tau T_i^\sigma(x) = r_i^{\sigma(i)\tau(i,\sigma(i))} + P_i^{\sigma(i)\tau(i,\sigma(i))} x .$$

We have

$$T(u + \alpha\nu) = {}^{\tau^*}T^{\sigma^*}(u + \alpha\nu) .$$

The dynamic programming principle, for one-player games, implies that

$$\inf_{\sigma} J_i^k(\sigma, \tau^*) = ({}^{\tau^*}T)_i^k(0) ,$$

where the infimum is taken over all strategies of player MIN. By (2.10), $\alpha \mapsto u + \alpha\nu$ is an invariant half-line of ${}^\tau T$, and so

$$\liminf_k \inf_{\sigma} \frac{1}{k} J_i^k(\sigma, \tau^*) = \lim_k \frac{1}{k} ({}^{\tau^*}T)_i^k(0) = \chi_i({}^{\tau^*}T) = \nu_i ,$$

showing that the second equality in (2.2) holds, with $v^U = \nu$. Considering T^{σ^*} instead of ${}^\tau T$ and arguing by duality, we deduce that the first equality in (2.2) also holds, so that σ^*, τ^* are uniform optimal policies, which implies that the uniform value exists. \square

Let us mention that the mean-payoff vector and the uniform value exist, more generally, when T is semialgebraic [BK76] or even definable in an o-minimal structure [BGV15]. However, the existence of $\chi(T)$ is not guaranteed in general: a recent result of Viger [Vig13] shows that the limit may not exist even with a compact action space and transition payments and probabilities that are continuous with respect to the actions.

Finally, observe that the existence of an invariant half-line $\alpha \mapsto u + \alpha\nu$ where $\nu = \lambda e$ is a constant vector is equivalent to the solvability of the ergodic equation $T(u) = u + \lambda e$. In the present setting, this implies that the solvability of the ergodic equation is equivalent to the mean-payoff vector being independent of the initial state, that is $\chi_i(T) = \chi_j(T)$ for every $i, j \in S$. Moreover, in this special case, the optimal policies σ^*, τ^* constructed in Theorem 2.1 are obtained by selecting minimizing and maximizing actions in the expression of $T(x)$ in (2.4), when x is replaced by u .

3. GENERIC UNIQUENESS OF THE BIAS VECTOR OF STOCHASTIC GAMES

3.1. Statement of the main result. Let us first recall some definitions. A *polyhedron* in \mathbb{R}^n is an intersection of finitely many closed half-spaces, a *face* of a polyhedron is an intersection of this polyhedron with a supporting half-space, and a *polyhedral complex* is a finite set \mathcal{K} of polyhedra satisfying the two following properties:

- (i) $P \in \mathcal{K}$ and F is a face of P implies that $F \in \mathcal{K}$;
- (ii) for all $P, Q \in \mathcal{K}$, $P \cap Q$ is a face of P and Q .

A polyhedron in \mathcal{K} is called a *cell* of the polyhedral complex. We refer to the textbook [DLRS10] for background on polyhedral complexes.

Also, a map over \mathbb{R}^n is said to be *piecewise affine* if \mathbb{R}^n can be covered by a finite union of polyhedra (with nonempty interior) on which its restriction is affine. In that case, the set of such polyhedra can be refined in a polyhedral complex. In [Ove02, AT07], it is shown that the piecewise affine functions are exactly the ones that are defined as in (2.4), i.e., the functions that can be written as a minimax over finite sets of affine functions.

Finally, we introduce the following definition, that extends the notion of (finite) ergodic Markov chain.

Definition 3.1. A stochastic game with finite state space and Shapley operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is *ergodic* if the ergodic equation (1.1) is solvable for all operators $g + T$ with $g \in \mathbb{R}^n$.

The authors have given in [AGH15] necessary and sufficient conditions for a perfect-information stochastic game with finite state space and bounded payment function to be ergodic, conditions that we partly recall in the next subsection. Note however that in the case of finite games, the existence of an invariant half-line for piecewise affine maps readily implies that the latter definition is equivalent to the fact that the mean-payoff vector is constant for all additive perturbation vectors of the transition payments.

We now state the main result of this paper, the proof of which is postponed to Subsection 3.3.

Theorem 3.2. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information. Assume that the game is ergodic. Then, the space*

\mathbb{R}^n can be covered by a polyhedral complex such that, for any additive perturbation vector $g \in \mathbb{R}^n$ in the interior of a full-dimensional cell, $g + T$ has a unique bias vector, up to an additive constant.

In particular, the set of perturbation vectors g for which $g + T$ has more than one bias vector, up to an additive constant, is included in a finite union of subspaces of codimension at least 1.

Remark 3.3. This perturbation theorem bears some conceptual similarity with results of weak KAM theory; we refer to the monograph by Fathi [Fat08] for more information. The latter theory deals with a class of one-player deterministic games with continuous time and space. In this setting, the bias vector u and the eigenvalue λ are solution of an ergodic Hamilton-Jacobi PDE $H(x, D_x u) = \lambda$ where the Hamiltonian $(x, p) \mapsto H(x, p)$ is convex in the adjoint variable p . One may consider the perturbation of a Hamiltonian by a potential, which amounts to replacing $H(x, p)$ by $H(x, p) + V(x)$, for some function V . This is similar to the replacement of the Shapley operator T by the perturbed Shapley operator $g + T$ in Theorem 3.2. As observed by Figalli and Rifford in [FR13, Th. 4.2], it follows from weak KAM theory results that under some assumptions, the solution u of $V(x) + H(x, D_x u) = \lambda$ is unique up to an additive constant for a generic function V . Theorem 3.2 shows that an analogous property is valid for finite two-player zero-sum stochastic games. We note however that Theorem 3.2 does not extend easily to the case of PDE, since zero-sum games correspond to Hamilton-Jacobi PDE with a *nonconvex* Hamiltonian (to which current weak KAM methods do not apply).

3.2. Nonlinear spectral theory. The purpose of this subsection is to present or extend some known results in nonlinear spectral theory that will be useful to prove Theorem 3.2, as well as further results in Section 4.

3.2.1. Recession operator and ergodicity conditions. To characterize the ergodicity of a perfect-information stochastic game with finite state space, we shall use, along the lines of [GG04, AGH15], the *recession operator* associated with the Shapley operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$. This operator is a self-map of \mathbb{R}^n defined by

$$(3.1) \quad \widehat{T}(x) := \lim_{\alpha \rightarrow +\infty} \frac{T(\alpha x)}{\alpha}, \quad x \in \mathbb{R}^n.$$

Its existence is not guaranteed in general, but it does exist when the game is finite, i.e., when T is piecewise affine (and more generally when the payment function is bounded). In this case, \widehat{T} is the Shapley operator of a modified version of the stochastic game represented by T where the transition payments are set to 0. Indeed, if T is given as in (2.4), then it is readily seen that the i th coordinate map of \widehat{T} is

$$\widehat{T}_i(x) = \min_{a \in A_i} \max_{b \in B_{i,a}} P_i^{ab} x, \quad x \in \mathbb{R}^n.$$

In this case, we also easily get that

$$\|T - \widehat{T}\|_\infty \leq \|r\|_\infty,$$

which implies in particular that the limit (3.1) defining \widehat{T} is uniform in x .

Observe that if \widehat{T} exists, then it inherits from T the additive homogeneity and the monotonicity properties. Furthermore, it is positively homogeneous, meaning that $\widehat{T}(\alpha x) = \alpha \widehat{T}(x)$ for every $\alpha \geq 0$. As a consequence, any vector proportional

to the unit vector of \mathbb{R}^n is a fixed point of \widehat{T} . We shall call such fixed points *trivial* fixed points. The following result relates the ergodicity of a game to the fixed points of the recession operator associated with its Shapley operator. Since it applies to games with bounded payment function, it deals a fortiori with the case of finite games.

Theorem 3.4 ([AGH15, Th. 3.1]). *Let Γ be a perfect-information stochastic game with finite state space and bounded payment function. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator. The following are equivalent:*

- (i) *the recession operator has only trivial fixed points;*
- (ii) *the mean-payoff vector exists and is constant for all additive perturbation vectors of the transition payments;*
- (iii) *the ergodic equation (1.1) has a solution for all Shapley operators $g + T$, $g \in \mathbb{R}^n$.*

3.2.2. Characterization of the ergodic constant. Let Γ be a finite stochastic game with perfect information, and let T be its Shapley operator. We shall make use of the reduced Shapley operator $T^\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$ associated with the policy $\sigma \in \mathcal{S}_p$ of player MIN, defined in (2.7). We also have, for all $x \in \mathbb{R}^n$,

$$(3.2) \quad T^\sigma(x) = \max_{\tau \in \mathcal{T}_p} (r^{\sigma\tau} + P^{\sigma\tau}x) ,$$

where $r^{\sigma\tau}$ is the vector in \mathbb{R}^n whose i th entry is defined by $r_i^{\sigma\tau} = r_i^{\sigma(i)\tau(i, \sigma(i))}$ and $P^{\sigma\tau}$ is the $n \times n$ stochastic matrix whose i th row is given by $P_i^{\sigma\tau} = P_i^{\sigma(i)\tau(i, \sigma(i))}$. Observe that T^σ is convex (componentwise), monotone and additively homogeneous.

If P is a $n \times n$ stochastic matrix, the *directed graph* associated with P is composed of the nodes $1, \dots, n$ and of the arcs (i, j) , $1 \leq i, j \leq n$, such that $P_{ij} > 0$. A *class* of the matrix P is a maximal set of nodes such that every two nodes in the set are connected by a directed path. A class is said to be *final* if every path starting from a node of this class remains in it. Let us denote by $\mathcal{M}(P)$ the set of invariant probability measures of P , i.e., the set of stochastic (column) vectors $m \in \mathbb{R}^n$ such that $m^\top P = m^\top$. Given a final class C of P , there is a unique invariant probability measure $m \in \mathcal{M}(P)$ the support of which is C , i.e., $\{1 \leq i \leq n \mid m_i > 0\} = C$. Moreover, the set $\mathcal{M}(P)$ is the convex hull of such measures. Since the number of final classes of P is finite, $\mathcal{M}(P)$ is a convex polytope.

