Notations:

$$v_{\mu,\nu} = (T_{\mu,\nu})^\infty$$

Algo PI standard: on choisit μ_{k+1} tel que

$$T_{\mu_{k+1}}(T_{\mu_k})^\infty = T(T_{\mu_k})^\infty$$

Alors, la preuve de la croissance de PI est:

$$\begin{aligned}(T_{\mu_{k+1}})^\infty - (T_{\mu_k})^\infty &= \sum_{i=0}^{\infty} (T_{\mu_{k+1}})^{i+1}(T_{\mu_k})^\infty - (T_{\mu_{k+1}})^i (T_{\mu_k})^\infty \\ &= \sum_{i=0}^{\infty} (T_{\mu_{k+1}})^i T(T_{\mu_k})^\infty - (T_{\mu_{k+1}})^i T_{\mu_k}(T_{\mu_k})^\infty \\ &\geq 0\end{aligned}$$

par monotonicité de $(T_{\mu_{k+1}})^i$.

La preuve de l'optimalité est: Supposons que $T_{\mu_{k+1}}(T_{\mu_k})^\infty = (T_{\mu_k})^\infty$. Alors,

$$(T_{\mu_k})^\infty = T(T_{\mu_k})^\infty.$$

Idée de l'algo: On a μ politique stationnaire du joueur MAX, ν meilleur réponse à μ:

$$(T_{\mu,\nu})^\infty = T_\mu^\infty.$$

On cherche un couple de politiques $\mu^m = (\mu_1, \mu_2, \ldots, \mu_m)$ et $\nu^m = (\nu_1, \nu_2, \ldots, \nu_m)$, un état x et $c \leq m$, tels que:

$$\begin{aligned}T_{\mu^c,\nu^c}\left[T^{m-c}(T_\mu)^\infty\right] &= T^c\left[T^{m-c}(T_\mu)^\infty\right], \\ \mathbb{1}_x' P_{\mu^c,\nu^c} &= \mathbb{1}_x', \\ \mathbb{1}_x' T_{\mu^c,\nu^c}\left[T^{m-c}(T_\mu)^\infty\right] &> \mathbb{1}_x'\left[T^{m-c}(T_\mu)^\infty\right].\end{aligned}$$

Croissance On a:

$$\begin{aligned}(T_\mu)^\infty &= T_\mu(T_\mu)^\infty \\ &\leq T(T_\mu)^\infty \\ &\ldots \\ &\leq T^{m-c}(T_\mu)^\infty \\ &\leq T^m(T_\mu)^\infty.\end{aligned}$$

Alors,

$$\begin{aligned}(T_{\mu^c,\nu^c})^\infty - (T_\mu)^\infty &\geq (T_{\mu^c,\nu^c})^\infty - T^{m-c}(T_\mu)^\infty \\ &= \sum_{i=0}^{\infty} (T_{\mu^c,\nu^c})^{i+1} T^{m-c}(T_\mu)^\infty - (T_{\mu^c,\nu^c})^i T^{m-c}(T_\mu)^\infty \\ &= \sum_{i=0}^{\infty} (T_{\mu^c,\nu^c})^i T^c T^{m-c}(T_\mu)^\infty - (T_{\mu^c,\nu^c})^i T^{m-c}(T_\mu)^\infty \\ &\geq 0\end{aligned}$$

par monotonicité de $(T_{\mu^c,\nu^c})^i$.

Propriétés du cycle trouvé Soit v une fonction quelconque. Pour tout $\bar{\mu^c}$ et $\bar{\nu^c}$.

$$
\begin{aligned}
(T_{\bar{\mu^c},\bar{\nu^c}})^{\infty} - v &= \sum_{i=0}^{\infty} (T_{\bar{\mu^c},\bar{\nu^c}})^{(i+1)} v - (T_{\bar{\mu^c},\bar{\nu^c}})^{i} v \\
&= \sum_{i=0}^{\infty} (T_{\bar{\mu^c},\bar{\nu^c}})^{i} (T_{\bar{\mu^c},\bar{\nu^c}}) v - (T_{\bar{\mu^c},\bar{\nu^c}})^{i} v \\
&= \sum_{i=0}^{\infty} (\gamma P_{\bar{\mu^c},\bar{\nu^c}})^{i} ((T_{\bar{\mu^c},\bar{\nu^c}}) v - v).
\end{aligned}
$$

