Probabilistic Aspects of Computer Science: Stochastic Games

Serge Haddad

LSV, ENS Paris-Saclay & CNRS & Inria

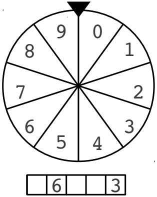
MPRI M1

Plan

The spinner game revisited

The player has to compose a five-digit number.

- The digits are randomly chosen by a spinner during five rounds.
- After every round (except the last one),
  the player chooses in which position he inserts the current digit.
- The goal of the player is to obtain the largest number as possible.



The presenter participates to the game.

- At any time but at most once, the presenter may switch the current digit with the previous one when their value difference is at most 2.
- The goal of the presenter is to obtain the smallest number as possible.

Introduction to stochastic games

A stochastic game (SG) is a finite transition system
where any state belongs to either player Max or Min.

The dynamic of the system is defined as follows.

- The player owning the current state chooses (possibly randomly) an enabled *action*.
- Then the environment randomly selects the next state.
  The distribution depends on the current state and the selected action.

There are several ways to define rewards.

Syntax of SG

A SG $\mathcal{G} \stackrel{\text{def}}{=} (S, \{A_s\}_{s \in S}, p)$ is defined by:

- $S = S_{\mathsf{Min}} \uplus S_{\mathsf{Max}}$, the finite set of states;
- For every state s, A_s, the finite set of actions enabled in s.
  $A \stackrel{\text{def}}{=} \bigcup_{s \in S} A_s$ is the whole set of actions.
- p, a mapping from $\{(s, a) \mid s \in S, a \in A_s\}$ to the set of distributions over S.
  $p(s'|s, a)$ denotes the probability to go from s to s' if a is selected.

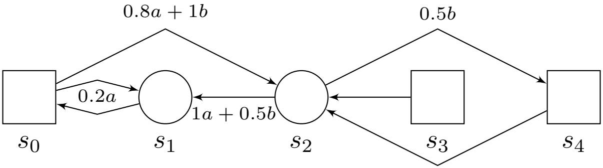
Histories.

A *history* $h \stackrel{\text{def}}{=} s_0 a_0 \ldots s_i a_i \ldots$ is a finite or infinite sequence alternating states and actions such that when s_{i+1} is defined $p(s_{i+1}|s_i, a_i) > 0$.

Illustration

A stochastic game is depicted as a labelled graph.

- States of player Max are represented by circles ($\bigcirc$).
- States of player Min are represented by squares ($\square$).
- An edge (s, s') is labelled by $\sum_{a \in A_s} p(s'|s,a)a$ (when non null).



From SG to DTMC

In order to obtain a stochastic process,

one needs to fix the non deterministic features of the SG.

A *strategy* of a player P is a mapping from histories ending in a state $s \in S_{\mathsf{P}}$ to a distribution over A_s.

Classes of strategies are defined depending on two criteria.

- the information used in the history.
  When a strategy only depends on the last state, it is called *memoryless*;
- the way the selection is performed.
  When a strategy deterministically selects its actions, it is called *pure*.

The DTMC $\mathcal{G}^{\sigma,\tau}$ is the behaviour of the SG $\mathcal{G}$

once strategies σ and τ of respectively Max and Min are chosen.

Its states are information used in strategies.

One denotes h the random infinite history and $\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G},s}$ (resp. $\mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}$)
the probability measure (the expectation operator) in $\mathcal{G}^{\sigma,\tau}$ when starting in s.

Illustration

Pure memoryless strategies.

- Let σ be the strategy of Max that selects b in s_2.
- Let τ be the strategy of Min that selects a in s_0.
- Then $\mathcal{G}^{\sigma,\tau}$ is depicted below.

Rewards for histories

Let $h = s_0 a_0 s_1 \dots$ be an infinite history

and $\mathit{Inf}(h) = \{s \mid \forall i \; \exists j > i \; s_j = s\}$, the set of states occurring infinitely often in h.

Discounted SG with rewards $r(s, a)$ in $[0, 1]$ and a discount $0 < \lambda < 1$.

$$r(h) = \sum_{n \in \mathbb{N}} \lambda^n r(s_n, a_n)$$

Mean Payoff SG with rewards $r(s, a)$ in $[0, 1]$.

$$r(h) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i<n} r(s_i, a_i)$$

Parity SG with integer priorities $pri(s)$.

Let $pri(h) = \max(pri(s) \mid s \in \mathit{Inf}(h))$ then $r(h) = \mathbb{1}_{pri(h) \text{ is even}}$

Priority SG with rewards $r(s)$ in $[0, 1]$ and unique integer priorities $pri(s)$.

Let $s_{\max} = \arg\max(pri(s) \mid s \in \mathit{Inf}(h))$ then $r(h) = r(s_{\max})$

Observation. Priority SG extend parity SG.

Problems for SG

- **Determinacy problem.**

Let s be a state of a SG $\mathcal{G}$.
Define $val^{\downarrow}_{\mathcal{G}}(s) = \sup_{\sigma} \inf_{\tau} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h))$ and $val^{\uparrow}_{\mathcal{G}}(s) = \inf_{\tau} \sup_{\sigma} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h))$
By construction, $val^{\downarrow}_{\mathcal{G}}(s) \leq val^{\uparrow}_{\mathcal{G}}(s)$.
Does $val^{\downarrow}_{\mathcal{G}}(s) = val^{\uparrow}_{\mathcal{G}}(s)$? Yes it is called the *value* of s in $\mathcal{G}$ and denoted $val_{\mathcal{G}}(s)$.

- **Existence of optimal strategies.**

Does there exist σ (resp. τ) such that:
$\inf_{\tau} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = val_{\mathcal{G}}(s)$ (resp. $\sup_{\sigma} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = val_{\mathcal{G}}(s)$)? Yes.

- **Classes of optimal strategies.**

How can σ and τ be chosen? Pure and memoryless.

- **Computational problems.**

What is the complexity of the associated decision problems?
in NP $\cap$ coNP for most of the SG.