Let us denote by $\bar{\chi}(T)$ the *upper mean payoff* of T , i.e., the greatest entry of the mean-payoff vector $\chi(T)$. We next give a characterization of $\bar{\chi}(T)$. Obviously, if T satisfies the ergodic equation (1.1), then $\chi(T)$ is a constant vector and the eigenvalue is $\lambda(T) = \bar{\chi}(T)$. In the sequel, we denote by $x \cdot y$ the standard scalar product in \mathbb{R}^n of two vectors x and y .

Lemma 3.5. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information Γ . Then the upper mean payoff of T is given by*

$$(3.3) \quad \bar{\chi}(T) = \min_{\sigma \in \mathcal{S}_p} \max \{m \cdot r^{\sigma\tau} \mid \tau \in \mathcal{T}_p, m \in \mathcal{M}(P^{\sigma\tau})\} .$$

Proof. First, observe that for all policies $\sigma \in \mathcal{S}_p$, we have $T \leq T^\sigma$, which yields, by monotonicity of the operators, $\chi(T) \leq \chi(T^\sigma)$, and in particular $\bar{\chi}(T) \leq \bar{\chi}(T^\sigma)$.

Considering an invariant half-line (2.6) of T , we know that there exist a vector $u \in \mathbb{R}^n$ such that $T(u) = u + \chi(T)$. Let $\sigma \in \mathcal{S}_p$ be a policy of player MIN such

that $T(u) = T^\sigma(u)$. Then, we have $T^\sigma(u) \leq u + \bar{\chi}(T)e$. Furthermore, we know by a Collatz-Wielandt formula (see [GG04, Prop. 1]) that, for any monotone and additively homogeneous map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, we have

$$\bar{\chi}(F) = \inf\{\mu \in \mathbb{R} \mid \exists x \in \mathbb{R}^n, F(x) \leq \mu e + x\}.$$

So, we deduce that $\bar{\chi}(T^\sigma) \leq \bar{\chi}(T)$, and finally that

$$\bar{\chi}(T) = \min_{\sigma \in \mathcal{S}_p} \bar{\chi}(T^\sigma).$$

Now we fix a policy σ of player MIN, and we let $\chi := \chi(T^\sigma)$ and

$$\mu^\sigma := \max\{m \cdot r^{\sigma\tau} \mid \tau \in \mathcal{T}_p, m \in \mathcal{M}(P^{\sigma\tau})\}.$$

Since T^σ has an invariant half-line with direction χ , there is a vector $v \in \mathbb{R}^n$ such that $T^\sigma(v + \alpha\chi) = v + (\alpha + 1)\chi$ for all $\alpha \geq 0$. In particular, for every policy $\tau \in \mathcal{T}_p$ we have

$$r^{\sigma\tau} + P^{\sigma\tau}v \leq T^\sigma(v) = v + \chi \leq v + \bar{\chi}(T^\sigma)e.$$

Multiplying this inequality by any $m \in \mathcal{M}(P^{\sigma\tau})$, we deduce that $\mu^\sigma \leq \bar{\chi}(T^\sigma)$.

Furthermore, since the germs of affine functions from \mathbb{R} to \mathbb{R} at infinity are totally ordered, there exists a policy $\tau \in \mathcal{T}_p$ such that

$$T^\sigma(v + \alpha\chi) = r^{\sigma\tau} + P^{\sigma\tau}(v + \alpha\chi)$$

for all α large enough. In particular, since the equality

$$(3.4) \quad v + (\alpha + 1)\chi = r^{\sigma\tau} + P^{\sigma\tau}(v + \alpha\chi)$$

holds for all α large enough, we get that $P^{\sigma\tau}\chi = \chi$. Thus, χ is an *harmonic vector* for the stochastic matrix $P^{\sigma\tau}$, and as such it is constant on any final class of $P^{\sigma\tau}$, and its maximum is attained on one of these final class (see [AG03, Lem. 2.9]). Let $m \in \mathcal{M}(P^{\sigma\tau})$ be the invariant probability measure associated with a final class C of $P^{\sigma\tau}$ such that $\chi_i = \bar{\chi}(T^\sigma)$ for all $i \in C$. Then, we deduce from (3.4) that $m \cdot r^{\sigma\tau} = \bar{\chi}(T^\sigma)$, which yields $\mu^\sigma \geq \bar{\chi}(T^\sigma)$ and finally $\mu^\sigma = \bar{\chi}(T^\sigma)$. \square

Let us mention that in (3.3), the set of invariant probability measures of $P^{\sigma\tau}$, $\mathcal{M}(P^{\sigma\tau})$, may be replaced by the set of its extreme points, denoted by $\mathcal{M}^*(P^{\sigma\tau})$, since it is a convex polytope. Note that $\mathcal{M}^*(P^{\sigma\tau})$ is the set of invariant probability measures, the support of which are the final classes of $P^{\sigma\tau}$.

3.2.3. Structure of the eigenspace. An ingredient of our approach is a result of [AG03] which describes the eigenspace of *one-player* Shapley operators T . We next recall this result. We assume that T arises from a game in which only player MAX has nontrivial actions, so that player MIN has only one possible policy. We also assume that T satisfies the ergodic equation (1.1) so that its eigenvalue $\lambda(T)$ is equal to the entries of the mean-payoff vector $\chi(T)$ which is constant. In this case, the representation of the upper-mean payoff $\bar{\chi}(T)$, hence of the eigenvalue $\lambda(T)$, in Lemma 3.5 simplifies as the dependency in σ can be dropped, and we arrive, with a trivial simplification of the notation, to

$$(3.5) \quad \lambda(T) = \max\{m \cdot r^\tau \mid \tau \in \mathcal{T}_p, m \in \mathcal{M}(P^\tau)\}.$$

It is shown in [AG03] that the dimension of the eigenspace of T is controlled by the number of *critical classes*. The latter can be defined through the notion of maximizing measures, which rely on *randomized* policies. For every state i , such a

policy τ assigns to every action $b \in B_i$ a probability $\tau(i, b)$ that this action is selected. This leads to the stochastic matrix P^τ with entries $P_{ij}^\tau = \sum_{b \in B_i} P_{ij}^b \tau(i, b)$, and to the payment vector r^τ with entries $r_i^\tau = \sum_{b \in B_i} r_i^b \tau(i, b)$. Then, the maximum in (3.5) is unchanged if it is taken over the set of randomized policies τ and of invariant probability measures m of the corresponding stochastic matrix P^τ (see [AG03, Prop. 7.2]). A measure m is *maximizing* if $\lambda(T) = m \cdot r^\tau$ for some randomized policy τ , and if $m \in \mathcal{M}^*(P^\tau)$, that is, the support of m is a final class of P^τ . A subset $I \subset \{1, \dots, n\}$ is a *critical class* if there exists a maximizing measure m whose support is I , i.e., $I = \{1 \leq i \leq n \mid m_i > 0\}$, and if I is a maximal element with respect to inclusion among all the subsets of $\{1, \dots, n\}$ which arise in this way. Note that critical classes are disjoint.

The following lemma gives a sufficient condition for the critical class to be unique. Note that the result does not require randomized stationary strategies.

Lemma 3.6. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a convex, monotone and additively homogeneous map. Suppose that the ergodic equation (1.1) is solvable. If there is a unique probability measure which attains the maximum in (3.5), then T has a unique critical class.*

Proof. Let C be a critical class of T . There exists a randomized policy τ such that C is a final class of P^τ and the unique invariant probability measure m with support C satisfies $\lambda = \lambda(T) = m \cdot r^\tau$.

Let u be an eigenvector. For every $i \in \{1, \dots, n\}$, we have $\lambda + u_i - r_i^\tau - P_i^\tau u \geq 0$. Furthermore, since $m^\top P^\tau = m^\top$, we also have

$$\sum_{i \in C} m_i (\lambda + u_i - r_i^\tau - P_i^\tau u) = \lambda + m \cdot u - m \cdot r^\tau - m \cdot (P^\tau u) = 0.$$

This yields that $\lambda + u_i - r_i^\tau - P_i^\tau u = 0$ for every $i \in C$.

Let τ' be a deterministic policy such that $\tau'(i) = b_i$ with $\tau(i, b_i) > 0$ for all indices i . Since P^τ can be written as a convex combination with positive coefficients of the $P^{\tau'}$ with such τ' , we deduce that C contains a final class C' of $P^{\tau'}$, and so there exists an invariant probability measure m' of $P^{\tau'}$ whose support C' is included in C . Similarly r^τ is a convex combination of the $r^{\tau'}$ with same coefficients as for P^τ . Hence, we deduce by a similar argument as above that $\lambda + u_i - r_i^{\tau'} - P_i^{\tau'} u = 0$ for every $i \in C$. Multiplying by m' , we obtain that $\lambda = m' \cdot r^{\tau'}$, which implies that m' attains the maximum in (3.5). Since such a probability measure is unique, m' and C' are the same for all τ' as above, hence $C = C'$ is the support of the unique probability measure attaining the maximum in (3.5). \square

We now describe the eigenspace of T .

Theorem 3.7 ([AG03, Th. 1.1]). *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a convex, monotone and additively homogeneous map. Suppose that the ergodic equation (1.1) is solvable. and let $C \subset \{1, \dots, n\}$ be the set of elements of critical classes. We denote by π_C the restriction map $\mathbb{R}^n \rightarrow \mathbb{R}^C$, $x \mapsto (x_i)_{i \in C}$. Then, the set of eigenvectors of T , denoted by $\mathcal{E}(T)$, satisfies the following properties:*

- (i) *Every element x of $\mathcal{E}(T)$ is uniquely determined by its restriction $\pi_C(x)$.*
- (ii) *The set $\pi_C(\mathcal{E}(T))$ is convex and its dimension is at most equal to the number of critical classes of T ; moreover, the latter bound is attained when T is piecewise affine.*

In particular, combined with Lemma 3.6, the above result yields the following.