Si x est tel que $1'_x(P_{\bar{\mu^c},\bar{\nu^c}}) = 1'_x$, alors:

$$
1'_x \left[(T_{\bar{\mu^c},\bar{\nu^c}})^{\infty} - v \right] = \frac{1'_x}{1-\gamma^c} \left[(T_{\bar{\mu^c},\bar{\nu^c}}) v - v \right],
$$

c'est-à-dire:

$$
1'_x (T_{\bar{\mu^c},\bar{\nu^c}})^{\infty} = \frac{1'_x [(T_{\bar{\mu^c},\bar{\nu^c}}) 0]}{1-\gamma^c}.
$$

Notons:

$$
\mathfrak{M}_c(x) = \{ \mu^c \ ; \ \mathbb{1}'_x P_{\mu^c} = 1'_x \}.
$$

On a:

$$
1'_x \left[\max_{\bar{\mu^c} \in \mathfrak{M}_c(x)} \min_{\bar{\nu^c}} (T_{\mu^c,\nu^c})^{\infty} \right] = \frac{1'_x [\max_{\bar{\mu^c} \in \mathfrak{M}_c(x)} \min_{\bar{\nu^c}} (T_{\bar{\mu^c},\bar{\nu^c}}) 0]}{1-\gamma^c}.
$$

Si les deux joueurs trouvent ensemble un cycle, cela signifie que le joueur MAX peut imposer le cycle à MIN si MIN reste sur le cycle limite. Sinon, MIN peut dégrader (réduire le bassin d'attraction du cycle) pour aller vers un autre cycle limite. S'il n'y a pas d'autre cycle limite, MIN ne peut qu'accepter le cycle proposé par MAX (et dans ce cas, on avance d'un coup !)

1 Algorithme en 2 phases

Notations

$$
\begin{aligned}
\forall \mu \in \Pi^c, \ \nu_{*,c}(\mu) &= \arg \min_{\nu \in \tilde{P}^c} \{ \nu ; v_{\mu} = v_{\mu,\nu} \} \\
\mathcal{C}_{x,c} &= \{ \mu \in \Pi^c \ ; \ \mathbb{1}'_x P_{\mu,\nu_{*,c}(\mu)} = \mathbb{1}'_x \}, \\
\forall v \in \mathbb{R}^n, \ \mathcal{G}_{x,c}(v) &= \{ \mu \in \mathcal{C}_{x,c} \ ; \ T_{\mu} v = T^c v \}, \\
\mathcal{V}_{x,c} &= \{ v \in \mathbb{R}^n \ ; \ \mathcal{G}_{x,c}(v) \in \mathcal{C}_{x,c} \}.
\end{aligned}
$$

Propriété fondamentale

Lemme 1. *Pour tout x, c, l'ensemble*

$$
\mathcal{G}_{x,c}(\mathcal{V}_{x,c}) = \{ \mathcal{G}_{x,c}(v) ; v \in \mathcal{V}_{x,c} \}
$$

contient au plus un élément.

Notons $v_{x,c} = v_{\mathcal{G}_{x,c}}(x)$.
Algorithme:
1) on calcule les $v_{x,c}$;
en résolvant un problème c-pas forcé à partir et à arriver en x.
il se peut que ce problème n'ait pas de solution possible (pas de chemin).
pour les noeuds max, on a une sur-estimation de la valeur v_*.
pour les noeuds min, on a une sous-estimation.
2) on en déduit v_* en utilisant le fait que

$$v_* = T^n w$$

avec

$$w(x) = O_{a,c} r(x,a) + \gamma v_{f(x,a),c}.$$

Preuve: $T^n w \leq v_*$.