A polynomial algorithm for the deterministic mean payoff game

Bruno Scherrer

March 21, 2022

Abstract

...

We consider an infinite-horizon game on a directed graph (X, E) between two players, MAX and MIN. For any vertex x, we write $E(x) = \{y; (x, y) \in E\}$ for the set of vertices that can be reached from x by following one edge and we assume $E(x) \neq \emptyset$. The set of vertices $X = \{1, 2, \ldots, n\}$ of the graph is partitionned into the sets X_+ and X_- of nodes respectively controlled by MAX and MIN. The game starts in some vertex x_0. At each time step, the player who controls the current vertex chooses a next vertex by following an edge. So on and so forth, the choices generate an infinitely long trajectory $(x_0, x_1, \ldots)$. In the mean payoff game, the goal of MAX is to maximize

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} r(x_t),$$

while that of MIN is to minimize

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} r(x_t).$$

As a proxy to solve the mean payoff game, our technical developments will mainly consider the γ-discounted payoff for some $0 \leq \gamma < 1$, where the goal of MAX is to maximize

$$(1-\gamma) \sum_{t=0}^{\infty} \gamma^t r(x_t)$$

while that of MIN is to minimize this quantity.

LITERATURE

1 Preliminaries

Let M and N be the set of stationary policies for MAX and MIN:

$$M = \{\mu : X_+ \to X \ ; \ \forall x \in X_+, \ \mu(x) \in E(x)\},$$
$$N = \{\nu : X_- \to X \ ; \ \forall x \in X_-, \ \nu(x) \in E(x)\}.$$

For any policies $\mu \in M$ and $\nu \in N$, let us write $P_{\mu,\nu}$ for the transition matrix induced by μ and ν:

$$\forall x \in X_+, \forall y \in X, \quad P_{\mu,\nu}(x, y) = \mathbb{1}_{\mu(x)=y},$$
$$\forall x \in X_-, \forall y \in X, \quad P_{\mu,\nu}(x, y) = \mathbb{1}_{\nu(x)=y}.$$

Seeing the reward $r : X \to 0, 1, \ldots, R$ and any function $v : X \to \mathbb{R}$ as vectors of $\mathbb{R}^n$, consider the following Bellman operators

$$\begin{aligned} T_{\mu,\nu}v &= (1-\gamma)r + \gamma P_{\mu,\nu}v, \\ T_\mu v &= \min_\nu T_{\mu,\nu}v, \\ \tilde{T}_\nu v &= \max_\mu T_{\mu,\nu}v, \\ Tv &= \max_\mu T_\mu v = \min_\nu \tilde{T}_\nu v. \end{aligned}$$

that are γ-contractions with respect to the max-norm $\|\cdot\|$, defined for all $u \in R^n$ as $\|u\| = \max_{x\in X} |u(x)|$. For any policies $\mu \in M$ and $\nu \in N$, the value $v_{\mu,\nu}(x)$ obtained by following policies μ and ν satisfies

$$v_{\mu,\nu} = (1-\gamma)\sum_{t=0}^{\infty}(\gamma P_{\mu,\nu})^t r = (1-\gamma)(I-\gamma P_{\mu,\nu})^{-1}r,$$

and is the only fixed point of the operator $T_{\mu,\nu}$. Given any policy μ for MAX, the minimal value that MIN can obtain

$$v_\mu = \min_\nu v_{\mu,\nu}$$

is the fixed point of the operator T_μ, and it is well known that any policy ν_+ for MIN such that $T_{\mu,\nu_+}v_\mu = T_\mu v_\mu = v_\mu$ is optimal. Symmetrically, given any policy ν for MIN, the maximal value that MAX can obtain

$$\tilde{v}_\mu = \max_\nu v_{\mu,\nu}$$

is the fixed point of $\tilde{T}_\nu$, and it is well known that any policy μ_+ for MAX such that $T_{\mu_+,\nu}v_\mu = \tilde{T}_\nu\tilde{v}_\nu = \tilde{v}_\nu$ is optimal. The optimal value

$$v_* = \max_\mu \min_\nu v_{\mu,\nu}$$

is the fixed point of the operator T. Let (μ_*, ν_*) be any pair of positional strategies such that $T_{\mu_*,\nu_*}v_* = Tv_*$. It is well-known that (μ_*, ν_*) is optimal.

2 Algorithm

Consider the following algorithm that iterates on policies of player MAX.

$$\begin{aligned} v_k &= T_{\mu_k}v_k, \\ T^n v_k &= T_{\vec{\mu}'_{k+1}}v_k \end{aligned}$$

3 Analysis

A local Bellman equation ? Our main technical result is the following observation.

