Jeux stochastiques à somme nulle: ergodicité, complexité et théorie de Perron-Frobenius non linéaire

Marianne Akian

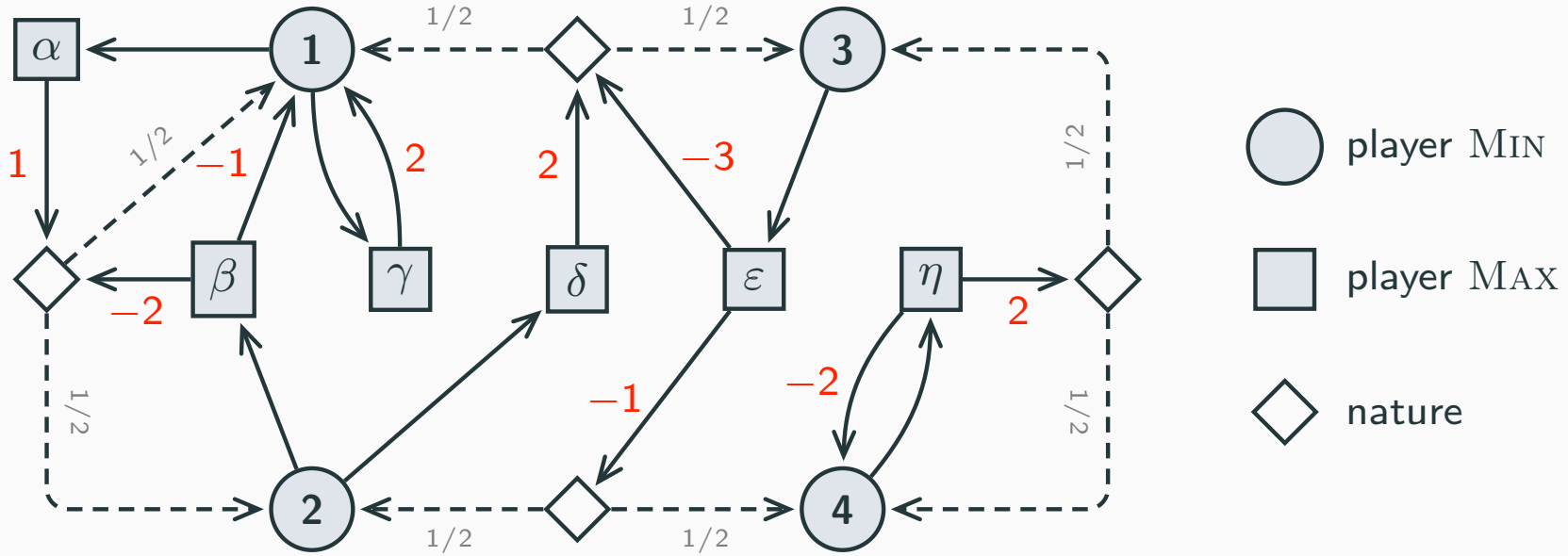
INRIA et CMAP, École polytechnique, CNRS, IP Paris

2ème Journée MAS-MODE 2022

INRIA Paris, 7 mars 2022

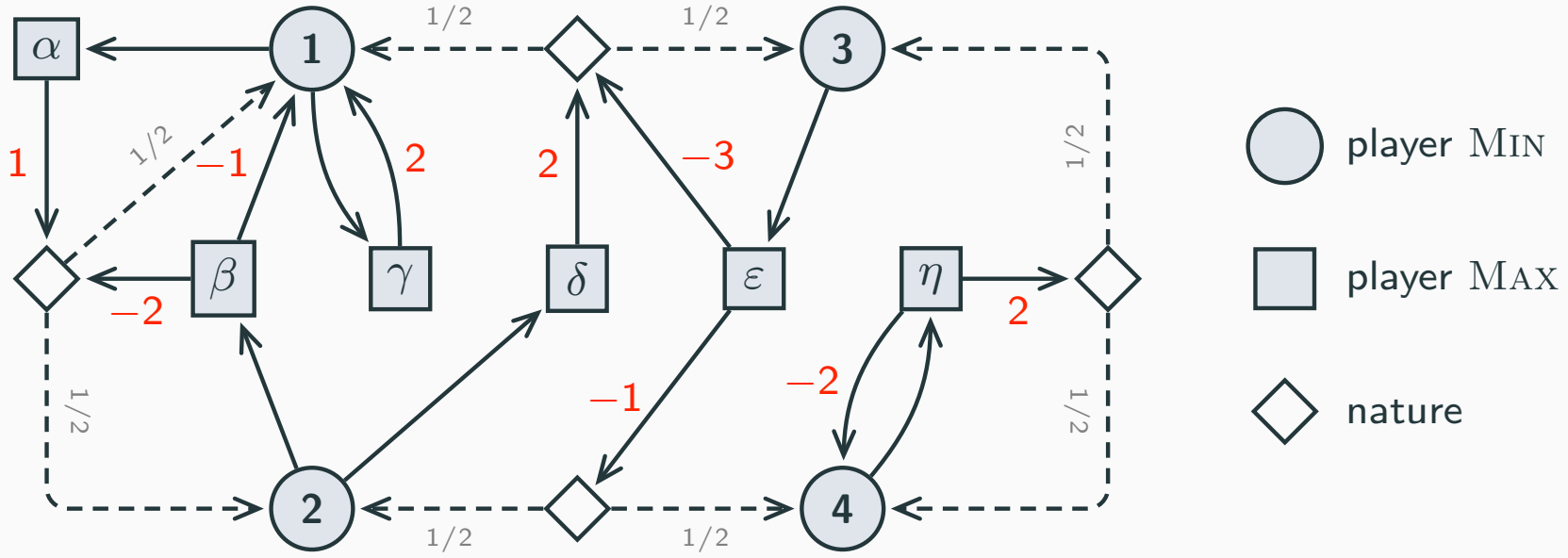
Travaux communs avec Stéphane Gaubert, Antoine Hochart, Omar Saadi, Zheng Qu, Julien Grand Clément, Jérémie Guillaud

Discrete time and state zero-sum stochastic games with perfect information



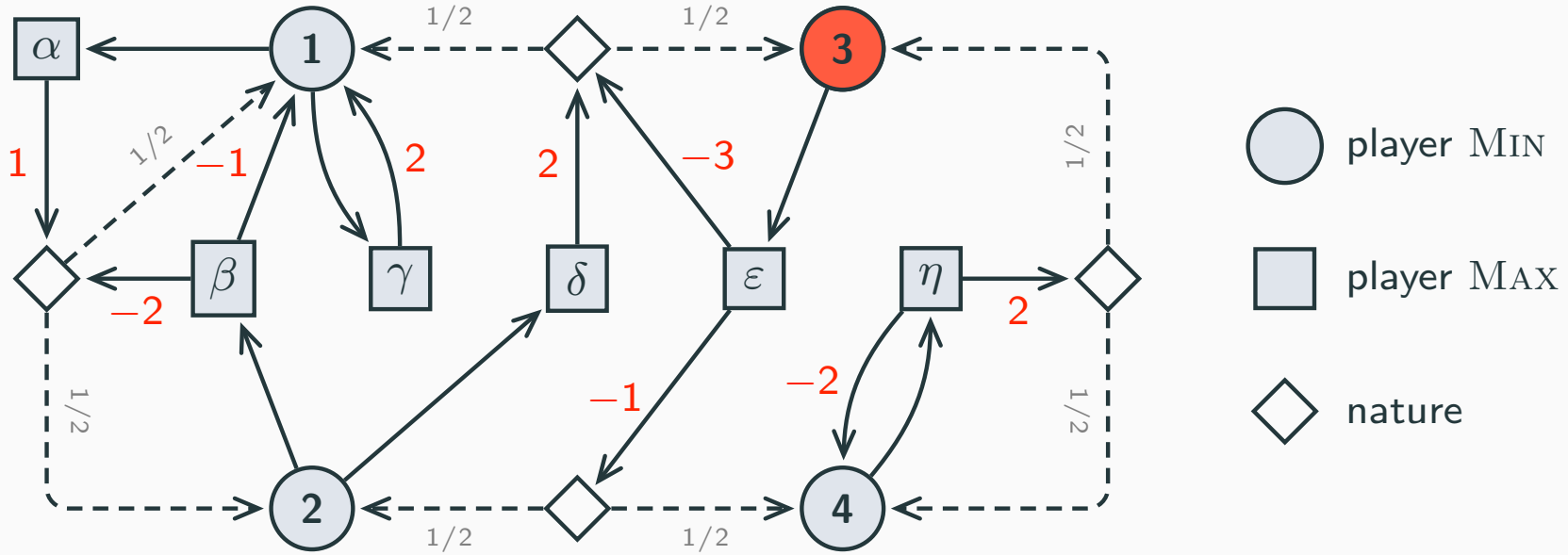
- Finite state space: $[n] = \{1, \ldots, n\}$.
- Finite action spaces: $\mathcal{A}_i$ (player MIN), $\mathcal{B}_i$ (player MAX)
- Transition payment from player MIN to player MAX: $r_i^{ab} \in \mathbb{R}$
- Transition probability: $P_i^{ab} = (P_{ij}^{ab})_{1 \leqslant j \leqslant n} \in \Delta([n])$.

Discrete time and state zero-sum stochastic games with perfect information



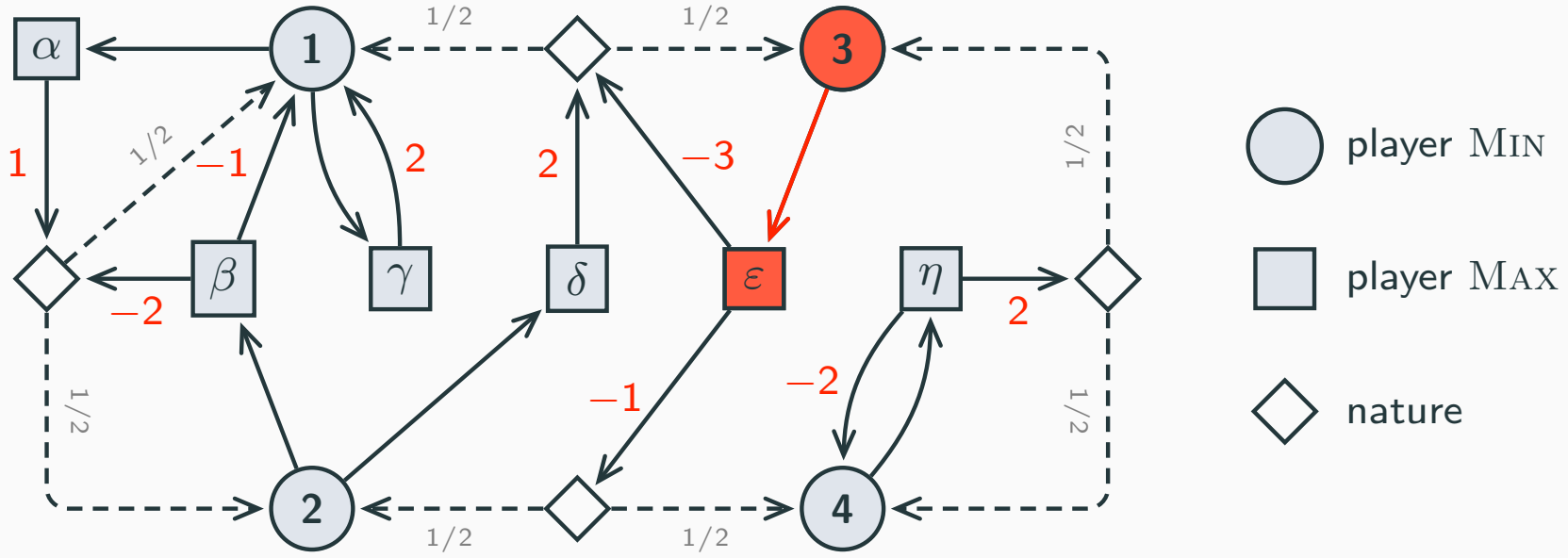
- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.

