Two-Player Perfect-Information Shift-Invariant Submixing Stochastic Games Are Half-Positional

Hugo Gimbert
CNRS, LaBRI, Bordeaux.

Meeting FREC, April 2014

Joint work with
Edon Kelmendi (LaBRI, INRIA Bordeaux-Sud-Ouest, Labex CPU)

arXiv : 1401.6575

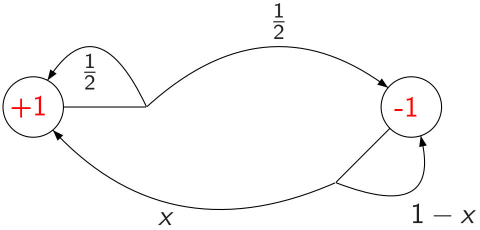
Stochastic games with perfect information

On the existence of ϵ-subgame perfect strategies

Games with shift-invariant and submixing payoff functions are half-positional

Stochastic games with perfect information

Markov chains (no players)



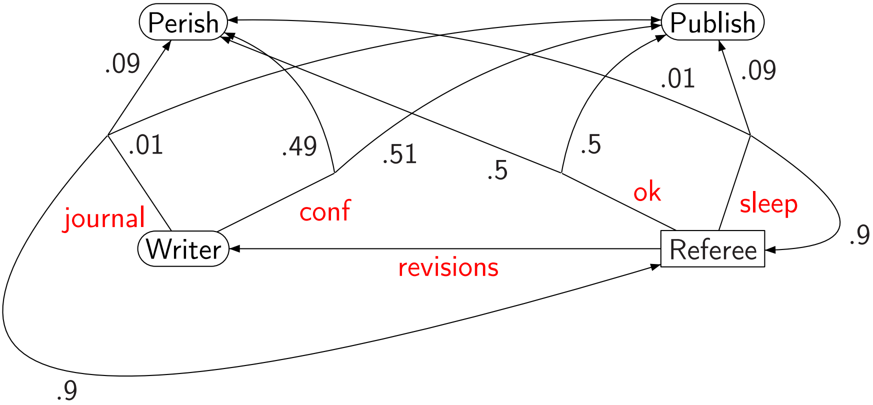
Mean-payoff: average payoff $\lim_n \frac{\sum_{i\leq n} r(s_i)}{n} = \frac{x-\frac{1}{2}}{x+\frac{1}{2}}$

Counter game: $\liminf \sum_{i\leq n} r(s_i) = \begin{cases} +\infty \text{ a.s. } & \text{if } x > \frac{1}{2} \\ -\infty \text{ a.s. } & \text{otherwise} \end{cases}$

Stochastic games with perfect information

Two players: Writer and Referee.

Play using actions.



Optimal strategy: Writer should send to conference and referee should write positive reviews.

Stochastic games with perfect information

Players 1 and 2
States S partitioned in S_1 and S_2, initial state s_0
Actions A partitioned in A_1 and A_2
Transitions $p : S \times A \to \Delta(S)$ where $\Delta(S)$ are probability distributions on S

Goal of player 1 choose actions such that the play $p \in S^\omega$ maximizes the payoff function $f : S^\omega \to \mathbb{R}$ (bounded and Borel-measurable)

Strategy for player 1 $\sigma : S^* S_1 \to A_1$
Strategy for player 2 $\tau : S^* S_2 \to A_2$
Probability measure $\mathbb{P}_{s_0}^{\sigma,\tau}$ on S^ω in the Markov chain induced by σ and τ

Goal of player 1 choose σ that maximizes $\mathbb{E}_{s_0}^{\sigma,\tau}[f]$

A fundamental theorem by Martin

Theorem [Martin 98] for every (concurrent) stochastic game where f is bounded and Borel-measurable,

$$\sup_\sigma \inf_\tau \mathbb{E}_{s_0}^{\sigma,\tau}[f] = \inf_\tau \sup_\sigma \mathbb{E}_{s_0}^{\sigma,\tau}[f] \ .$$

and this defines the value $val(s_0)$ of s_0.

