Solving Strong Stackelberg Equilibrium in Stochastic Games

Víctor Bucarey López
vbucarey@ing.uchile.cl
INRIA – LILLE

December 2017

Joint Collaboration…

- Fernando Ordoñez
  Departamento Ingeniería Industrial Universidad de Chile.
- Eugenio Della Vecchia
  FCEIA - Universidad Nacional de Rosario.
- Alain Jean Marie
  INRIA - Research center of Sophia-Antipolis Méditerranée

Project DyGaMe STIC AmSud

Stackelberg Game

Leader
π

$\longrightarrow$

Follower
$\gamma(\pi)$

STRONG STACKELBERG EQUILIBRIUM

- Leader commits to a payoff maximizing strategy.
- Follower best responds
- Follower breaks ties in favor of the leader

Example

Player A : Leader → $\mathcal{A} = \{a_1, a_2\}$

Actions

Player B: Follower → $\mathcal{B} = \{b_1, b_2\}$

		b_1	b_2
x	a_1	(10, -10)	$(-5, 6)$
1-x	a_2	(-8, 4)	$(6, -4)$

If B plays b_1 leader's reward will be
10x+-8(1-x)

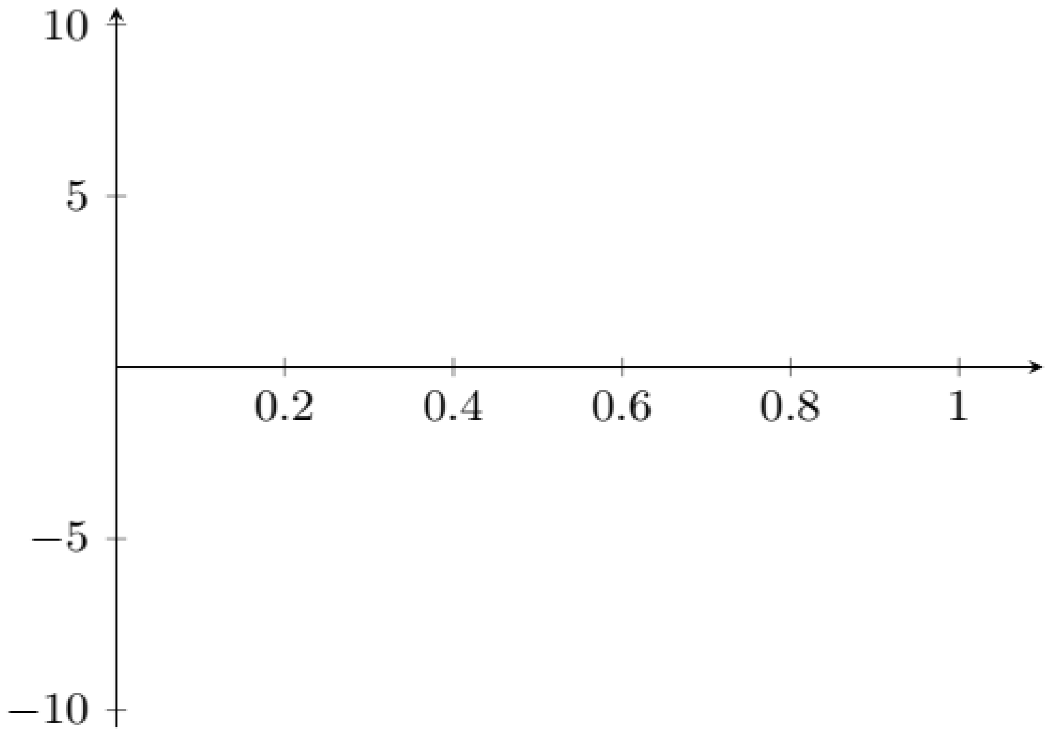
If B plays b_1 his expected reward will be
-10x+4(1-x)

If B plays b_2 leader's reward will be
-5x+6(1-x)

If B plays b_2 his expected reward will be
6x+-4(1-x)

Example

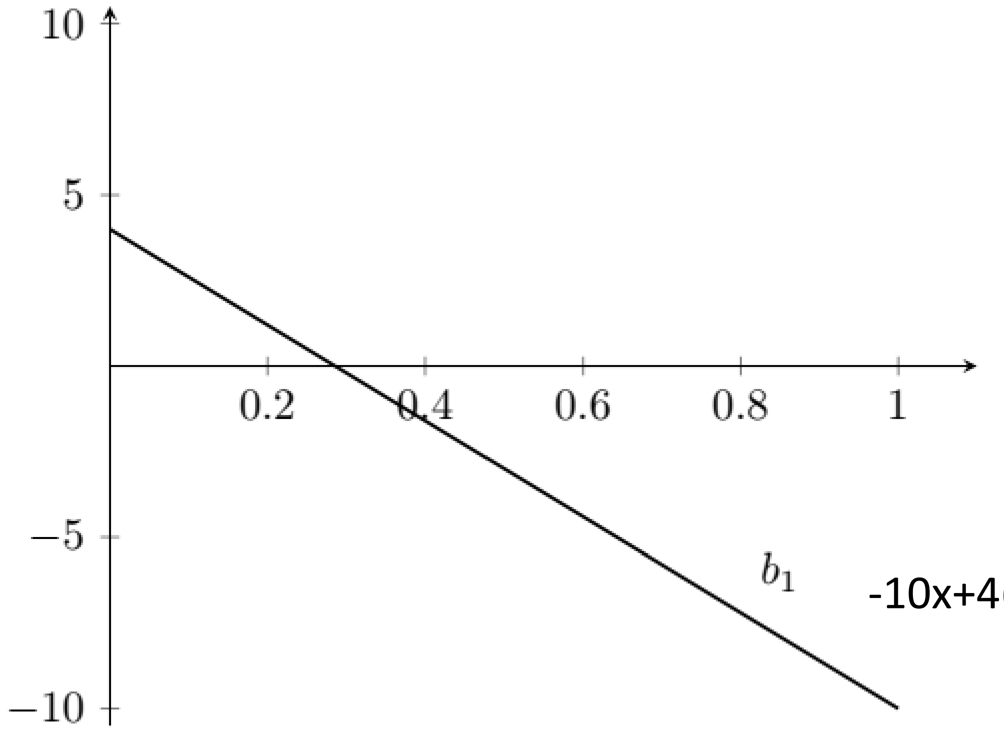
		b_1	b_2
x	a_1	(10,-10)	$(-5, 6)$
1-x	a_2	(-8,4)	$(6, -4)$



Follower

Example

		b_1	b_2
x	a_1	(10,-10)	$(-5, 6)$
1-x	a_2	(-8,4)	$(6, -4)$



-10x+4(1-x)

Follower

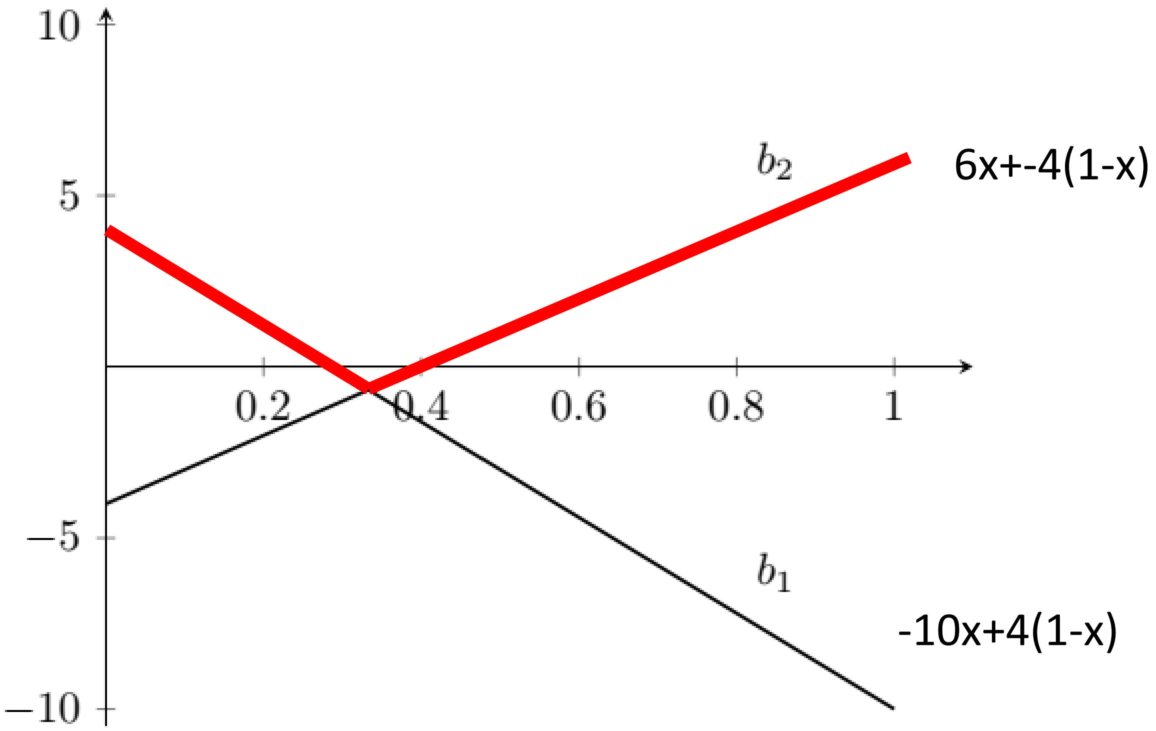
Example

		b_1	b_2
x	a_1	(10,-10)	$(-5, 6)$
1-x	a_2	(-8,4)	$(6, -4)$

Best Response:

$$g(x) = \begin{cases} b_2 & if\ x > \dfrac{1}{3} \\ b_1 & if\ x < \dfrac{1}{3} \end{cases}$$