Discrete time and state zero-sum stochastic games with perfect information



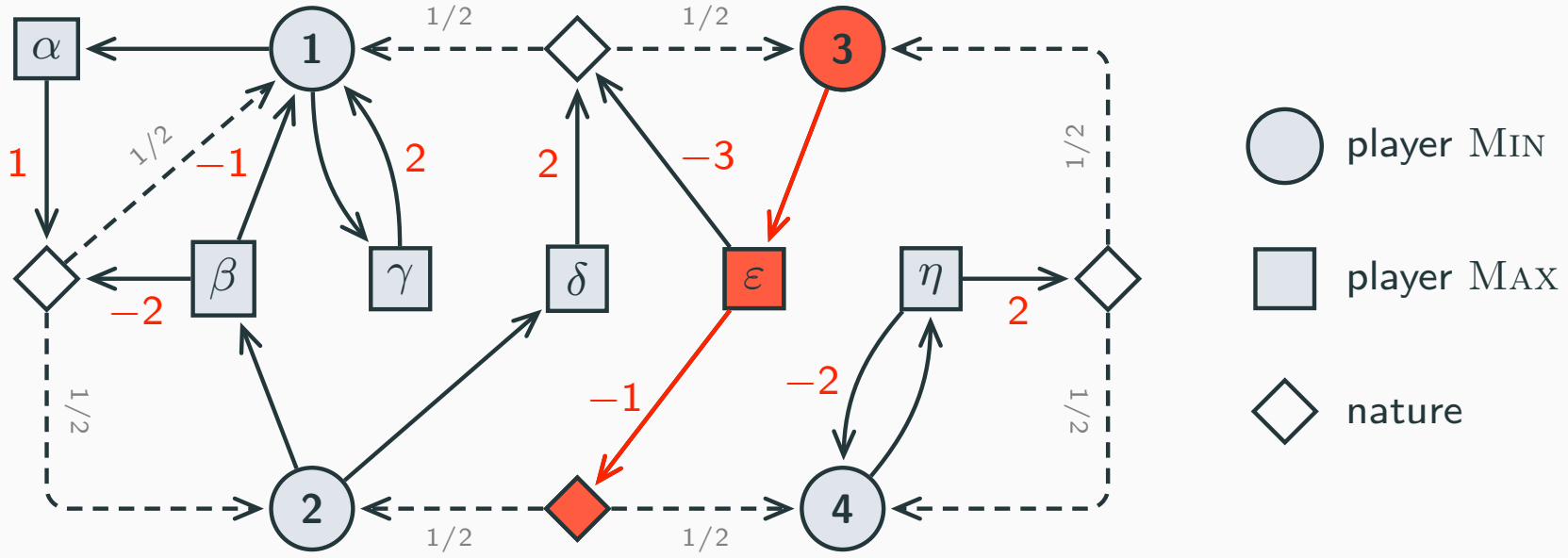
- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.

Discrete time and state zero-sum stochastic games with perfect information



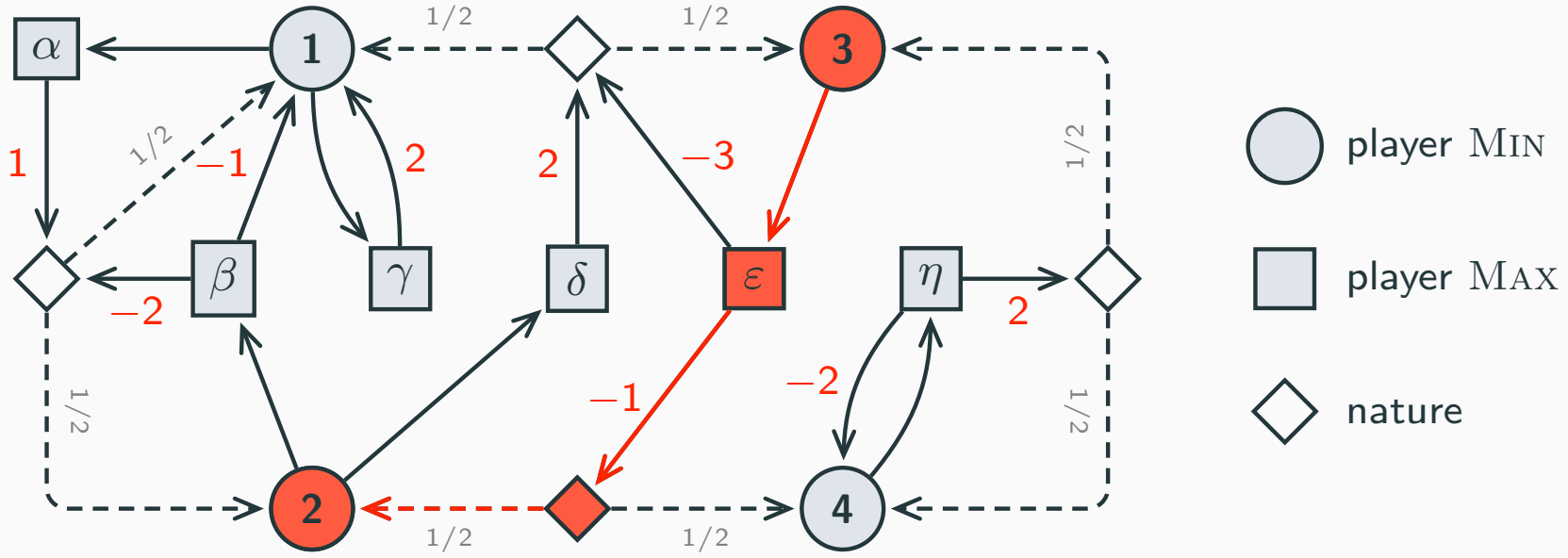
- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.

Discrete time and state zero-sum stochastic games with perfect information

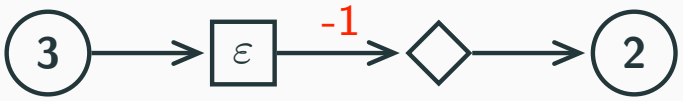


- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.

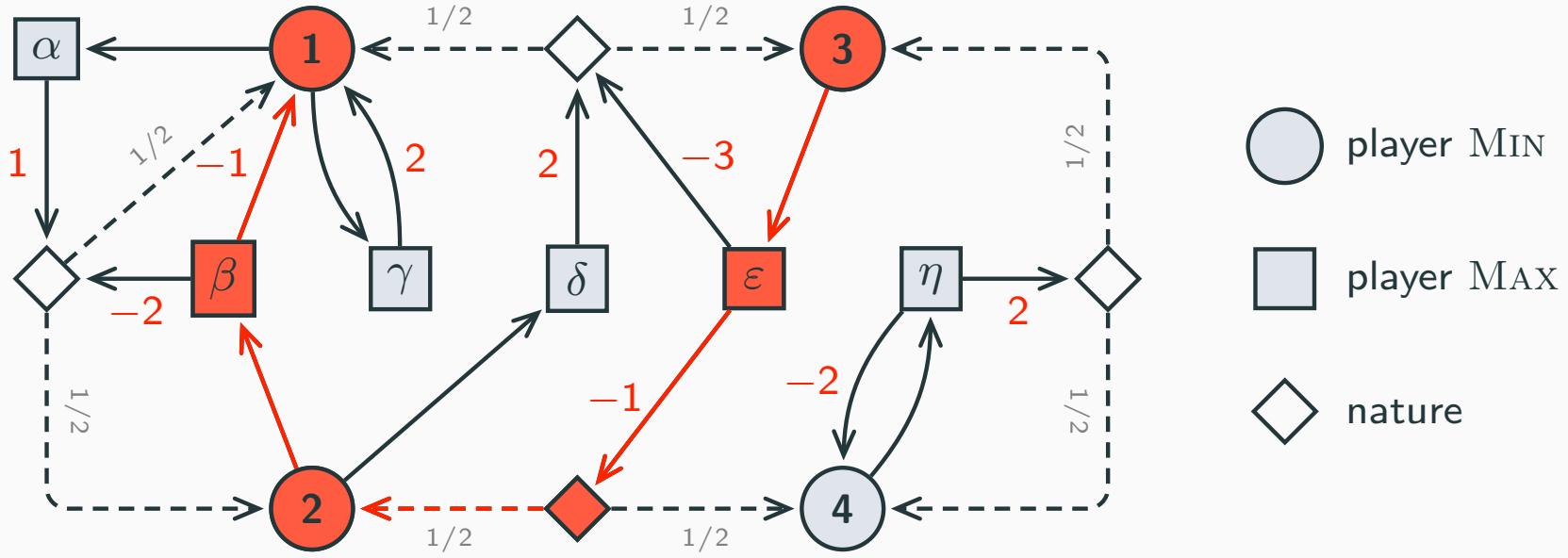
Discrete time and state zero-sum stochastic games with perfect information



- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.



Discrete time and state zero-sum stochastic games with perfect information



- A play: $(i_\ell, a_\ell, b_\ell)_{\ell\in\mathbb{N}}$ with $i_\ell \in [n]$, $a_\ell \in \mathcal{A}_{i_\ell}$, $b_\ell \in \mathcal{B}_{i_\ell}$ following some strategies σ and π of the two players.

Discrete time and state zero-sum stochastic games with perfect information

- 0-player (all $\mathcal{A}_i$ and $\mathcal{B}_i$ are reduced to one element): Markov chains.
- 1-player case (all $\mathcal{A}_i$, or all $\mathcal{B}_i$, are reduced to one element): Markov Decision Processes or a discrete time and state stochastic optimal control problems.

Discrete time and state zero-sum stochastic games with perfect information

- 0-player (all $\mathcal{A}_i$ and $\mathcal{B}_i$ are reduced to one element): Markov chains.
- 1-player case (all $\mathcal{A}_i$, or all $\mathcal{B}_i$, are reduced to one element): Markov Decision Processes or a discrete time and state stochastic optimal control problems.
- A *strategy* σ of Player $\textsc{Min}$ is a map which assigns to every history $(i_0, a_0, b_0, \ldots, b_{k-1}, i_k)$ an action $a_k \in \mathcal{A}_{i_k}$.
- It is *positional* if it only depends on the last state i_k of the history. In that case it is also called a *policy*.

Discrete time and state zero-sum stochastic games with perfect information

- 0-player (all $\mathcal{A}_i$ and $\mathcal{B}_i$ are reduced to one element): Markov chains.
- 1-player case (all $\mathcal{A}_i$, or all $\mathcal{B}_i$, are reduced to one element): Markov Decision Processes or a discrete time and state stochastic optimal control problems.
- A *strategy* σ of Player $\textsc{Min}$ is a map which assigns to every history $(i_0, a_0, b_0, \ldots, b_{k-1}, i_k)$ an action $a_k \in \mathcal{A}_{i_k}$.
- It is *positional* if it only depends on the last state i_k of the history. In that case it is also called a *policy*.
- A *strategy* π) of Player $\textsc{Max}$ is a map which assigns to every history $(i_0, a_0, b_0, \ldots, b_{k-1}, i_k, a_k)$ an action $b_k \in \mathcal{B}_{i_k}$.
- It is *positional* if it only depends on the last state and action (i_k, a_k) of the history. In that case it is also called a *policy*.