The $\leq$ inequality is trivial.

Corollary: for every $\epsilon > 0$ player 1 has an ϵ-optimal strategy σ_ϵ such that:

$$\inf_\tau \mathbb{E}_{s_0}^{\sigma_\epsilon,\tau}[f] \geq val(s_0) - \epsilon.$$

Mean-payoff games [Mertens-Neyman 81].

How do the ϵ-optimal strategies look like? Can we compute them?

Computing payoffs (1)

Parity games

One-counter stochastic games (Brazdil, Brozek, Etessami): each state s is labelled with $r(s) \in \{0, -1, +1\}$.
Payoff after play $p = s_0 s_1 s_2 \cdots$ is 0 or 1 depending whether:

$$\text{Unboundedness: } \limsup \sum_{i \leq n} r(s_i) = +\infty$$

$$\text{Divergence: } \liminf \sum_{i \leq n} r(s_i) = +\infty$$

Computing payoffs (2)

Mean-payoff games and variants: each state s is labelled with $r(s) \in \mathbb{R}$. Payoff after play $p = s_0 s_1 s_2 \cdots$ is:

Mean-payoff game
$$f_{mean}(p) = \limsup \frac{1}{n} \sum_{i \leq n} r(s_i)$$

Positivity game
$$\begin{cases} 1 \text{ if } f_{mean}(p) > 0 \\ 0 \text{ otherwise.} \end{cases}$$

Generalized mean-payoff game
$$\begin{cases} 1 \text{ if } \exists k,\, f_{mean,k}(p) > 0 \\ 0 \text{ otherwise,} \end{cases}$$

where each $f_{mean,k}$ is associated to a different reward mapping $r_k : S \to \mathbb{R}$.

Positionality results

Theorem [several authors[1]]: in all these games player 1 has optimal positional strategies. Also player 2 except in the generalized mean-payoff game.

Positional strategy: $\sigma : S \to A$ instead of $\sigma : S^* \to A$.

The generalized mean-payoff game is half-positional, others are positional.

Motivation: positional strategies = main ingredient of many algorithms for stochastic games.

[1] Brazdil, Brozek, Chaterjee, Doyen, Etessami, G., Henzinger, Kelmendi, Raskin, Zielonka + unconvicing proofs by Gilette, Liggett and Lippman

Tools to prove positionality

Fix a payoff function f.

Theorem: [G., Zielonka 07] if every one-player game equipped with f or $-f$ is positional then every two-player game as well.

Theorem: [G. 07 [2]] if f is prefix-independent and submixing then every one-player game equipped with f is positional.

prefix-independent: $\forall u \in S^*, \forall p \in S^\omega, f(up) = f(p)$

submixing: $\forall u_1, v_1, u_2, v_2, \ldots \in S^+$,

$$f(u_1 v_1 u_2 v_2 \cdots) \leq \max\{f(u_1 u_2 \cdots), f(v_1 v_2 \cdots)\} \ .$$

Used successfully to study one-counter games.

What about half-positional games?.

Theorem: [Kopczynski, 06] if a game is deterministic, prefix-independent and submixing then it is half-positional.

[2] Warsaw, -25° Celsius.

Tool to prove half-positionality

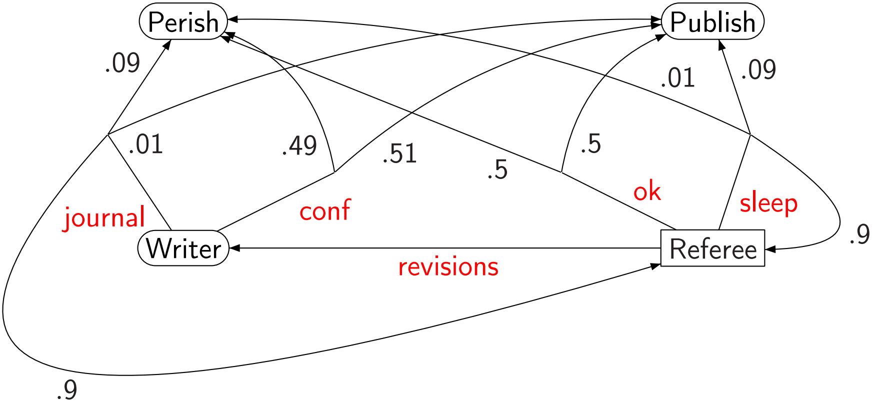
Fix a payoff function f.

