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

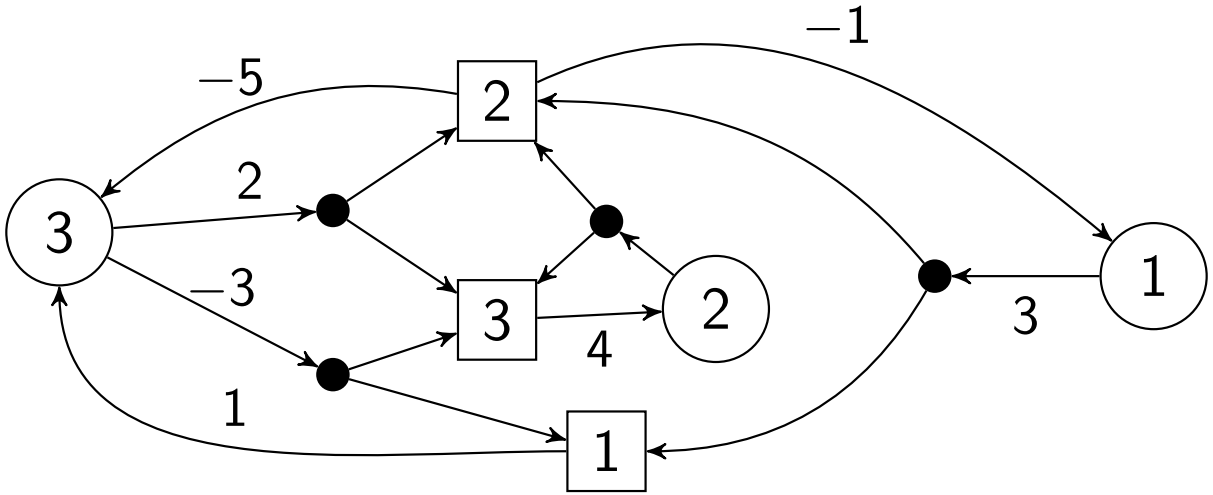
Xavier Allamigeon,[1] Stéphane Gaubert,[1] Ricardo D. Katz,[2]
Mateusz Skomra[3]

[1]INRIA and CMAP, École polytechnique, IP Paris, CNRS, France
[2]CONICET-CIFASIS, Argentina
[3]LAAS-CNRS, Université de Toulouse, CNRS, France

19th October 2022

Section I: Stochastic mean payoff games and entropy games

Stochastic mean payoff games (SMPGs)

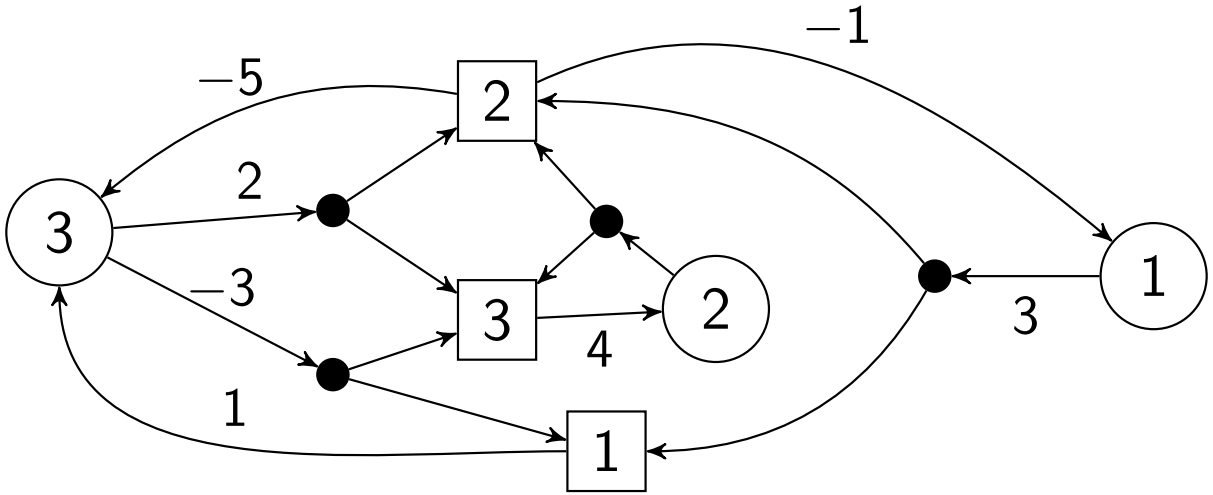


There are two players, Max and Min, who move a pawn on a directed graph $\mathcal{G} = (V, E)$ as above. Min decides the move at square states (here $V_{\mathrm{Min}} = \{\boxed{1}, \boxed{2}, \boxed{3}\}$) and Max decides the move at circle states (here $V_{\mathrm{Max}} = \{\textcircled{1}, \textcircled{2}, \textcircled{3}\}$).