Corollary 3.8. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a convex, monotone and additively homogeneous map. Suppose that the ergodic equation (1.1) is solvable. If there is a unique probability measure which attains the maximum in (3.5), then T has a unique eigenvector up to an additive constant.*

We refer the reader to [AG03] for more background on critical classes, which admit several characterizations and can be computed in polynomial time when the game is finite. We only provide here a simple illustration in order to understand the latter theorem.

Example 3.9. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be such that

$$T_1(x) = \max \left\{ x_1, \frac{x_1 + x_2}{2} \right\}, \quad T_2(x) = \max \left\{ -3 + x_2, \frac{x_1 + x_2}{2} \right\}.$$

We have $T(0) = 0$, which shows in particular that the upper mean payoff is 0. If player MAX chooses, when in state 1, the action corresponding to the first term in the expression of T_1 , and when in state 2, the action corresponding to the second term in the expression of T_2 , we arrive at the transition matrix

$$P = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$$

which has the invariant probability measure $m = (1, 0)^\top$. This measure attains the maximum in (3.5). However, its support $I = \{1\}$ is not a critical class for it is not maximal with respect to inclusion. Indeed, if player MAX chooses instead, when in state 1, the action corresponding to the second term in the expression of T_1 , we arrive at the transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

which has the invariant probability measure $m = (1/2, 1/2)^\top$. This measure attains the maximum in (3.5) and its support $I = \{1, 2\}$ is maximal with respect to inclusion. Hence, $I = \{1, 2\}$ is the unique critical class. It follows from Theorem 3.7 that 0 is the unique eigenvector of T , up to an additive constant.

Another ingredient is a variant of a result of Bruck [Bru73], concerning the topology of fixed-point sets of nonexpansive maps. We now consider a *two-player* Shapley operator T such that the ergodic equation (1.1) is solvable, and denote by

$$\mathcal{E}(T) := \{u \in \mathbb{R}^n \mid T(u) = \lambda e + u\}$$

the set of eigenvectors of T (recall that the eigenvalue λ is unique).

Theorem 3.10 (Compare with [Bru73, Th. 2]). *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a monotone and additively homogeneous map. Assume that the ergodic equation (1.1) is solvable. Then, the set of eigenvectors $\mathcal{E}(T)$ is a retract of \mathbb{R}^n by a sup-norm nonexpansive map, meaning that $\mathcal{E}(T) = p(\mathbb{R}^n)$ where p is a sup-norm nonexpansive self-map of \mathbb{R}^n such that $p = p^2$. In particular, $\mathcal{E}(T)$ is arcwise connected.*

Proof. The result of Bruck [Bru73, Th. 2] shows that, under some compactness conditions, the fixed-point set of a nonexpansive self-map of a Banach space is a retract of the whole space by a nonexpansive map.

Assume now that T is monotone, additively homogeneous, and admits an eigenvector for the eigenvalue λ . Then, the eigenspace $\mathcal{E}(T)$ coincides with the fixed-point

set of the map $x \mapsto -\lambda e + T(x)$. The latter map is sup-norm nonexpansive and satisfies the condition of [Bru73, Th. 2], and so, $\mathcal{E}(T)$ is a nonexpansive retract of \mathbb{R}^n . \square

Remark 3.11. The retraction p in Theorem 3.10 can be chosen to be monotone and additively homogeneous. This can actually be shown by elementary means, following a construction in the proof of [GG04, Lem. 3]. Indeed, we may assume without loss of generality that $\lambda = 0$, and consider $q(x) := \lim_{k \rightarrow \infty} \inf_{\ell \geq k} T^\ell(x)$, which is finite because every orbit of a nonexpansive map that admits a fixed point must be bounded. Since T is monotone and continuous, we get $T(q(x)) \leq q(x)$, and so, $p(x) := \lim_{k \rightarrow \infty} T^k(q(x))$, which is the limit of a nonincreasing and bounded sequence, exists and is finite. The map p is easily shown to be monotone and additively homogeneous and to satisfy $p = p^2$.

3.3. Proof of Theorem 3.2. Let T be the Shapley operator of a finite stochastic game with perfect information Γ which is assumed to be ergodic. Let $\sigma \in \mathcal{S}_p$ be a policy of player MIN. We define the real map $\lambda^\sigma(\cdot)$ on \mathbb{R}^n by

$$(3.6) \quad \lambda^\sigma(g) := \max\{m \cdot (g + r^{\sigma\tau}) \mid \tau \in \mathcal{T}_p, m \in \mathcal{M}^*(P^{\sigma\tau})\},$$

where $\mathcal{M}^*(P)$ denotes the set of extreme points of the convex polytope $\mathcal{M}(P)$, that is, the set of invariant probability measures associated with the final classes of the stochastic matrix P . The fact that $\mathcal{M}^*(P^{\sigma\tau})$ is a set of probability measures yields that λ^σ is monotone and additively homogeneous, hence sup-norm nonexpansive (and continuous). Furthermore, since the set of policies \mathcal{T}_p of player MAX is finite, as well as all the sets $\mathcal{M}^*(P^{\sigma\tau})$, then the map λ^σ is piecewise affine.

We now define the polyhedral complex \mathcal{C}^σ covering \mathbb{R}^n , the full-dimensional cells of which are precisely the maximal polyhedra on which the piecewise affine map λ^σ coincides with an affine map.

Let Q be a cell of \mathcal{C}^σ with full dimension. We claim that if a vector g is in the interior of Q , then the set of eigenvectors of the reduced one-player Shapley operator $F := g + T^\sigma$ is either empty or reduced to a line. To see this, it suffices to observe that the measure m attaining the maximum in (3.6) is unique for all g in the interior of Q and independent of the choice of g in this interior. Indeed, if m is such a measure, then $m \in \mathcal{M}^*(P^{\sigma\tau})$ for some $\tau \in \mathcal{T}_p$ and for $d = m \cdot r^{\sigma\tau} \in \mathbb{R}$, we have $\lambda^\sigma(g') \geq m \cdot g' + d$, for all $g' \in \mathbb{R}^n$, with equality at g . Since $g' \mapsto \lambda^\sigma(g')$ is an affine map on Q , and g is in the interior of Q , the equality $\lambda^\sigma(g') = m \cdot g' + d$ holds for all $g' \in Q$ and so m must coincide with the linear part of the affine map, which is independent of g (and unique). We deduce from Corollary 3.8 that $\mathcal{E}(F)$, if it is nonempty, is reduced to a line of direction e .

Consider now the polyhedral complex \mathcal{C} obtained as the refinement of all the complexes \mathcal{C}^σ . This complex still covers \mathbb{R}^n and has cells with nonempty interior. Let g be a perturbation vector in the interior of a full-dimensional cell of \mathcal{C} . Since the game Γ is ergodic, $\mathcal{E}(g+T)$ is not empty. Let u be an eigenvector of $g+T$. According to (2.8), there is a policy $\sigma \in \mathcal{S}_p$ of player MIN such that $g + T(u) = g + T^\sigma(u)$. Hence u is also an eigenvector of $g + T^\sigma$. So, there is a finite family Σ^* of \mathcal{S}_p such that $\mathcal{E}(g+T) = \bigcup_{\sigma \in \Sigma^*} \mathcal{E}(g+T^\sigma)$. Moreover, we have proved that for any policy $\sigma \in \Sigma^*$, the eigenspace $\mathcal{E}(g+T^\sigma)$ is reduced to a line. Thus, $\mathcal{E}(g+T)$ is composed of a finite union of lines which all have the same direction, e . Consider now the hyperplane orthogonal to the unit vector, $H := \{x \in \mathbb{R}^n \mid x \cdot e = 0\}$, and let π denote the orthogonal projection on H . Then, $\pi(\mathcal{E}(g+T)) = \mathcal{E}(g+T) \cap H$ is finite.

However, by Theorem 3.10, $\mathcal{E}(g + T)$ is connected. Then, the set $\pi(\mathcal{E}(g + T))$ is also connected, and since it is finite, it must be reduced to a point. It follows that $g + T$ has a unique eigenvector, up to an additive constant. \square

3.4. Example. We conclude this section by an example illustrating Theorem 3.2. Consider the following Shapley operator defined on \mathbb{R}^3 (here we use \wedge and \vee instead of min and max, respectively, and we recall that the addition has precedence over them):

$$T(x) = \begin{pmatrix} \frac{1}{2}(x_1 + x_3) \wedge 1 + \frac{1}{2}(x_1 + x_2) \\ 2 + \frac{1}{2}(x_1 + x_3) \wedge (1 + \frac{1}{2}(x_1 + x_2) \vee -2 + x_3) \\ -3 + \frac{1}{2}(x_1 + x_3) \vee 1 + x_3 \end{pmatrix}.$$

It can be proved, using Theorem 3.4, that the ergodic equation (1.1) is solvable for every perturbation vector $g \in \mathbb{R}^3$. Figure 1 shows the intersection of the hyperplane $\{g \in \mathbb{R}^3 \mid g_3 = 0\}$ and the polyhedral complex introduced in Theorem 3.2.

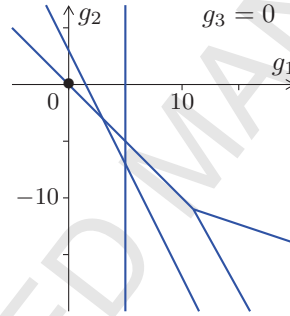


FIGURE 1. The polyhedral complex introduced in Theorem 3.2. The bias vector of the operator $g + T$ is unique, up to an additive constant, if the perturbation vector g belongs to the interior of a maximal dimensional cell of this complex.

Let us detail what happens in the neighborhood of $g = 0$, point in which $g + T$ fails to have a unique eigenvector. Note that in the neighborhood of $g = 0$, the eigenvalue of $g + T$ remains 1.