Plan

Presentation

2 Pure Memoryless Determinacy

Discounted Games

Mean Payoff Games

Priority (and Parity) Games

Computational issues

Complexity results

From Mean Payoff Games to Discounted Games

From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

A generic method

Let σ^* be a strategy of Player Max and τ^* be a strategy of Player Min that fulfill for all s:

$$\inf_{\tau} \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau}(r(h)) = \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau^*}(r(h)) = \sup_{\sigma} \mathbf{E}_{\mathcal{G},s}^{\sigma,\tau^*}(r(h))$$

Then the game is determined and σ^* and τ^* are optimal strategies.

Proof.

$$\sup_{\sigma} \inf_{\tau}(\mathbf{E}_{\mathcal{G},s}^{\sigma,\tau}(r(h))) \geq \inf_{\tau}(\mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau}(r(h)))$$

$$= \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau^*}(r(h)) = \sup_{\sigma}(\mathbf{E}_{\mathcal{G},s}^{\sigma,\tau^*}(r(h)))$$

$$\geq \inf_{\tau} \sup_{\sigma}(\mathbf{E}_{\mathcal{G},s}^{\sigma,\tau}(r(h)))$$

So $\mathcal{G}$ is determined and σ^* and τ^* are optimal strategies in $\mathcal{G}$.

Plan

Presentation

2. Pure Memoryless Determinacy
   - Discounted Games

   Mean Payoff Games

   Priority (and Parity) Games

Computational issues

Complexity results

From Mean Payoff Games to Discounted Games

From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

A contracting operator

Let L be the mapping from $\mathbb{R}^S$ to $\mathbb{R}^S$ defined by:

$$L(\mathbf{v})[s] \stackrel{\text{def}}{=} \max \left(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \mid a \in A_s \right) \text{ when } s \in S_{\text{Max}}$$

$$L(\mathbf{v})[s] \stackrel{\text{def}}{=} \min \left(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \mid a \in A_s \right) \text{ when } s \in S_{\text{Min}}$$

L "selects" the best decision rule for the owner of s
in a game that stops at time 1 including a terminal reward $\lambda \mathbf{v}$.

Properties of L.

L is Lipschitz-continuous with Lipschitz constant equal to $\lambda < 1$.

Thus L admits a unique fixed-point denoted $\mathbf{v}^*_\lambda$.

Pure memoryless strategies

Let σ^* be a strategy of player Max that selects in $s \in S_{\mathsf{Max}}$ some a_s such that:

$$r(s, a_s) + \lambda \sum_{s' \in S} p(s'|s, a_s)\mathbf{v}^*_\lambda[s'] = \mathbf{v}^*_\lambda[s]$$

Let τ^* be a strategy of player Min that selects in $s \in S_{\mathsf{Min}}$ some a_s such that:

$$r(s, a_s) + \lambda \sum_{s' \in S} p(s'|s, a_s)\mathbf{v}^*_\lambda[s'] = \mathbf{v}^*_\lambda[s]$$

Pure memoryless determinacy

• Let $\mathbf{v}_n$ be the infimum of the expected discounted rewards up to time n in $\mathcal{G}^{\sigma^*,\tau}$ against an arbitrary strategy τ of player Min. Then:

$$\mathbf{v}_n[s] \geq \mathbf{v}^*_\lambda[s] - \frac{\lambda^n}{1-\lambda}$$

Proof by induction on n.

Inductive step: $s \in S_{\mathsf{Max}}$.

$\mathbf{v}_{n+1}[s] = r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}_n[s'] \geq r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)(\mathbf{v}^*_\lambda[s'] - \frac{\lambda^n}{1-\lambda})$
$= r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}^*_\lambda[s'] - \frac{\lambda^{n+1}}{1-\lambda} = \mathbf{v}^*_\lambda[s] - \frac{\lambda^{n+1}}{1-\lambda}$

Inductive step: $s \in S_{\mathsf{Min}}$. Let a be the action selected by τ.

$\mathbf{v}_{n+1}[s] = r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}_n[s'] \geq r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)(\mathbf{v}^*_\lambda[s'] - \frac{\lambda^n}{1-\lambda})$
$= r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}^*_\lambda[s'] - \frac{\lambda^{n+1}}{1-\lambda} \geq \mathbf{v}^*_\lambda[s] - \frac{\lambda^{n+1}}{1-\lambda}$

• Let $\mathbf{w}_n$ be the infimum of the expected discounted rewards up to time n in $\mathcal{G}^{\sigma,\tau^*}$ against an arbitrary strategy σ of player Max.
By a similar reasoning $\mathbf{w}_n[s] \leq \mathbf{v}^*_\lambda[s] + \frac{\lambda^n}{1-\lambda}$.

Thus the game is determined with value $\mathbf{v}^*_\lambda$ and σ^* and τ^* are optimal.

Plan

Presentation

2 Pure Memoryless Determinacy

Discounted Games

- Mean Payoff Games

Priority (and Parity) Games

Computational issues

Complexity results

From Mean Payoff Games to Discounted Games

From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

From mean payoff to discounted games

Let $\mathcal{G}$ be a mean payoff game and $\mathcal{G}_\lambda$ the discounted version with discount λ.

Pick some increasing sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \lambda_n = 1$.

Let σ_n and τ_n be pure memoryless optimal strategies for $\mathcal{G}_{\lambda_n}$.

Since there are only finite such strategies,
some strategies σ^* and τ^* must occur *simultaneously* infinitely often.

By considering a subsequence, one assumes that σ^* and τ^* are optimal for all $\mathcal{G}_{\lambda_n}$.

A property of σ^* and τ^*

There exists n_0 such that for all $\lambda \geq \lambda_{n_0}$, σ^* and τ^* are optimal in $\mathcal{G}_\lambda$.

Proof by contradiction.