The states marked with dots, denoted V_{Nat}, are controlled by Nature, who is not a rational player.

In our setting, the graph is tripartite: players move in order Min $\rightarrow$ Max $\rightarrow$ Nature $\rightarrow$ Min $\rightarrow$ Max $\rightarrow$ Nature $\rightarrow$ ...

Stochastic mean payoff games (SMPGs)

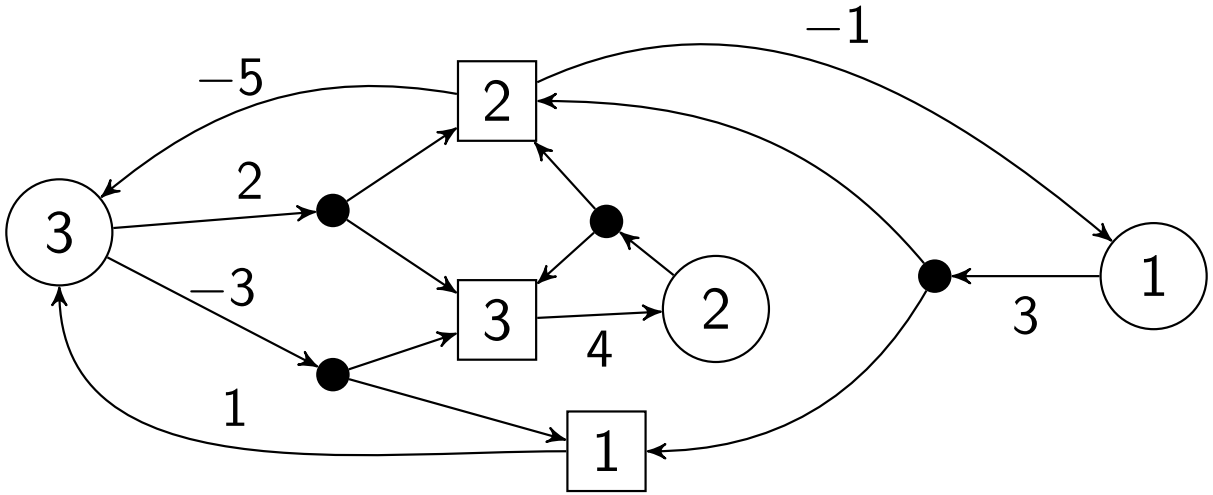


A **strategy** of player Min is a function that, to any finite history of the play (path taken by the pawn) that ends at a node controlled by Min, associates the next state. Analogously for player Max.

In a SMPG, we suppose that Nature makes decision at random, according to a (fixed) probability distribution at any given state. (In the figure: take $(1/2, 1/2)$ at any state.)

The numbers on the edges indicate the amount of money that Min pays to Max when the pawn goes through a given edge. In our setting: rewards are integer.

Stochastic mean payoff games (SMPGs)



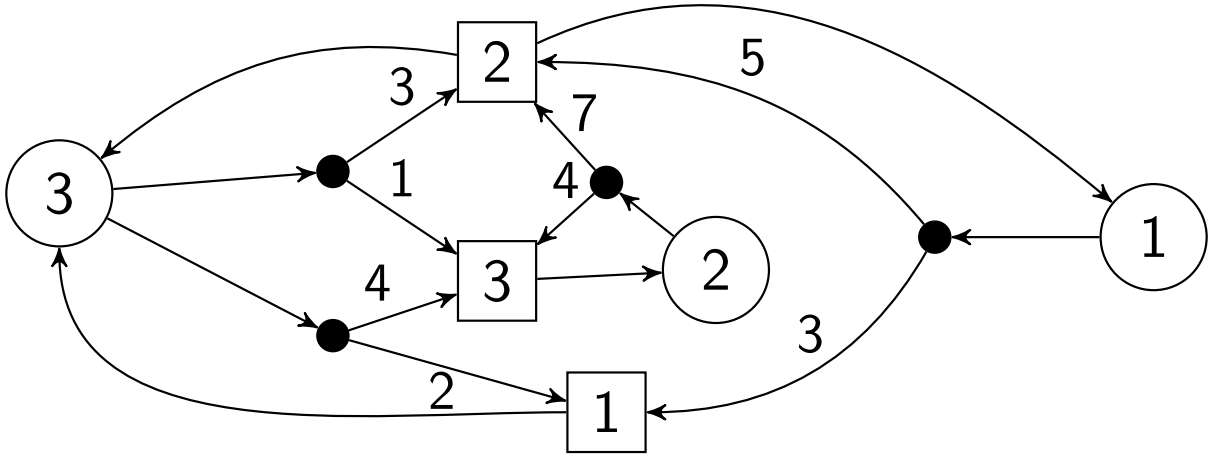
We suppose that the game is played for *infinite* time. In this case, the sum of payoffs may be undefined.