- If $g_1 + g_2 = 0$, the eigenvectors of $g + T$ are defined by

$$x_1 = x_2 + 2g_1, \quad -3 + g_2 \leq x_2 - x_3 \leq -2 - g_1.$$

- If $g_1 + g_2 > 0$, the unique eigenvector, up to an additive constant, is

$$(-2 + 2g_1, -2 + 2g_1 + 2g_2, 0).$$

- If $g_1 + g_2 < 0$, the unique eigenvector, up to an additive constant, is

$$(-3 + 2g_1 + g_2, -3 + g_2, 0).$$

4. APPLICATION TO POLICY ITERATION

We finally apply our results to show that policy iteration combined with a perturbation scheme can solve degenerate stochastic games.

4.1. Hoffman-Karp policy iteration. Let us recall the notation of Sections 2 and 3. We denote by Γ a finite stochastic game with perfect information. The (finite) set of policies of player MIN is denoted by \mathcal{S}_p , i.e., an element of \mathcal{S}_p is a map $\sigma : S \rightarrow \bigcup_{i \in S} A_i$ such that $\sigma(i) \in A_i$ for every state $i \in S$. The (finite) set of policies of player MAX is denoted by \mathcal{T}_p , i.e., an element of \mathcal{T}_p is a map $\tau : K_A \rightarrow \bigcup_{i \in S, a \in A_i} B_{i,a}$ such that $\tau(i, a) \in B_{i,a}$ for every state $i \in S$ and action $a \in A_i$. Finally, recall that for $\sigma \in \mathcal{S}_p$ and $\tau \in \mathcal{T}_p$, $P^{\sigma\tau}$ denotes the $n \times n$ stochastic matrix whose i th row is given by $P_i^{\sigma\tau} = P_i^{\sigma(i)\tau(i,\sigma(i))}$.

When $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the Shapley operator (2.4) of a finite stochastic game with perfect information, Hoffman and Karp [HK66] have introduced a policy iteration algorithm, which takes the description of the game as the input and returns the eigenvalue λ and an eigenvector u of T , i.e., a solution $(\lambda, u) \in \mathbb{R} \times \mathbb{R}^n$ of the ergodic equation $T(u) = \lambda e + u$. Also, optimal stationary strategies for both players in the mean-payoff game can be derived from the output of the algorithm.

This algorithm and more generally policy iteration procedures are a standard general way for solving mean-payoff stochastic games. In worst-case scenario, they require an exponential number of iterations, see [Fri11]. However, no polynomial-time algorithm is known to solve mean-payoff games. In fact, it is an important open question to know whether there exists one, since this problem, or any problem involving a polynomial-time equivalent perfect-information two-player zero-sum stochastic game (such as discounted stochastic games, simple stochastic games, parity games, see [AM09]) is one of the few problems that belong to the complexity class $\text{NP} \cap \text{coNP}$, see [Con92].

It is convenient here to state an abstract, slightly more general, version of the Hoffman-Karp algorithm, described in terms of the operators T and T^σ (Algorithm 1).

Algorithm 1: Policy iteration, compare with [HK66]

input : Shapley operator T of perfect-information finite stochastic game.
output: eigenvalue λ and eigenvector u of T .

```

1 initialization: select an arbitrary policy  $\sigma_0 \in \mathcal{S}_p$ 
2 repeat
3   compute an eigenpair  $(\lambda^k, v^k)$  of  $T^{\sigma_k}$ 
4   improve the policy  $\sigma_k$  in a conservative way: select a policy  $\sigma_{k+1} \in \mathcal{S}_p$ 
     such that  $T^{\sigma_{k+1}}(v^k) = T(v^k)$  with, for every state  $i \in S$  such that
      $T_i^{\sigma_k}(v^k) = T_i(v^k)$ ,  $\sigma_{k+1}(i) = \sigma_k(i)$ 
5 until  $\sigma_{k+1} = \sigma_k$ 
6 return  $\lambda^k$  and  $v^k$ 
```

We assume Algorithm 1 is interpreted in exact arithmetics (the vectors v^k have rational coordinates and the λ^k are rational numbers). To implement Step 3, we may call any oracle able to compute the eigenvalue and an eigenvector of a one-player stochastic game. In the original approach of Hoffman and Karp, the oracle consists in applying the same policy iteration algorithm for the one-player game with fixed policy σ_k . The proof of Hoffman and Karp shows that Algorithm 1 is valid under a restrictive assumption.

Theorem 4.1 (Corollary of [HK66]). *Algorithm 1 terminates and is correct if for all choices of policies σ and τ of the two players, the corresponding transition matrix $P^{\sigma\tau}$ is irreducible.*

Indeed, it is easy to see that the sequence $(\lambda^k)_k$ of Algorithm 1 is nonincreasing, that is, $\lambda^{k+1} \leq \lambda^k$ for all iterations k . The irreducibility assumption was shown to imply that the latter inequalities are always strict, which entails the finite time convergence (each policy yields a unique well defined eigenvalue, these eigenvalues constitute a decreasing sequence, and there are finitely many policies).

However, the assumption that *all* the stochastic matrices $P^{\sigma\tau}$ be irreducible is way too strong to guarantee that Algorithm 1 is properly posed. Indeed, to execute the algorithm, it suffices that at every iteration k the operator T^{σ_k} admits an eigenvalue and an eigenvector, which is the case in particular if for all policies σ , the graph obtained by taking the *union* of the edge sets of all the graphs associated to $P^{\sigma\tau}$ for the different choices of τ is strongly connected (the existence of the eigenvalue and eigenvector, in this generality, goes back to Bather [Bat73], see also [GG04, AGH15] for a more general discussion). In particular, the irreducibility assumption of Hoffman and Karp is essentially never satisfied for deterministic games, whereas the condition involving the union of the edge sets is satisfied by relevant classes of deterministic games.

It should be noted that Algorithm 1 may, in general, lead to degenerate iterations, in which $\lambda^{k+1} = \lambda^k$. As shown by an example in [ACTDG12, Sec. 6], this may lead the algorithm to cycle when the bias vector is not unique. This difficulty was solved first in the deterministic framework in [CTGG99], where it was shown that cycling can be avoided by enforcing a special choice of the bias vector, obtained by a nonlinear projection operation. This approach was then extended to the stochastic framework in [CTG06, ACTDG12]. As a special case of these results, we get that policy iteration is correct and does terminate under much milder conditions than in Theorem 4.1.

Theorem 4.2 (Corollary of [CTG06, Th. 7]). *Algorithm 1 terminates and is correct if for each choice of policy σ of player MIN, the operator T^σ has an eigenvalue and a unique eigenvector, up to an additive constant.*

We next show that the conditions of Theorem 4.2 are satisfied for generic payments, and conclude that nongeneric instances can still be solved by the Hoffman-Karp algorithm, after an effective perturbation of the input.

4.2. Generic termination of policy iteration. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information. The following assumption guarantees that Algorithm 1 is well posed for any additive perturbation of T .

Assumption A. For any policy $\sigma \in S_p$ of player MIN, the one-player game with Shapley operator T^σ is ergodic in the sense of Definition 3.1, meaning that for all perturbation vectors $g \in \mathbb{R}^n$, the operator $g + T^\sigma$ has an eigenvalue (and an eigenvector).

This assumption is much milder than the original assumption of Hoffman and Karp, requiring all the transition matrices $P^{\sigma\tau}$ to be irreducible (see Theorem 4.1). Also, since $\widehat{T} = \min_{\sigma \in S_p} \widehat{T}^\sigma$, it readily follows from Theorem 3.4 that Assumption A implies that the original two-player game is ergodic. Moreover, we shall see in the

next subsection that one can always transform a game (in polynomial time) by a “big M ” trick in such a way that Assumption A becomes satisfied.

By using the arguments of Section 3, we now show that under Assumption A, Algorithm 1 terminates for a generic perturbation of the payments.

Theorem 4.3. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information satisfying Assumption A. Then, the space \mathbb{R}^n can be covered by a polyhedral complex such that for each additive perturbation vector $g \in \mathbb{R}^n$ in the interior of a full-dimensional cell, Algorithm 1 terminates after a finite number of steps and gives an eigenpair of $g + T$.*

Proof. Consider the same complex \mathcal{C} as in Section 3 and let g be a perturbation vector in the interior of a full-dimensional cell of \mathcal{C} . It follows from the proof of Theorem 3.2 (see Subsection 3.3) that for any policy $\sigma \in \mathcal{S}_p$, the eigenvector of $g + T^\sigma$, which exists according to Assumption A, is unique up to an additive constant. Hence, at each step k of Algorithm 1, the bias vector v^k of $g + T^{\sigma_k}$ is unique up to an additive constant. The conclusion follows from Theorem 4.2. \square

We next provide an explicit perturbation g , depending on a parameter ε , for which the policy iteration algorithm applied to $g + T$ is valid. We shall see in the next subsection that ε can be instantiated with a polynomial number of bits, in such a way that the original unperturbed problem is solved.

Before giving this explicit perturbation scheme, let us mention that the polyhedral complex \mathcal{C} introduced in the proof of Theorem 3.2 (Subsection 3.3) can be constructed for any perfect-information finite stochastic game, whether it is ergodic or not. Recall indeed that \mathcal{C} is obtained as a refinement of all the regions where the maps $g \mapsto \lambda^\sigma(g)$ with $\sigma \in \mathcal{S}_p$, defined in (3.6), are affine. In particular, we do not assume ergodicity in the next two propositions.

Proposition 4.4. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information. Then, there exists $\varepsilon_0 > 0$ such that all perturbation vectors $g_\varepsilon := (\varepsilon, \varepsilon^2, \dots, \varepsilon^n)$ with $0 < \varepsilon < \varepsilon_0$ are in the interior of the same full-dimensional cell of the polyhedral complex \mathcal{C} introduced in Subsection 3.3.*

Proof. The cells of the polyhedral complex \mathcal{C} introduced in Subsection 3.3 that are not full-dimensional are included in an arrangement of a finite number of hyperplanes. The real curve $\varepsilon \mapsto g_\varepsilon = (\varepsilon, \dots, \varepsilon^n)$ cannot cross a given hyperplane in this arrangement more than n times (otherwise, a polynomial of degree n would have strictly more than n roots). We deduce that there is a value $\varepsilon_0 > 0$ such that the restriction of the curve $\varepsilon \mapsto g_\varepsilon$ to the open interval $(0, \varepsilon_0)$ crosses no hyperplane of the arrangement. Therefore, it must stay in the interior of a full-dimensional cell of the complex \mathcal{C} . \square

The following result is a refinement of the previous one.