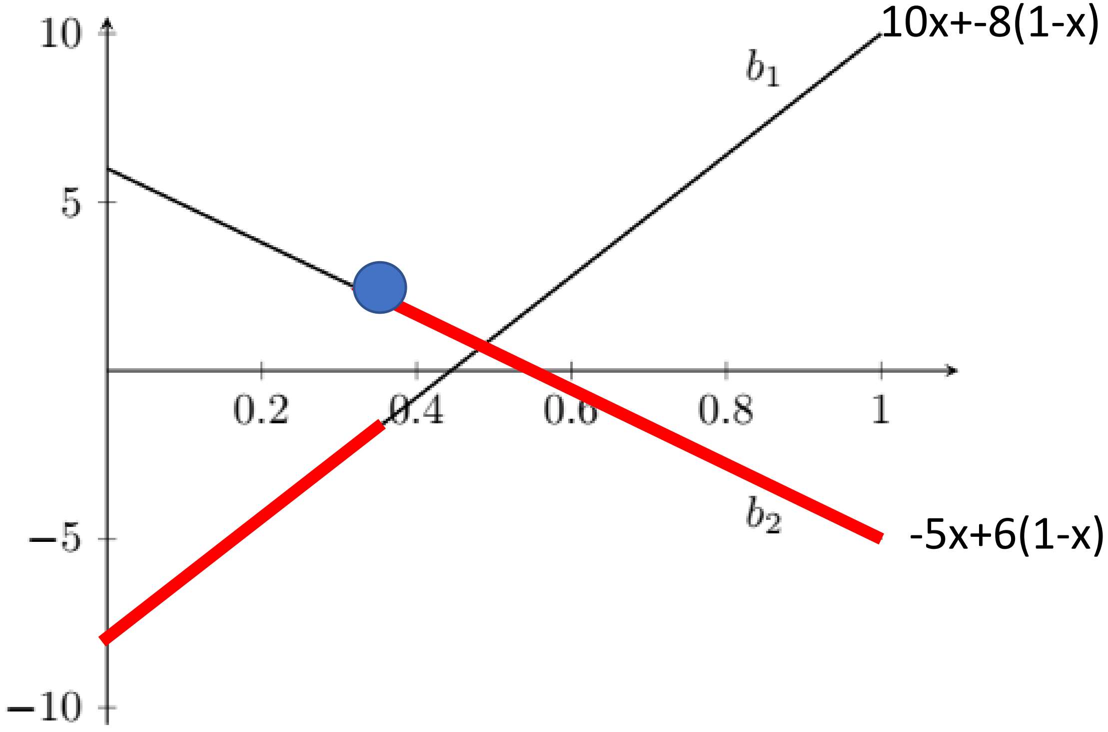
$x = 1/3$?



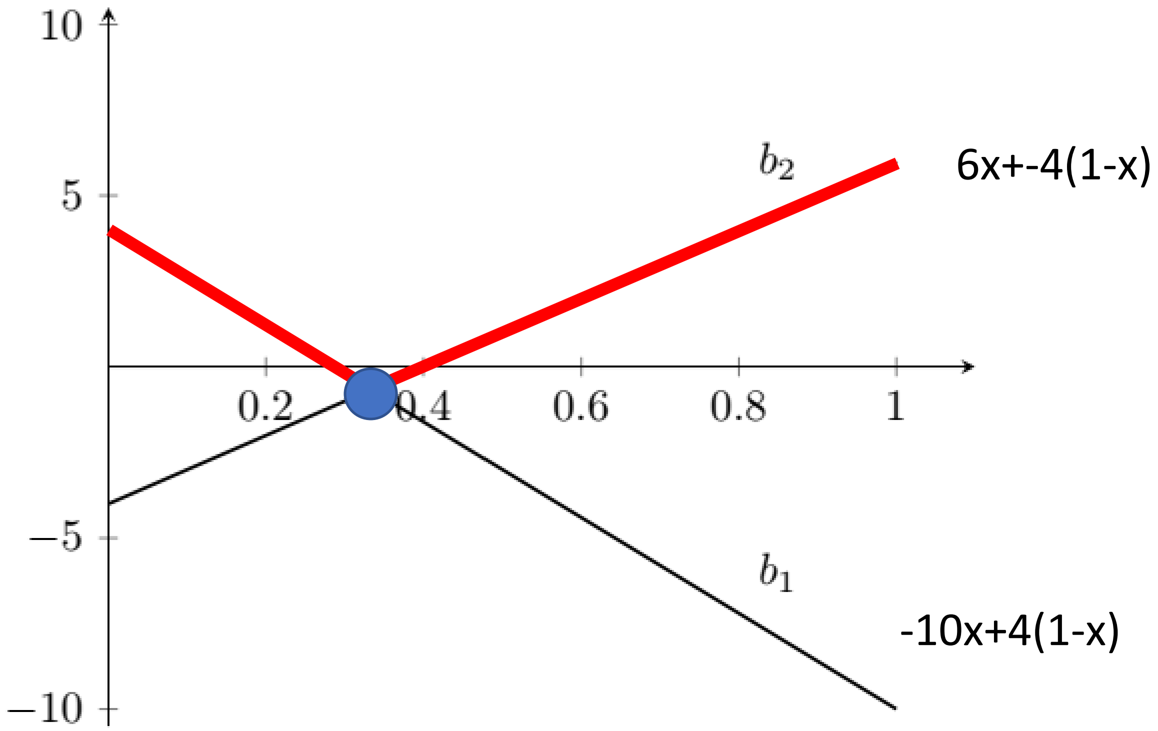
Follower

Example

		b_1	b_2
x	a_1	(10, -10)	$(-5, 6)$
1-x	a_2	(-8, 4)	$(6, -4)$



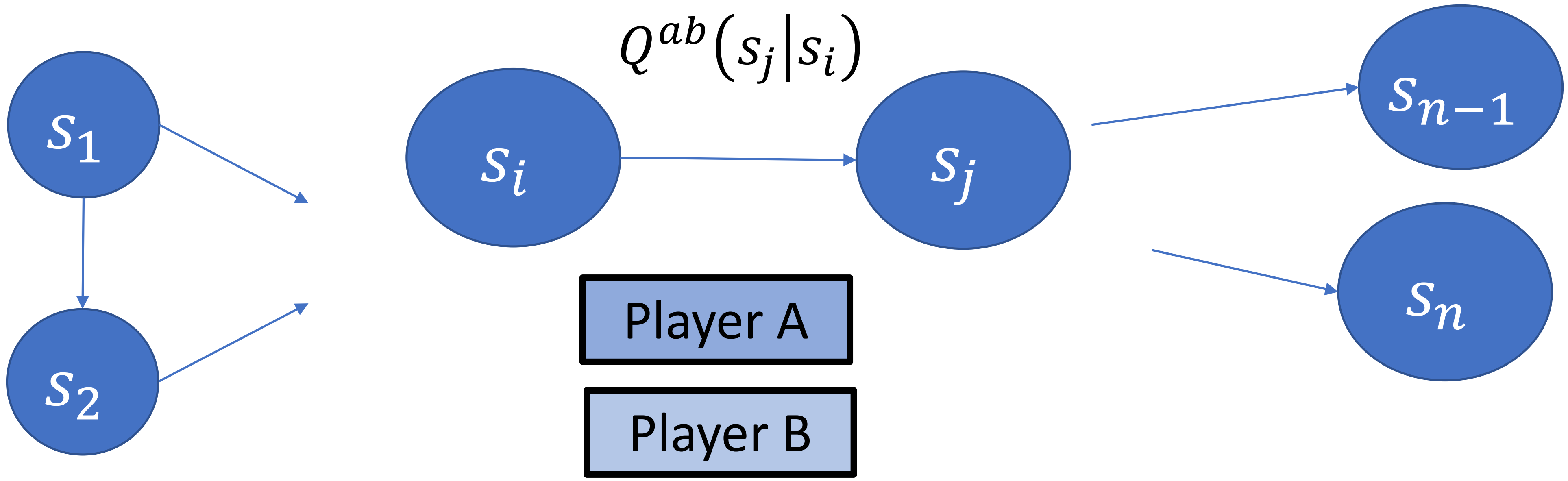
Leader



Follower

Stochastic Games

$$\mathcal{G} = (\mathcal{S}, \mathcal{A}, \mathcal{B}, r_A, r_B, Q, \beta_A, \beta_B, \tau)$$



Payoffs:

$$r_A = \{r_A^{ab}(s_i)\}$$
$$r_B = \{r_B^{ab}(s_i)\}$$

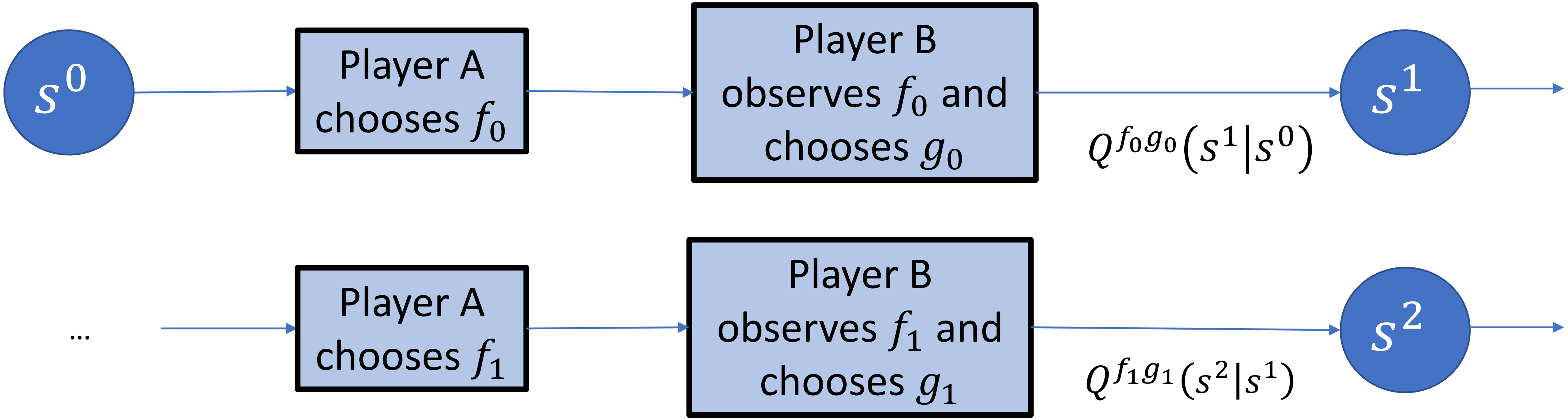
Discount Factors:

$$\beta_A, \beta_B \in [0,1)$$

Time horizon

$$\tau \in \mathbb{N} \cup \{+\infty\}$$

Stochastic Games



Feedback Policies:

$$\pi = \pi(s,t) = \{f_1, \dots, f_\tau\}$$

Stationary Policies:

$$\pi = \pi(s) = \{f, \dots, f\}$$

Framework

- **General Objectives**
  - Existence and characterization of value functions
  - Existence of equilibrium strategies
  - Algorithms to compute them