Assume there exists some increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ and $\lambda_{n_k} < \mu_k < \lambda_{n_{k+1}}$ such that for all k,

there exist $s \in S$ and pure memoryless strategies σ_k and τ_k fulfilling:

- either $\mathbf{E}^{\sigma_k,\tau_k}_{\mathcal{G}_{\mu_k},s}(r(h)) > \mathbf{E}^{\sigma^*,\tau_k}_{\mathcal{G}_{\mu_k},s}(r(h))$;
- or $\mathbf{E}^{\sigma_k,\tau_k}_{\mathcal{G}_{\mu_k},s}(r(h)) < \mathbf{E}^{\sigma_k,\tau^*}_{\mathcal{G}_{\mu_k},s}(r(h))$.

For pure memoryless strategies σ and τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h))$ is a rational function of λ.

Define:

$$f_s(\lambda) = \prod_{\substack{\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) \neq \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h)) \\ \sigma,\sigma',\tau,\tau' \text{ pure memoryless}}} \mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h))$$

Then some f_s would have an infinite number of zeroes.

Pure memoryless determinacy

Let us denote the random history $h = s_0 a_0 s_1 \dots$
Consider the MDP $\mathcal{G}_{\tau^*}$ obtained by using strategy τ^* for player Min.
σ^* is a Blackwell policy in $\mathcal{G}_{\tau^*}$. So it is optimal for mean payoff reward:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}_{\mathcal{G},s}^{\sigma,\tau^*}(r(s_i,a_i)) \leq \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau^*}(r(s_i,a_i))$$

Using a similar reasoning, one gets for all s and τ:

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau}(r(s_i,a_i)) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}_{\mathcal{G},s}^{\sigma^*,\tau^*}(r(s_i,a_i))$$

So $\mathcal{G}$ is determined and σ^* and τ^* are optimal strategies in $\mathcal{G}$.

Plan

Presentation

2. Pure Memoryless Determinacy
   - Discounted Games
   - Mean Payoff Games
   - Priority (and Parity) Games

Computational issues

- Complexity results
- From Mean Payoff Games to Discounted Games
- From Discounted Games to Parity Games
- From Parity Games to Mean Payoff Games

Scheme of the proof

A state s is *absorbing* if $A_s = \{a\}$ for some a and $p(s|s,a) = 1$.

Observe that the priority of an absorbing state is irrelevant.

A state s is *vanishing* if for all s' and $a \in A_{s'}$, $p(s|s',a) = 0$.

A state is *relevant* if it is neither absorbing nor vanishing.

The proof is done by induction on the number of relevant states.

The basis case

When there is no relevant state, all strategies are memoryless.

The value of an absorbing state s is $r(s)$.

The value of a vanishing state s belonging to Max (resp. Min) is:

$$\max_{a \in A_s} \sum_{s'} p(s'|s,a) r(s') \qquad \text{(resp. } \min_{a \in A_s} \sum_{s'} p(s'|s,a) r(s'))$$

and a corresponding pure strategy is some:

$$\arg\max_{a \in A_s} \sum_{s'} p(s'|s,a) r(s') \qquad \text{(resp. } \arg\min_{a \in A_s} \sum_{s'} p(s'|s,a) r(s'))$$

Interlude

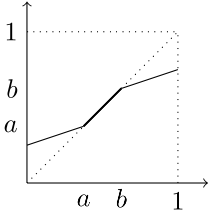
Let f be a function from $[0,1]$ to $[0,1]$ that fulfills:

- f is non decreasing;
- f is 1-Lipschitz: $|f(x) - f(x')| \leq |x - x'|$.

The set of fixed points of f is a non empty interval $[a,b]$.

Denoting $f^\infty(x) = \lim_{n\to\infty} f^{(n)}(x)$:

- for all $x < a$, $f^\infty(x) = a$ and $f(x) > x$;
- for all $a \leq x \leq b$, $f^\infty(x) = x$;
- for all $b < x$, $f^\infty(x) = b$ and $f(x) < x$.



Building the inductive step

Let $\mathcal{G}$ be a stochastic game with s the relevant state with maximal priority.
We consider all rewards for s and denote $\mathcal{G}_v$ the game $\mathcal{G}$ with $r(s) = v$.

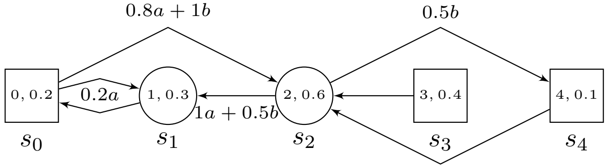
We define the game $\mathcal{G}'_v$ as follows.

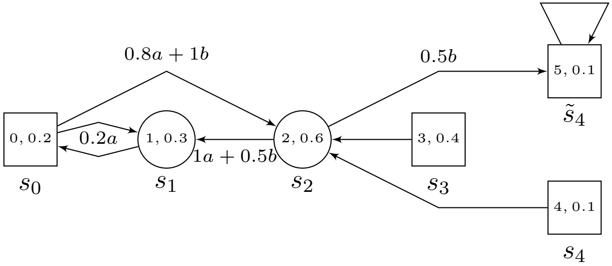
- Add an absorbing state $\tilde{s}$ with reward v.
- Redirect all incoming transitions in s to $\tilde{s}$:
  $p'(\tilde{s}|s', a) = p(s|s', a)$ and $p'(s|s', a) = 0$.

Since s is vanishing in $\mathcal{G}'_v$, it has less relevant states than $\mathcal{G}_v$.
So the induction applies.

One denotes by $f_t(v)$, $val_{\mathcal{G}'_v}(t)$ the value of state t in $\mathcal{G}'_v$.
By construction, f_t is non decreasing and $|f_t(v) - f_t(v')| \leq |v - v'|$.

Illustration





Analysis of $\mathcal{G}'_v$ (1)

Proposition.

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$.
Assume $v < f_s(v)$.
Then there exists $\varepsilon > 0$ such that given any strategy τ of Min:
the probability to reach $\tilde{s}$ from s in $\mathcal{G}'^{\sigma_v,\tau}_v$ is bounded by $1-\varepsilon$.

Proof.

Otherwise by a family of strategies τ_n reaching $\tilde{s}$ with probability at least $1-\frac{1}{n}$
Min can ensure that $f_s(v) \leq v$.

