Towards a strongly polynomial algorithm for deterministic payoff games ?

Bruno Scherrer*

April 1, 2022

Abstract

Given a zero-sum two-player γ-discounted deterministic game with n states, we try to build an algorithm that is polynomial on n (and independent of γ).

Consider a zero-sum two-player γ-discounted game with n states and m transitions, and its corresponding Bellman operators:

$$\begin{aligned} T_{\mu,\nu}v &= r + \gamma P_{\mu,\nu}v, \\ T_{\mu}v &= \min_{\nu \in N} T_{\mu,\nu}v, \\ \tilde{T}_{\nu} &= \max_{\mu \in M} T_{\mu,\nu}v, \\ Tv &= \max_{\mu \in M} T_{\mu}v = \min_{\nu \in N} \tilde{T}_{\nu}v. \end{aligned}$$

It is well-known that the optimal value v_* is the only fixed point of T and that any pair of stationary policies (μ_*, ν_*) such that $T_{\mu_*,\nu_*}v_* = v_*$ form a pair of optimal policies.

We shall consider non-stationary policies $\vec{\mu} = (\mu_1, \dots, \mu_\ell) \in M^\ell$ and $\vec{\nu} = (\nu_1, \dots, \nu_\ell) \in N^\ell$. The operators above can be extended straightforwardly to this kind of policies:

$$\begin{aligned} P_{\vec{\mu},\vec{\nu}} &= P_{\mu_1,\nu_1} \dots P_{\mu_\ell,\nu_\ell}, \\ T_{\vec{\mu},\vec{\nu}}v &= T_{\mu_1,\nu_1} \dots T_{\mu_\ell,\nu_\ell}v, \\ T_{\vec{\mu}}v &= \min_{\vec{\nu} \in N^\ell} T_{\vec{\mu},\vec{\nu}}v = T_{\mu_1} \dots T_{\mu_\ell}v, \\ \tilde{T}_{\vec{\nu}}v &= \max_{\vec{\mu} \in M^\ell} T_{\vec{\mu},\vec{\nu}}v = \tilde{T}_{\nu_1} \dots \tilde{T}_{\mu_\ell}v, \\ T^\ell v &= \max_{\vec{\mu} \in M^\ell} T_{\vec{\mu}}v = \min_{\vec{\nu} \in N^\ell} \tilde{T}_{\vec{\nu}}v. \end{aligned}$$

For any stationary policy μ or ν, we shall write μ^ℓ and ν^ℓ for their non-stationary clones $(\mu, \mu, \dots, \mu)$ and $(\nu, \nu, \dots, \nu)$.

Let

$$I = \{ \ (i,j) \ ; 1 \leq i \leq j \leq n \ \}.$$

For any non-stationary policies $\vec{\mu} = (\mu_1, \dots, \mu_n) \in M^n$ and $\vec{\nu} = (\nu_1, \dots, \nu_n) \in N^n$, for all $(i,j) \in I$, we shall write $\vec{\mu}_i^j$ and $\vec{\nu}_i^j$ for the sub-policies:

$$\begin{aligned} \vec{\mu}_i^j &= \mu_i \mu_{i+1} \dots \mu_{j-1} \mu_j, \\ \vec{\nu}_i^j &= \nu_i \nu_{i+1} \dots \nu_{j-1} \nu_j. \end{aligned}$$

*INRIA, Université de Lorraine, bruno.scherrer@inria.fr

1 A Policy Iteration algorithm

Take an arbitrary stationary policy μ_0. Initialize the set $C_0 = \emptyset$. We shall describe how we compute C_{k+1} and μ_{k+1} from C_k and μ_k.

Let v_k be the value of μ_k against its best adversary:

$$v_k = T_{\mu_k} v_k = \min_{\nu} T_{\mu_k,\nu} v_k.$$

Compute the set of policies that avoid the cycles of C_k:

$$M_{C_k} = \{\ \vec{\mu} \in M^n\ ;\ \forall \vec{\nu} \in N^n,\ \forall (x,c) \in C_k,\ \forall (i,j) \in I,\ \mathbb{1}_x P_{\vec{\mu}_i^j, \vec{\nu}_i^j} \neq \mathbb{1}_x \}.$$

Identify policies $\vec{\mu} \in M_{C_k} \cup \{(\mu_k)^n\}$ and $\vec{\nu} \in N^n$ such that

$$\max_{\vec{\mu}' \in M_{C_k} \cup \{(\mu_k)^n\}} T_{\vec{\mu}'} v = T_{\vec{\mu},\vec{\nu}} v.$$

If $T_{\vec{\mu},\vec{\nu}} v_k = v_k$, stop (and output μ_k).