We consider the **mean-payoff** criterion: Max wants to maximize the quantity

$$g_{i_0}(\sigma,\tau) := \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}_{\sigma,\tau}(r_{i_0 i_1} + \cdots + r_{i_{N-1} i_N}),$$

where (σ,τ) is a pair of strategies (giving a probability distribution on possible paths of length N), r_{kl} are the rewards, and $i_0 \in V_{\mathrm{Min}}$ is the initial state.

Entropy games (EG)

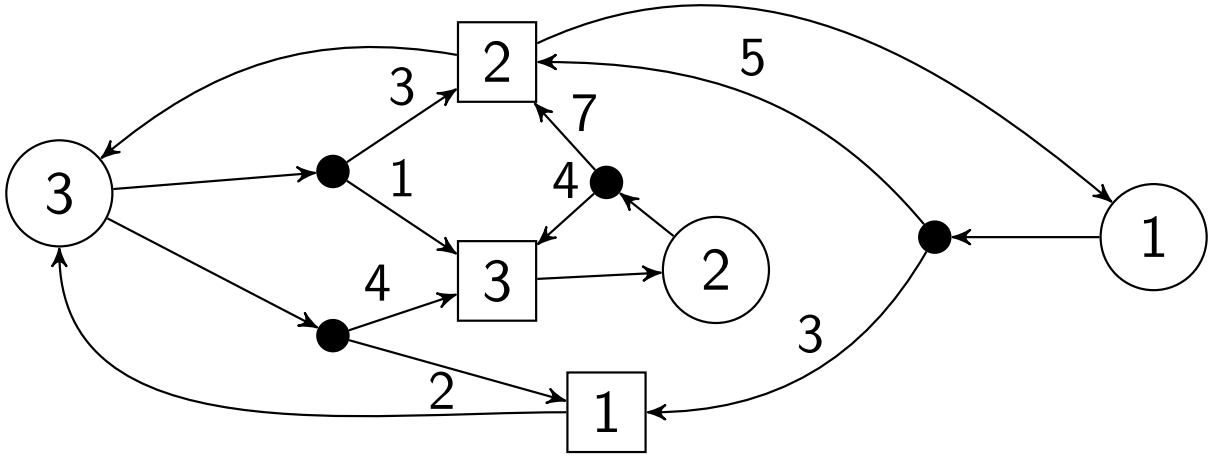


In an entropy game, Min is called "Despot", Max is called "Tribune", and Nature is called "People".

Nature is nondeterministic – we do not assume anything about its behavior.

Each edge controlled by People is equipped with a positive integer called *multiplicity*. There are no weights on other edges.

Entropy games (EG)



Given (σ, τ) we denote by $R_{i_0}^N(\sigma, \tau)$ the number of paths (counting multiplicities) that the pawn can make in horizon N, starting from $i_0 \in V_{\mathrm{Min}}$.

Tribune wants to maximize

$$g_{i_0}(\sigma, \tau) := \limsup_{N \to \infty} (R_{i_0}^N(\sigma, \tau))^{1/N} .$$

The logarithm of this quantity is the **topological entropy**.

Existence of optimal strategies

Theorem (Liggett and Lippman, 1969; Akian, Gaubert, Grand-Clément, et al., 2019)

For every SMPG and every EG, there exists a couple of strategies $(\bar{\sigma}, \bar{\tau})$ *and a vector* $\eta \in \mathbb{R}^{V_{\mathrm{Min}}}$ *such that the inequality*

$$g_i(\bar{\sigma}, \tau) \leq \eta_i \leq g_i(\sigma, \bar{\tau})$$

is true for all states $i \in V_{\mathrm{Min}}$ *and all policies* (σ, τ)*. The vector* η *is called the* ***value*** *of the game, and the strategies* $(\bar{\sigma}, \bar{\tau})$ *are called* ***optimal****. The value is unique.*

Furthermore, there exist optimal strategies which are **memoryless**, i.e., they depend only on the current position of the pawn. Memoryless strategies are called **policies**.