- **State of the art**
  - For finite horizon, Stackelberg equilibrium in stochastic games via Dynamic programming.
  - Mathematical programming approach to compute stationary values.

Our contribution

- We define suitable Dynamic Programming operators.
- We used it to characterize value functions and to prove existence and unicity of stationary values forming a Strong Stackelberg Equilibrium for a family of problems.
- We define Value Iteration and Policy Iteration for this family and prove its convergence.
- We prove via counterexample that this methodology is not always applicable for the general case.

Stackelberg Equilibrium in Stochastic Games

$$(\pi, \gamma)$$

Value functions

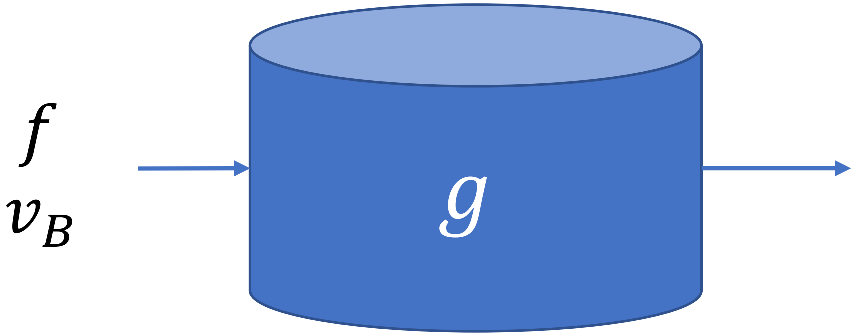
$$v_A^{\pi,\gamma}(s) = \mathbb{E}^{\pi,\gamma}(s)\left[\sum_{t=0}^{\tau} \beta_A^t r_A^{A_t B_t}(S_t)\right]$$

$$v_B^{\pi,\gamma}(s) = \mathbb{E}^{\pi,\gamma}(s)\left[\sum_{t=0}^{\tau} \beta_B^t r_B^{A_t B_t}(S_t)\right]$$

Stackelberg Equilibrium: (π^*, γ^*)

$$v_A^{\pi^*,\gamma^*}(s) = \max\ v_A^{\pi,\gamma^*}(s)$$
$$\gamma^* \in argmax\ v_B^{\pi,\gamma}(s)$$

Best response functional



$$\begin{pmatrix} b_1(s_1) \\ b_2(s_2) \\ \vdots \\ b_n(s_n) \end{pmatrix} \in \mathcal{B}^{|S|}$$

Given a stationary policy and future values, g computes the best actions to perform in each state.

$$g(f, v_B)(s) = \mathrm{argmax}_{\mathrm{b} \in \mathcal{B}} \sum_{a \in \mathcal{A}} f(a, s) \left[r_B^{ab}(s) + \beta_B \sum_{z \in S} Q^{ab}(z|s) v_B(z) \right]$$

- **Myopic follower strategies (MFS)**

$$g(f, v_B) = g(f)$$

MFS case

$$v_A \in \mathbb{R}^{|S|} \longrightarrow \boxed{T_A^f} \longrightarrow \begin{pmatrix} v_A{}'(s_1) \\ v_A{}'(s_2) \\ \vdots \\ v_A{}'(s_n) \end{pmatrix} \in \mathbb{R}^{|S|}$$

For a fixed stationary policy f, T_A^f computes the value of applyng this policy.

$$T_A^f(v_A)(s) = \sum_{a \in \mathcal{A}} f(a, s) \left[r_B^{ag(f)(s)}(s) + \beta_B \sum_{z \in S} Q^{ag(f)(s}(z|s) v_A(z) \right]$$

MFS case

$$v_A \in \mathbb{R}^{|S|} \longrightarrow \boxed{T_A} \longrightarrow \begin{pmatrix} v_A{}'(s_1) \\ v_A{}'(s_2) \\ \vdots \\ v_A{}'(s_n) \end{pmatrix} \in \mathbb{R}^{|S|}$$

T_A computes the value of the best policy for the leader.

$$T_A(v_A)(s) = \max_{f \in \mathbb{P}(\mathcal{A})} T_A^f(v_A)(s)$$

MFS case

- **Theorem 1.**

a) T_A^f, T_A are monotone.

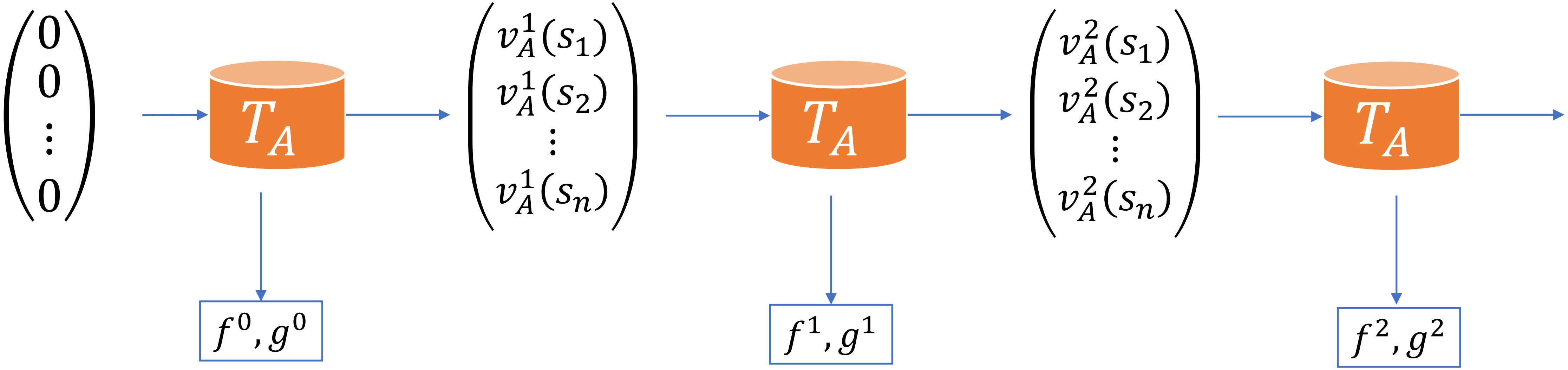
b) For any stationary strategy f, the operator T_A^f is a contraction on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ of modulus β_A.

c) The operator T_A is a contraction on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ of modulus β_A.

- **Theorem 2.**

There exists a equilibrium value function v_A^* and it is the unique solution of $v_A^* = T_A(v_A^*)$. Moreover, the pair f^* and g(f^*) which maximizes the RHS of (1) are the equilibrium strategies.

Value Iteration algorithm



Repeat until convergence

Theorem 3.
The sequence of value functions v_A^n converges to v_A^*. Furthermore, v_A^* is the fixed point of T_A with the following bound:

$$\|v_A^* - v_A^n\| \leq \beta_A^n \frac{\|r_A\|_\infty}{1-\beta_A}$$

Value Iteration algorithm

Algorithm 1 Value function iteration: Infinite horizon

Require: $\varepsilon > 0$

Initialize with $n = 1$, $v_A^0(s) = 0$ for every $s \in \mathcal{S}$ and $v_A^1 = T_A(v_A^0)$

while $||v_A^n - v_A^{n-1}||_\infty > \varepsilon$ **do**

  Compute v_A^{n+1} by

$$v_A^{n+1}(s) = T_A(v_A^n)(s) \ .$$

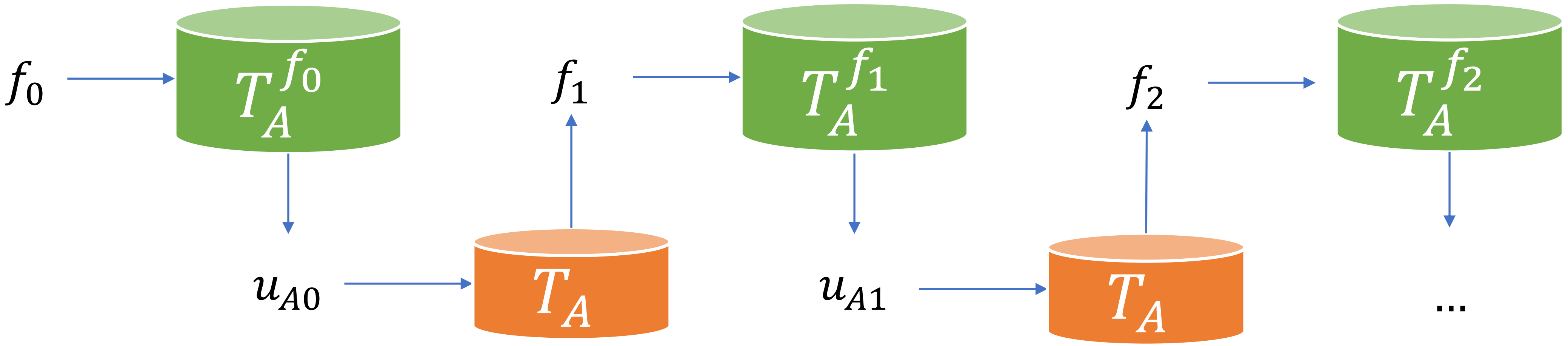
  Finding f^* and $g^*(f)$ at stage n.