Proposition 4.5. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a finite stochastic game with perfect information. Then, there exist $\varepsilon_1 > 0$, policies $\sigma \in \mathcal{S}_p$ and $\tau \in \mathcal{T}_p$, and an invariant probability measure $m^{\sigma\tau}$ of the stochastic matrix $P^{\sigma\tau}$ such that for all $\varepsilon \in [0, \varepsilon_1]$, the upper mean payoff of $g_\varepsilon + T$ is given by $\bar{\chi}(g_\varepsilon + T) = m^{\sigma\tau} \cdot (g_\varepsilon + r^{\sigma\tau})$.*

Proof. Recall that the upper mean payoff of $g + T$ is given by

$$\begin{aligned} \bar{\chi}(g + T) &= \min_{\sigma \in \mathcal{S}_p} \lambda^\sigma(g) \\ \text{with } \lambda^\sigma(g) &= \max\{m \cdot (g + r^{\sigma\tau}) \mid \tau \in \mathcal{T}_p, m \in \mathcal{M}^*(P^{\sigma\tau})\} . \end{aligned}$$

By construction of the polyhedral complex \mathcal{C} in Subsection 3.3, in the interior of a full-dimensional cell, each piecewise affine map $g \mapsto \lambda^\sigma(g)$ coincides with a unique affine map, but $g \mapsto \bar{\chi}(g + T)$ need not be affine. Hence, we can refine the complex \mathcal{C} into a complex \mathcal{C}' such that the latter piecewise affine map also coincides with a unique affine map on each cell of full dimension. The same proof as Proposition 4.4 leads to the existence of a parameter ε_1 such that all the perturbations g_ε with $\varepsilon \in (0, \varepsilon_1)$ lie in the interior of the same full-dimensional cell of \mathcal{C}' . Let Q be this cell. By construction of the latter complex, there exists a policy $\sigma \in \mathcal{S}_p$ such that $\bar{\chi}(g + T) = \lambda^\sigma(g)$ for all $g \in Q$, and for that policy σ there exists a policy $\tau \in \mathcal{T}_p$ and $m^{\sigma\tau} \in \mathcal{M}^*(P^{\sigma\tau})$ such that $\lambda^\sigma(g) = m^{\sigma\tau} \cdot (g + r^{\sigma\tau})$ for all $g \in Q$. \square

It follows that solving the game with Shapley operator $g_\varepsilon + T$ for ε small enough entails a solution of the original game.

Proposition 4.6. *If T satisfies Assumption A, then there exists $\varepsilon_1 > 0$ (same as in Proposition 4.5) such that Algorithm 1 terminates for any input $g_\varepsilon + T$ with $\varepsilon \in (0, \varepsilon_1)$. Furthermore, for any policy σ satisfying $\lambda(g + T) = \lambda(g + T^\sigma)$, we also have $\lambda(T) = \lambda(T^\sigma)$.*

Proof. First note that here, since the game is ergodic, we have $\bar{\chi}(g + T) = \lambda(g + T)$ for all perturbations $g \in \mathbb{R}^n$.

Following Proposition 4.5, the parameter ε_1 is such that all perturbations g_ε with $0 < \varepsilon < \varepsilon_1$ lie in the interior of the same full-dimensional cell of the complex \mathcal{C}' (which is a refinement of \mathcal{C} , see the proof of Proposition 4.5). Denote by Q this cell. The termination of Algorithm 1 is then a straightforward consequence of Theorem 4.3.

By definition of the polyhedral complex \mathcal{C}' , the piecewise affine map $g \mapsto \lambda(g + T)$ is affine when restricted to Q . Furthermore, for any policy $\sigma \in \mathcal{S}_p$, we have either $\lambda(g + T) = \lambda^\sigma(g) = \lambda(g + T^\sigma)$ for all g in Q , or $\lambda(g + T) < \lambda^\sigma(g) = \lambda(g + T^\sigma)$ for all g in the interior of Q . Hence the result. \square

4.3. Complexity issues. In this subsection, we show that computing the upper mean payoff of a Shapley operator (a fortiori the eigenvalue if it exists) is polynomial-time reducible to the computation of the eigenvalue of a Shapley operator for which Algorithm 1 terminates. This fact is a direct consequence of Theorem 4.12 below. To do so, we shall need explicit bounds on the perturbation parameter ε .

We first explain how the general case can be reduced to the situation in which Assumption A holds. To that purpose, let us introduce for any real number $M \geq 0$, the map $R_M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose i th coordinate is given by

$$[R_M(x)]_i = \max \left\{ x_i, \max_{1 \leq j \leq n} (-M + x_j) \right\} , \quad x \in \mathbb{R}^n .$$

It is convenient to introduce Hilbert's seminorm on \mathbb{R}^n , defined by

$$\|x\|_H = \max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i .$$

Observe that R_M is a projection on the set $\{x \in \mathbb{R}^n \mid \|x\|_H \leq M\}$, meaning that $R_M^2 = R_M$ and

$$R_M(x) = x \iff \|x\|_H \leq M .$$

Lemma 4.7. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a perfect-information finite stochastic game, and let $M \geq 0$. Then, $T \circ R_M$ has an eigenvalue.*

Proof. First, note that the recession operator of \widehat{R}_M is given by

$$\widehat{R}_M(x) = (\max x) e , \quad x \in \mathbb{R}^n ,$$

where $\max x := \max_{1 \leq i \leq n} x_i$. Second, it has been noted in Subsection 3.2 that the limit (3.1) defining \widehat{T} is uniform in x . Hence, we get that $\widehat{T \circ R_M} = \widehat{T} \circ \widehat{R}_M$. Thus, using the properties of recession operators, we have, for any vector $x \in \mathbb{R}^n$,

$$\widehat{T \circ R_M}(x) = \widehat{T} \circ \widehat{R}_M(x) = \widehat{T}((\max x) e) = (\max x) e .$$

This proves that the only fixed points of $\widehat{T \circ R_M}$ are trivial fixed points. The conclusion follows from Theorem 3.4. \square

Given a perfect-information finite stochastic game Γ with Shapley operator T , the operator $T \circ R_M$ can be interpreted as the Shapley operator of another perfect-information finite stochastic game with state space S . In this game, at each step, if the current state is $i \in S$, player MIN start by choosing an action $a \in A_i$. Then, player MAX chooses an action $b \in B_{i,a}$ which gives rise to a transition payment r_i^{ab} and a state, j , is chosen by nature with probability $[P_i^{ab}]_j$ and announced to the players. Finally, player MAX chooses the next state to be either j with no additional payment, or to be any other state k with an additional payment of $-M$. In other words, player MAX has the option of teleporting himself to any other state, by accepting a penalty M .

Note that, since player MIN has the same action space in the latter game as in the game Γ , the sets of her stationary strategies in both games are identical. Then, for a fixed policy σ of player MIN, the one-player Shapley operator $(T \circ R_M)^\sigma$ is equal to $T^\sigma \circ R_M$, and we get from Lemma 4.7 the following result.

Corollary 4.8. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a perfect-information finite stochastic game. Then, $T \circ R_M$ satisfies Assumption A.*

In the modified game with Shapley operator $T \circ R_M$, player MAX makes, at each step, the final decision about the next state, provided an additional cost of M . The following result shows that if this cost is large enough, then player MAX cannot do better, in the long run, than in the game Γ .

Lemma 4.9. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a perfect-information finite stochastic game. Then, there exists a positive constant M_0 such that for any $M > M_0$, the eigenvalue of $T \circ R_M$ is equal to the upper mean payoff $\bar{\chi}(T)$.*

Proof. First, note that $R_M(x) \geq x$ for all $x \in \mathbb{R}^n$. Hence, by monotonicity of T , we deduce that $T \circ R_M \geq T$, which yields $\chi(T \circ R_M) \geq \chi(T)$. Since $T \circ R_M$ has an eigenvalue, denoted by $\lambda(T \circ R_M)$, then we have $\lambda(T \circ R_M) \geq \bar{\chi}(T)$.

Second, we know that T has an invariant half-line with direction $\chi(T)$. So there exists a vector $u \in \mathbb{R}^n$ such that $T(u) = u + \chi(T)$. Now let $M_0 := \|u\|_H$. For every $M > M_0$, we have

$$T \circ R_M(u) = T(u) = u + \chi(T) \leq u + \bar{\chi}(T)e .$$

By application of a Collatz-Wielandt formula (see [GG04]), we know that the eigenvalue of $T \circ R_M$ is given by

$$\lambda(T \circ R_M) = \inf\{\mu \in \mathbb{R} \mid \exists u \in \mathbb{R}^n, T \circ R_M(u) \leq u + \mu e\} .$$

Hence $\lambda(T \circ R_M) \leq \bar{\chi}(T)$. \square

We shall need a technical bound on invariant probability measures of stochastic matrices arising from strategies. We state it here for an arbitrary irreducible stochastic matrix.

