A strongly polynomial algorithm for mean payoff games

Bruno Scherrer

March 19, 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 $X = \{1, 2, \dots, n\}$ 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 $(x_0, x_1, \dots)$. We shall 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 $r : X \to 0, 1, \dots, R$ and any function $v : X \to \mathbb{R}$ as vectors of $\mathbb{R}^n$, consider the following Bellman operators

$$T_{\mu,\nu}v = (1-\gamma)r + \gamma P_{\mu,\nu}v,$$
$$Tv = \max_{\mu}\min_{\nu} T_{\mu,\nu}v.$$

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 $T_{\mu,\nu}$. 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 $T_{\mu_*,\nu_*} v_* = T v_*$. It is well-known that (μ_*, ν_*) is optimal.

2 Algorithm

Solve the n-step problem with terminal cost, i.e. identify a set of strategies $\mu_1, \dots, \mu_n$ and $\nu_1, \dots, \nu_n$ such that:

$$T^n 0 = T_{\mu_1,\nu_1} \dots T_{\mu_n,\nu_n} 0$$

Theorem 1. *For any state x, let p_x and c_x be the smallest integers such that*

$$\mathbb{1}_x P_{\mu_1,\nu_1} \dots P_{\mu_{p_x},\nu_{p_x}} = \mathbb{1}_x P_{\mu_1,\nu_1} \dots P_{\mu_{p_x+c_x},\nu_{p_x+c_x}}.$$

Then

$$v_*(x) = T_{\mu_1,\nu_1} \dots T_{\mu_{p_x},\nu_{p_x}} (T_{\mu_{p_x+1},\nu_{p_x+1}} \dot{(} T_{\mu_{p_x+1},\nu_{p_x+1}})^{\infty} 0.$$

Proof. Assume MIN uses $\nu_1, \dots, \nu_n$ to play n steps against the optimal policy μ_* of MAX from x. Consider the $n+1$ vertices visited:

$$x_0 = x,\ x_1,\ x_2,\ \dots,\ x_n.$$

Since there are n different vertices, by the pigeonhole principle, there necessarily exists $0 \leq p < p + c \leq n$ such that $x_p = x_{p+c}$.

Now, assume that against μ_*, MIN uses the strategy $\bar{\nu} = \nu_1, \dots, \nu_p, (\nu_{p+1} \dots \nu_{p+c})^{\infty}$ The trajectory is made of a path followed by a cycle of length c that is repeated infinitely often:

$$\underbrace{x_0 = x,\ x_1,\ x_2,\ \dots, x_{i-1}}_{\text{path}},\ \underbrace{x_i,\ x_{i+1},\ \dots,\ x_{j-1}}_{\text{cycle}},\ \underbrace{x_i,\ x_{i+1},\ \dots, x_{j-1}}_{\text{cycle}},\ \dots$$

The value of this game satisfies for any w,

$$\begin{aligned}
v_{\mu_*,\bar{\nu}}(x) - w(x) &= \mathbb{1}_x (T_{\mu_*,\vec{\nu}_p \vec{\nu}_c} (T_{\mu_*,\vec{\nu}_c})^{\infty} w - w) \\
&= \mathbb{1}_x T_{\mu_*,\vec{\nu}_p \vec{\nu}_c} 0 + \gamma^j \mathbb{1}_{x_i} \sum_{k=0}^{\infty} [(T_{\mu_*,\vec{\nu}_c})^{k+1} w - T_{\mu_*,\vec{\nu}_c})^k w] \\
&= \mathbb{1}_x T_{\mu_*,\vec{\nu}_p \vec{\nu}_c} w + \gamma^j \mathbb{1}_{x_i} \sum_{k=0}^{\infty} \gamma^{(j-i)k} (P_{\mu_*,\vec{\nu}_c})^k (T_{\mu_*,\vec{\nu}_c} w - w) \\
&= \mathbb{1}_x T_{\mu_*,\vec{\nu}_p \vec{\nu}_c} w + \frac{\gamma^j}{1 - \gamma^{j-i}} \mathbb{1}_{x_i} (T_{\mu_*,\vec{\nu}_c} w - w) \\
&\leq \mathbb{1}_x \tilde{T}_{\vec{\nu}_p \vec{\nu}_c} w + \frac{\gamma^j}{1 - \gamma^{j-i}} \mathbb{1}_{x_i} (\tilde{T}_{\vec{\nu}_c} w - w).
\end{aligned}$$

Taking $w = \tilde{T}_{\vec{\nu}_{p'}} v$, we obtain

$$\begin{aligned}
v_{\mu_*,\bar{\nu}}(x) - [\tilde{T}_{\vec{\nu}_{p'}} v](x) &\leq \mathbb{1}_x(\tilde{T}_{\vec{\nu}_p \vec{\nu}_c} \tilde{T}_{\vec{\nu}_{p'}} v - T_{\vec{\nu}_{p'}} v) + \frac{\gamma^j}{1-\gamma^{j-i}} \mathbb{1}_{x_i}(\tilde{T}_{\vec{\nu}_c} \tilde{T}_{\vec{\nu}_{p'}} v - \tilde{T}_{\vec{\nu}_{p'}} v) \\
&= \mathbb{1}_x(\tilde{T}_{\vec{\nu}_p \vec{\nu}_c \vec{\nu}_{p'}} v - T_{\vec{\nu}_{p'}} v) + \frac{\gamma^j}{1-\gamma^{j-i}} \mathbb{1}_{x_i}(\tilde{T}_{\vec{\nu}_c \vec{\nu}_{p'}} v - \tilde{T}_{\vec{\nu}_{p'}} v) \\
&= \mathbb{1}_x(T^n v - T^{n-j} v) + \frac{\gamma^j}{1-\gamma^{j-i}} \mathbb{1}_{x_i}(T^{n-i} v - T^{n-j} v) \\
&\leq \mathbb{1}_x(T^n v - v) + \frac{\gamma^j}{1-\gamma^{j-i}} \mathbb{1}_x(T^n v - v) \\
&\leq \frac{\epsilon}{1-\gamma},
\end{aligned}$$

where we eventually used the facts that $T^n v - v \leq \epsilon$, $j \geq 1$ and $j - i \geq 1$. The result follows by the facts that $v_*(x) = v_{\mu_*,\nu_*}(x) \leq v_{\mu_*,\bar{\nu}}(x)$ and $T^n v \geq T^{n-j} v = \tilde{T}_{\vec{\nu}_{p'}} v$. □