  $n := n + 1$

end while

return Stationary Stackelberg policies $\pi^* = \{f^*, \ldots\}$ and $\gamma^* = \{g^*, \ldots\}$

Policy Iteration algorithm



Repeat until convergence

Theorem 4.
The sequence of functions $u_{A,n}$ verifies $u_{A,n} \uparrow v_A^*$.
Furthermore, if for any $n \in \mathbb{N}$, $u_{A,n} = u_{A,n+1}$ then it is true that $u_{A,n} = v_A^*$.

Policy Iteration algorithm

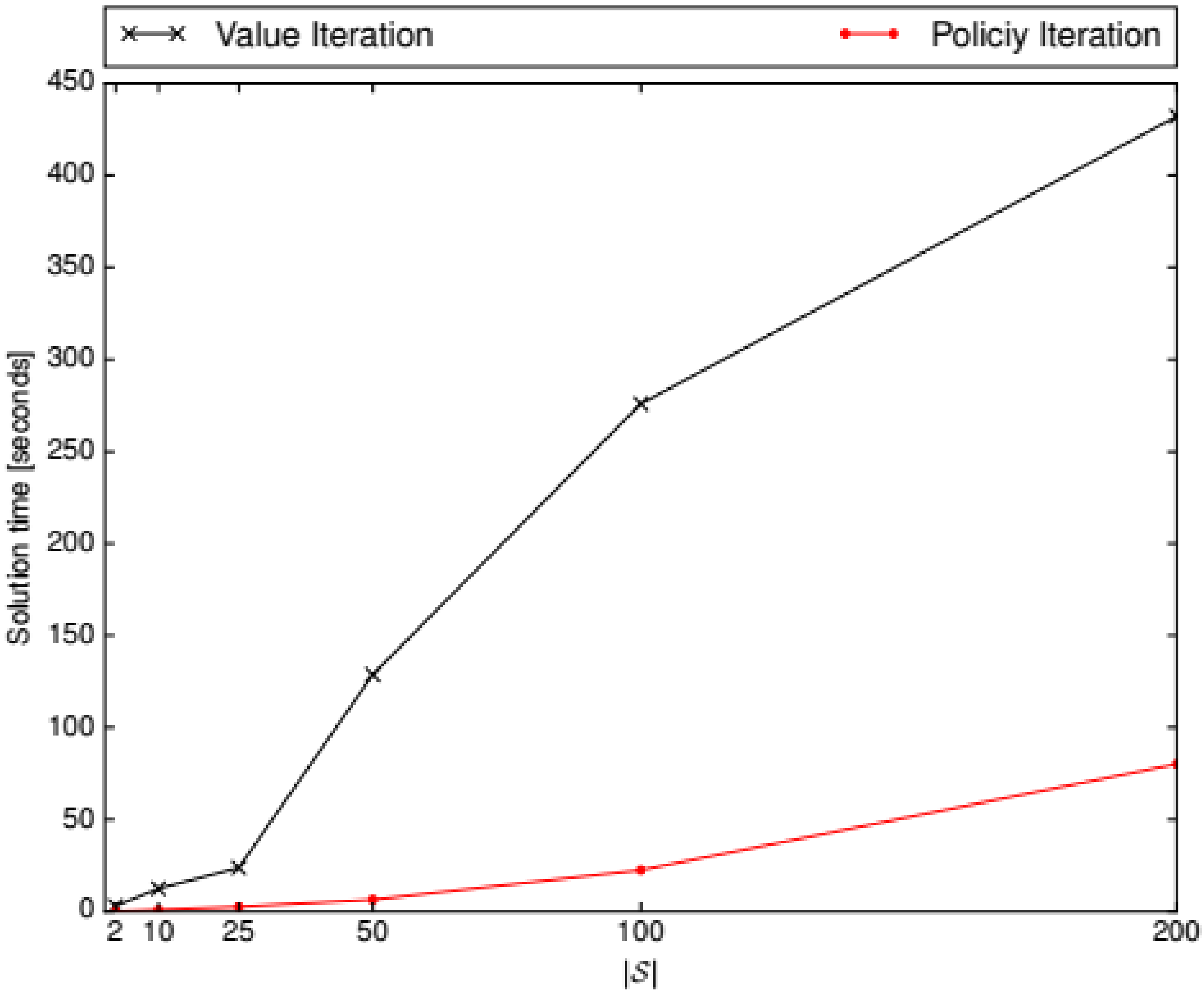
Algorithm 2 Policy Iteration (PI)

Choose a stationary Stackelberg pair $(f_0, g(f_0))$.
while $||u_{A,n} - u_{A,n+1}|| > \varepsilon$ **do**
    Evaluation Phase: Find $u_{A,n}$ fixed point of the operator $T_A^{f_n}$.
    Improvement Phase: Find a strategy f_{n+1} such that

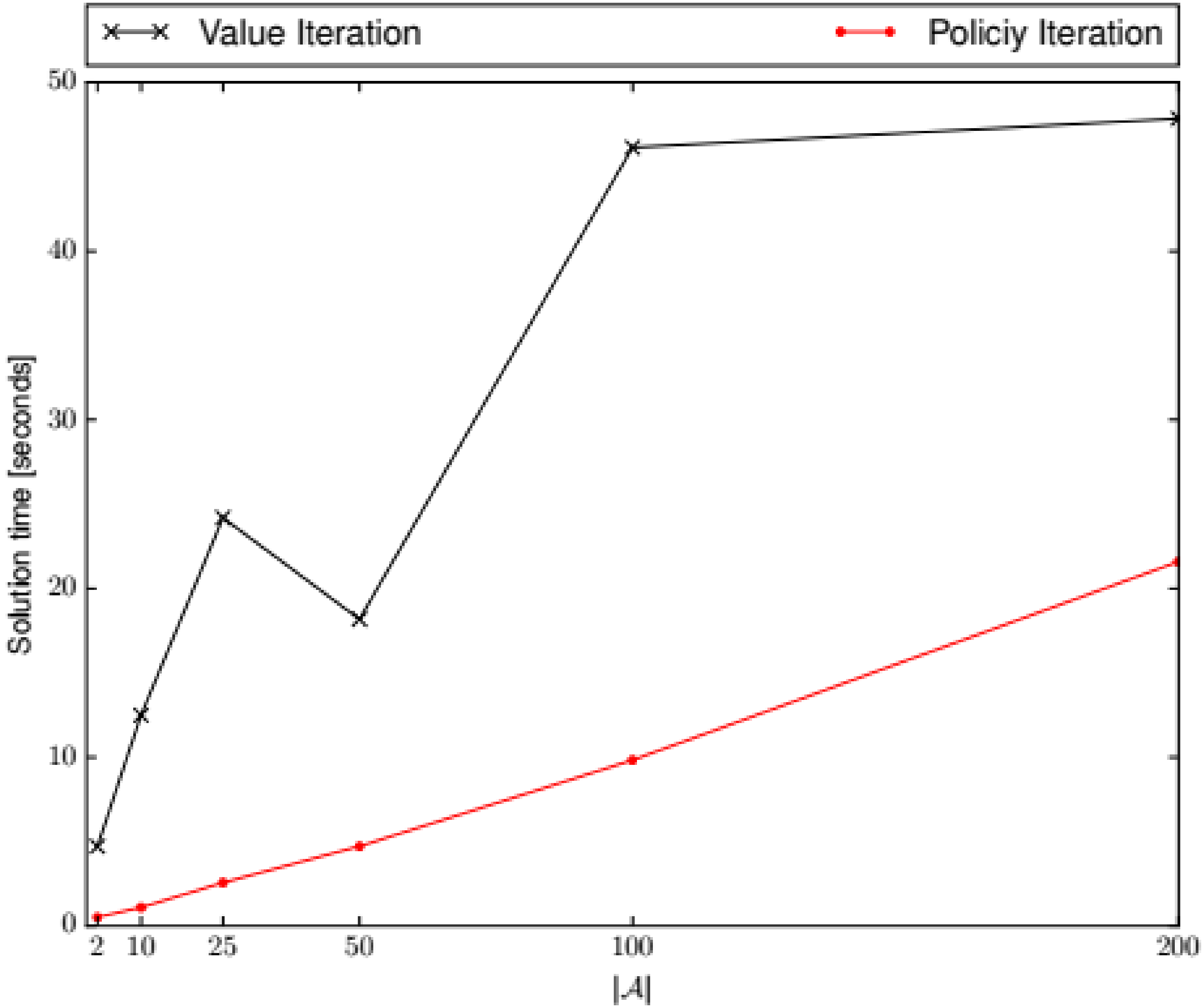
$$T_A^{f_{n+1}}(u_{A,n}) = T_A(u_{A,n}) \ .$$

    n:= n+1
end while
return Stationary Stackelberg policies $\pi^* = \{f^*, \ldots\}$ and $\gamma^* = \{g(f^*), \ldots\}$

