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, I try to build an algorithm that is polynomial on n (and independent of γ). Here, I describe a contraction property that is 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 $\vec{\mu} = (\mu_1, \ldots, \mu_\ell) \in M^\ell$ and $\vec{\nu} = (\nu_1, \ldots, \nu_\ell) \in N^\ell$. The operators above can be extended straightforwardly to this kind of policies:

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

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

Let

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

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

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

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

Let

$$C = \{ \ (x, c) \ ; \ x \in X, \ 1 \leq c \leq n \ \}.$$

For all $(x, c) \in C$, consider the following sets of policies:

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

Observe that

$$\bigcup_{(x,c) \in C} M_{x,c} = M.$$

For all $(x, c) \in C$, consider the following thresholds:

$$v_{x,c} = \max_{\mu \in M_{x,c}} \min_{\nu} v_{\mu,\nu}(x).$$

An improvement step Assume we are given μ_k. Let v_k be the value of μ_k against its best adversary:

$$v_k = T_{\mu_k} v_k = \min_{\vec{\nu} \in N^n} T_{\mu_k, \vec{\nu}} v_k.$$

We compute $\vec{\mu}$ and $\vec{\nu} \in N^n$ such that

$$T^n v = T_{\vec{\mu},\vec{\nu}} v.$$

One can finally "project" the policy $\vec{\mu} \in M^n$ to a stationary policy $\mu_{k+1} \in M$ that is at least as good as $\vec{\mu}$ by choosing any policy $\mu_{k+1} \in M$ that satisfies

$$T_{\mu_{k+1}} w_k = T w_k,$$

where

$$w_k = \max_{1 \leq \ell \leq n} T_{\vec{\mu}_\ell^n} v_{\vec{\mu}}$$

and $v_{\vec{\mu}}$ is the value of $\vec{\mu}$ against its optimal opponent $\vec{\nu}'$:

$$v_{\vec{\mu}} = T_{\vec{\mu}} v_{\vec{\mu}} = T_{\vec{\mu},\vec{\nu}'} v_{\vec{\mu}} = v_{\vec{\mu},\vec{\nu}'}.$$

Monotonicity One can see that $v_k \leq T v_k$, by monotonicity of the operator T, we have

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

We know that $\vec{\mu}$ is better than μ_k since:

$$\begin{aligned} v_{\vec{\mu}} - v_k &= v_{\vec{\mu},\vec{\nu}'} - v_k \\ &= (I - \gamma^n P_{\vec{\mu},\vec{\nu}'})^{-1} (T_{\vec{\mu},\vec{\nu}'} v_k - v_k) \\ &= (I - \gamma^n P_{\vec{\mu},\vec{\nu}'})^{-1} (\tilde{T}_{\vec{\nu}'} v_k - v_k) \\ &\geq (I - \gamma^n P_{\vec{\mu},\vec{\nu}'})^{-1} (\underbrace{T^n v_k - v_k}_{\geq 0}) \\ &\geq 0. \end{aligned}$$

“Strong” contraction There necessarily exists a state x and an integer $1 \leq \ell \leq n$, such that

$$\mathbb{1}_x (P_{\vec{\mu},\vec{\nu}'})^{\ell} = \mathbb{1}_x.$$

Now, there necessarily exist $(i,j) \in I$ and y such that:

$$\mathbb{1}_x P_{\vec{\mu}_1^{i-1},\vec{\nu}_1^{i-1}} = \mathbb{1}_x P_{\vec{\mu}_1^{j},\vec{\nu}_1^{j}} = \mathbb{1}_y = \mathbb{1}_y P_{\vec{\mu}_i^{j},\vec{\nu}_i^{j}}.$$

Let $c = j - i + 1$.

Lemma 1. *With the notations above,*

$$v_{\vec{\mu}}(x) - v_k(x) \geq \frac{1}{n^2}(v_{y,c} - v_k(y)).$$

Corollary 1. *We have the following contraction towards $v_{y,c}$:*

$$v_{y,c} - v_{k+1}(x) \leq \left(1 - \frac{1}{n^2}\right)(v_{y,c} - v_k(x)).$$

Proof. For the state x mentionned above, we have:

$$\begin{aligned}
v_{\vec{\mu}}(x) - v_k(x) &= \mathbb{1}_x (I - (\gamma^n P_{\vec{\mu},\vec{\nu}'})^{\ell})^{-1} (T^n v_k - v_k) \\
&\geq \frac{1}{1-\gamma^{n\ell}} \mathbb{1}_x (T^n v_k - v_k) = \frac{1}{1-\gamma^{n\ell}} \mathbb{1}_x (T_{\vec{\mu},\vec{\nu}} v_k - v_k).
\end{aligned}$$

Write $w = T^{n-j} v_k$. By equation (1), we have

$$\begin{aligned}
\mathbb{1}_x (T^n v_k - v_k) &\geq \mathbb{1}_y (T^{n-i-1} v_k - T^{n-j} v_k) \\
&= \mathbb{1}_y (T^c w - w) \\
&= \min_{\vec{\nu}'' \in N^c} \mathbb{1}_y (T_{\vec{\mu}_i^j,\vec{\nu}''} w - w) \\
&= \min_{\vec{\nu}'' \in N_{y,c}(\vec{\mu}_i^j)} \mathbb{1}_y (T_{\vec{\mu}_i^j,\vec{\nu}''} w - w) \\
&= \max_{\vec{\mu}' \in M^c} \min_{\vec{\nu}'' \in N_{y,c}(\vec{\mu}'} \mathbb{1}_y (\tilde{T}_{\vec{\nu}''} w - w) \qquad (2)
\end{aligned}$$

because of the choice of y.

Finally, observe that

$$\begin{aligned}
v_{y,c} - v_k(y) &= \max_{\mu \in M_{x,c}} \min_{\nu'' \in N} \mathbb{1}_y (v_{\mu,\nu''} - v_k) \\
&= \max_{\mu \in M_{x,c}} \min_{\vec{\nu}'' \in N_{x,c}(\mu^c)} \mathbb{1}_y (v_{\mu^c,\vec{\nu}''} - v_k) \\
&= \max_{\mu \in M_{x,c}} \min_{\vec{\nu}'' \in N_{x,c}(\mu^c)} \mathbb{1}_y (I - \gamma^c P_{\mu^c,\vec{\nu}''})^{-1} (T_{\mu^c,\nu''} v_k - v_k) \\
&= \max_{\mu \in M_{x,c}} \min_{\nu'' \in N_{x,c}(\mu^c)} \frac{1}{1-\gamma^c} \mathbb{1}_y (T_{\mu^c,\nu''} v_k - v_k) \\
&\leq \max_{\vec{\mu}' \in M^c} \min_{\nu'' \in N_{x,c}(\vec{\mu}'} \frac{1}{1-\gamma^c} \mathbb{1}_y (T_{\mu^c,\nu''} v_k - v_k). \qquad (3)
\end{aligned}$$

The result is obtained by combining Equations (2) and (3).

□