Discrete time and state zero-sum stochastic games with perfect information

- 0-player (all $\mathcal{A}_i$ and $\mathcal{B}_i$ are reduced to one element): Markov chains.
- 1-player case (all $\mathcal{A}_i$, or all $\mathcal{B}_i$, are reduced to one element): Markov Decision Processes or a discrete time and state stochastic optimal control problems.
- A *strategy* σ of Player $\textsc{Min}$ is a map which assigns to every history $(i_0, a_0, b_0, \ldots, b_{k-1}, i_k)$ an action $a_k \in \mathcal{A}_{i_k}$.
- It is *positional* if it only depends on the last state i_k of the history. In that case it is also called a *policy*.
- A *strategy* π) of Player $\textsc{Max}$ is a map which assigns to every history $(i_0, a_0, b_0, \ldots, b_{k-1}, i_k, a_k)$ an action $b_k \in \mathcal{B}_{i_k}$.
- It is *positional* if it only depends on the last state and action (i_k, a_k) of the history. In that case it is also called a *policy*.
- Under positional strategies, the sequence of states $(i_\ell)_{\ell \geqslant 0}$ is a Markov chain.

The value of the game and the Dynamic Programming operator

- **Payoff** of the k-stage game with initial state i:
$$J_i^k(\sigma,\pi) = \mathbb{E}_{i,\sigma,\pi}\left[\sum_{\ell=0}^{k-1} r_{i_\ell}^{a_\ell b_\ell}\right].$$
- **Value** of the k-stage game with initial state i:
$$v_i^k = \min_\sigma \max_\pi J_i^k(\sigma,\pi) = \max_\pi \min_\sigma J_i^k(\sigma,\pi),$$
with min and max taken over all strategies σ and π of Players MIN and MAX.

The value of the game and the Dynamic Programming operator

- **Payoff** of the k-stage game with initial state i:
$$J_i^k(\sigma,\pi) = \mathbb{E}_{i,\sigma,\pi}\left[\sum_{\ell=0}^{k-1} r_{i_\ell}^{a_\ell b_\ell}\right].$$
- **Value** of the k-stage game with initial state i:
$$v_i^k = \min_\sigma \max_\pi J_i^k(\sigma,\pi) = \max_\pi \min_\sigma J_i^k(\sigma,\pi),$$
with min and max taken over all strategies σ and π of Players MIN and MAX.
- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be the **Shapley (Bellman) operator** of the game (MDP):
$$F_i(x) := \min_{a\in\mathcal{A}_i} \max_{b\in\mathcal{B}_i} \left(r_i^{ab} + P_i^{ab} x\right) \qquad i \in [n],\ x \in \mathbb{R}^n.$$
- **Dynamic programming** principle (Shapley 53):
the value v_i^k of the k-stage game with initial state i is given recursively by
$$v_i^0 = 0 \quad \text{and} \quad v_i^{k+1} = F_i(v^k).$$
- Moreover the optimal actions in $F_i(v^k)$ give optimal strategies of the game that are positional.

Mean-payoff games

- **Mean-payoff** of the infinite horizon game with initial state i:
$$J_i^\infty(\sigma,\pi) = \limsup_{k\to\infty} \frac{1}{k}\mathbb{E}_{i,\sigma,\pi}\left[\sum_{\ell=0}^{k-1} r_{i_\ell}^{a_\ell b_\ell}\right] \quad .$$
- **Value** of the mean-payoff game with initial state i:
$$v_i^\infty = \min_\sigma \max_\pi J_i^\infty(\sigma,\pi) = \max_\pi \min_\sigma J_i^\infty(\sigma,\pi),$$
with min and max taken over all strategies σ and π of Players $\textsc{Min}$ and $\textsc{Max}$.
- **Mean payoff vector**:
$$\chi(F) = \lim_{k\to\infty} \frac{v^k}{k} = \lim_{k\to\infty} \frac{F^k(0)}{k} \quad .$$
- **Existence problem for v^∞ and $\chi(F)$**:
difficult in general, studied by Bewley and Kohlberg, Mertens and Neyman, Rosenberg and Sorin, Renault, Vigeral,...;

Properties of the Shapley operator

- F is monotone (order preserving): $x \leqslant y \;\Rightarrow\; F(x) \leqslant F(y)$;
- F is additively homogeneous: $F(x+\lambda 1) = F(x) + \lambda 1, \quad \lambda \in \mathbb{R}$;
- F is sup-norm nonexpansive: $\|F(x) - F(y)\|_\infty \leqslant \|x - y\|_\infty$;

- F is piecewise affine when the action spaces are finite.
- Then, (Kohlberg, 80) F has an invariant half-line:

$$\exists x, \nu \in \mathbb{R}^n \quad \forall t > 0 \quad F(x + t\nu) = x + (t+1)\nu.$$

- This implies $\chi(F) := \lim_{k\to\infty} \frac{v^k}{k} = v^\infty = \nu$.

Properties of the Shapley operator

- F is monotone (order preserving): $x \leqslant y \;\Rightarrow\; F(x) \leqslant F(y)$;
- F is additively homogeneous: $F(x+\lambda 1) = F(x) + \lambda 1, \quad \lambda \in \mathbb{R}$;
- F is sup-norm nonexpansive: $\|F(x) - F(y)\|_\infty \leqslant \|x - y\|_\infty$;

- F is piecewise affine when the action spaces are finite.
- Then, (Kohlberg, 80) F has an invariant half-line:

$$\exists x, \nu \in \mathbb{R}^n \quad \forall t > 0 \quad F(x + t\nu) = x + (t+1)\nu.$$

- This implies $\chi(F) := \lim_{k\to\infty} \frac{v^k}{k} = v^\infty = \nu$.
- Let $w_\gamma = F(\gamma v_\gamma)$ be the value of the discounted game. Then, the following holds true

$$\chi(F) := \lim_{k\to\infty} \frac{v^k}{k} = v^\infty = \lim_{\gamma\to 1^-} (1-\gamma) w_\gamma \ ,$$

  - if F is semi-algebraic (Beyley, Kohlberg, 76), (Neyman, 03);
  - if F is definable in a o-minimal structure (Bolte, Gaubert, Vigeral, 14).

The ergodic equation

- Assume that $\rho \in \mathbb{R}$ and $v \in \mathbb{R}^n$ satisfy the **ergodic equation** associated to F:

$$F(v) = \rho + v.$$

- Then $v_i^\infty = \chi(F)_i = \rho$ for all $i \in [n]$.
- Moreover the optimal actions in $F(v)$ give **optimal stationary positional strategies of the game**.
- λ is the **ergodic constant** (or **additive eigenvalue**): $\chi(F) = \lambda 1$.
  u is a **bias vector** (or **additive eigenvector**).

This talk

1. When does the ergodic constant/eigenvalue exist, that is the ergodic equation has a solution?
2. Is the bias vector unique, up to an additive constant?
3. What does "the game is ergodic" mean?
4. Algorithms to solve the ergodic equation ?
5. Or to find the optimal stationnary policies?
6. Complexity ?
7. Can we compute the mean payoff vector $\chi(F)$, or $\max_i \chi_i(F)$, or $\min_i \chi_i(F)$ in general ?

Nonlinear Perron-Frobenius theory

- Let $\mathrm{Log} : \mathbb{R}^n \to \mathbb{R}^n_+$ applies log componentwise, and $\mathrm{Exp} = \mathrm{Log}^{-1}$, consider

$$G = \mathrm{Exp} \circ F \circ \mathrm{Log} : \mathbb{R}^n_+ \to \mathbb{R}^n_+ \ .$$

  Then, G is a positively homogeneous and monotone if and only if F is additively homogeneous and monotone.

- $F(u) = \lambda 1 + u$ if and only if $\exp \lambda$ and $\mathrm{Exp}\, u$ are **the Perron eigenvalue and a Perron eigenvector of** G.

- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be additively homogeneous and monotone, then it is the Shapley operator of a game, possibly with infinite action spaces (Kolokoltsov, 92), or a deterministic game (Rubinov, Singer, 01), (Gunawardena, Sparrow, 03).

- Then, if $G : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is positively homogeneous and monotone, it is the Shapley operator of a multiplicative (deterministic) game.

Ergodic games (A.,Gaubert, Hochart)

Theorem (0-player case: Ergodicity of Markov chains)

Let P be a Markov matrix. T.F.A.E.:

1. *Every vector $\eta \in \mathbb{R}^n$ such that $P\eta = \eta$ is constant;*
2. *For every vector $g \in \mathbb{R}^n$, the following Cesaro limit is a constant vector:*

$$\lim_{k\to\infty} \frac{1}{k}(g + Pg + \cdots + P^{k-1}g) \ ;$$

3. *For every vector $g \in \mathbb{R}^n$, there exists a solution $(\lambda, u) \in \mathbb{R} \times \mathbb{R}^n$ to the ergodic equation*

$$g + Pu = \lambda 1 + u \ ;$$

4. *The directed graph associated to P has only one final class;*
5. *The matrix P has only one invariant measure (a stochastic row vector $m \in \mathbb{R}^{1\times n}$ such that $mP = m$);*
6. *For a given vector $g \in \mathbb{R}^n$, a solution $(\lambda, u) \in \mathbb{R} \times \mathbb{R}^n$ to the ergodic equation is unique up to an additive constant.*

Then, the Markov matrix P is ergodic.

Existence of an ergodic constant using the recession function

- Consider the *recession function:*

$$\widehat{T} : x \in \mathbb{R}^n \mapsto \widehat{T}(x) = \lim_{\rho \to +\infty} \frac{T(\rho x)}{\rho} \ , \tag{1}$$

- When the action spaces are finite, $\widehat{T}$ exists and is given by:

$$[\widehat{T}(x)]_i = \min_{a \in A_i} \max_{b \in B_{ia}} \left(P_i^{ab} x \right) \qquad i \in [n],\ x \in \mathbb{R}^n.$$

- All constant vectors are fixed points of $\widehat{T}$.
- (Gaubert, Gunawardena, 04): If $\widehat{T}$ has only trivial (constant) fixed points, then T has an additive eigenvalue.

Existence of an ergodic constant using the recession function

- Consider the *recession function:*

$$\widehat{T} : x \in \mathbb{R}^n \mapsto \widehat{T}(x) = \lim_{\rho \to +\infty} \frac{T(\rho x)}{\rho} \ , \tag{1}$$

- When the action spaces are finite, $\widehat{T}$ exists and is given by:

$$[\widehat{T}(x)]_i = \min_{a \in A_i} \max_{b \in B_{ia}} \left(P_i^{ab} x \right) \qquad i \in [n],\ x \in \mathbb{R}^n.$$

- All constant vectors are fixed points of $\widehat{T}$.
- (Gaubert, Gunawardena, 04): If $\widehat{T}$ has only trivial (constant) fixed points, then T has an additive eigenvalue.
- Given a *perturbation vector* $g \in \mathbb{R}^n$, $g + T$ is the operator of the game with transition payments $r_i^{ab} + g_i$. We have $\widehat{g + T} = \widehat{T}$.

Theorem (A., Gaubert, Hochart, 15)

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the Shapley operator of a zero-sum game with finite state space and bounded transition payments. T.F.A.E.:

1. *the recession function $\widehat{T}$ has only trivial (constant) fixed points;*
2. *the ergodic equation $g + T(u) = \lambda 1 + u$ has a solution for all perturbation vectors $g \in \mathbb{R}^n$;*
3. *for all state-dependent perturbations $g \in \mathbb{R}^n$ of the transition payments, the mean payoff vector $\chi(g + T)$ does exist and is a constant vector;*

1. $\implies$ 2. $\iff$ 3. follow from (Gaubert, Gunawardena, 04).