Theorem: [G., Kelmendi 14] if f is prefix-independent and submixing then every two-player game equipped with f is half-positional.

Generalizes both [Kopczynski 06] and [G. 07].

Covers all examples and many more.

ϵ-subgame-perfect strategies



Optimal strategy for referee:
if Referee starts he writes a positive review.
If Writer starts and sends to journal (suboptimal) referee writes negative review (suboptimal).

ϵ-subgame-perfect strategies

Play ϵ-optimally whatever happens.

Notation: Strategy σ, finite play $p \in S^*$,

$$\sigma[p] : q \mapsto \sigma(pq) \ .$$

Definition:Strategy σ is ϵ-subgame-perfect if $\forall p \in S^*, \sigma[p]$ is ϵ-optimal.

Theorem: [G., Kelmendi 14] In every prefix-independent game, both players have ϵ-subgame-perfect strategies.

Weakness. Given a strategy σ for Player 1, a finite play $p \in S(AS)^*$ is a σ-weakness if $\sigma[p]$ is not 2ϵ-optimal.

Factorizing plays according to weaknesses. Every infinite play $q \in S(AS)^\omega$ can be factorized uniquely as a finite or infinite sequence $q = p_0 p_1 p_2 \ldots$ such that $\forall p_n$,

1. p_n finite $\implies$ p_n is a σ-weakness,
2. p_n finite $\implies$ no strict prefix of p_n is a σ-weakness,
3. p_n infinite $\implies$ no prefix of p_n is a σ-weakness,

and this factorization can be computed online.

The reset strategy

Definition each time a σ-weakness occurs, we reset the memory.

$$\hat{\sigma}(p_0 p_1 \ldots p_n) = \sigma(p_n) \quad .$$

Lemma: if σ is ϵ-optimal and consistent then the reset strategy $\hat{\sigma}$ is 2ϵ-subgame-perfect

Consistent strategies

A strategy is consistent if whenever it plays action a in state s,

$$val(s) = \sum_t p(s,a,t)val(t) \ .$$

Lemma: player 1 has a consistent ϵ-optimal strategy.

Lemma: if σ is consistent and T is a stopping time with respect to $(S_n)_{n\in\mathbb{N}}$ then

$$\mathbb{E}[\lim_n val(S_{\min(n,T)})] \geq val(s) \ .$$

Lemma: σ ϵ-optimal, then $\exists \mu > 0$ such that for all consistent τ,

$$\mathbb{P}_s^{\sigma,\tau}(\exists n, S_0 \cdots S_n \text{ is a } \sigma\text{-weakness}) \leq 1 - \mu.$$

Only finitely many weaknesses occur with the reset strategy

Lemma: If σ is ϵ-optimal,

$$\mathbb{P}_s^{\hat{\sigma},\tau}(\exists n \in \mathbb{N}, \text{ there is no } \sigma\text{-weakness after date } n) = 1 \ .$$

Corollary: If σ is ϵ-optimal then $\hat{\sigma}$ is 2ϵ-subgame-perfect.

Theorem: [G., Kelmendi 14] if f is prefix-independent and submixing then every two-player game equipped with f is half-positional.

Proof: by induction on the size of the arena.
The existence of an ϵ-subgame-perfect strategies τ_0 and τ_1 for player 2 in the subgames G_0 and G_1 is used to build a pair (σ_0, τ_{01}) of optimal strategies in G.

Conclusion

New generic result about half-positionality of games.

Next:
games with compact action spaces.
strategies with simple memory structures.