For every x, there exists a minimal pair $(i_x, j_x) \in I$ and y_x such that the trajectory first reach a loop of length $c_x = j_x - i - x + 1$ involving y_x, i.e. such that

$$\mathbb{1}_x P_{\vec{\mu}_1^{i_x - 1}, \vec{\nu}_1^{i_x - 1}} = \mathbb{1}_x P_{\vec{\mu}_1^{j_x}, \vec{\nu}_1^{j_x}} = \mathbb{1}_{y_x} = \mathbb{1}_y P_{\vec{\mu}_{i_x}^{j_x}, \vec{\nu}_{i_x}^{j_x}}.$$

We take

$$C_{k+1} = C_k \cup \{(y_x, c_x)\ ;\ \mathbb{1}_{y_x} (T_{\vec{\mu}_{i_x}^n, \vec{\nu}_{i_x}^n} v_k - T_{\vec{\mu}_{j_x+1}^n, \vec{\nu}_{j_x+1}^n} v_k) = 0\}.$$

2 Analysis of the 1-player case

Let us first consider the situation of a 1-player game (where $N = \nu$). We shall omit all references to ν for clarity.

3 Analysis of the 2-player case

3.1 Monotonicity

We begin by a monotonicity property:

Lemma 1. *For all k, and all $1 \leq i \leq j \leq n$,*

$$v_{k+1} \geq w_{k,i,j} \geq v_k.$$

Proof. Since $v_k \leq T v_k$, by monotonicity of the operator T, we have

$$T^n v_k \geq T^{n-1} v_k \geq \cdots \geq T v_k \geq v_k.$$

Therefore, for every $1 \leq i \leq j \leq n$, writing $c = j - i + 1$, we have for any

$$\begin{aligned} w_{k,i,j} - v_k &\geq v_{ij} - T^{n-j} v_k \\ &= (I - \gamma^c P_{\vec{\mu}_i^j, \nu_i^j})^{-1} (T_{\vec{\mu}_i^j, \vec{\nu}_i^j} T^{n-j} v_k - T^{n-j} v_k) \\ &= (I - \gamma^c P_{\vec{\mu}_i^j, \nu_i^j})^{-1} (\underbrace{T^{n-i+1} v_k - T^{n-j} v_k}_{\geq 0}) \\ &\geq 0. \end{aligned}$$

We deduce that $w_k \geq v_k$.

Now take any $1 \leq i \leq j \leq n$ and $c = j-i+1$. To finish the proof, we are going to prove that $v_{k+1} \geq w_{k,i,j}$. By monotonicity of $T_{\mu_{k+1}}$, we have for all $i \leq \ell \leq j$,

$$T_{\mu_{k+1}} w_k \geq T_{\mu_{k+1}} T_{\vec{\mu}_\ell^j} w_{k,i,j}.$$

□

3.2 Strong contraction

Consider the following sets of policies:

$$\begin{aligned} N_{x,c}(\mu) &= \{\ \vec{\nu} \in N^c\ ;\ \mathbb{1}_x P_{\mu^c,\vec{\nu}} = \mathbb{1}_x \}, \\ M_{x,c} &= \{\ \mu\ ;\ \ \arg\min v_{\mu,\nu} \cap N_{x,c}(\mu) \neq \emptyset \}. \end{aligned}$$

Observe that

$$\bigcup_{x,c} M_{x,c} = M.$$

For every x, there exist i_x, j_x, c_x such that $1 \leq i_x < j_x \leq n$, and

$$\mathbb{1}_x P_{\vec{\mu}_{i_x}^{j_x}, \vec{\nu}_{i_x}^{j_x}} = \mathbb{1}_x.$$

Take a $\mu \in M_{x,c}$. Then

$$\begin{aligned} v_\mu(x) - v(x) &= \min_\nu v_{\mu,\nu}(x) - v(x) \\ &= \min_{\nu \in N_{x,c}(\mu)} v_{\mu,\nu}(x) - v(x) \\ &= \min_{\nu \in N_{x,c}(\mu)} \mathbb{1}_x (I - (\gamma P_{\mu,\nu})^c)^{-1} (T^c_{\mu,\nu} v - v) \\ &= \min_{\nu \in N_{x,c}(\mu)} \frac{1}{1-\gamma^c} \mathbb{1}_x (T^c_{\mu,\nu} v - v). \end{aligned}$$

When running $\vec{\mu}$ against its adversary, there exists a x such that

$$v_{\vec{\mu}}(x) - v(x) \leq \frac{1}{n(1-\gamma)} \mathbb{1}_x (T^n v - v)$$

For all policies such $\mu_+ \in M_x(v)$,

$$\begin{aligned} v_{\mu_+}(x) - v_{\vec{\mu}}(x) &= v_{\mu_+}(x) - v(x) + v(x) - v_{\vec{\mu}}(x) \\ &\leq (1 - \frac{2}{n})(v_{\mu_+}(x) - v(x)) \end{aligned}$$