Existence of an ergodic constant in terms of graphs

- **Sufficient conditions:** One player case: Bather, 1973. Two player case: Gaubert, Gunawardena, 04, Cavazos-Cadena, Hernández-Hernández, 10.
- A *dominion* of some player of a zero-sum payment free game is a nonempty subset Δ of states such that, from any initial position $i \in \Delta$, this player can force the state to remain almost surely in Δ at each stage.

Existence of an ergodic constant in terms of graphs

- **Sufficient conditions:** One player case: Bather, 1973. Two player case: Gaubert, Gunawardena, 04, Cavazos-Cadena, Hernández-Hernández, 10.
- A *dominion* of some player of a zero-sum payment free game is a nonempty subset Δ of states such that, from any initial position $i \in \Delta$, this player can force the state to remain almost surely in Δ at each stage.

Theorem (A., Gaubert, Hochart, 15)

Let $\widehat{T} : \mathbb{R}^n \to \mathbb{R}^n$ be the Shapley operator of a payment free zero-sum game with finite state space. T.F.A.E.:

1. *$\widehat{T}$ has only trivial (constant) fixed points;*
2. *The two players MIN and MAX have no disjoint dominions I and J in the game.*

These conditions depend only on the support of the transition probabilities P_i^{ab} (that is the set $\{(i,a,b,j) \mid P_{ij}^{ab} > 0\}$).

Existence of an ergodic constant in terms of graphs

- **Sufficient conditions:** One player case: Bather, 1973. Two player case: Gaubert, Gunawardena, 04, Cavazos-Cadena, Hernández-Hernández, 10.
- A *dominion* of some player of a zero-sum payment free game is a nonempty subset Δ of states such that, from any initial position $i \in \Delta$, this player can force the state to remain almost surely in Δ at each stage.

Theorem (A., Gaubert, Hochart, 15)

Let $\widehat{T} : \mathbb{R}^n \to \mathbb{R}^n$ be the Shapley operator of a payment free zero-sum game with finite state space. T.F.A.E.:

1. *$\widehat{T}$ has only trivial (constant) fixed points;*
2. *The two players MIN and MAX have no disjoint dominions I and J in the game.*

These conditions depend only on the support of the transition probabilities P_i^{ab} (that is the set $\{(i,a,b,j) \mid P_{ij}^{ab} > 0\}$).

+ characterization in terms of hypergraphs of the same game.

The general case using an abstract game

- The **additive slice spaces** of $f : \mathbb{R}^n \to \mathbb{R}^n$ are (Gaubert, Gunawardena, 04):

$$\mathcal{S}_\alpha^\beta(f) := \{x \in \mathbb{R}^n \mid \alpha 1 + x \leqslant f(x) \leqslant \beta 1 + x\}\,, \quad \alpha, \beta \in \mathbb{R}\ .$$

- The **Hilbert's seminorm** on $\mathbb{R}^n$ is defined for $x \in \mathbb{R}^n$ as

$$\|x\|_{\mathsf{H}} := \inf\{\beta - \alpha \mid \alpha, \beta \in \mathbb{R},\ \alpha 1 \leqslant x \leqslant \beta 1\} = \max_{i,j=1,\ldots,n}(x_i - x_j)\ .$$

Theorem (Gaubert, Gunawardena, 04)

If T has at least one slice $\mathcal{S}_\alpha^\beta(T)$ bounded in Hilbert seminorm, then T has an additive eigenvalue.

Theorem (A., Gaubert, Hochart, 20)

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be monotone and additively homogeneous. T.F.A.E.

1. *All the additive slice spaces of T are bounded in Hilbert's seminorm;*
2. *for all $g \in \mathbb{R}^n$, there exists $(\lambda, u) \in \mathbb{R} \times \mathbb{R}^n$ such that $g + T(u) = \lambda 1 + u$;*
3. *for all $g \in \mathbb{R}^n$, the mean payoff vector $\chi(g + T)$ does exist and is a constant vector;*
4. *the two players do not have disjoint dominions in the game $\Gamma_\infty(T)$.*

One can also consider general uniform perturbations of T, and prove directly 1. $\iff$ 2. $\iff$ 3. using accretive operators (Hochart, 19).

The game $\Gamma_\infty(T)$

- The state space is $[n]$.
- The action spaces A_i and B_i of players $\textsc{Min}$ and $\textsc{Max}$ are subsets of $2^{[n]}$.
- At state i, a possible action of player $\textsc{Min}$ is a subset I of $[n]$ such that

$$\lim_{\alpha \to +\infty} T_i(\alpha 1_{I^c}) < +\infty.$$

- And a possible action of player $\textsc{Max}$ is a subset J of $[n]$ such that

$$\lim_{\alpha \to -\infty} T_i(\alpha 1_{J^c}) > -\infty.$$

- Then, the state at the next stage is chosen in $I \cap J$, with uniform probability, that is $P_{ik}^{IJ} = 1/|I \cap J|$ if $k \in I \cap J$ and 0 otherwise.
- No additive payoff: $r_i^{IJ} = 0$.

In the game $\Gamma_\infty(T)$:

- One can choose $I = [n]$
- The game is well posed: $I \cap J \neq \emptyset$.
- $\textsc{Min}$ and $\textsc{Max}$ play at the same time.

In the game $\Gamma_\infty(T)$:

- One can choose $I = [n]$
- The game is well posed: $I \cap J \neq \emptyset$.
- $\textsc{Min}$ and $\textsc{Max}$ play at the same time.
- Δ is a dominion of some player of $\Gamma_\infty(T)$ if he can choose Δ in each state $i \in \Delta$.
- If the transition payments are bounded, then dominions of $\Gamma_\infty(T)$ and of $\Gamma(\widehat{T})$ are the same.

Example

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$T(x) = \begin{pmatrix} \sup\limits_{0<p\leqslant 1} \left\{ \log p + p(x_2 \wedge x_3) + (1-p)x_1 \right\} = h((x_2 \wedge x_3) - x_1) + x_1 \\ \inf\limits_{0<p\leqslant 1} \left\{ -\log p + px_3 + (1-p)x_1 \right\} = -h(x_1 - x_3) + x_1 \\ x_3 \end{pmatrix},$$

with $h(z) = -1 - \log(-z)$ for $z \leqslant -1$, and $h(z) = z$ for $z \geqslant -1$.

- The set of actions of player $\textsc{Min}$ in $\Gamma_\infty(T)$ are
  - in state 1: $\{1,2\}$, $\{1,3\}$, $\{1,2,3\}$;
  - in state 2: $\{3\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$;
  - in state 3: $\{3\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$.
- The set of actions of player $\textsc{Max}$ in $\Gamma_\infty(T)$ are
  - in state 1: $\{2,3\}$, $\{1,2,3\}$;
  - in state 2: $\{1,3\}$, $\{1,2,3\}$;
  - in state 3: $\{3\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$.

- The dominions of $\textsc{Min}$ in $\Gamma_\infty(T)$ are $\{3\}$, $\{1,3\}$, $\{2,3\}$ and $\{1,2,3\}$.
- The dominions of $\textsc{Max}$ are $\{3\}$ and $\{1,2,3\}$.
- So the dominion condition is not satisfied: every two dominions of $\textsc{Min}$ and $\textsc{Max}$, respectively, have a nonempty intersection.
- So the ergodic equation is solvable for all operators $g + T$ with $g \in \mathbb{R}^3$.
- The recession operator is given by

$$\hat{T}(x) = \begin{pmatrix} x_1 \vee (x_2 \wedge x_3) \\ x_1 \wedge x_3 \\ x_3 \end{pmatrix}.$$

- Any vector $x = (\alpha, 0, 0)$ with $\alpha \geqslant 0$ is a fixed point of $\hat{T}$, so the sufficient condition appearing in the first characterization (A., Gaubert, Hochart, 15) is not satisfied.

Uniqueness of eigenvectors: a local game

Assume $u \in \mathbb{R}^n$ is an additive eigenvector of T. Define the local game $\Gamma_u(T)$ by:

- The state space is $[n]$.
- The action spaces A_i and B_i of players $\textsc{Min}$ and $\textsc{Max}$ are subsets of $2^{[n]}$.
- At state i, a possible action of player $\textsc{Min}$ is a subset I of $[n]$ such that

$$\exists \varepsilon > 0, \quad \forall \alpha \in [0, \varepsilon], \quad T_i(u + \alpha 1_{I^c}) = T_i(u).$$

- And a possible action of player $\textsc{Max}$ is a subset J of $[n]$ such that

$$\exists \varepsilon > 0, \quad \forall \alpha \in [0, \varepsilon], \quad T_i(u - \alpha 1_{J^c}) = T_i(u).$$

- Then, the state at the next stage is chosen in $I \cap J$, with uniform probability, that is $P_{ik}^{IJ} = 1/|I \cap J|$ if $k \in I \cap J$ and 0 otherwise.
- No additive payoff: $r_i^{IJ} = 0$.

Theorem (A., Gaubert, Hochart, 20)

An eigenvector u of T is unique, up to an additive constant, if and only if the players do not have disjoint dominions in the game $\Gamma_u(T)$.

On the algorithms for solving the ergodic equation (A. Gaubert, Qu, Saadi)

To solve the ergodic equation $F(v) = \rho + v$, usual methods are

- Relative value iterations: $v_{k+1} = F(v_k) - [F(v_k)]_c$, $\rho_{k+1} = [F(v_k)]_c$ (for some $c \in [n]$).
- They converge under acyclicity conditions only.
- Policy iterations for ergodic problems.
- They converge, but the complexity may be exponential.
- Linear Programming formulation.
- For 1-player games only.

To solve the ergodic equation $F(v) = \rho + v$, usual methods are

- Relative value iterations: $v_{k+1} = F(v_k) - [F(v_k)]_c$, $\rho_{k+1} = [F(v_k)]_c$ (for some $c \in [n]$).
- They converge under acyclicity conditions only.
- Policy iterations for ergodic problems.
- They converge, but the complexity may be exponential.
- Linear Programming formulation.
- For 1-player games only.

Recent progresses in complexity of value iteration and policy iteration for discounted MDP or games.

Transforming an ergodic problem into a discounted infinite horizon problem, one can obtain convergence rate and complexity results.

Infinite horizon discounted games

- **Payoff** of the infinite horizon discounted game with initial state i and a discount factor $\gamma < 1$:

$$J_i^{\infty}(\sigma, \pi) = \mathbb{E}_{i,\sigma,\pi}\left[\sum_{k=0}^{\infty} \gamma^k r_{i_k}^{a_k, b_k}\right] \quad .$$

- The value v of the discounted game is solution of $v = F^{(\gamma)}(v) := F(\gamma v)$.
- $F^{(\gamma)}$ is contracting for the sup-norm with contraction factor γ.

Infinite horizon discounted games

- **Payoff** of the infinite horizon discounted game with initial state i and a discount factor $\gamma < 1$:

$$J_i^{\infty}(\sigma, \pi) = \mathbb{E}_{i,\sigma,\pi}\left[\sum_{k=0}^{\infty} \gamma^k r_{i_k}^{a_k, b_k}\right] \quad .$$

- The value v of the discounted game is solution of $v = F^{(\gamma)}(v) := F(\gamma v)$.
- $F^{(\gamma)}$ is contracting for the sup-norm with contraction factor γ.
- The **value iterations** coincide with fixed point iterations: $v^{k+1} = F^{(\gamma)}(v^k)$. They converge geometrically towards the solution v with factor γ:

$$\lim_{k \to \infty} \|v^k - v\|^{1/k} \leqslant \gamma \quad .$$

Complexity of value iterations

- Let $E := \{(i, a, b) \mid i \in [n], a \in \mathcal{A}_i, b \in \mathcal{B}_i\}$ be the set of state-actions and $R = \max_{(i,a,b)\in E} |r_i^{ab}|$. Then the time to obtain an error $\leqslant \epsilon$ is (Tseng, 1990):

$$\mathcal{O}\left(n|E| \frac{\log\left(\frac{R}{(1-\gamma)\epsilon}\right)}{1-\gamma} \right)$$

  This is pseudopolynomial.