$\square$

Consequence for $\mathcal{G}_v$.

When $v < f_s(v)$, for all strategy τ of Min
the probability to visit infinitely often s in $\mathcal{G}^{\sigma_v,\tau}_v$ is null.

Analysis of $\mathcal{G}'_v$ (2)

Proposition.

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$.
Assume $v \leq f_s(v)$.
Let Div be the event: h does not reach $\tilde{s}$. Then for all strategy τ of Min:
(when defined) $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) \geq f_s(v)$

Proof.

$f_s(v) \leq \mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)) = \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(Div)\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) + (1 - \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(Div))v$

So $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) \geq f_s(v)$.

$\square$

Consequence for $\mathcal{G}_v$.

Let R_n be the event: h visits s exactly n times.

If $v \leq f_s(v)$ then for all strategy τ of Min:

(when defined) $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) \geq f_s(v)$.

A first lower bound

Proposition.

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$.
If $v \leq f_s(v)$ then for all strategy τ of Min:
$f_s(v) \leq \mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h))$.

Proof. Let R_∞ be the event: h visits s infinitely often.

$$\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)) = \sum_n \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_n)\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) + \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_\infty)v$$

Recall that $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) \geq f_s(v)$.

Now:

- either $f_s(v) = v$ and thus $\mathbf{E}^{\sigma_v,\tau}_s(r(h)) \geq f_s(v)$;
- or $f_s(v) > v$ and implying $\mathbf{Pr}^{\sigma_v,\tau}(R_\infty) = 0$ implying $\mathbf{E}^{\sigma_v,\tau}_s(r(h)) \geq f_s(v)$.

$\square$

A second lower bound (1)

Proposition.

There exists a pure memoryless strategy σ of Max in $\mathcal{G}_v$ such that:

1. σ is optimal in $\mathcal{G}'_{f_s^\infty(v)}$;
2. for all τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,s}(r(h)) \geq f_s^\infty(v)$;
3. for all t, for all τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)) \geq f_t(f_s^\infty(v))$.

Proof.

• **Proof of 1,2: Case $f_s(v) \leq v$.**

A pure memoryless optimal strategy $\sigma_{f_s^\infty(v)}$ in $\mathcal{G}'_{f_s^\infty(v)}$ ensures for s a value $f_s^\infty(v)$ in $\mathcal{G}_{f_s^\infty(v)}$ thus also in $\mathcal{G}_v$.

• **Proof of 1,2: Case $v < f_s(v)$.**

A pure memoryless optimal strategy σ_v in $\mathcal{G}'_v$ ensures for s a value $f_s(v)$ in $\mathcal{G}_v$.

Since for all τ $\mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_\infty) = 0$, σ_v ensures a value $f_s(v)$ in $\mathcal{G}_{v'}$ for any v'.

A second lower bound (2)

Proof continued.

Let us note $a = f_s^\infty(v)$ the least fixed point of f_s.
Observe that $v < f_s(v)$ is equivalent to $v < a$.

There is a finite number of pure memoryless strategies.
Consider a strategy σ such for all $\varepsilon > 0$ there is some $a - \varepsilon < v < a$ with $\sigma_v = \sigma$.

Thus σ ensures for s a value a in all $\mathcal{G}_{v'}$.

Since σ is optimal in $\mathcal{G}'_{v'}$ for v' as close as possible to a, σ is optimal in $\mathcal{G}'_a$.

• **Proof of 3.**
Since σ is optimal in $\mathcal{G}'_{f_s^\infty(v)}$, for all τ, $f_t(f_s^\infty(v)) \le \mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f_s^\infty(v)},t}(r(h))$

Let R be the event h reaches $\tilde{s}$. Then:

$$\begin{aligned}
\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f_s^\infty(v)},t}(r(h)) = & \quad (1-\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_{f_s^\infty(v)},t}(R))\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f_s^\infty(v)},t}(r(h)|R^c) + \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_{f_s^\infty(v)},t}(R) f_s^\infty(v) \\
\le & \quad (1-\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_v,t}(R))\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)|R^c) + \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_v,t}(R)\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_{f_s^\infty(v)},t}(r(h)|R) \\
= & \quad \mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h))
\end{aligned}$$

Pure memoryless determinacy

Proposition.

There exists a pure memoryless strategy τ of Min in $\mathcal{G}_v$ such that:

- τ is optimal in $\mathcal{G}'_{f_s^\infty(v)}$;
- for all σ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,s}(r(h)) \leq f_s^\infty(v)$.
- for all t, for all σ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)) \leq f_t(f_s^\infty(v))$.

Proof by a similar reasoning.

Thus pure memoryless determinacy is established
and the value of t in $\mathcal{G}_v$ is $f_t(f_s^\infty(v))$.

Plan

Plan

Presentation

Pure Memoryless Determinacy

Discounted Games

Mean Payoff Games

Priority (and Parity) Games

3 Computational issues

- Complexity results

From Mean Payoff Games to Discounted Games

From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

Mean payoff and discounted games

Consider the following decision problem.

Input. A mean payoff or discounted game $\mathcal{G}$ and a value v.

Output. Is $val_{\mathcal{G}}(s) \geq v$?

This problem is in NP**.**

Guess a pure memoryless strategy σ of Max.

Build the MDP $\mathcal{G}_\sigma$.

Minimize (in polynomial time) the objective o.

Answer yes if $o \geq v$.

This problem is in coNP**.**

Guess a pure memoryless strategy τ of Min.

Build the MDP $\mathcal{G}_\tau$.

Maximize (in polynomial time) the objective o.

Answer no if $o < v$.

Value iteration for discounted game (1)

The algorithm

$\mathbf{v} \leftarrow \mathbf{0}$

For i **from** 1 to n **do** $\mathbf{v} \leftarrow L(\mathbf{v})$ // n is precomputed.