Complexity questions

Theorem (Condon, 1992; Asarin et al., 2016)

Given a SMPG or an EG and a number $\alpha \in \mathbb{Q}$, the problem of deciding if $\max_i \eta_i < \alpha$ belongs to **NP** $\cap$ **coNP**.

- Neither problem is known to be in **P**.
- For SMPG: open for >30 years, generalizes **parity games**.
- Closely related: find optimal policies of both players.
- For EG: we do not know how to solve Tribune-free games.

Theorem (Gimbert and Horn, 2008)

Simple stochastic games (subclass of SMPG) with fixed number of significant Nature states can be solved in polynomial time.

Theorem (Akian, Gaubert, Grand-Clément, et al., 2019)

EGs with fixed number of significant Despot states can be solved in polynomial time.

"Significant" = at least two outgoing edges.

Our contribution

Theorem

For a SMPG, we can find its top class (all states of maximal value) and a pair of optimal policies on top class in $O(|V|^4|E|WM^{3K})$ complexity.

Notation: K is the number of significant Nature states, M is the common denominator of all the probabilities, and W is the highest absolute value of any reward.

Note that the bound is pseudopolynomial for fixed K.

Theorem

EGs with fixed number of People states can be solved in pseudopolynomial time.

To do so, we develop a unified approach to bound complexity of **value iteration**, which is (arguably) the simplest nontrivial algorithm for solving these games.

Section II: Shapley operators and value iteration

Shapley operators of SMPGs

Consider the **Shapley operator** $F : \mathbb{R}^{V_{\mathrm{Min}}} \to \mathbb{R}^{V_{\mathrm{Min}}}$:

$$(F(x))_i := \min_{(i,s)\in E}\Big(r_{is} + \max_{(s,l)\in E}\big(r_{sl} + \sum_{j\in V_{\mathrm{Min}}} p_{lj}x_j\big)\Big),$$

where r_{is}, r_{sl} are the payoffs obtained by going from $i \in V_{\mathrm{Min}}$ to $s \in V_{\mathrm{Max}}$ and from s to $l \in V_{\mathrm{Nat}}$, and p_{lj} is the probability of going from l to $j \in V_{\mathrm{Min}}$.

Lemma (folklore)

$F^N(0)$ is the value vector of a SMPG that lasts N turns.

Theorem (Mertens and Neyman, 1981)

*The value is equal to the **escape rate** $\chi(F) = \lim_N F^N(0)/N$.*

Observation

F is monotone, $x \le y \implies F(x) \le F(y)$;
F is additively homogenous, $F(\lambda + x) = \lambda + F(x)$ for $\lambda \in \mathbb{R}$.

Shapley operators of EGs

Consider the operator $T : \mathbb{R}_{>0}^{V_{\mathrm{Min}}} \to \mathbb{R}_{>0}^{V_{\mathrm{Min}}}$

$$(T(x))_i := \min_{(i,s)\in E} \max_{(s,l)\in E} \sum_{j\in V_{\mathrm{Min}}} m_{lj} x_j$$

and let $F(x) := \log \circ T \circ \exp$ (coordinatewise log and exp).

Theorem (Akian, Gaubert, Grand-Clément, et al., 2019)

$T^N(1,\dots,1)$ is the value of an EG that lasts N turns. Moreover, the escape rate of F exists and is equal to the logarithm of the value vector.

Observation

F is monotone and additively homogenous.

Lemma (Akian, Gaubert, Grand-Clément, et al., 2019)

Given any $\epsilon > 0$ we can build an approximation oracle for F, $\|F(x) - \tilde{F}(x)\|_\infty \leq \epsilon$.

Collatz–Wielandt certificates

Our approach is based on **Collatz–Wielandt certificates**.
Suppose that $u \in \mathbb{R}^V_{\mathrm{Min}}$ satisfies $\lambda + u \leq F(u)$ for $\lambda \in \mathbb{R}$. Then,

$$N\lambda + u \leq F^N(u) \quad \text{for all } N \geq 1$$

and so $\min \chi_i \geq \lambda$. Analogously, $\lambda + u \geq F(u)$ gives $\max \chi_i \leq \lambda$.