- Recent progress: (Sidford, Wang, Wu and Ye, 2018) constructed a Variance Reduced Value Iteration algorithm that gives, with probability greater that $1-\delta$, an $\epsilon-$optimal solution to the discounted Markov Decision Process problem in the sublinear time:

$$\tilde{\mathcal{O}}\left(|E| \frac{R^2}{(1-\gamma)^4 \epsilon^2} \log(\frac{1}{\delta}) \right) \ .$$

Policy iterations for discounted games

Denote by $\Sigma := \{\sigma : i \in [n] \mapsto \sigma_i \in \mathcal{A}_i\}$ and $\Pi := \{\pi : i \in [n] \mapsto \pi_i \in \mathcal{B}_i\}$ the sets of policies, and for $\sigma \in \Sigma$ and $\pi \in \Pi$, define:

$$M^{(\sigma\pi)} = (\gamma P_{ij}^{\sigma_i \pi_i})_{ij=1,\dots,n}, \quad \text{and } r^{(\sigma\pi)} = (r_i^{\sigma_i \pi_i})_{i=1,\dots,n} \ ,$$

and

$$F^{(\gamma,\sigma,\pi)}(v) = r^{(\sigma\pi)} + M^{(\sigma\pi)} v \quad v \in \mathbb{R}^n \ .$$

Then, $F^{(\gamma)}$ can be written as:

$$F^{(\gamma)}(v) = \min_{\sigma \in \Sigma} F^{(\gamma,\sigma)}(v) \ , \text{ with } F^{(\gamma,\sigma)}(v) := \max_{\pi \in \Pi} F^{(\gamma,\sigma\pi)}(v) \ , \quad v \in \mathbb{R}^n \ ,$$

where minima and maxima are for the partial order of $\mathbb{R}^n$.

Policy iterations for discounted games

(Howard, 1960) for 1-player games, (Denardo, 1967) for 2-player games.

Given an initial policy $\sigma^0 \in \Sigma$, apply successively the two following steps for $s \geqslant 0$ until $\sigma^{s+1} = \sigma^s$:

1. Compute the fixed point v^s of $F^{(\gamma,\sigma^s)}$;
2. *Improve the policy*: choose an optimal policy for v^s, that is $\sigma^{s+1} \in \Sigma$ such that $F^{(\gamma)}(v^s) = F^{(\gamma,\sigma^{s+1})}(v^s)$
   with $\sigma^{s+1} = \sigma^s$ as soon as this is possible.

Policy iterations for discounted games

(Howard, 1960) for 1-player games, (Denardo, 1967) for 2-player games.

Given an initial policy $\sigma^0 \in \Sigma$, apply successively the two following steps for $s \geqslant 0$ until $\sigma^{s+1} = \sigma^s$:

1. Compute the fixed point v^s of $F^{(\gamma,\sigma^s)}$;
2. *Improve the policy*: choose an optimal policy for v^s, that is $\sigma^{s+1} \in \Sigma$ such that $F^{(\gamma)}(v^s) = F^{(\gamma,\sigma^{s+1})}(v^s)$
   with $\sigma^{s+1} = \sigma^s$ as soon as this is possible.

Step 1 is solved by using Policy iteration for the (one-player) game with fixed policy σ^s, which constructs $v^{s,l}$ and $\pi^{s,l}$ from $\pi^{s,0}$.

Complexity of policy iterations for discounted games

- The sequence $(v^s)_{s \geqslant 0}$ is nonincreasing and is stationary after a finite time (at most $|\Sigma|$), thus converges to the solution v of $v = F^{(\gamma)}(v)$.
- (Friedmann, 2009) showed a 2-player deterministic game problem with $\gamma \simeq 1$ and an exponential number of iterations.
- (Fearnley, 2010) and (Andersson, 2009) showed the same for a 1-player stochastic game.

Complexity of policy iterations for discounted games

(Ye, 2011) showed that Policy iteration algorithm and Simplex algorithm solve 1-player discounted games with fixed discount factor $\gamma < 1$ in *strongly polynomial* time.

(Hansen, Miltersen and Zwick, 2011) extended and improved this result to Policy iteration algorithm for 2-player games. They show that the number of iterations $s_{\max}$ (to obtain stationarity) satisfies:

$$s_{\max} = \mathcal{O}(\frac{m}{1-\gamma} \log \frac{n}{1-\gamma}),$$

with $m = |E|$ the *total number of state-actions*.

(Scherrer, 2013) gave a better bound in the one-player case.

(Feinberg, Huang, 2013): Same for a one-player game with mean-payoff, and a state i_0 such that $P^a_{i,i_0} \geqslant 1-\gamma$, for all $i \in [n],\ a \in \mathcal{A}_i$.

Reduction from ergodic to discounted games

In the ergodic equation $\rho + v = F(v)$, the addive eigenvalue ρ is unique, but the additive eigenvector v is not unique, since $\alpha + v$ is also a solution.

Lemma

If for all the Markov matrices $M^{\sigma\pi}$, the state c is accessible from every state, or equivalently $M^{\sigma\pi}$ has a unique recurrent (final) class and c belongs to it, then by fixing $v_c = 0$, the solution (ρ, v) becomes unique.

Reduction from ergodic to discounted games

In the ergodic equation $\rho + v = F(v)$, the addive eigenvalue ρ is unique, but the additive eigenvector v is not unique, since $\alpha + v$ is also a solution.

Lemma

If for all the Markov matrices $M^{\sigma\pi}$, the state c is accessible from every state, or equivalently $M^{\sigma\pi}$ has a unique recurrent (final) class and c belongs to it, then by fixing $v_c = 0$, the solution (ρ, v) becomes unique.

We reduce the ergodic equation $\rho + v = F(v)$ with $v_c = 0$, to a fixed point problem:

$$H(w) = w \ ,$$

where H is a contracting operator, and compute the contracting factor.

Definition

Let $u \in \mathbb{R}^n$ be a positive vector ($u \gg 0$), i.e. for all $i \in \{1, \cdots, n\}$, $u_i > 0$.

We define the *weighted sup-norm* $\|\cdot\|_u$ by :

$$\|x\|_u = \max_{1 \leqslant i \leqslant n} \frac{|x_i|}{u_i}, \quad \forall x \in \mathbb{R}^n \ .$$

Definition

Let $u \in \mathbb{R}^n$ be a positive vector ($u \gg 0$), i.e. for all $i \in \{1, \cdots, n\}$, $u_i > 0$.

We define the *weighted sup-norm* $\|\cdot\|_u$ by :

$$\|x\|_u = \max_{1 \leqslant i \leqslant n} \frac{|x_i|}{u_i}, \quad \forall x \in \mathbb{R}^n \ .$$

- Weighted sup-norms used in Stochastic Shortest Path problem analysis.
- First used by Bertsekas and Tsitsiklis [1991] for the mean payoff MDP.
- Here, an explicit weighted sup-norm and contraction under it are specified.
- This uses nonlinear Perron-Frobenius theory (Nussbaum,Mallet-Paret, 1998), (Nussbaum, LAA 1986), (A., Gaubert, Nussbaum, arXiv 2011).

Definition

Let $u \in \mathbb{R}^n$ be a positive vector ($u \gg 0$), i.e. for all $i \in \{1, \cdots, n\}$, $u_i > 0$.

We define the *weighted sup-norm* $\|\cdot\|_u$ by :

$$\|x\|_u = \max_{1 \leqslant i \leqslant n} \frac{|x_i|}{u_i}, \quad \forall x \in \mathbb{R}^n \ .$$

- Weighted sup-norms used in Stochastic Shortest Path problem analysis.
- First used by Bertsekas and Tsitsiklis [1991] for the mean payoff MDP.
- Here, an explicit weighted sup-norm and contraction under it are specified.
- This uses nonlinear Perron-Frobenius theory (Nussbaum,Mallet-Paret, 1998), (Nussbaum, LAA 1986), (A., Gaubert, Nussbaum, arXiv 2011).

Definition

The *Collatz-Wielandt number* of a monotone positively homogenous map $f : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is $\mathrm{cw}(f) := \inf\{\lambda > 0 \mid \exists u \gg 0; f(u) \leqslant \lambda u\}$.

- Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$, we define its *Clarke recession function* $\hat{f} : \mathbb{R}^n \to (\mathbb{R} \cup \{+\infty\})^n$ by:

$$\hat{f}(y) = \sup_{s>0,\, x\in\mathbb{R}^n} \frac{f(x+sy) - f(x)}{s} \ .$$

- The Clarke recession function is positively homogeneous and convex.

Theorem (A., Gaubert, Qu, Saadi, 2019)

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a monotone function, $u \gg 0$ be a positive vector, and $\lambda \in \mathbb{R}_+$. We have $\hat{f}(u) \leqslant \lambda u$ if and only if *the function f is λ-Lipschitz in the weighted sup-norm $\|\cdot\|_u$:*

$$\forall x, y \in \mathbb{R}^n, \qquad \|f(x) - f(y)\|_u \leqslant \lambda \|x - y\|_u \ .$$

Then, $\mathrm{cw}(\hat{f})$ is the best Lipschitz constant of f for a weighted sup-norm.

Application to Shapley operators

- For the Shapley operator $H : \mathbb{R}^n \to \mathbb{R}^n$:

$$H_i(v) = \min_{a \in \mathcal{A}_i} \max_{b \in \mathcal{B}_i} \left\{ r_i^{ab} + M_i^{ab} v \right\}, \quad i \in [n],\ v \in \mathbb{R}^n \ .$$

- We associate the "max-max" operator $H^{\max} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$H_i^{\max}(y) = \max_{a \in \mathcal{A}_i,\ b \in \mathcal{B}_i} \{ M_i^{ab} y \}, \ \forall i \in [n], y \in \mathbb{R}^n \ .$$

- We have

$$\hat{H}(y) \leqslant H^{\max}(y), \ \forall y \in \mathbb{R}^n \ .$$

- Therefore, $\mathrm{cw}(\hat{H}) \leqslant \mathrm{cw}(H^{\max})$.
- Equality holds when all actions (a, b) are useful.
- By Perron-Frobenius results, we have

$$\mathrm{cw}(H^{\max}) = \max_{\sigma \in \Sigma, \pi \in \Pi} \rho(M^{\sigma\pi}) \ .$$

First hitting times

Definition

For a Markov chain X_k with transition matrix P, and states i, j, we denote:

$$\mathcal{T}_{ij}(P) := \mathbb{E}[\inf\{k \geqslant 1 \mid X_k = j\} \mid X_0 = i]$$

the expected first hitting time of state j of the Markov chain X_k with initial state i.

Then, $\mathcal{T}_{ic}(P) < +\infty$ for all $i \in [n]$ if and only if P has a unique recurrent (final) class and c belongs to it.

Definition

For any matrix $P \in \mathbb{R}^{n \times n}$, we denote by $P_{(c)} \in \mathbb{R}^{n \times n}$ the matrix obtained from P by replacing the column c of P with zeros.