Lemma 4.10. *Let P be a $n \times n$ irreducible stochastic matrix whose entries are rational numbers with numerators and denominators bounded by an integer D . Then, the entries of the invariant probability measure of P are rational numbers whose least common denominator is bounded by*

$$n^{n/2} D^{n^2}$$

Proof. The invariant probability measure m of P is the unique solution of the linear system

$$(4.1) \quad \begin{cases} (I - P^\top) m = 0 \\ e^\top m = 1 \end{cases} ,$$

where I is the identity matrix. Note that one row of the subsystem $(I - P^\top) m = 0$ is redundant since we are dealing with stochastic vectors and $e^\top m = 1$. Then, by deleting this row, and by multiplying every row of the latter subsystem by all the denominators of the coefficients appearing in this row, we arrive at a Cramer linear system with integer coefficients of absolute value less than D^n , and with unit coefficients on the last row. Solving this system by Cramer's rule, we obtain that the entries of m are rational numbers whose denominators divide the determinant of the system. Using Hadamard's inequality for determinants, we deduce that these denominators are bounded by

$$((n-1)(D^n)^2 + 1)^{n/2} \leq n^{n/2} D^{n^2} . \quad \square$$

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of a perfect-information finite stochastic game Γ with state space $\{1, \dots, n\}$. We just showed that the upper mean payoff of T can be recovered from the eigenvalue of the operator $T \circ R_M$ for a suitable large M . The latter operator satisfies Assumption A, and so, we can in principle apply Algorithm 1 to it. However, to do so in a way which leads to a polynomial-time transformation of the input, it is convenient to introduce the following modified Shapley operator $T_M : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ given, for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, by

$$T_M(x, y) := (T(y), R_M(x)) .$$

Note that we have

$$(T_M)^2(x, y) = \begin{pmatrix} T \circ R_M(x) \\ R_M \circ T(y) \end{pmatrix} .$$

The following immediate lemma shows that one can recover the bias vectors and the eigenvalue of $T \circ R_M$ from those of T_M .

Lemma 4.11. *If v is a bias vector of the operator $T \circ R_M$ with eigenvalue λ , then $(v, R_M(v) - (\lambda/2)e)$ is a bias vector of the operator T_M with eigenvalue $\lambda/2$, and all bias vectors of T_M arise in this way.* \square

The operator T_M is the dynamic programming operator of a game, denoted by Γ_M , with state space $\{1, \dots, 2n\}$. In each state $i \in \{1, \dots, n\}$, the actions, the payments and the transition function are the same as in Γ , except that the next state is labeled by an element of $\{n+1, \dots, 2n\}$ instead of $\{1, \dots, n\}$. Moreover, in each state $i \in \{n+1, \dots, 2n\}$, player MIN has only one possible action, while player MAX chooses the next state j among $\{1, \dots, n\}$ with a cost M if $i - j \neq n$. In particular, the policies of player MIN in the two games Γ and Γ_M are in one-to-one correspondence, and to simplify the presentation, we shall use the same notation for these policies. Hence, we shall write $(T_M)^\sigma(x, y) = (T^\sigma(y), R_M(x))$ for such a policy σ .

We saw in Lemma 4.7 that the operator $T \circ R_M$ has an eigenvalue. The same is true for the operator T_M by Lemma 4.11. Moreover, the same conclusion applies to the operator $(T_M)^\sigma$ for any policy σ , so that T_M satisfies Assumption A. We know that for $\varepsilon > 0$ small enough, the perturbed operator $g_\varepsilon + T \circ R_M$ has a unique bias vector, up to an additive constant, where $g_\varepsilon = (\varepsilon, \dots, \varepsilon^n)$. This leads to considering, for $\varepsilon > 0$,

$$T_{M,\varepsilon} := (g_\varepsilon, 0) + T_M .$$

Theorem 4.12. *Let Γ be a perfect-information finite stochastic game whose transition payments and probabilities are rational numbers with numerators and denominators bounded by an integer $D \geq 2$. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the Shapley operator of Γ . If*

$$M > 4n^{n/2} D^{n^2+1} \quad \text{and} \quad 0 < \varepsilon < \frac{1}{n^n D^{2n(n+1)}} ,$$

then the upper mean payoff of T can be recovered from $T_{M,\varepsilon} = (g_\varepsilon, 0) + T_M$, in the sense that for any policy σ of player MIN such that $\lambda(T_{M,\varepsilon}) = \lambda((T_{M,\varepsilon})^\sigma)$, we have $\bar{\chi}(T) = \bar{\chi}(T^\sigma)$. Furthermore, such a policy can be obtained by applying Algorithm 1 with the input $T_{M,\varepsilon}$.

Proof. Let $g \in [0, 1]^n$, and fix a policy σ of player MIN. In the game Γ_M , consider a policy of player MAX such that, when in state $i \in \{1, \dots, n\}$, he chooses action $b_i \in B_{i,\sigma(i)}$, and when in state $i \in \{n+1, \dots, 2n\}$, he chooses the next state to be $j(i) \in \{1, \dots, n\}$. Then, the transition matrix associated with that choice of policy is the following $2n \times 2n$ block matrix:

$$(4.2) \quad \begin{pmatrix} 0 & P^{\sigma\tau} \\ Q & 0 \end{pmatrix} ,$$

where $\tau \in \mathcal{T}_p$ is a policy of player MAX in the game Γ such that $\tau(i, \sigma(i)) = b_i$ for each state i , and where Q is the $n \times n$ stochastic matrix whose coefficients are $Q_{ij} = 1$ for $j = j(i+n)$ and 0 otherwise.

Let $(m, m') \in \mathbb{R}^n \times \mathbb{R}^n$ be an invariant probability measure of the stochastic matrix (4.2). The vectors m and m' satisfy in particular

$$(4.3) \quad m^\top P^{\sigma\tau} = m'^\top , \quad m'^\top Q = m^\top , \quad m \cdot e + m' \cdot e = 1 .$$

We are interested in the eigenvalue of the perturbed one-player Shapley operator $(g, 0) + (T_M)^\sigma$. Hence, following formula (3.5), we consider the quantity

$$\gamma := m \cdot (g + r^{\sigma\tau}) - M \sum_{\substack{1 \leq i \leq n \\ Q_{ii}=0}} m'_i .$$

If for every index i in the support of m' , we have $Q_{ii} = 1$, then we deduce from the second equality in (4.3) that $m' = m$. This yields that $2m$ is an invariant probability measure of $P^{\sigma\tau}$, and that

$$\gamma = m \cdot (g + r^{\sigma\tau}) \leq \frac{1}{2} \bar{\chi}(g + T^\sigma) .$$

Note that the equality is attained in the above inequality for some policy τ and some invariant probability measure m .

If there is an index i in the support of m' such that $Q_{ii} = 0$, then we have

$$\gamma \leq m \cdot (g + r^{\sigma\tau}) - M m'_i \leq 1 + \max_{i,a,b} r_i^{ab} - K_1 M ,$$

where K_1 is a positive constant such that $K_1 \leq m'_i$. Note that K_1 can be chosen independently of M and of the particular choices of the policies. Then, taking $M > M_0 := (K_1)^{-1}(1 + (3/2)\|r\|_\infty)$, we obtain that

$$\gamma < \frac{1}{2} \min_{i,a,b} r_i^{ab} \leq \frac{1}{2} \bar{\chi}(g + T^\sigma) .$$

Thus, we have proved that, for all policies σ of player MIN and for all $g \in [0, 1]^n$, we have

$$\lambda((g, 0) + (T_M)^\sigma) = \frac{1}{2} \bar{\chi}(g + T^\sigma) ,$$

as soon as $M > M_0$. In particular, the choice of the parameter ε such that $T_{M,\varepsilon}$ is a generic instance only relies on T (Proposition 4.5).

We now fix some $M > M_0$. Consider, for policies σ, σ' of player MIN and τ, τ' of player MAX, two distinct pairs $(m, d) \neq (m', d')$, where $m \in \mathcal{M}^*(P^{\sigma\tau})$, $m' \in \mathcal{M}^*(P^{\sigma'\tau'})$, $d := m \cdot r^{\sigma\tau}$ and $d' := m' \cdot r^{\sigma'\tau'}$. We need to compare the affine maps $g \mapsto m \cdot g + d$ and $g \mapsto m' \cdot g + d'$ along the curve $\varepsilon \mapsto g_\varepsilon$ with $\varepsilon \in (0, 1)$.

Assume first that $d = d'$. Then $m \neq m'$ and we can select the smallest index i such that $m_i \neq m'_i$. Note that since m and m' are stochastic vectors, we necessarily have $i < n$ and we also have the existence of another index j such that $i < j \leq n$ and $m_j \neq m'_j$. Without loss of generality, we may assume that $m_i - m'_i > 0$. Let $K_2 \in \mathbb{R}$ be such that $0 < K_2 < m_i - m'_i$. Then, for any positive parameter $\varepsilon < K_2 n^{-1}$, we have

$$\begin{aligned} (m \cdot g_\varepsilon + d) - (m' \cdot g_\varepsilon + d') &= (m_i - m'_i)\varepsilon^i + \sum_{i < j \leq n} (m_j - m'_j)\varepsilon^j \\ &> K_2 \varepsilon^i - n\varepsilon^{i+1} = \varepsilon^i (K_2 - n\varepsilon) > 0 . \end{aligned}$$

Assume now that $d \neq d'$, say $d > d'$, and let $K_3 \in \mathbb{R}$ be such that $0 < K_3 < d - d'$. Then, for any positive parameter $\varepsilon < K_3$, we have

$$(m \cdot g_\varepsilon + d) - (m' \cdot g_\varepsilon + d') > \varepsilon^n - \varepsilon + K_3 > 0 .$$

Note that we can choose the positive constants K_2 and K_3 independently of $\sigma, \sigma', \tau, \tau', m$ and m' . Hence, the above arguments show that the set of polynomial functions $\varepsilon \mapsto m \cdot (g_\varepsilon + r^{\sigma\tau})$ with $\sigma \in \mathcal{S}_p$, $\tau \in \mathcal{T}_p$, and $m \in \mathcal{M}^*(P^{\sigma\tau})$, is totally

ordered if ε is restricted to the interval $(0, \min\{K_2 n^{-1}, K_3\})$. Thus, the parameter ε_1 of Proposition 4.5 may be taken equal to $\min\{K_2 n^{-1}, K_3\}$.