Select σ and τ the optimal policies w.r.t. 1-step horizon and value $\lambda\mathbf{v}$

Denote:

- $\mathbf{P}$ the transition matrix of $\mathcal{G}_{\sigma,\tau}$;
- $\mathbf{r}$ defined by $\mathbf{r}[s] = r(s, a_s)$ where a_s is selected by σ or τ.

Then $val_{\mathcal{G}}[s] = ((\mathbf{Id} - \lambda\mathbf{P})^{-1}\mathbf{r})[s]$.

Value iteration for discounted game (2)

Analysis of the game

Let σ and τ be some pure memoryless policies and the DTMC $\mathcal{G}^{\sigma,\tau}$.
Denote $\mathbf{P}$ its transition matrix and $\mathbf{r}$ as above.
Then $\mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))$ is the unique solution $(\mathbf{Id} - \lambda\mathbf{P})X = \mathbf{r}$.

So:

- Compute β the product of the denominators of the probabilities and rewards occurring in $\mathcal{G}$ and λ in polynomial time.
- Rewrite all values (including $1 - \lambda a$ for appropriate a's) as $\frac{\alpha}{\beta}$.
- So β is an upper bound of the $|\alpha|$'s.
- Omit β without changing the equation system.
- Denote $B = |S|!\beta^{|S|}$.
- Then any $\mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))[s]$ can be written as $\frac{c}{d}$ for some $d \leq B$.

Value iteration for discounted games (3)

Which value for n?

L the contracting operator fulfills $\|val_{\mathcal{G}} - L^n(\mathbf{0})\| \leq \frac{\lambda^n}{1-\lambda}$.

Select n such that: $\frac{\lambda^n}{1-\lambda} < \frac{1}{2B^2}$.

Let σ and τ be the policies returned by the algorithm. They fulfill:
$\|val_{\mathcal{G}} - \mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))\| \leq \frac{2\lambda^n}{1-\lambda} < \frac{1}{B^2}$ *(reasoning as in MDP)*

Different values provided by two pairs of strategies differ form at least $\frac{1}{B^2}$.
So: $val_{\mathcal{G}} = \mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))$.

Analysis

Write $\lambda = \frac{p}{q}$. Then $\log_2(\frac{1}{\lambda}) \geq \log_2(1 + \frac{1}{p}) \geq \frac{1}{p}$ and $\log_2(\frac{1}{1-\lambda}) \leq \log_2(q)$.

So $n > p(\log_2(q) + 2\log_2(B) + 1)$ implies $\frac{\lambda^n}{1-\lambda} < \frac{1}{2B^2}$.

The value problem of a discounted game is in PTIME with *unary* λ or with $\lambda = \frac{1}{q}$.

Plan

Presentation

Pure Memoryless Determinacy

Discounted Games

Mean Payoff Games

Priority (and Parity) Games

3 Computational issues

Complexity results

- From Mean Payoff Games to Discounted Games

From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

Looking for large λ (1)

Scheme of the reduction.

- Computation of λ_∞ such that optimal strategies are Blackwell strategies;
- Solving the discounted game;
- Deduce the values of the mean payoff game from the optimal strategies

Computation of λ_∞ by analysis of the zeroes of $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h))$.

- $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h)) = (\mathbf{Id} - \lambda\mathbf{P})^{-1}\mathbf{r} - (\mathbf{Id} - \lambda\mathbf{P'})^{-1}\mathbf{r'}$
  for some $\mathbf{P}$, $\mathbf{P'}$, $\mathbf{r}$, $\mathbf{r'}$ with items occurring in $\mathcal{G}$.
- Let M be the product of denominators occurring in values of $\mathcal{G}$.
  and $X = 1 - \lambda$ with X in $]0, \frac{1}{2}]$.
- The items of $\mathbf{Id} - (1-X)\mathbf{P}$, $\mathbf{Id} - (1-X)\mathbf{P'}$, $\mathbf{r}$ and $\mathbf{r'}$ can be written as $aX + b$ with numerators of a and b bounded by M and denominator M.
- Looking for zeroes one may omit the common denominator.

Looking for large λ (2)

- $(\mathbf{Id} - (1-X)\mathbf{P})^{-1}\mathbf{r} - (\mathbf{Id} - (1-X)\mathbf{P'})^{-1}\mathbf{r'} = \frac{N}{D} - \frac{N'}{D'}$ with $N, D, N', D' \in \mathbb{Z}[X]$.

- Using Cramer's rule the coefficients of $ND' - N'D$ are bounded by:

$$R = 2n(n!)^4 M^{2n}$$

- Let $P \in \mathbb{Z}[X]$ whose coefficients are bounded by R. Then the smallest (if any) root of P in $]0, \frac{1}{2}]$ is at least $\frac{1}{2R}$.

- Thus an upper bound of λ_∞ is $1 - \frac{1}{2R+1}$.

- Since R has a polynomial size w.r.t. the size of $\mathcal{G}$ this reduction is polynomial.

Plan

Presentation

Pure Memoryless Determinacy

Discounted Games

Mean Payoff Games

Priority (and Parity) Games

3 Computational issues

Complexity results

From Mean Payoff Games to Discounted Games

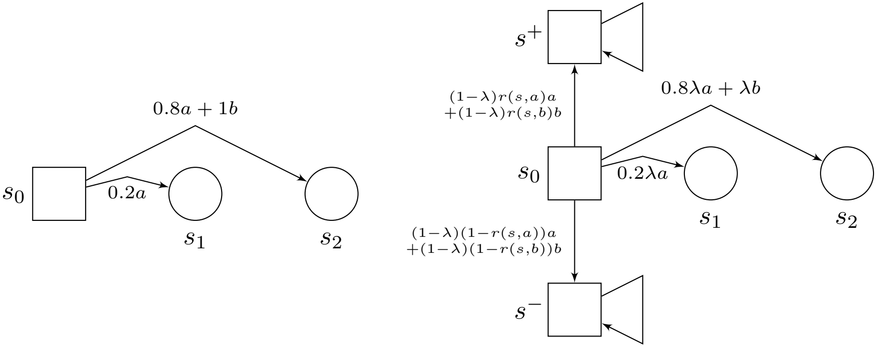
- From Discounted Games to Parity Games

From Parity Games to Mean Payoff Games

Reduction to reachability games

Let $\mathcal{G}$ a game with discount λ.