Main assumption

There exists a state $c \in S$ accessible under all policies σ, π, that is $M^{\sigma\pi}$ has a unique final class and c belongs to it.

Lemma

The following assertions are equivalent:

- $\mathcal{T}_{ic} := \max_{\sigma\in\Sigma,\pi\in\Pi} \mathcal{T}_{ic}(P^{\sigma\pi}) < +\infty, \quad \forall i \in S;$
- *For all* $(\sigma, \pi) \in \Sigma \times \Pi$, $P^{\sigma\pi}$ *has a unique final class, and the state* c *is common to each of these classes;*
- $\max_{\sigma\in\Sigma,\pi\in\Pi} \rho(P^{\sigma\pi}_{(c)}) < 1.$
- *There is a unique vector* $\varphi^\star$ *solution of the equation:*

$$\varphi^\star = 1 + \max_{\sigma\in\Sigma,\pi\in\Pi} [P^{\sigma\pi}_{(c)}\varphi^\star] =: F_{(c)i}(\varphi^\star) \ ,$$

Under these assumptions, we have $\varphi^\star_i = \mathcal{T}_{ic}$, *for all* $i \in [n]$.

- Let $\varphi \in \mathbb{R}^n_+$ be such that

$$\varphi_i \geqslant 1 + \max_{a,b}[P^{ab}_{(c)i}\varphi] = F_{(c)i}(\varphi), \quad \forall i \in [n] \ .$$

- We associate to φ, the scalar $\lambda_\varphi = 1 - 1/\|\varphi\|_\infty \in [0,1)$.
- Then, $F_{(c)}$ is a λ_φ-contraction under the weighted sup-norm $\|\cdot\|_\varphi$.
- We define a new operator $H^\varphi : \mathbb{R}^n \to \mathbb{R}^n$, such that for $i \in [n]$ and $w \in \mathbb{R}^n$:

$$H^\varphi_i(w) = \min_{a \in A_i} \max_{b \in B_{i,a}} \left\{ \varphi_i^{-1} r_i^{ab} + w_c(1 - \varphi_i^{-1}) + \sum_{j \neq c} \varphi_i^{-1} P^{ab}_{ij} \varphi_j (w_j - w_c) \right\}.$$

Proposition

The operator H^φ is monotone and $\lambda_\varphi-$Liptschitz in the sup-norm $\|\cdot\|_\infty$.

$$H_i^{\varphi}(w) = \min_{a \in A_i} \max_{b \in B_{i,a}} \left\{ \varphi_i^{-1} r_i^{ab} + w_c(1 - \varphi_i^{-1}) + \sum_{j \neq c} \varphi_i^{-1} P_{ij}^{ab} \varphi_j (w_j - w_c) \right\},$$

Theorem

The non-linear eigenproblem

$$\rho 1 + v = F(v), \qquad \rho \in \mathbb{R}, v \in \mathbb{R}^n; v_c = 0 \ ,$$

can be reduced to the fixed point problem:

$$H^{\varphi}(w) = w \ ,$$

where $w \in \mathbb{R}^n$ *is such that* $w = \rho 1 + \varphi^{-1} v$, *and also* $\rho = w_c$ *and* $v = \varphi(w - w_c 1)$.

Deflation technique applied to policy iterations

We apply this change of variable to in policy iterations as a proof argument: policy iterations for a mean-payoff games is then equivalent to policy iterations for a discounted game.

Theorem (A., Gaubert, 2013)

Let us fix $K > 0$ and a state c. The policy iteration algorithm for the class of 2-player mean-payoff games such that

$$\mathcal{T}_{ic}(M^{(\sigma\pi)}) \leqslant K \quad \forall \sigma \in \Sigma,\ \pi \in \Pi,\ i \in [n]$$

is strongly polynomial. More precisely, the number of external iterations $s_{\max}$ satisfies:

$$s_{\max} \leqslant (m_1 - n)(1 + \lfloor \frac{\log(K)}{\log(K/(K-1))} \rfloor) = \mathcal{O}((m_1 - n)K \log K),$$

with $m_1 =$ the total number of actions of the first player.

Deflated Value Iteration

- **Preprocessing:**
  Compute maximal first hitting time vector:

$$\varphi^\star = 1 + \max_{\sigma\in\Sigma,\pi\in\Pi}[P_{(c)}^{\sigma\pi}\varphi^\star] := F_{(c)}(\varphi^\star) \ ,$$

  to get a super-eigenvector φ satisfying:

$$\varphi_i \geqslant 1 + \max_{a,b}[P_{(c)i}^{ab}\varphi], \quad \forall i \in [n] \ .$$

  and the constant $\lambda = \lambda_\varphi = 1 - 1/\|\varphi\|_\infty$.
- **Solve the discounted fixed point problem:**

$$H^\varphi(w) = w \ ,$$

  to deduce $\rho = w_c$ and $v = \varphi(w - w_c 1)$ solutions of the non-linear eigenproblem:

$$\rho 1 + v = F(v), \qquad \rho \in \mathbb{R}, v \in \mathbb{R}^n; v_c = 0 \ ,$$

Deflated Value Iteration: deterministic algorithm

Theorem

The deterministic Deflated Value Iteration algorithm finds a solution (ρ, v) *of the mean payoff problem such that* $|\rho - \rho^\star| \leqslant \epsilon$, *and* $\|v - v^*\|_\infty \leqslant \frac{5\epsilon}{1-\lambda}$ *in time complexity:*

$$O\left(\frac{n|E|}{1-\lambda}\log\left(\frac{R}{(1-\lambda)\epsilon}\right)\right) \ .$$

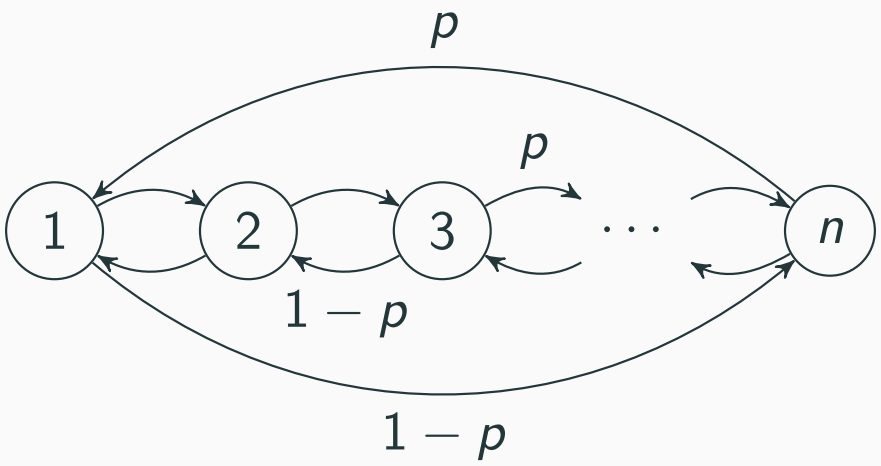
To be compared with the complexity of VI in the discounted case:

$$O\left(\tfrac{n|E|}{1-\gamma}\log\left(\tfrac{R}{(1-\gamma)\epsilon}\right)\right) \ .$$

Our conditions of convergence are typically less demanding than for Relative Value Iteration (D.J. White [1963]).