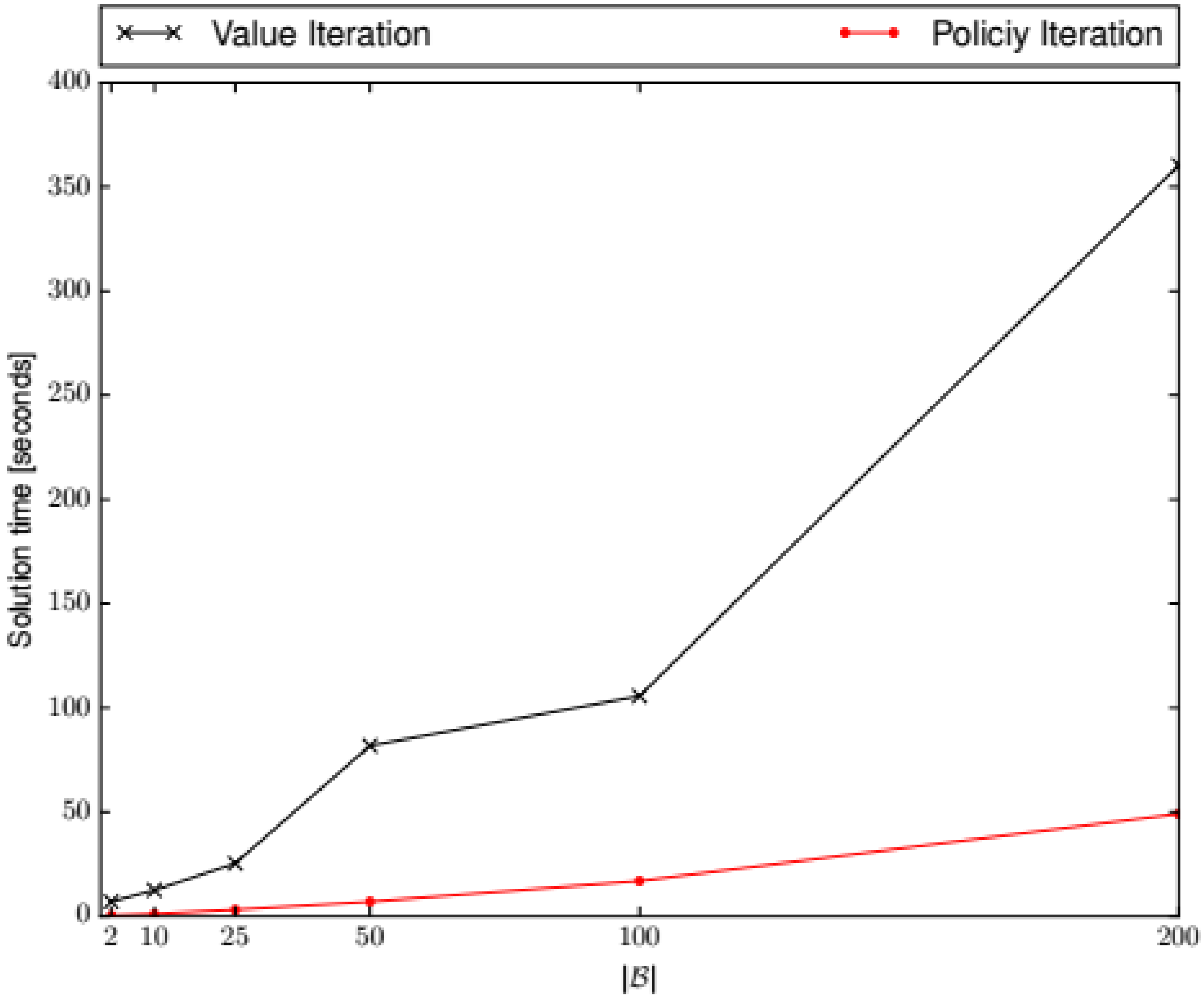
Computational Results



Computational Results



Computational Results



General case

$v_i \in \mathbb{R}^{|S|}$ → T_i^f → $\begin{pmatrix} v_i'(s_1) \\ v_i'(s_2) \\ \vdots \\ v_i'(s_n) \end{pmatrix} \in \mathbb{R}^{|S|}$

$(v_A, v_B) \in \mathbb{R}^{|S|\times|S|}$ → T → $(v'_A, v'_B) \in \mathbb{R}^{|S|\times|S|}$

$$(T(v_A, v_B))(s) = \left(\max_{f\in\mathbb{P}(\mathcal{A}_s),} T_A^{f,g(f,v_B)}(v_A)(s),\ T_B^{f^*,g(f^*,v_B)}(v_B)(s) \right)$$

General case

Algorithm 3 Value Iteration (VI): Finite horizon for the general ca

Initialize with $v_A^{\tau+1}(s) = v_B^{\tau+1}(s) = 0$ for every $s \in \mathcal{S}$
for $t = \tau, \ldots, 0$, and for every $s \in \mathcal{S}$ **do**
    Solve

$$(v_A^t(s), v_B^t(s)) = T(v_A^{t+1}, v_B^{t+1})(s) \quad \forall s \in \mathcal{S}$$

    Finding f_t^* and g_t^* SSE strategies at stage t.
end for
return Stackelberg policies $\pi^* = \{f_0^*, \ldots, f_\tau^*\}$ and $\gamma^* = \{g_0^*, \ldots, g_\tau^*\}$

This algorithm returns an Strong Stackelberg Equilibrium in **feedback policies** for the τ-horizon problem.

What about **stationary policies**?

Example

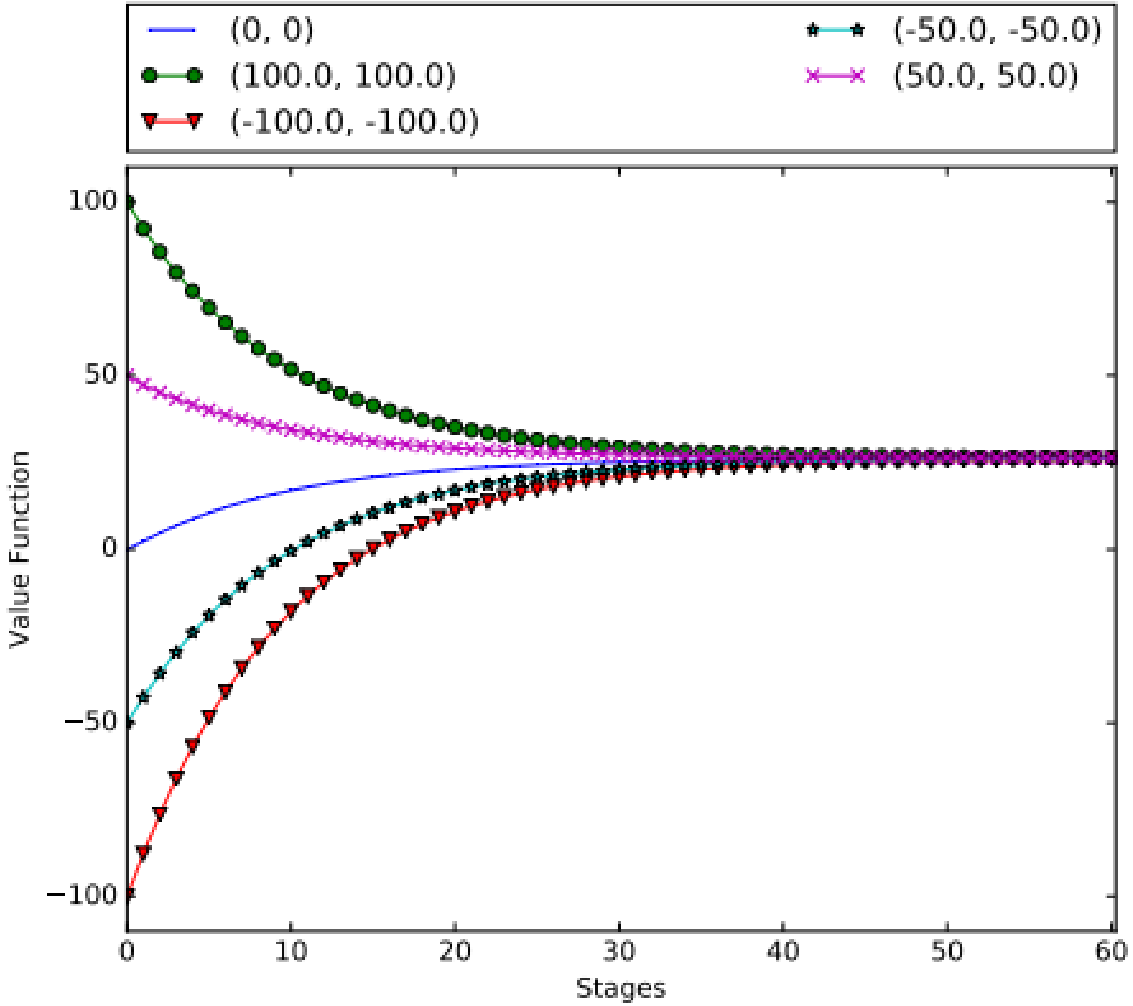
	b_1	b_2
a_1	$(\frac{1}{2}, \frac{1}{2})$ (10,-10)	$(0, 1)$ $(-5, 6)$
a_2	$(\frac{1}{4}, \frac{3}{4})$ (-8,4)	$(1, 0)$ $(6, -4)$

State s_1

	b_1	b_2
a_1	$(\frac{1}{2}, \frac{1}{2})$ (7,-5)	$(0, 1)$ $(-1, 6)$
a_2	$(\frac{1}{4}, \frac{3}{4})$ (-3,10)	$(1, 0)$ $(2, -10)$

State s_2

$$\beta_A = \beta_B = 0.9$$



Example

	b_1	b_2
a_1	(1, 0) / (1,-1)	(0, 1) / (0, 1)
a_2	(0, 1) / (-1,1)	(0, 1) / $(-1, -1)$