One builds $\mathcal{G}_\lambda$ with additional states s^+ and s^- and reachability target s^+.



Then for all s, σ and τ: $\mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = (1-\lambda)\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(h \text{ reaches } s^+)$

Plan

Presentation

Pure Memoryless Determinacy

Discounted Games

Mean Payoff Games

Priority (and Parity) Games

3. Computational issues

   Complexity results

   From Mean Payoff Games to Discounted Games

   From Discounted Games to Parity Games

   - From Parity Games to Mean Payoff Games

Scheme of the reduction

A two-step reduction

- Computing the states s for which $val_{\mathcal{G}}(s) \in \{0, 1\}$;
- Reduction to a mean-payoff game once these states are computed.

Let us call a game *pure* if there is no randomness in the game.

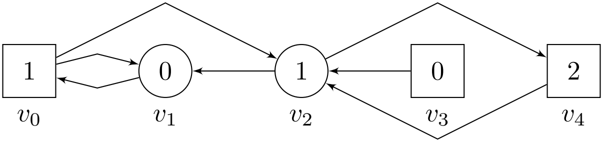
Computation of the states s for which $val_{\mathcal{G}}(s) \in \{0, 1\}$

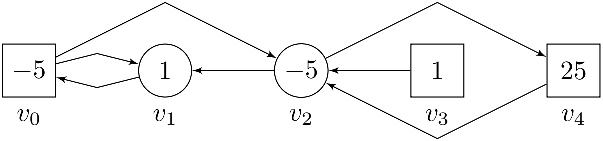
- Reduction to a pure parity game;
- Reduction from a pure parity game to a pure mean-payoff game.

From pure parity to mean payoff games

The value of a pure parity game $\mathcal{G}$ belongs to $\{0, 1\}$.

When $pri(s) = x$ in $\mathcal{G}$, $r(s, a) = (-m)^x$ in $\mathcal{G}'$ with $m = |S|$.





We claim that the mean payoff game $\mathcal{G}'$ fulfills $val_{\mathcal{G}'}(s) > 0$ iff $val_{\mathcal{G}}(s) = 1$.

Correctness of the reduction

- Let σ be a pure optimal strategy of Player Max in $\mathcal{G}$ and τ' a pure optimal strategy of Player Min in $\mathcal{G}'$.

$\mathcal{G}^{\sigma,\tau'}$ is a graph where any vertex has exactly one successor.

From s one reaches a circuit. Let p be the maximal priority occurring in the circuit.

If p is even then the $\mathbf{E}^{\sigma,\tau'}_{\mathcal{G}',s}(r(h)) \geq m^p - (m-1)m^{p-1} > 0$.

Thus $val_{\mathcal{G}}(s) = 1$ implies $val_{\mathcal{G}'}(s) > 0$.

- Let τ be a pure optimal strategy of Player Min in $\mathcal{G}$ and σ' a pure optimal strategy of Player Max in $\mathcal{G}'$.

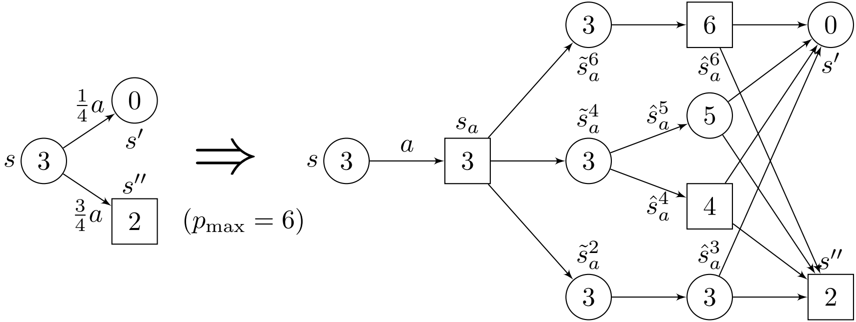
$\mathcal{G}^{\sigma',\tau}$ is a graph where any vertex has exactly one successor.

From s one reaches a circuit. Let p be the maximal priority occurring in the circuit.

If p is odd then $\mathbf{E}^{\sigma',\tau}_{\mathcal{G}',s}(r(h)) \leq -m^p + (m-1)m^{p-1} < 0$;

Thus $val_{\mathcal{G}}(s) = 0$ implies $val_{\mathcal{G}'}(s) < 0$.

From parity to pure parity games



Let $p_{\max}$ be the maximal priority assumed to be even w.l.o.g.

For all $s \in S$ with $pri(s) = p$ and $a \in A_s$:

- Add to S_{Max}: $\tilde{s}_a^q$ with $q \geq p - 1$ and q even and $\hat{s}_a^q$ with $q \geq p$ and q odd;
- Add to S_{Min}: s_a and $\hat{s}_a^q$ with $q \geq p$ and q even.

$$pri(s_a) = pri(\tilde{s}_a^q) = p \text{ and } pri(\hat{s}_a^q) = q.$$

The set of edges is:

- (s, s_a) and $(s_a, \tilde{s}_a^q)$;
- $(\tilde{s}_a^q, \hat{s}_a^q)$ and $(\tilde{s}_a^q, \hat{s}_a^{q+1})$ *when defined*;
- $(\hat{s}_a^q, s')$ when $p(s'|s, a) > 0$.

Property of E (1)

Let E' (resp. O') the winning set of player Max (resp. Min) in $\mathcal{G}'$.
Let $E = E' \cap S$ and $O = O' \cap S$.

$$\boxed{\text{For all } s \in E,\ val_{\mathcal{G}}(s) = 1.}$$

Proof.