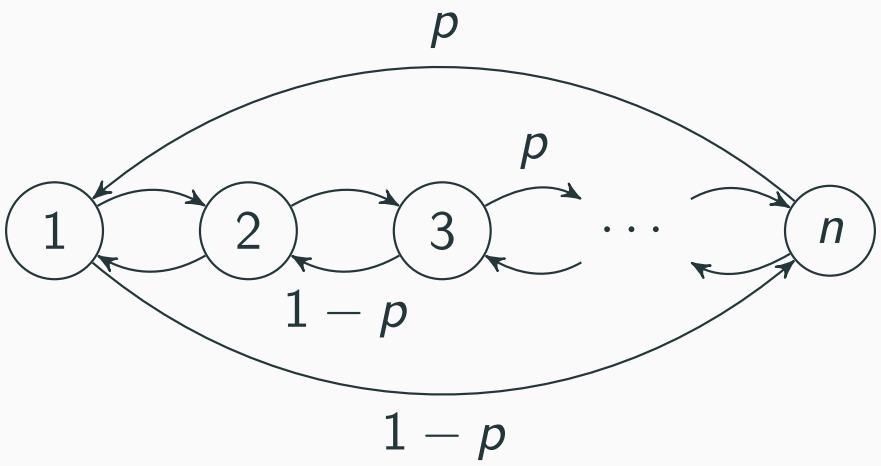
Simulations

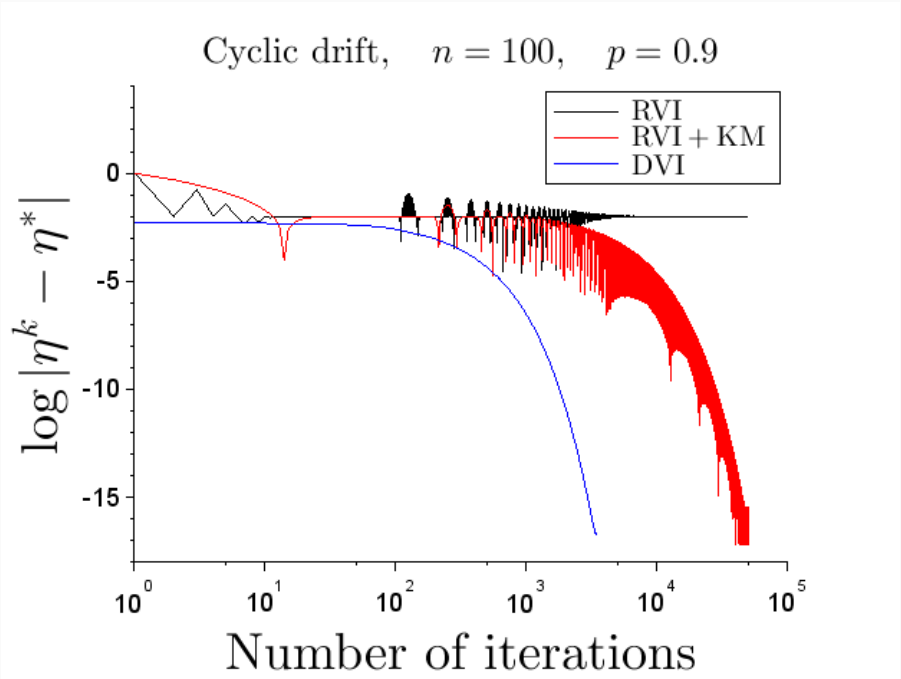
Cyclic drift

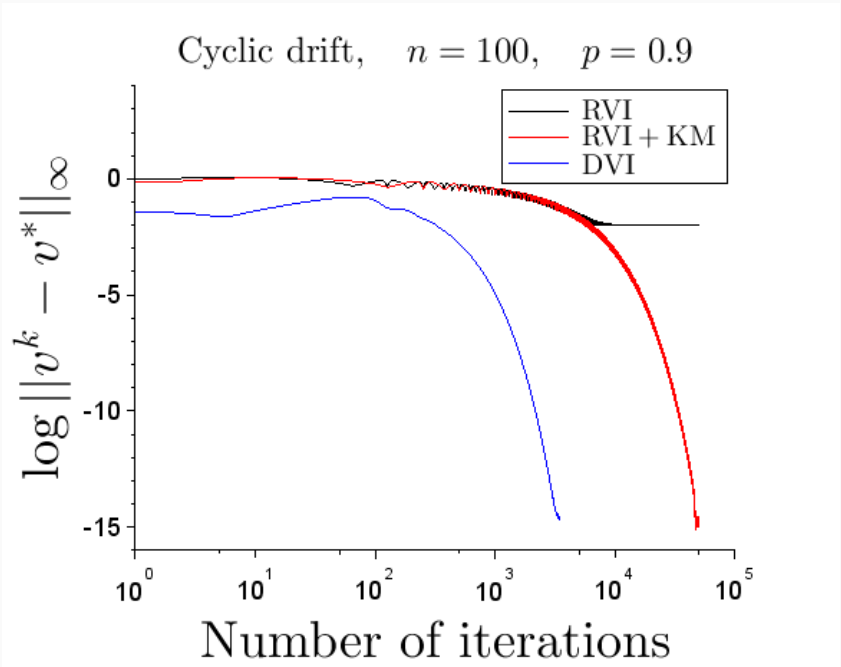


Simulations

Cyclic drift







Variance Reduced Value Iteration for structured stochastic games

We consider Shapley operators written as

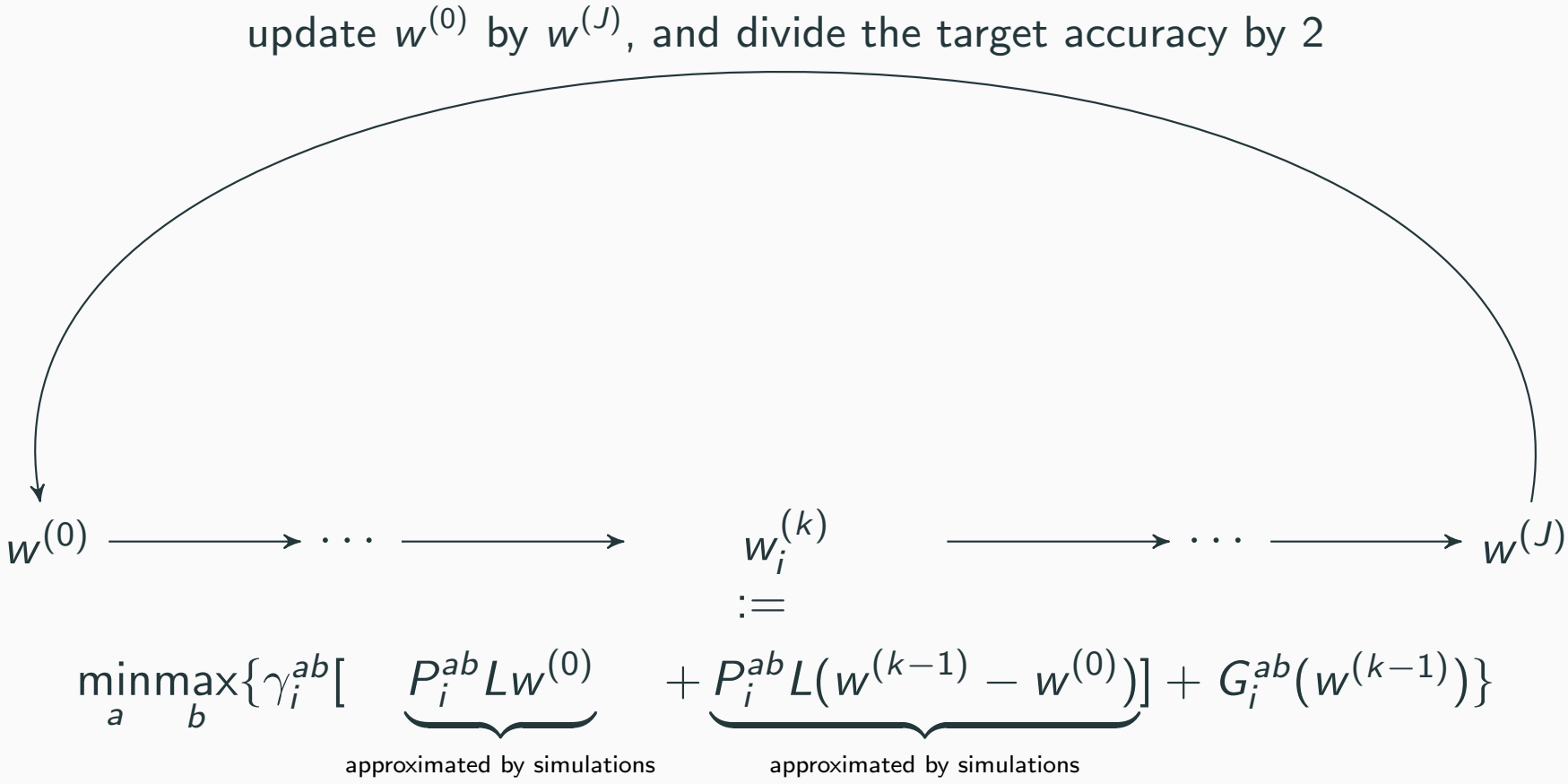
$$H_i(w) = \min_{a \in A_i} \max_{b \in B_{i,a}} \left\{ \gamma_i^{ab} P_i^{ab} L w + H_i^{ab}(w) \right\}, \forall i \in [n].$$

Here $L \in \mathbb{R}^{n \times n}$ is a sparse operator such that for all $w \in \mathbb{R}^n$, Lw can be computed in $O(|S|)$. H_i^{ab} is a sparse affine operator such that $H_i^{ab}(w)$ can be computed in $O(1)$ for all $w \in \mathbb{R}^n$. The matrices $P^{\sigma\pi}$ are sub-Markovian for each couple of policies $(\sigma, \pi) \in \Sigma \times \Pi$.

The problem that we want to solve is:

$$H(w) = w \ .$$

Applying the variance reduction method of Sidford et al. to structured stochastic games



Variance Reduced Deflated Value Iteration

Theorem ((AGQS, 2019), Quasilinear)

With probability $1-\delta$, *we find* (ρ, v) *such that* $|\rho-\rho^\star| \leqslant \epsilon$ *and* $\|v-v^*\|_\infty \leqslant \frac{5\epsilon}{1-\lambda}$ *in time:*

$$\tilde{\mathcal{O}}\left(\left(n|E| + \frac{|E|}{(1-\lambda)^5}\right)\log(\frac{R}{\epsilon})\log(\frac{1}{\delta})\right) \ .$$

To be compared with (Sidford, Wang, Wu and Ye, 2018): with probability greater than $1-\delta$, they find an $\epsilon-$optimal solution to the discounted Markov Decision Process problem in the quasilinear time:

$$\tilde{\mathcal{O}}\left(\left(n|E| + \frac{|E|}{(1-\gamma)^3}\right)\log(\frac{R}{\epsilon})\log(\frac{1}{\delta})\right) \ .$$

Theorem ((AGQS, 2019), Sublinear)

With probability $1-\delta$, *we find* (ρ, v) *such that* $|\rho - \rho^\star| \leqslant \epsilon$ *and* $\|v - v^*\|_\infty \leqslant \frac{5\epsilon}{1-\lambda}$ *in time:*

$$\tilde{O}\left(|E|\left[\frac{R^2}{(1-\lambda)^4\epsilon^2} + \frac{1}{(1-\lambda)^6}\right]\log(\frac{1}{\delta})\right) \ .$$

To be compared with (Sidford, Wang, Wu and Ye, 2018): with probability greater than $1-\delta$, they find an $\epsilon-$optimal solution to the discounted Markov Decision Process problem in the sublinear time:

$$\tilde{\mathcal{O}}\left(|E|\frac{R^2}{(1-\gamma)^4\epsilon^2}\log(\frac{1}{\delta})\right) \ .$$

Multiplicative games and entropy games (A. Gaubert, Grand Clément, Guillaud, 19)

Multiplicative games

- Replace additive transition payments $r_i^{ab} \in \mathbb{R}$ by multiplicative positive transition payments $\gamma_i^{ab} \in \mathbb{R}_+$
- **Payoff** of the k-stage game with initial state i:
$$J_i^k(\sigma,\pi) = \mathbb{E}_{i,\sigma,\pi}\left[\prod_{\ell=0}^{k-1} \gamma_{i_\ell}^{a_\ell b_\ell}\right].$$
- **Geometrical mean-payoff** of the infinite horizon multiplicative game with initial state i:
$$J_i^\infty(\sigma,\pi) = \limsup_{k\to\infty} \left(J_i^k(\sigma,\pi)\right)^{\frac{1}{k}} \quad .$$
- For $k \in \mathbb{N} \cup \{\infty\}$, the **Value** of the k-stage game with initial state i is:
$$V_i^k = \min_\sigma \max_\pi J_i^k(\sigma,\pi) = \max_\pi \min_\sigma J_i^k(\sigma,\pi),$$
with min and max taken over all strategies σ and π of Players MIN and MAX.
- This is also called Risk sensitive control or growth optimality of branching Markov Decision Chains (Howard, Matheson, 72), (Sladky, 76), (Rothblum, Whittle, 82), (Rothblum, 84), (Zijm, 87), (Fleming, Hernandez-Hernandez, 97), (Anantharam, Borkar, 17).

- The **Shapley (Bellman) operator** of the multiplicative game (MDP) is $G : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ such that

$$G_i(x) := \min_{a \in \mathcal{A}_i} \max_{b \in \mathcal{B}_i} \left(\gamma_i^{ab} P_i^{ab} x \right) \qquad i \in [n],\ x \in \mathbb{R}^n.$$

- With Log glasses, we get $F = \mathrm{Log} \circ G \circ \mathrm{Exp} : \mathbb{R}^n \to \mathbb{R}^n$.
- **Dynamic programming** principle: the value V_i^k of the k-stage game with initial state i is given by

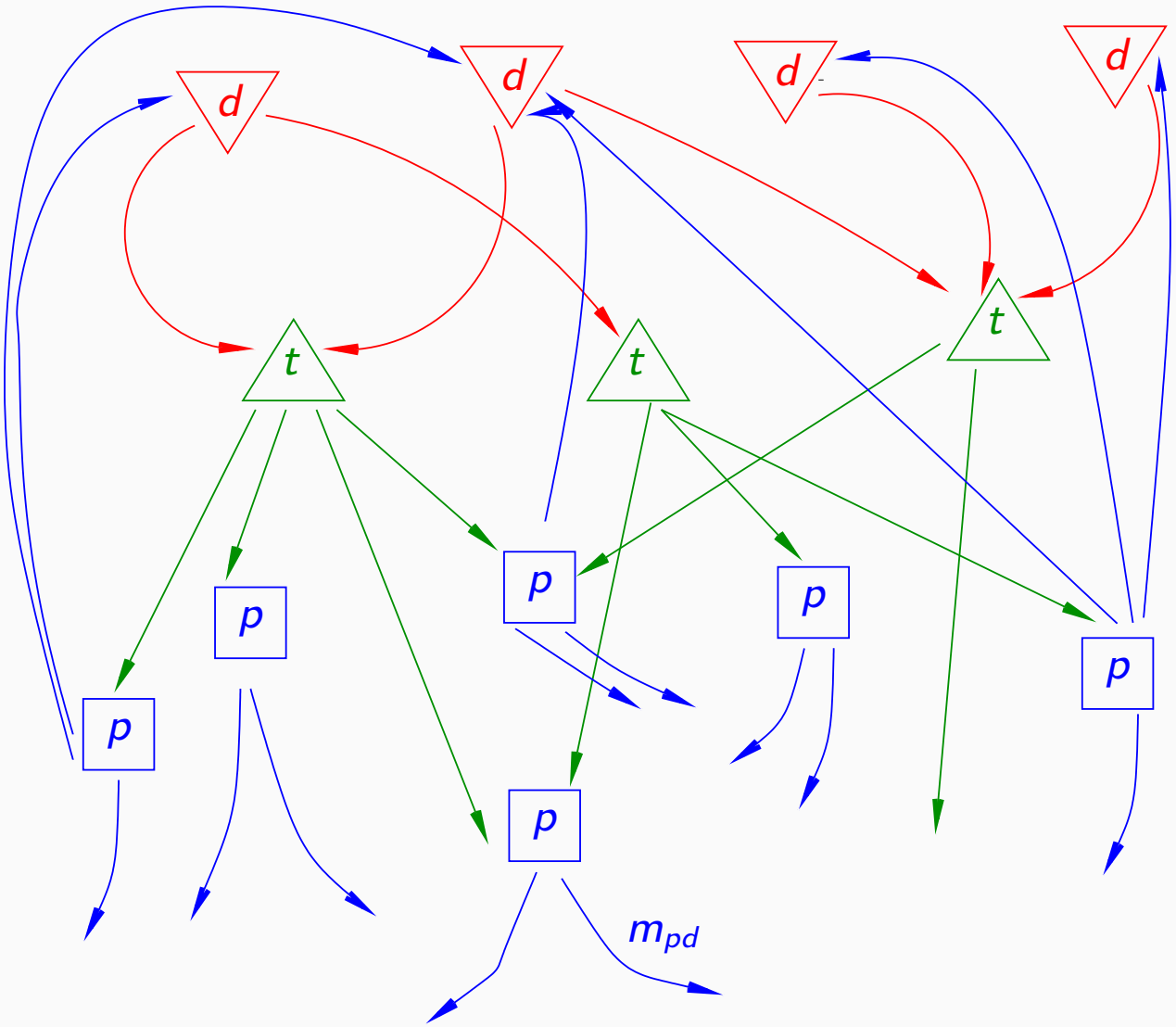
$$V_i^0 = 0 \quad \text{and} \quad V_i^{k+1} = G_i(V^k).$$