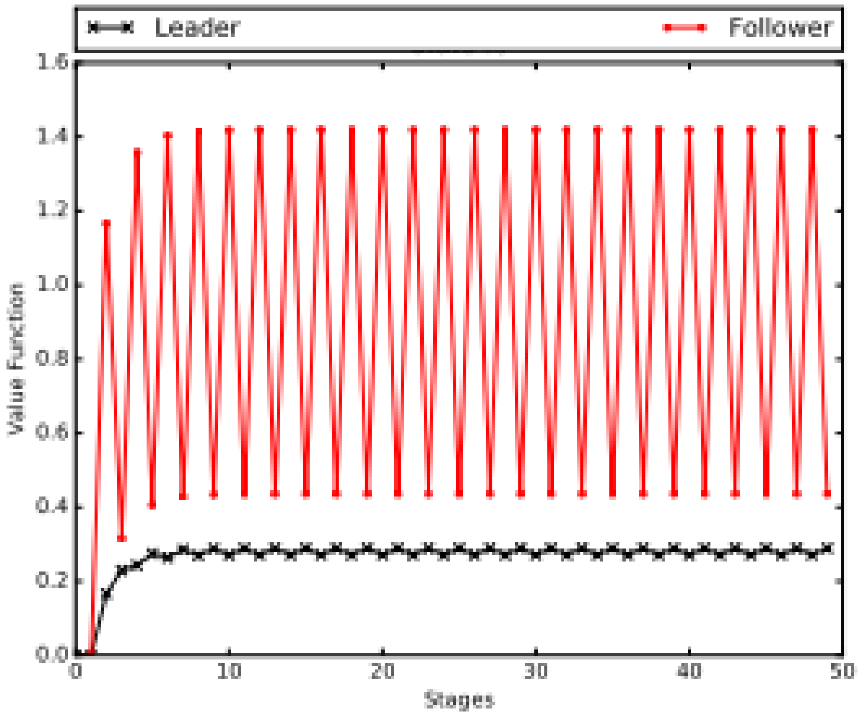
State s_1

	b_1	b_2
a_1	(0, 1) / (-1,0)	(1, 0) / (0, 1)
a_2	(1, 0) / (0,1)	(0, 1) / $(1, -1)$

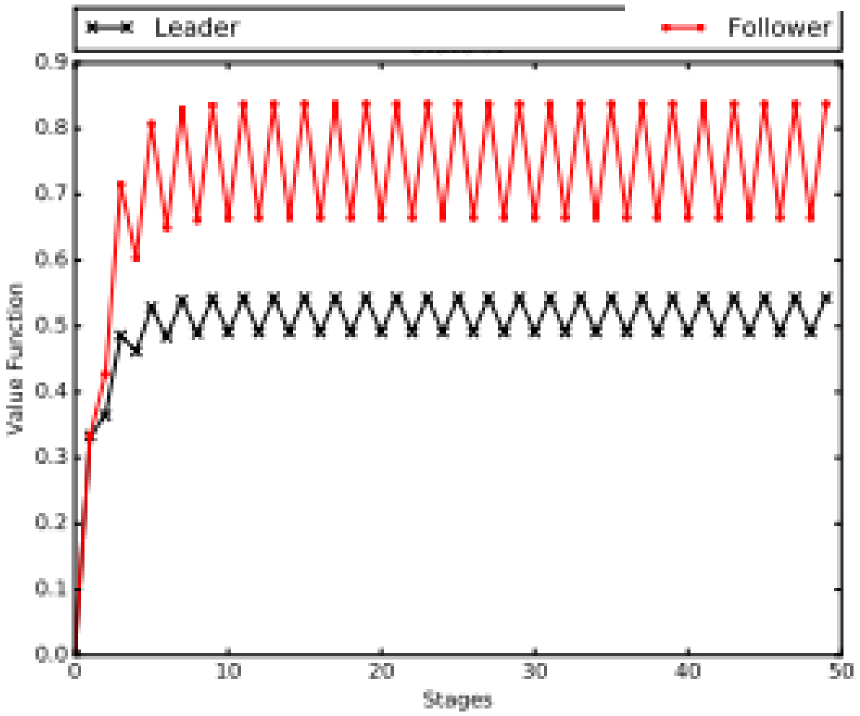
State s_2

$$\beta_A, \beta_B = \frac{1}{2}$$

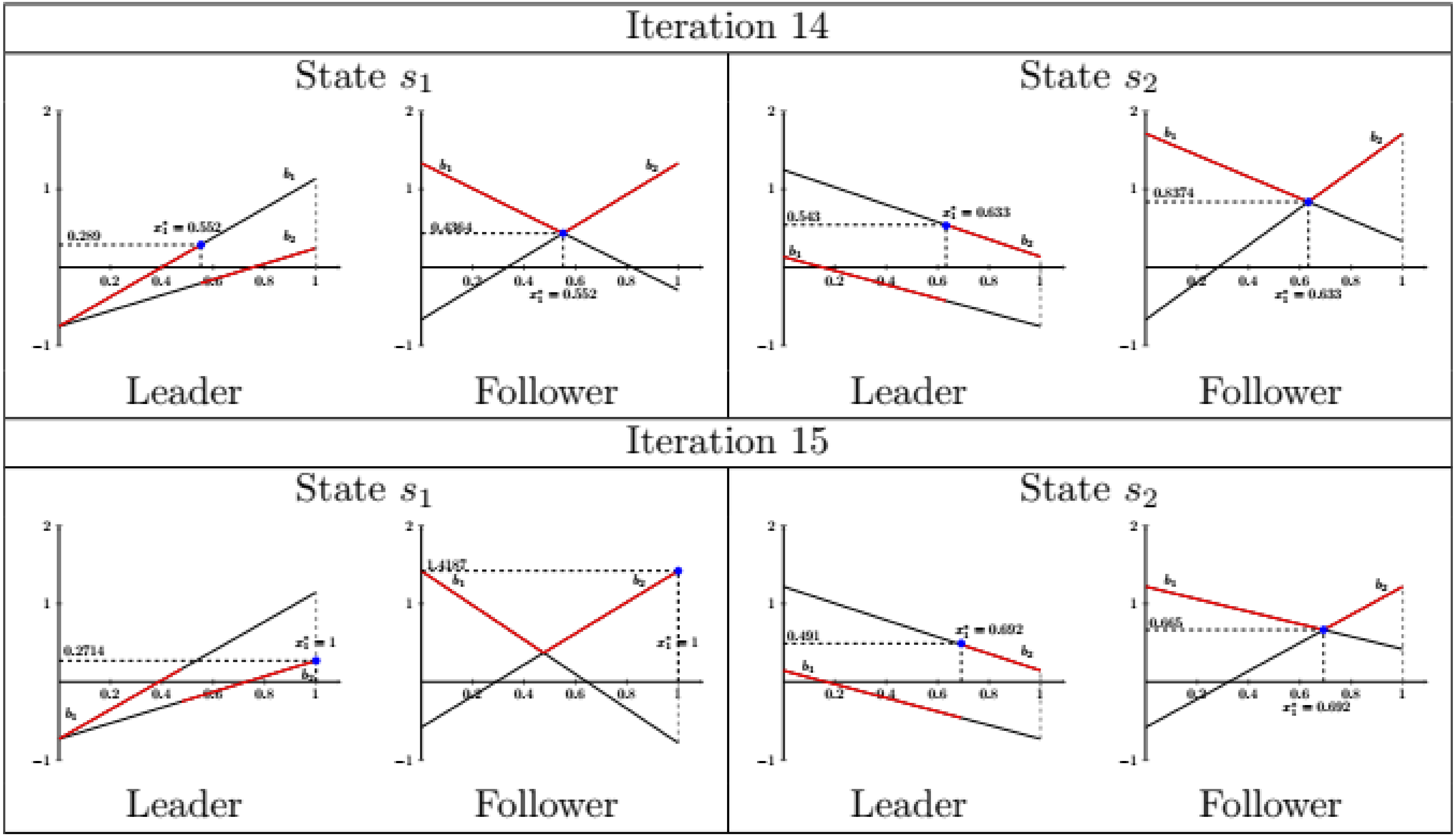
Example



State s_1



State s_2



Iteration 14
State s_1
0.289
$x_1^* = 0.552$
Leader
0.4364
$x_1^* = 0.552$
Follower
State s_2
0.543
$x_1^* = 0.633$
Leader
0.8374
$x_1^* = 0.633$
Follower
Iteration 15
State s_1
0.2714
$x_1^* = 1$
Leader
1.4187
$x_1^* = 1$
Follower
State s_2
0.491
$x_1^* = 0.692$
Leader
0.665
$x_1^* = 0.692$
Follower

Algorithm 4 VI modified: Infinite horizon for the general case

Initialize with $n = 0$, $v_A^0(s) = v_B^0(s) = 0$ for every $s \in \mathcal{S}$.

for $n = 1, \cdots, MAX\_IT$ **do**

  Find the pair (v_A^n, v_B^n) by

$$(v_A^n, v_B^n)(s) = T(v_A^{n-1}, v_B^{n-1})(s) \; .$$

  Finding f^* and g^* SSE strategies at stage $n-1$.

  if $(v_A^n, v_B^n) = (v_A^{n-1}, v_B^{n-1})$ **then**

    return (v_A^n, v_B^n) fixed point of T.

  end if

  if $||(v_A^n, v_B^n) - (v_A^{n-1}, v_B^{n-1})|| > 2\frac{\beta^{n-1}}{1-\beta}||(r_A, r_B)||$ **then**

    return UNDEFINED 1.