Lemma 1. *For any v, such that $v \leq Tv$, find $\nu_1, \ldots, \nu_n$ be a set of policies such that*

$$T^n v = \tilde{T}_{\nu_1} \ldots \tilde{T}_{\nu_n} v.$$

Take any starting state x. By the pigeonhole principle, there necessarily exist i, c, y such that $0 \leq i < i+c \leq n$ and

$$\mathbb{1}_y = \mathbb{1}_x P_{\mu_*,\nu_1} \ldots P_{\mu_*,\nu_i} = \mathbb{1}_x P_{\mu_*,\nu_1} \ldots P_{\mu_*,\nu_{i+c}},$$

for which we have

$$\mathbb{1}_x(v_* - T^n v) \leq \mathbb{1}_x(T^n v - v) + \frac{\gamma}{1-\gamma}\mathbb{1}_y(T^n v - v).$$

Remark 1.

Proof. First, observe that by the monotonicity of T, and since $v \le Tv$, we have

$$v \le Tv \le T^2 v \le \cdots \le T^n v.$$

Let $\nu_1, \dots, \nu_n$ be a set of policies such that

$$T^n v = \tilde{T}_{\nu_1} \dots \tilde{T}_{\nu_n} v.$$

Take any starting state x. By the pigeonhole principle, there necessarily exist i, c, y such that $0 \le i < i + c \le n$ and

$$\mathbb{1}_y = \mathbb{1}_x P_{\mu_*,\nu_1} \dots P_{\mu_*,\nu_i} = \mathbb{1}_x P_{\mu_*,\nu_1} \dots P_{\mu_*,\nu_{i+c}}.$$

Consider a play where MAX uses μ_* and MIN uses the policy $\nu = \nu_1 \dots \nu_i (\nu_{i+1} \dots \nu_{i+c})^\infty$: the trajectory formed by this play is a path of length i followed by an infinitely repeated cycle of length c. By a telescoping argument, we have for any w,

$$\begin{aligned}
v_{\mu_*,\nu}(x) - w(x) &= \mathbb{1}_x (T_{\mu_*,\nu_1} \dots T_{\mu_*,\nu_{i+c}} w - w) + \gamma^{i+c} \mathbb{1}_y (I - \gamma^c P_{\mu_*,\nu_i} \dots P_{\mu_*,\nu_{i+c}})^{-1} (T_{\mu_*,\nu_{i+1}} \dots T_{\mu_*,\nu_{i+c}} w - w) \\
&= \mathbb{1}_x (T_{\mu_*,\nu_1} \dots T_{\mu_*,\nu_{i+c}} w - w) + \frac{\gamma^{i+c}}{1-\gamma^c} \mathbb{1}_y (T_{\mu_*,\nu_{i+1}} \dots T_{\mu_*,\nu_{i+c}} w - w) \\
&\le \mathbb{1}_x (\tilde{T}_{\nu_1} \dots \tilde{T}_{\nu_{i+c}} w - w) + \frac{\gamma^{i+c}}{1-\gamma^c} \mathbb{1}_y (\tilde{T}_{\nu_{i+1}} \dots \tilde{T}_{\nu_{i+c}} w - w)
\end{aligned}$$

Taking $w = \tilde{T}_{\nu_{i+c+1}} \dots T_{\nu_n} v$, and recalling the definition of $\nu_1, \dots, \nu_n$, we obtain

$$\begin{aligned}
v_{\mu_*,\nu}(x) - T^n v(x) &\le v_{\mu_*,\nu}(x) - w(x) \\
&\le \mathbb{1}_x (T^n v - v) + \frac{\gamma^{i+c}}{1-\gamma^c} \mathbb{1}_y (T^{n-i} v - v) \\
&\le \mathbb{1}_x (T^n v - v) + \frac{\gamma}{1-\gamma} \mathbb{1}_y (T^n v - v).
\end{aligned}$$

The result follows by noting that $v_*(x) \le v_{\mu_*,\nu}(x)$. □

$$\begin{aligned}
v_{k+1} - v_k &= (I - \gamma^n P_{\vec{\mu}_{k+1},\vec{\nu}_{k+1}})^{-1} (T_{\vec{\mu}_{k+1},\vec{\nu}_{k+1}} v_k - v_k) \\
&= \frac{1}{1-\gamma^n} (T_{\vec{\mu}_{k+1},\vec{\nu}_{k+1}} v_k - v_k) \\
&\ge \frac{1}{1-\gamma^n} (T_{\vec{\mu}_{k+1}} v_k - v_k) \\
&= \frac{1}{1-\gamma^n} (T^n v_k - v_k) \\
&\ge \frac{1-\gamma}{1-\gamma^n} (v_* - T^n v_k) \\
&\ge \frac{1}{n} (v_* - T^n v_k) \\
&= \frac{1}{n} (v_* - T_{\vec{\mu}_{k+1},\vec{\nu}_{k+1}} v_k) \\
&\ge \frac{1}{n} (v_* - v_{k+1}).
\end{aligned}$$

As a consequence:

$$\begin{aligned}
v_* - v_{k+1} &= v_* - v_k - (v_{k+1} - v_k) \\
&\le \left(1 - \frac{1}{n}\right) (v_* - v_k).
\end{aligned}$$