- Then $\lim_{k\to\infty} (V^k)^{\frac{1}{k}} = \mathrm{Exp}(\chi(F))$.
- So the existence of V^∞ is related to the existence of $\chi(F)$.

Example: entropy games (A. Gaubert, Grand Clément, Guillaud, 19)

- Entropy games were introduced by Eugene Asarin, Julien Cervelle, Aldric Degorre, Cătălin Dima, Florian Horn and Victor Kozyakin, (STACS 2016): the payment is a topological entropy.
- Special case of joint spectral radius problem.
- We consider a variant of entropy games which is equivalent to a geometrical mean payoff game in which
  - Player MIN is called "Despot" and wishes to minimize the freedom of "People";
  - Player MAX is called "Tribune" and wishes to maximize the freedom of "People";
  - Player "Nature" is called "People".
  - Freedom is represented by a topological entropy.

The model



Despot plays

Tribune plays

People plays

The model

- The graph of the game has 3 kinds of nodes: D, T, P.
- a set of arcs $E \subset (D \times T) \cup (T \times P) \cup (P \times D)$
- arcs from D to T and from T to P have weights 1 and arcs $(p, d) \in P \times D$ have weight m_{pd}.
- an action is a following node in the graph.
- Given an initial state $i \in D$ and strategies σ, π of Despot and Tribune, the *payoff* in horizon k (that Despot pays to Tribune) is

  $J_i^k(\sigma, \pi) =$ sum of weights of paths of length $3k$ (corresponding to k steps of the game), starting from initial state i.
- This is a multiplicative game in which the multiplicative positive transition payments and transition probabilities are

$$\gamma_d^{tp} = \gamma(p) := \sum_{d' \in D} m_{pd'}$$

$$P_{dd'}^{tp} = q_{pd'} := m_{pd'}/\gamma(p) \ .$$

- Our variant of the *entropy game* is the associated geometrical mean payoff game:

$$J_i^{\infty}(\sigma,\pi) = \limsup_{k\to\infty} \left(J_i^k(\sigma,\pi)\right)^{\frac{1}{k}} \quad .$$

- Its value V^{∞} is such that

$$V_i^{\infty} = \min_{\sigma}\max_{\pi} J_i^{\infty}(\sigma,\pi) = \max_{\pi}\min_{\sigma} J_i^{\infty}(\sigma,\pi).$$

- In (Asarin et al.) variant, People also chooses the initial state i (paths with any initial state are considered).

Results (A. Gaubert, Grand Clément, Guillaud, 19)

- The value $V^k = (V_d^k)_{d\in D}$ of the entropy game in horizon k does exist and satisfies $V^0 \equiv 1$, $V^k = G(V^{k-1})$, $k \geqslant 1$ and there exists optimal positional strategies σ^* and π^* of Despot and Tribune.
- Using Fenchel-Legendre transform, we obtain, for a general multiplicative game:

$$F_i(x) = \min_{a\in\mathcal{A}_i} \max_{b\in\mathcal{B}_{i,a}} \max_{\vartheta\in\Delta_n} \left(\log(\gamma_i^{ab}) - S(\vartheta, P_i^{ab}) + \sum_j \theta_j x_j \right) \qquad i \in [n],\ x \in \mathbb{R}^n,$$

where Δ_n is the set of probability measures ϑ on $[n]$, and S is the *relative entropy* or *Kullback-Leibler divergence*:

$$S(\vartheta; m) := \sum_{j\in[n]} \vartheta_j \log(\vartheta_j/m_j) \ .$$

- For the entropy game, this is rewritten:

$$F_d(x) = \min_{t\in T,\ (d,t)\in E} \max_{p\in P,\ (t,p)\in E} \max_{\vartheta\in\Delta_D} \left(-S(\vartheta, m_p) + \sum_{d\in D} \vartheta_d x_d \right) \ ,$$

where m_p is the measure on D given by $(m_{pd})_{d\in D}$ with $m_{pd} = 0$ if $(p,d) \notin E$.

Theorem

The infinite horizon entropy game has a value and it has optimal positional strategies (σ^*, π^*)*, meaning:*

$$V_d^\infty(\sigma^*, \pi) \leqslant V_d^\infty \leqslant V_d^\infty(\sigma, \pi^*) \ .$$

Moreover, for all initial states d*,*

$$V_d^\infty = \lim_{k\to\infty} (V_d^k)^{1/k} \ .$$

Idea of proof: $\log V_d^\infty$ is the value of the mean payoff game with Kullback-Leibler rewards.

The corresponding dynamic programming operator F is definable in a o-minimal structure.

By (Bolte, Gaubert, Vigeral, 14), the value of the mean payoff game exists.

- An o-minimal structure (van den Dries) consists, for each integer n, of a family of *definable* subsets of $\mathbb{R}^n$.
- definable sets must be closed under: Boolean operations, projection map from $\mathbb{R}^n$ to $\mathbb{R}^{n-1}$, under the lift, meaning if $A \subset \mathbb{R}^n$ is definable, then $A \times \mathbb{R} \subset \mathbb{R}^{n+1}$ and $\mathbb{R} \times A \subset \mathbb{R}^{n+1}$ are also definable.
- Definable subsets of $\mathbb{R}$ are precisely finite unions of intervals.
- A function f from $\mathbb{R}^n$ to $\mathbb{R}^k$ is definable if its graph is definable.
- The *real exponential field* $\mathbb{R}_{\mathsf{alg},\exp}$ is a o-minimal structure (Wilkie). The definable sets are the *subexponential sets*, i.e., the images under the projection maps $\mathbb{R}^{n+k} \to \mathbb{R}^n$ of the *exponential sets* of $\mathbb{R}^{n+k}$, $\{x \mid P(x_1, \ldots, x_{n+k}, e^{x_1}, \ldots, e^{x_{n+k}}) = 0\}$ where P is a real polynomial.
- The Shapley operator F of the Kullback-Leibler game is definable in this structure, since $G = \mathrm{Exp} \circ F \circ \mathrm{Log}$ is piecewise affine.

The Collatz-Wielandt theorem

Theorem (Nussbaum 86 and Gaubert Gunawardena 2004)

The value $\bar{V}^\infty := \max_d V_d^\infty$ *of the original entropy game (in which People chooses initial state) coincides with the Collatz-Wielandt number*

$$\begin{aligned} \mathrm{cw}(G) &:= \inf\{\lambda > 0 \mid \exists X \in \operatorname{int} \mathbb{R}_+^D,\ G(X) \leqslant \lambda X\} \\ &= \max\{\lambda > 0 \mid \exists X \in \mathbb{R}_+^D \setminus \{0\},\ G(X) = \lambda X\}\ . \end{aligned}$$

An entropy game is *Despot-free* if in each state $d \in D$, Despot has only one possible action $\sigma(d)$.

Corollary

The logarithm of the value $\bar{V}^{\infty}$ of a Despot-free entropy game is given by

$$\begin{aligned} &\inf \mu \\ &(\mu, x) \in \mathbb{R} \times \mathbb{R}^D, \textit{satisfying} \\ &\mu + x_d \geqslant \log\Big(\sum_{d' \in D} m_{p,d'} e^{x_{d'}}\Big) \; \forall d \in D, p \in P \textit{ s.t. } (\sigma(d), p) \in E \quad . \end{aligned}$$

This can also be derived from Anantharam and Borkar (2016), dual of Donsker Varadhan formula, similar ideas in Chen and Han (FSTTCS 2014).

We assume from now that the weights $m_{p,d}$ are integers.

By *solving strategically* an (extended) entropy game, we mean, finding a pair of optimal policies (stationary positional strategies).

Then, the value is an algebraic number.

Theorem

Despot-free entropy games can be solved strategically in polynomial time.

Corollary

Despot-free entropy games in which Despot has a fixed number of significant positions can be solved strategically in polynomial time.

Elements of proof (Despot-free case):

- The value in a state $t \in T$ is the max of the values of strongly connected components to which t has access (Zijm, 87).
- In the irreducible case, the value $\lambda^* > 0$ is such that $\log \lambda^*$ is solution of the bounded convex programming / Collatz-Wielandt formulation:

$$\inf \mu$$
$$(\mu, x) \in \mathcal{K} \ .$$

  where $\mathcal{K}$ is a bounded convex subset of $(\mu, x) \in \mathbb{R} \times \mathbb{R}^D$.
- Use the *ellipsoid method* (Grötschel, Lovász and A. Schrijver 81) to compute an ϵ-approximate solution λ of the convex program.
- Let M^π be the matrix of the dynamic programming operator obtained when the strategy of Tribune is fixed and equal to π.
- Perron-Frobenius techniques show that $\lambda \leqslant \rho(M^\pi)$ for some π.
- The differences between the different values of the $\rho(M^\pi)$ is bounded below by a rational number $\eta_{\mathsf{sep}} > 0$, whose number of bits is polynomially bounded in the size of the input (Rump 79).
- Fixing the precision ϵ smaller than η_{sep}, we get $\rho(M^\pi) = \lambda^*$, hence π is optimal.

Summary

This talk: a survey of results on mean-payoff zero-sum games with the use of Perron-Frobenius techniques:

- **Ergodicity of zero-sum games, or general Shapley operators:** combinatorial characterization of existence of eigenvectors, valid for any perturbation of payments, and a similar characterization of **uniqueness** of eigenvectors (up to an additive constant).
- **Open:** Continuous time or infinite state space cases.
- **Complexity** results for policy and value iterations using a deflation technique, when the game has a state accessible from any state for all policies.
- **Open:** complexity of multichain mean-payoff games.
- **Study of multiplicative mean-payoff games**: existence of the value of games and polynomial complexity in the 1-player case.
- **Open:** complexity of the 2-player case.

Any questions ?