Moreover, such certificates exist:

Theorem (Akian, Gaubert, and Guterman, 2012)

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a monotone and additively homogeneous operator that has an escape rate χ. Moreover, suppose that χ_i does not depend on the choice of i. Then,

$$\begin{aligned}\chi_1 &= \sup\{\lambda \in \mathbb{R} : \exists u \in \mathbb{R}^n, \lambda + u \leq F(u)\} \\ &= \inf\{\lambda \in \mathbb{R} : \exists u \in \mathbb{R}^n, \lambda + u \geq F(u)\}\,.\end{aligned}$$

Condition number

We introduce a condition number based on Collatz–Wielandt certificates.

For $x \in \mathbb{R}^n$ denote $\mathbf{t}(x) := \max_i x_i$, $\mathbf{b}(x) := \min_i x_i$, and $\|x\|_{\mathrm{H}} := \mathbf{t}(x) - \mathbf{b}(x)$ (*Hilbert seminorm*).

Definition (Condition number)

For any $\delta > 0$ we put $R_\delta := \max\{R_\delta^+, R_\delta^-\}$, where

$$R_\delta^+ := \inf\{\|u\|_{\mathrm{H}} \colon F(u) \leq \chi_1 + \delta + u\}\,,$$
$$R_\delta^- := \inf\{\|u\|_{\mathrm{H}} \colon \chi_1 - \delta + u \leq F(u)\}\,.$$

Approximating constant value

We have the following approximation algorithm for any $\delta > 0$.

Algorithm to approximate a constant value

Suppose that the value χ is constant. Let $\tilde{F}$ be such that $\|F(x) - \tilde{F}(x)\|_\infty \leq \delta/8$. Compute $\tilde{F}(0), \tilde{F}^2(0), \ldots, \tilde{F}^N(0)$ until

$$\frac{\mathbf{t}(\tilde{F}^N(0)) - \mathbf{b}(\tilde{F}^N(0))}{N} \leq (3/4)\delta\,.$$

Then, the value belongs to the interval

$$[-\delta/8 + \mathbf{b}(\tilde{F}^N(0))/N, \delta/8 + \mathbf{t}(\tilde{F}^N(0))/N]\,,$$

which is of width at most δ.

Theorem

The algorithm is correct and stops in at most $\lceil 8R_{\delta/8}/\delta \rceil$ *iterations.*

Deciding if value is constant

Algorithm to detect if value is constant

Let $\tilde{F}$ be such that $\|F(x) - \tilde{F}(x)\|_\infty \leq \delta/8$. Compute $\tilde{F}(0), \tilde{F}^2(0), \ldots, \tilde{F}^N(0)$ until

$$\frac{\mathbf{t}(\tilde{F}^N(0)) - \mathbf{b}(\tilde{F}^N(0))}{N} \leq (3/4)\delta$$

or $N = 1 + \lceil 8R_{\delta/8}/\delta \rceil$. In the first case, declare that the value is constant. Otherwise, take any i such that $\tilde{F}^N(0)_i = \mathbf{b}(\tilde{F}^N(0))$ and declare that this state does not belong to the top class, $\chi_i < \mathbf{t}(\chi)$.

Theorem

The algorithm above is correct as soon as

- $\mathbf{t}(\chi) = \mathbf{b}(\chi)$ *or* $\mathbf{t}(\chi) - \mathbf{b}(\chi) > \delta$,
- $R_{\delta/8}$ *is the condition number of the operator on top class,*
- F *satisfies an additional technical assumption (which is true for SMPGs and EGs).*

Finding top class

Given i such that $\chi_i < \mathbf{t}(\chi)$, we can "extend" it to a group of states that does not belong to the top class. By repeatedly removing such groups from the graph, we can find the top class.

Theorem

Let $\delta > 0$ be a number such that $\mathbf{t}(\chi^\Gamma) - \mathbf{b}(\chi^\Gamma) > \delta$ for all subgames Γ that strictly contain the top class. Then, the top class can be found by making at most $n^2 + n\lceil 8R_\epsilon/\delta \rceil$ calls to an oracle that approximates F to precision $\epsilon := \delta/8$.

For EGs, we can go even further.

Algorithm for solving entropy games

Find the top class. Remove it from the graph and repeat on the smaller game.

Note: removing the top class does **not** work for SMPGs.

Complexity estimates for SMPGs

Notation: K is the number of significant Nature states, M is the common denominator of all the probabilities, and W is the highest absolute value of any reward.