  end if

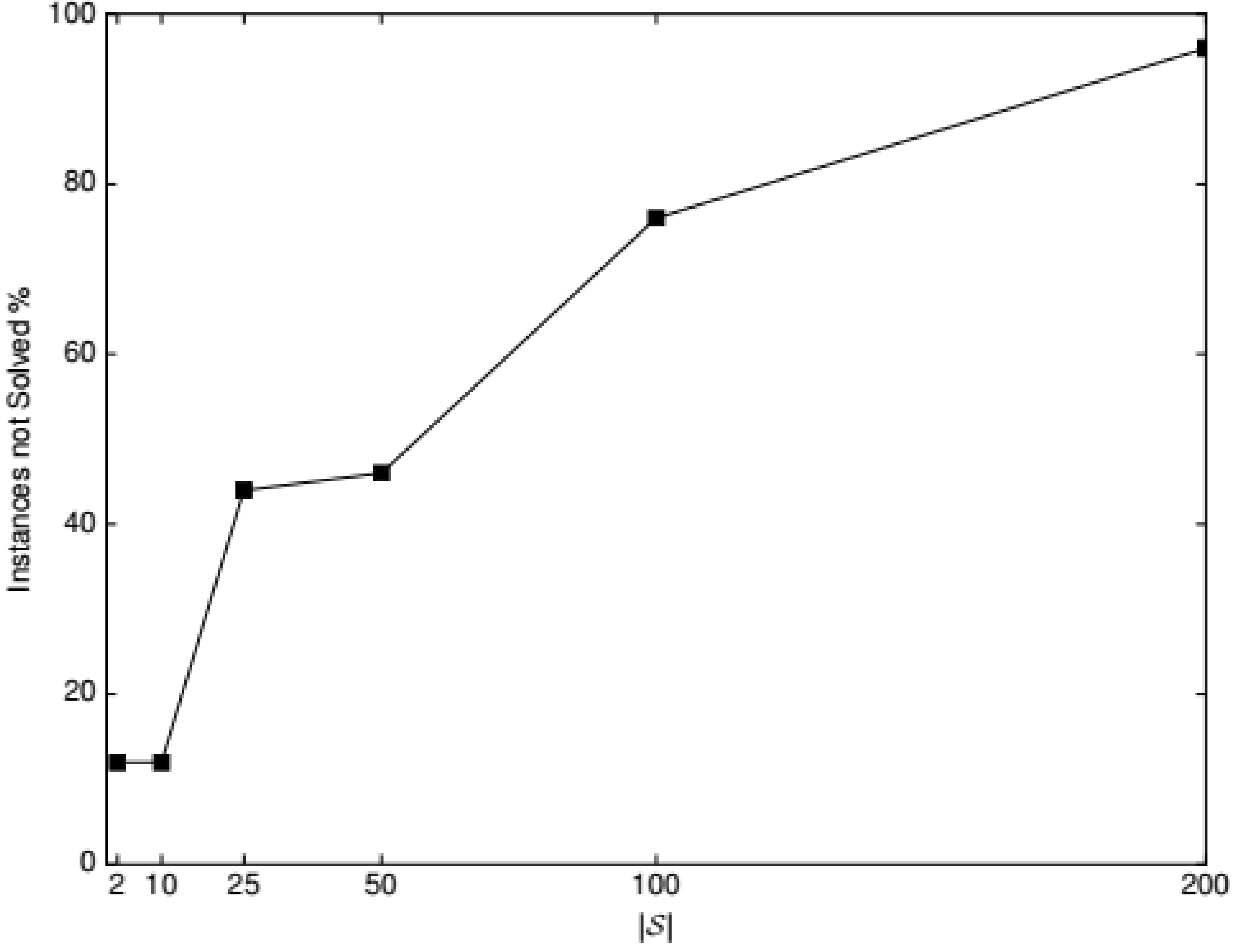
end for

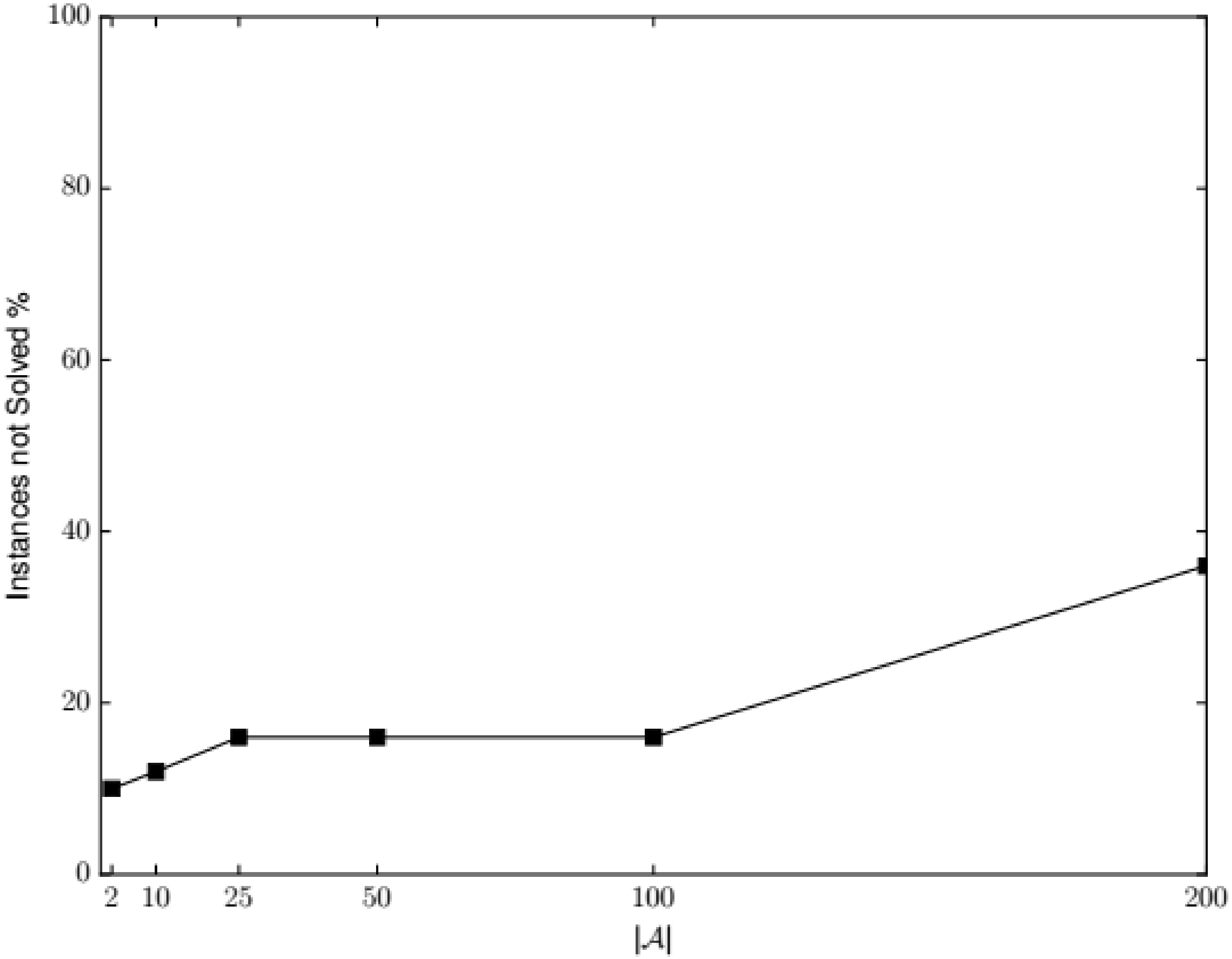
return UNDEFINED 2.

We found an equilibrium.

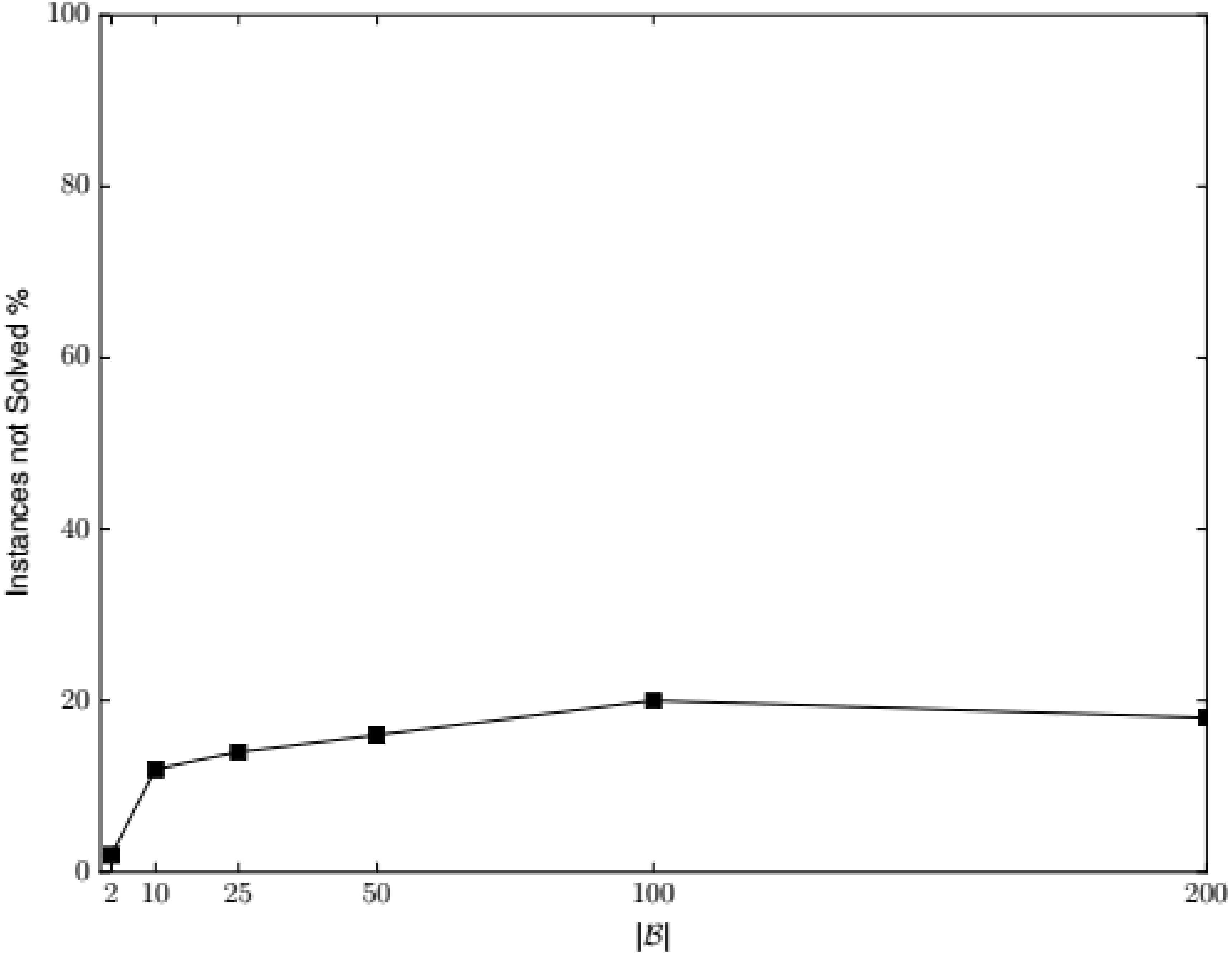
There is no geometrical decreasing.