To complete the proof, we next explain how to instantiate the constants K_1 to K_3 . Let us start with K_2 . It is a lower bound on the absolute values of the differences between two distinct entries (with same index) of invariant probability measures associated with the transition matrices of Γ . These differences are of the form $|p_1/q_1 - p_2/q_2| \geq 1/(q_1 q_2)$, where p_1, p_2, q_1, q_2 are integers. By Lemma 4.10, we know that $q_1, q_2 \leq n^{n/2} D^{n^2}$, and so one can choose

$$K_2 = \frac{1}{n^n D^{2n^2}} .$$

Likewise, K_3 is a lower bound on the absolute values of the differences between two distinct scalar products $m \cdot r^{\sigma\tau}$. Let $m \in \mathcal{M}(P^{\sigma\tau})$. It follows from Lemma 4.10 that the i th entry of m can be written as $m_i = p_i/q$ where p_i is an integer and $q \leq n^{n/2} D^{n^2}$ is an integer independent of i . Since every entry of $r^{\sigma\tau}$ has a denominator at most D , it follows that $m \cdot r^{\sigma\tau}$ is a rational number with denominator at most $n^{n/2} D^{n^2} D^n$. Therefore, the difference between two distinct values of $m \cdot r^{\sigma\tau}$ is at least $K_3 = (n^n D^{2n^2} D^{2n})^{-1}$, and so

$$\min\{K_2 n^{-1}, K_3\} = \frac{1}{n^n D^{2n^2}} \min\left\{\frac{1}{n}, \frac{1}{D^{2n}}\right\} = \frac{1}{n^n D^{2n(n+1)}} .$$

Finally the constant K_1 is a lower bound for the positive entries of the restrictions m' of the invariant probability measures (m, m') of the transition matrices (4.2) arising in the game Γ_M . A direct application of Lemma 4.10 provides the following coarse bound:

$$K_1 = \frac{1}{(2n)^n D^{4n^2}} .$$

This bound can be improved by noting that the matrices (4.2) have a particular structure. Indeed, if (m, m') is an invariant probability measure of (4.2), then it satisfies (4.3), and so $2m'$ is an invariant probability measure of the matrix $QP^{\sigma\tau}$, the entries of which are entries of $P^{\sigma\tau}$ (since Q has only one nonzero entry in each row and this entry is equal to 1). Applying Lemma 4.10 to the matrix $QP^{\sigma\tau}$, we obtain the following lower bound for the positive entries of m' :

$$K_1 = \frac{1}{2n^{n/2} D^{n^2}} .$$

Hence,

$$M_0 \leq (1 + (3/2)D) 2n^{n/2} D^{n^2} \leq 4n^{n/2} D^{n^2+1} . \quad \square$$

An important special case to which the method of Theorem 4.12 can be applied concerns *deterministic* mean-payoff games [GKK88, ZP96]. The input of such games can be described, as in [AGG12], by means of two matrices $A, B \in (\mathbb{Z} \cup \{-\infty\})^{m \times n}$. The corresponding Shapley operator can be written as

$$(4.4) \quad T_i(x) = \min_{1 \leq j \leq m} (-A_{ji} + \max_{1 \leq k \leq n} (B_{jk} + x_k)) , \quad x \in \mathbb{R}^n , \quad 1 \leq i \leq n .$$

The corresponding game is played by moving a token on a graph in which n nodes, denoted by $1, \dots, n$, belong to player MIN, whereas m other nodes, denoted by $1', \dots, m'$, belong to player MAX. In state $i \in \{1, \dots, n\}$, player MIN can move the token to a state $j \in \{1', \dots, m'\}$ such that $A_{ji} \neq -\infty$, receiving A_{ji} . In state j ,

player MAX can move the token to a state $k \in \{1, \dots, n\}$ such that $B_{jk} \neq -\infty$, receiving B_{jk} . We assume that the matrix B has no identically infinite row, and that the matrix A has no identically infinite column, meaning that each player has at least one available action in each state. Then, the modified operator $T \circ R_M$ corresponds to the matrix B_M in which infinite entries of B are replaced by $-M$, and the operator $g + T \circ R_M$ arises by subtracting the constant g_i to every entry in the i th column of A .

Theorem 4.13. *Let T denote the Shapley operator (4.4) of a deterministic mean-payoff game, with integer payoffs bounded in absolute value by $D \geq 2$. Then, for*

$$M > 4nD \quad \text{and} \quad 0 < \varepsilon < 1/n^3 ,$$

the policy iteration Algorithm 1 applied to the operator $g_\varepsilon + T \circ R_M$ terminates, and we can compute the upper mean payoff of T from any policy σ of player MIN such that $\lambda(g_\varepsilon + T \circ R_M) = \lambda(g_\varepsilon + T^\sigma \circ R_M)$, in the sense that $\bar{\chi}(T) = \bar{\chi}(T^\sigma)$.

Proof. We adapt the proof of Theorem 4.12 to the case of deterministic transition matrices. In that special case, every invariant probability measure is uniform on its support, hence its positive entries are bounded below by $1/n$ if the state space has cardinality n . Since the constant K_1 is a lower bound for the positive entries of the invariant probability measures of the transition matrices in Γ_M , one can choose $K_1 = 1/(2n)$, and then

$$M_0 \leq (2n)(1 + (3/2)D) \leq 4nD .$$

The constant K_2 , which is a lower bound on the absolute values of the differences between two distinct entries of invariant probability measures of transition matrices in Γ , can be chosen as $K_2 = 1/n^2$. As for K_3 , it is a lower bound on the absolute values of the differences between two distinct values of $m \cdot r^{\sigma\tau}$. Since the payments are integers, every scalar product $m \cdot r^{\sigma\tau}$ is a rational number whose denominator divides the denominator of the positive entries of m which are themselves bounded by n . Hence, one can choose $K_3 = 1/n^2$, and the parameter ε must be lower than

$$\min\{K_2 n^{-1}, K_3\} = 1/n^3 . \quad \square$$

Remark 4.14. One step in Algorithm 1 consists in computing an eigenpair (λ^k, v^k) of the reduced Shapley operator T^{σ_k} obtained by fixing the strategy σ_k of player MIN. This is a simpler problem which can be solved by several known methods. We may apply, for instance, a similar policy iteration algorithm to T^{σ_k} , iterating this time in the space of policies τ of player MAX. In this way, for each choice of τ , we arrive at an operator of the form $T^{\sigma_k, \tau}(x) = g + Px$, where P is a stochastic matrix which cannot in general be assumed to be irreducible. However, for one-player problems, a classical version of policy iteration, the multichain policy iteration introduced by Howard [How60] and Denardo and Fox [DF68], does allow one to determine (λ^k, v^k) (without genericity conditions). Moreover, in the special case of deterministic games, the vector v^k is known to be a tropical eigenvector and λ^k is a tropical eigenvalue. The tropical eigenpair can be computed by direct combinatorial algorithms, see e.g. the discussion in [CTGG99].

Remark 4.15. Theorem 4.13 should be compared with the other known perturbation scheme, relying on vanishing discount. This method requires the computation of a

fixed point of the operator $x \mapsto T(\alpha x)$ for $0 < \alpha < 1$ sufficiently close to one. It is known that, for deterministic mean-payoff games, if the discount factor α is chosen so that

$$\alpha > 1 - \frac{1}{4(n+m)^3 D} ,$$

where D denotes the maximal absolute value of a finite entry A_{ij} or B_{ij} , then, the solution of the mean-payoff problem can be derived from the solution of the discounted problem, see [ZP96, Sec. 5]. The latter can be obtained by policy iteration (which terminates without any nondegeneracy conditions in the discounted case). Applying Algorithm 1 to the map $g_\varepsilon + T$ requires solving linear systems in which the matrix is independent of ε . If this is done by calling the Denardo-Fox algorithm to solve one-player problem (see previous remark), these systems are well conditioned. By comparison, vanishing discount requires the inversion of a matrix which becomes singular as $\alpha \rightarrow 1$. In particular, if policy iteration is interpreted in floating-point arithmetics, vanishing discount based perturbations may lead to numerical instabilities or to overflows, whereas the present additive perturbation scheme is insensitive to this pathology, because it only perturbs the right-hand sides of the linear systems to be solved.

Remark 4.16. The present approach allows one to compute the *upper mean payoff*, i.e., the maximum of the mean payoff over all initial states. This leads to no loss of expressivity since it follows from known reductions that this problem is polynomial time equivalent to solving a mean-payoff game in which the initial state is fixed: combine the results of Appendix C in the extended version of [AGS16] (especially Lemma C.2 and Corollary C.3) with the reductions in [AM09]. An alternative route to compute the mean payoff of a given initial state, avoiding the use of such reductions, would be to extend the present perturbation scheme to the “multichain” version of policy iteration, discussed in [CTG06, ACTDG12].

REFERENCES

- [ABG13] M. Akian, R. B. Bapat, and S. Gaubert, *Max-plus algebras*, Handbook of Linear Algebra, Second Edition (Leslie Hogben, ed.), Discrete Mathematics and its Applications, Chapman & Hall/CRC, Boca Raton, FL, 2013.
- [ACTDG12] M. Akian, J. Cochet-Terrasson, S. Detournay, and S. Gaubert, *Policy iteration algorithm for zero-sum multichain stochastic games with mean payoff and perfect information*, [arXiv:1208.0446](https://arxiv.org/abs/1208.0446), 2012.
- [AG03] M. Akian and S. Gaubert, *Spectral theorem for convex monotone homogeneous maps, and ergodic control*, Nonlinear Anal. **52** (2003), no. 2, 637–679.
- [AGG12] M. Akian, S. Gaubert, and A. Guterman, *Tropical polyhedra are equivalent to mean payoff games*, Internat. J. Algebra Comput. **22** (2012), no. 1, 1250001, 43.
- [AGH14] M. Akian, S. Gaubert, and A. Hochart, *Generic uniqueness of the bias vector of mean payoff zero-sum games*, 53rd IEEE Conference on Decision and Control (Los Angeles, California, USA), December 2014, pp. 1581–1587.
- [AGH15] ———, *Ergodicity conditions for zero-sum games*, Discrete Contin. Dyn. Syst. **35** (2015), no. 9, 3901–3931.
- [AGS16] X. Allamigeon, S. Gaubert, and M. Skomra, *Solving generic nonarchimedean semi-definite programs using stochastic game algorithms*, Proceedings of the 41st International Symposium on Symbolic and Algebraic Computation (ISSAC) (Wterlo, Ontario, Canada), July 2016, extended version: [arXiv:1603.06916](https://arxiv.org/abs/1603.06916), pp. 31–38.
- [AGW09] M. Akian, S. Gaubert, and C. Walsh, *The max-plus Martin boundary*, Doc. Math. **14** (2009), 195–240.