• Let σ be a pure memoryless optimal strategy of Max in $\mathcal{G}'$.
We claim that in the MDP $\mathcal{G}^\sigma$, one never leaves E.
Assume there exists $s \in E$ and $a \in A_s^\sigma$ such that $p(s'|s, a) > 0$ and $s' \in O$.
In $\mathcal{G}'$ (after possibly selecting a),

- in s_a, Min could select $\tilde{s}_a^{p_{\max}}$;
- and in $\hat{s}_a^{p_{\max}}$ Min could select $s' \in O$, a contradiction.

• Let τ be a pure memoryless optimal strategy of Min in the MDP $\mathcal{G}^\sigma$.
Consider $\mathcal{M}$ the Markov chain $\mathcal{G}^{\sigma,\tau}$ restricted to E.

Property of E (2)

Proof (continued).

Assume there exists $\mathcal{C}$ a terminal s.c.c. of $\mathcal{M}$

whose maximal priority is odd, say $2r+1$ for state s_0.

Let τ' be (partially) defined as follows. For all $s \in \mathcal{C} \cap S_{\mathsf{Min}}$, $\tau'(s) = \tau(s)$.

Let $\mathcal{C}^\bullet = \{s_a \mid s \in \mathcal{C} \cap S \; a \in A_s$ is selected by σ or $\tau\}$.

For all $s_a \in \mathcal{C}^\bullet$:

- $\tau'(s_a) = \tilde{s}_a^{2r}$;
- if $\sigma(\tilde{s}_a^{2r}) = \hat{s}_a^{2r}$ then $\tau'(\hat{s}_a^{2r}) = s'$
  with s' minimizing the distance to s_0 in $\mathcal{G}^{\sigma,\tau}$.

Consider in $\mathcal{G}'$ the set of states $S^* = \mathcal{C} \cup \mathcal{C}^\bullet \cup \{\tilde{s}_a^{2r}, \sigma(\tilde{s}_a^{2r}) \mid s_a \in \mathcal{C}^\bullet\}$.

Observe that for all $t \in S^*$, $pri(t) \leq 2r+1$.

Every state in S^* has exactly one successor defined by σ or τ' still in S^*.

Consider any circuit in the induced graph:

- either some state $\hat{s}_a^{2r+1}$ occurs in the circuit;
- or s_0 occurs in the circuit.

Thus $S^* \cap E' = \emptyset$ which contradicts the definition of $\mathcal{M}$.

Property of O (1)

$\boxed{\text{For all } s \in O,\ val_{\mathcal{G}}(s) < 1.}$

Proof.

Let τ be a pure memoryless optimal strategy of Min in $\mathcal{G}'$ and the MDP $\mathcal{G}^{\tau}$.

Let σ be a pure memoryless optimal strategy of Max in $\mathcal{G}^{\tau}$ and the DTMC $\mathcal{G}^{\sigma,\tau}$.

Let $\mathcal{H}$ be the graph over S', the set of vertices, defined by:

- If $s \in S_{\mathsf{Max}}$ (resp. $t \in S'_{\mathsf{Min}}$) then $(s, s_{\sigma(s)})$ (resp. $(t, \tau(t))$ is an edge;
- for other t, any edge (t, t') of $\mathcal{G}'$ is an edge.

Let $s_0 \in O$ belonging to a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$.

By construction, $\mathcal{C} \subseteq O'$ and the maximal priority in $\mathcal{C}$ is odd.

• We prove by induction that for all s reachable from s_0 in $\mathcal{G}^{\sigma,\tau}$, $s \in \mathcal{C}$.

Let $a \in A_s$ be selected either by σ or τ.

Then in s_a, τ does not select $\tilde{s}_a^{p_{\max}}$. Otherwise $\hat{s}_a^{p_{\max}}$ would belong to $\mathcal{C}$.

Let $\tilde{s}_a^{2\ell}$ be selected by τ.

Then $\hat{s}_a^{2\ell+1}$ belongs to $\mathcal{C}$ and so all s' with $p(s'|s,a) > 0$ belongs to $\mathcal{C}$.

Thus in $\mathcal{G}^{\sigma,\tau}$, s_0 belongs to a terminal s.c.c. with all states in O.

Property of O (2)

Proof (continued).

- We claim that for all $s \in O$, there is a positive probability in $\mathcal{G}^{\sigma,\tau}$ to reach a state $s' \in O$ such that s' belongs to a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$.

We prove it by induction on the length of a path from s along O' to some $s' \in O$ of a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$.

Assume the shortest path starts by $s s_a \tilde{s}_a^r \hat{s}_a^\ell s'$.
for some a selected either by σ or τ, and some r and some ℓ.

Then $p(s'|s,a) > 0$.

Thus, for all $s \in O$ there is a positive probability in $\mathcal{G}^{\sigma,\tau}$ to reach a terminal s.c.c. with all states in O.

Property of O (3)

Proof (continued).

Assume there exists $\mathcal{C}$ a terminal s.c.c. of $\mathcal{G}^{\sigma,\tau}$ with all states in O whose maximal priority is even, say $2r$ for state s_0.

Let σ' be (partially) defined as follows. For all $s \in \mathcal{C} \cap S_{\mathsf{Max}}$, $\sigma'(s) = \sigma(s)$.

Let $\mathcal{C}^\bullet = \{s_a \mid s \in \mathcal{C} \cap S \ a \in A_s \text{ is selected by } \sigma \text{ or } \tau\}$.

For all $s_a \in \mathcal{C}^\bullet$:

- If $\tau(s_a) = \tilde{s}_a^{2\ell}$ with $\ell \geq r$ then $\sigma'(\tilde{s}_a^{2\ell}) = \hat{s}_a^{2\ell}$;
- If $\tau(s_a) = \tilde{s}_a^{2\ell}$ with $\ell < r$ then $\sigma'(\tilde{s}_a^{2\ell}) = \hat{s}_a^{2\ell+1}$ and $\sigma'(\hat{s}_a^{2\ell+1}) = s'$ with s' minimizing the distance to s_0 in $\mathcal{G}^{\sigma,\tau}$.

Consider in $\mathcal{G}'$ the set of states $S^* = \mathcal{C} \cup \mathcal{C}^\bullet \cup \{\tau(s_a), \sigma'(\tau(s_a)) \mid s_a \in \mathcal{C}^\bullet\}$.