Proposition

For a constant value game, the denominator of $\chi \in \mathbb{Q}$ is at most $|V|M^K$ and $\|R_\delta\|_{\mathrm{H}} \leq 4|V|WM^K$ for any $\delta > 0$.

Theorem

We can find the top class and a pair of optimal policies on top class in $O(|V|^4|E|WM^{3K})$ complexity.

An algorithm of Boros et al. (2019) solves the same problem in $O(|V|^6|E|WK2^K M^{4K} + |V|^3|E|W \log W)$ complexity.

They also show how an oracle to top class + an oracle to solving deterministic MPGs can be used to solve general SMPGs in $\mathrm{poly}(|V|, \log W)W(|V|KM)^{O(K)}$ complexity.

Complexity estimates of EGs

Notation: P is the number of People states, W is the highest multiplicity of an edge.

Lemma

Every e^{χ_i} is an algebraic number of degree at most P. In particular, if $\mathbf{t}(\chi) \neq \mathbf{b}(\chi)$, then $\mathbf{t}(\chi) - \mathbf{b}(\chi) > \mathrm{poly}(|V|, W)^{-P^2}$ for some polynomial poly.

Lemma

We have $\|R_\delta\|_{\mathrm{H}} \leq 1200|V|^2(|V| \log W - \log \delta)$ for all $0 < \delta < 1$.

Theorem

We can solve EGs in $\mathrm{poly}(|V|, |E|, W)^{P^2}$ complexity.

Here, "solve" means find optimal policies of both players and express the value at each state as a unique root of a univariate polynomial that belongs to some interval.

Questions for further study

- Can Tribune-free entropy games be solved in (pseudo)polynomial time? What about games with limited number of *significant* People states?
- What can we get by applying this approach to SMPGs with imperfect information or other mean-payoff games with more complicated strategy sets?
- Are there policy iteration algorithms that achieve similar bounds for EGs and SPMGs? (We have such algorithms for SSGs.)
- How to exploit the properties of the graph to get better algorithms?
- Can we simplify the approach of Boros et al. (2019) showing that general stochastic mean payoff games are solvable in pseudopolynomial time when the number of significant Nature states is fixed?

Thank you for your attention

X. Allamigeon et al. “Universal Complexity Bounds Based on Value Iteration and Application to Entropy Games”. In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, 110:1–110:20

References I

[AGG12] M. Akian, S. Gaubert, and A. Guterman. "Tropical polyhedra are equivalent to mean payoff games". In: *Int. J. Algebra Comput.* 22.1 (2012), 125001 (43 pages).

[Aki+19] M. Akian, S. Gaubert, J. Grand-Clément, and J. Guillaud. "The Operator Approach to Entropy Games". In: *Theor. Comp. Sys.* 63.5 (July 2019), pp. 1089–1130. ISSN: 1432-4350.

[All+] X. Allamigeon, S. Gaubert, R. D. Katz, and M. Skomra. "Universal Complexity Bounds Based on Value Iteration and Application to Entropy Games". In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, 110:1–110:20.

References II

[Asa+16] E. Asarin, J. Cervelle, A. Degorre, C. Dima, F. Horn, and V. Kozyakin. "Entropy Games and Matrix Multiplication Games". In: *Proceedings of the 33rd International Symposium on Theoretical Aspects of Computer Science (STACS)*. Vol. 47. LIPIcs. Leibniz Int. Proc. Inform. Wadern: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016, 11:1–11:14.

[Bor+19] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. "A Pseudo-Polynomial Algorithm for Mean Payoff Stochastic Games with Perfect Information and Few Random Positions". In: *Inform. and Comput.* 267 (2019), pp. 74–95.

[Con92] A. Condon. "The Complexity of Stochastic Games". In: *Inform. and Comput.* 96.2 (1992), pp. 203–224.

References III

[GH08] H. Gimbert and F. Horn. "Simple stochastic games with few random vertices are easy to solve". In: *Proceedings of the 11th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS)*. Vol. 4962. Lecture Notes in Comput. Sci. Springer, 2008, pp. 5–19.

[LL69] T. M. Liggett and S. A. Lippman. "Stochastic Games with Perfect Information and Time Average Payoff". In: *SIAM Rev.* 11.4 (1969), pp. 604–607.

[MN81] J.-F. Mertens and A. Neyman. "Stochastic games". In: *Internat. J. Game Theory* 10.2 (1981), pp. 53–66.