- [AM09] D. Andersson and P. B. Miltersen, *The complexity of solving stochastic games on graphs*, Algorithms and computation, Lecture Notes in Comput. Sci., vol. 5878, Springer, Berlin, 2009, pp. 112–121.
- [AT07] C. D. Aliprantis and R. Tourky, *Cones and duality*, Graduate Studies in Mathematics, vol. 84, American Mathematical Society, Providence, RI, 2007.
- [Bap98] R. B. Bapat, *A max version of the Perron-Frobenius theorem*, Proceedings of the Sixth Conference of the International Linear Algebra Society (Chemnitz, 1996), vol. 275/276, 1998, pp. 3–18.
- [Bat73] J. Bather, *Optimal decision procedures for finite Markov chains. Part II. Communicating systems*, Adv. Appl. Prob. **5** (1973), 521–540.
- [BCOQ92] F. L. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester, 1992.
- [BGV15] J. Bolte, S. Gaubert, and G. Vigeral, *Definable zero-sum stochastic games*, Math. Oper. Res. **40** (2015), no. 1, 171–191.
- [BK76] T. Bewley and E. Kohlberg, *The asymptotic theory of stochastic games*, Math. Oper. Res. **1** (1976), no. 3, 197–208.
- [BR14] M. Bourque and T. E. S. Raghavan, *Policy improvement for perfect information additive reward and additive transition stochastic games with discounted and average payoffs*, J. Dyn. Games **1** (2014), no. 3, 347–361.
- [Bru73] R. E. Bruck, *Properties of fixed-point sets of nonexpansive mappings in Banach spaces*, Trans. Amer. Math. Soc. **179** (1973), 251–262.
- [But10] P. Butkovič, *Max-linear systems: theory and algorithms*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London, 2010.
- [CG79] R. Cuninghame-Green, *Minimax algebra*, Lecture Notes in Economics and Mathematical Systems, vol. 166, Springer-Verlag, Berlin-New York, 1979.
- [Con92] A. Condon, *The complexity of stochastic games*, Inform. and Comput. **96** (1992), no. 2, 203–224.
- [CT80] M. G. Crandall and L. Tartar, *Some relations between nonexpansive and order preserving mappings*, Proc. Amer. Math. Soc. **78** (1980), no. 3, 385–390.
- [CTG06] J. Cochet-Terrasson and S. Gaubert, *A policy iteration algorithm for zero-sum stochastic games with mean payoff*, C. R. Math. Acad. Sci. Paris **343** (2006), no. 5, 377–382.
- [CTGG99] J. Cochet-Terrasson, S. Gaubert, and J. Gunawardena, *A constructive fixed point theorem for min-max functions*, Dynam. Stability Systems **14** (1999), no. 4, 407–433.
- [DF68] E. V. Denardo and B. L. Fox, *Multichain Markov renewal programs*, SIAM J. Appl. Math. **16** (1968), 468–487.
- [DLRS10] J. A. De Loera, J. Rambau, and F. Santos, *Triangulations*, Algorithms and Computation in Mathematics, vol. 25, Springer-Verlag, Berlin, 2010.
- [EM79] A. Ehrenfeucht and J. Mycielski, *Positional strategies for mean payoff games*, Internat. J. Game Theory **8** (1979), no. 2, 109–113. MR 83d:90246
- [Eve57] H. Everett, *Recursive games*, Contributions to the theory of games, vol. 3, Annals of Mathematics Studies, no. 39, Princeton University Press, Princeton, N. J., 1957, pp. 47–78.
- [Fat08] A. Fathi, *Weak KAM theorem in Lagrangian dynamics*, Tenth preliminary version. Available on line., 2008.
- [FR13] A. Figalli and L. Rifford, *Aubry sets, Hamilton-Jacobi equations, and the Mañé conjecture*, Geometric analysis, mathematical relativity, and nonlinear partial differential equations, Contemp. Math., vol. 599, Amer. Math. Soc., Providence, RI, 2013, pp. 83–104.
- [Fri11] O. Friedmann, *An exponential lower bound for the latest deterministic strategy iteration algorithms*, Log. Methods Comput. Sci. **7** (2011), no. 3, 3:19, 42.
- [GG04] S. Gaubert and J. Gunawardena, *The Perron-Frobenius theorem for homogeneous, monotone functions*, Trans. Amer. Math. Soc. **356** (2004), no. 12, 4931–4950 (electronic).

- [GKK88] V. A. Gurvich, A. V. Karzanov, and L. G. Khachiyan, *Cyclic games and finding minimax mean cycles in digraphs*, Zh. Vychisl. Mat. i Mat. Fiz. **28** (1988), no. 9, 1407–1417, 1439.
- [GM77] M. Gondran and M. Minoux, *Valeurs propres et vecteurs propres dans les dioïdes et leur interprétation en théorie des graphes*, Bull. Direction Études Recherches Sér. C Math. Informat. (1977), no. 2, i, 25–41.
- [HK66] A. J. Hoffman and R. M. Karp, *On nonterminating stochastic games*, Management Sci. **12** (1966), 359–370.
- [How60] R. A. Howard, *Dynamic programming and Markov processes*, The Technology Press of M.I.T., Cambridge, Mass.; John Wiley & Sons, Inc., New York-London, 1960.
- [KM97] V. N. Kolokoltsov and V. P. Maslov, *Idempotent analysis and its applications*, Mathematics and its Applications, vol. 401, Kluwer Academic Publishers Group, Dordrecht, 1997.
- [Koh74] E. Kohlberg, *Repeated games with absorbing states*, Ann. Statist. **2** (1974), 724–738.
- [Koh80] ———, *Invariant half-lines of nonexpansive piecewise-linear transformations*, Math. Oper. Res. **5** (1980), no. 3, 366–372.
- [KY92] L. A. Kontoror and S. Yu. Yakovenko, *Nonlinear semigroups and infinite horizon optimization*, Idempotent analysis, Adv. Soviet Math., vol. 13, Amer. Math. Soc., Providence, RI, 1992, pp. 167–210.
- [LL69] T. M. Liggett and S. A. Lippman, *Stochastic games with perfect information and time average payoff*, SIAM Rev. **11** (1969), 604–607.
- [LR17] Rida Laraki and Jérôme Renault, *Acyclic gambling games*, [arXiv:1702.06866](https://arxiv.org/abs/1702.06866), 2017.
- [MN81] J.-F. Mertens and A. Neyman, *Stochastic games*, Internat. J. Game Theory **10** (1981), no. 2, 53–66.
- [MPN02] J. Mallet-Paret and R. D. Nussbaum, *Eigenvalues for a class of homogeneous cone maps arising from max-plus operators*, Discrete Contin. Dyn. Syst. **8** (2002), no. 3, 519–562.
- [NS03] A. Neyman and S. Sorin, *Stochastic games and applications*, Nato Science Series C, vol. 570, Springer Netherlands, 2003.
- [Ovc02] S. Ovchinnikov, *Max-min representation of piecewise linear functions*, Beiträge Algebra Geom. **43** (2002), no. 1, 297–302.
- [Pur95] A. Puri, *Theory of hybrid systems and discrete event systems*, Phd thesis, 1995.
- [Ren11] J. Renault, *Uniform value in dynamic programming*, J. Eur. Math. Soc. (JEMS) **13** (2011), no. 2, 309–330.
- [Ren12] Jérôme Renault, *The value of repeated games with an informed controller*, Math. Oper. Res. **37** (2012), no. 1, 154–179. MR 2891152
- [Rom67] I. V. Romanovsky, *Optimization of stationary control of a discrete deterministic process*, Cybernetics **3** (1967), no. 2, 52–62 (English).
- [RS01] D. Rosenberg and S. Sorin, *An operator approach to zero-sum repeated games*, Israel J. Math. **121** (2001), 221–246.
- [Sha53] L.S. Shapley, *Stochastic games*, Proc. Natl. Acad. Sci. U. S. A. **39** (1953), 1095–1100.
- [Sor02] S. Sorin, *A first course on zero-sum repeated games*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 37, Springer-Verlag, Berlin, 2002. MR 1890574
- [Sor04] ———, *Asymptotic properties of monotonic nonexpansive mappings*, Discrete Event Dyn. Syst. **14** (2004), no. 1, 109–122.
- [ST13] B. Sturmfels and N. M. Tran, *Combinatorial types of tropical eigenvectors*, Bull. Lond. Math. Soc. **45** (2013), no. 1, 27–36.
- [Vig13] G. Vigeral, *A zero-sum stochastic game with compact action sets and no asymptotic value*, Dyn. Games Appl. **3** (2013), no. 2, 172–186.
- [Zil16] B. Ziliotto, *A tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games*, Mathematics of Operations Research **41** (2016), no. 4, 1522–1534.
- [ZP96] U. Zwick and M. Paterson, *The complexity of mean payoff games on graphs*, Theoret. Comput. Sci. **158** (1996), no. 1-2, 343–359.

CMAF, INRIA, ECOLE POLYTECHNIQUE, CNRS, UNIVERSITÉ PARIS-SACLAY, 91128 PALAISEAU, FRANCE

E-mail address: `marianne.akian@inria.fr`

E-mail address: `stephane.gaubert@inria.fr`

TOULOUSE SCHOOL OF ECONOMICS, UNIVERSITÉ TOULOUSE I, 31015 TOULOUSE, FRANCE

E-mail address: `antoine.hochart@polytechnique.edu`