Every state in S^* has exactly one successor defined by σ' or τ still in S^*.

Consider the maximal priority of any circuit in the induced graph:

- either its is 2ℓ for some $\ell \geq r$ and state $\hat{s}_a^{2\ell}$;
- or it is $2r$ with s_0 occuring in the circuit.

Thus $S^* \cap O' = \emptyset$ which contradicts the definition of $\mathcal{C}$.

Interlude

Let $\mathcal{M}$ be an irreducible Markov chain
with m states and minimum positive transition probability δ.

Then for all $s \in S$,

$$\pi_\infty(s) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{Pr}(X_i = s) \geq \frac{1}{m}\delta^{m-1}$$

Proof.

Consider s_0, a state with maximal Cesaro-limit probability $\pi_\infty(s_0) \geq \frac{1}{m}$.

In the DTMC, there is a path of length $\ell \leq m-1$ from s_0 to s.

Thus:

$$\mathbf{Pr}(X_{i+\ell} = s) \geq \delta^\ell \mathbf{Pr}(X_i = s_0) \geq \delta^{m-1}\mathbf{Pr}(X_i = s_0)$$

Implying:

$$\pi_\infty(s) \geq \pi_\infty(s_0)\delta^{m-1} \geq \frac{1}{m}\delta^{m-1}$$

From parity to mean payoff game

Let $\mathcal{G}$ be a parity game with m states and δ minimal positive probability:
Define:

$$S_i = \{s \mid val_{\mathcal{G}} = i\} \text{ for } i \in \{0, 1\}$$

$\mathcal{G}'$ the mean payoff game with same structure as $\mathcal{G}$ is defined by:

- For all $s \in S_1$ and $a \in A_s$, $r(s, a) = 1$;
- For all $s \in S_0$ and $a \in A_s$, $r(s, a) = -1$;
- For all $s \notin S_0 \cup S_1$ with $p = pri(s)$ and $a \in A_s$, $r(s, a) = (\frac{-2m}{\delta^{m-1}})^p$;

Observation. This reduction is performed in polynomial time.

Then, for all $s \in S$

$$val_{\mathcal{G}'}(s) = 2val_{\mathcal{G}}(s) - 1$$

Correctness of the reduction (1)

Proof. $val_{\mathcal{G}'}(s) \geq 2val_{\mathcal{G}}(s) - 1$

- Let σ (resp. τ) be a pure optimal strategy of Player Max (resp. Min) in $\mathcal{G}$ and σ' (resp. τ') be a pure optimal strategy of Player Max (resp. Min) in $\mathcal{G}'$.

Observations

- Under strategy σ (resp. τ), the game never leaves S_1 (resp. S_0).
- $val_{\mathcal{G}}(s) \leq 1 - \mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } S_0)$ since by combining τ and τ', Min ensures a value no more than $1 - \mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } S_0)$.
- Let $\mathcal{C}$ be a terminal s.c.c. of $\mathcal{G}^{\sigma,\tau'}$. Then:
  - either $S_1 \cap \mathcal{C} \neq \emptyset$, since σ never leaves S_1 $\mathcal{C} \subseteq S_1$ and thus $val_{\mathcal{G}'}(t) = 1$ for all $t \in \mathcal{C}$;
  - either $S_0 \cap \mathcal{C} \neq \emptyset$ since for all $s \in \mathcal{C}$, $\mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } S_0) = 1$ implying $val_{\mathcal{G}}(s) = 0$, thus $\mathcal{C} \subseteq S_0$ and $val_{\mathcal{G}'}(t) = -1$ for all $t \in \mathcal{C}$;
  - or $\mathcal{C} \cap (S_0 \cup S_1) = \emptyset$ with all state $t \in \mathcal{C}$ fulfilling $0 < val_{\mathcal{G}}(t) < 1$.

Let us denote $\mathcal{C}_0$ the union of the terminal s.c.c. included in S_0.

Correctness of the reduction (2)

Proof (continued).

Let $\mathcal{C}$ be terminal s.c.c. that fufills $\mathcal{C} \cap (S_0 \cup S_1) = \emptyset$.

Thus $z \in \mathcal{C}$, a vertex with maximal priority, fulfills $p \stackrel{\text{def}}{=} pri(z)$ is even.

- When $p = 0$, for all $t \in \mathcal{C}$, $r(t, a) = 1$. So one immediately gets $\mathbf{E}^{\sigma,\tau'}_{\mathcal{G},t}(r(h)) = 1$.
- When $p > 0$, the contribution of z to the mean payoff reward is at least:

$$\frac{1}{m}\delta^{m-1}(\frac{2m}{\delta^{m-1}})^p = 2(\frac{2m}{\delta^{m-1}})^{p-1}$$

The accumulated contribution of all $t \in \mathcal{C} \setminus \{z\}$ is at least: $-(\frac{2m}{\delta^{m-1}})^{p-1}$.

So for all $t \in \mathcal{C}$, $\mathbf{E}^{\sigma,\tau'}_{\mathcal{G},t}(r(h)) \geq (\frac{2m}{\delta^{m-1}})^{p-1} \geq 1$. Thus:

$$\begin{aligned}
val_{\mathcal{G}'}(s) \geq & \quad -\mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } \mathcal{C}_0) + (1 - \mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } \mathcal{C}_0)) \\
= & \quad 1 - 2\mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } \mathcal{C}_0) \\
\geq & \quad 1 - 2\mathbf{Pr}^{\sigma,\tau'}_{\mathcal{G},s}(h \text{ reaches } \mathcal{S}_0) \\
\geq & \quad 1 - 2(1 - val_{\mathcal{G}}(s)) \\
= & \quad 2val_{\mathcal{G}}(s) - 1
\end{aligned}$$

One gets $val_{\mathcal{G}'}(s) \leq 2val_{\mathcal{G}}(s) - 1$ by a similar reasoning about $\mathcal{G}^{\sigma',\tau}$.