We get tired of finding an equilibrium.


100
80
60
40
20
0
Instances not Solved %
2
10
25
50
100
200
|S|

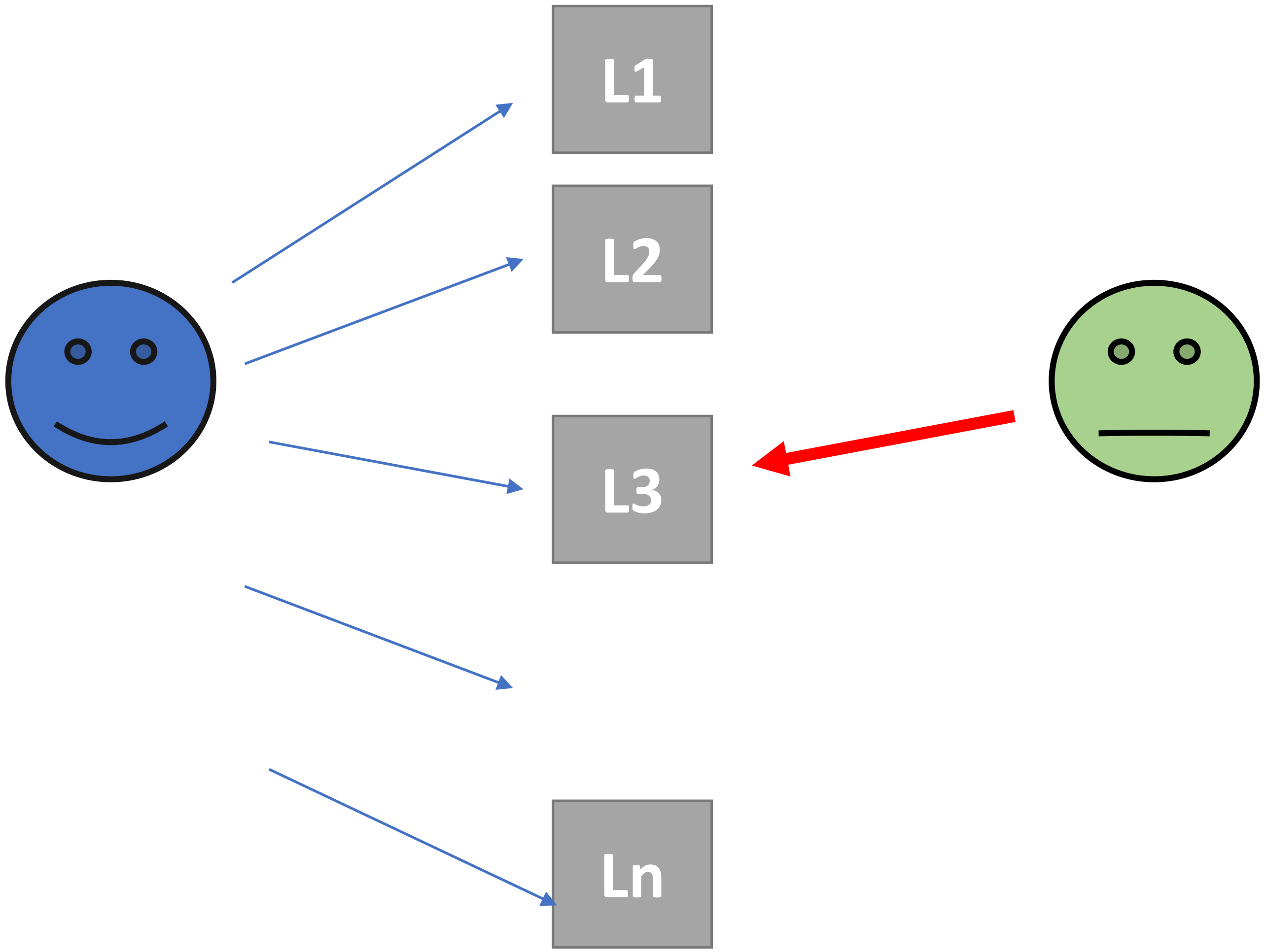


100
80
60
40
20
0
Instances not Solved %
2
10
25
50
100
200
$|\mathcal{A}|$



100
80
60
40
20
0
2
10
25
50
100
200
Instances not Solved %
$|\mathcal{B}|$

Security Games

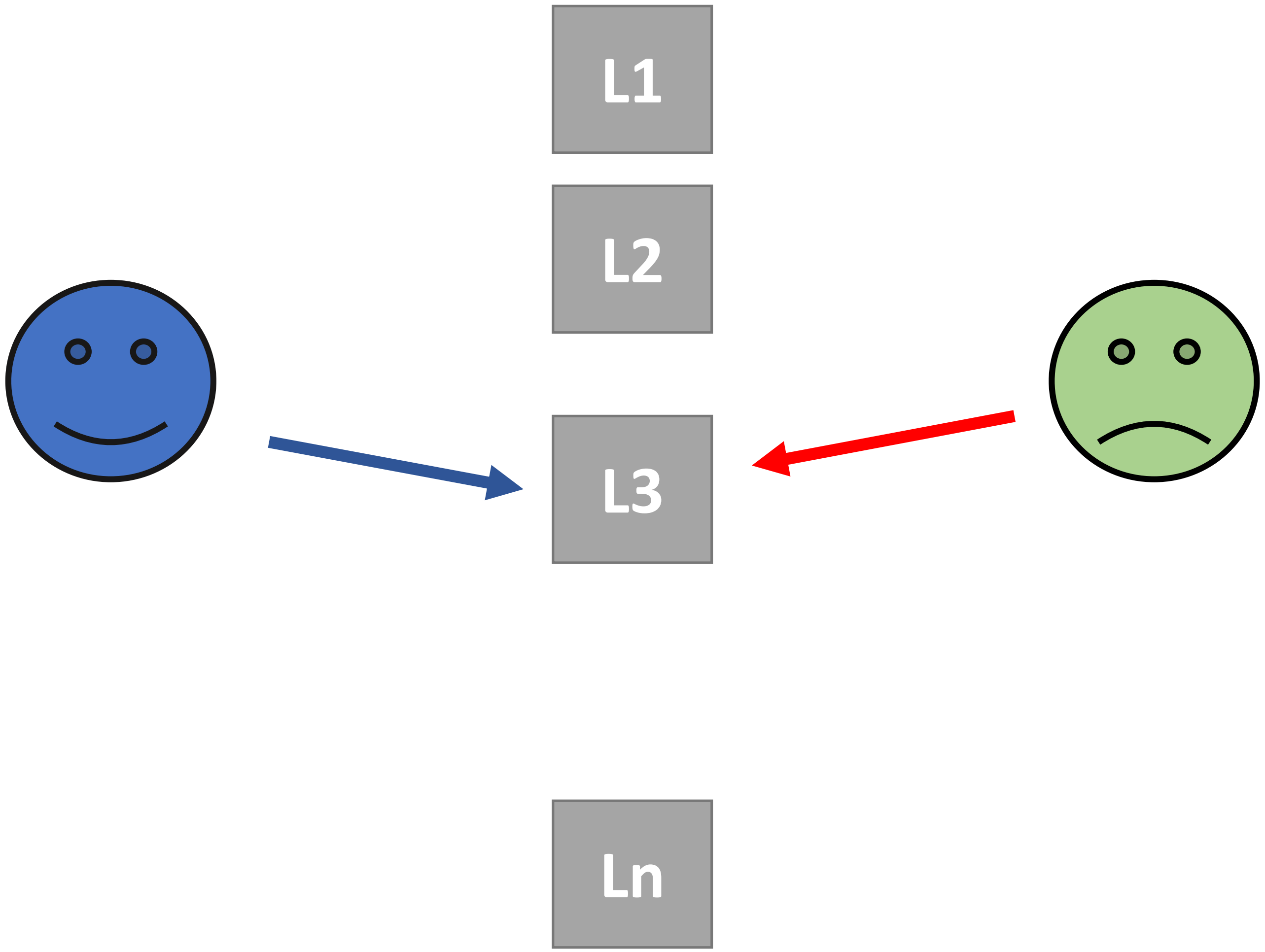


Leader = DEFENDER

Follower = ATTACKER

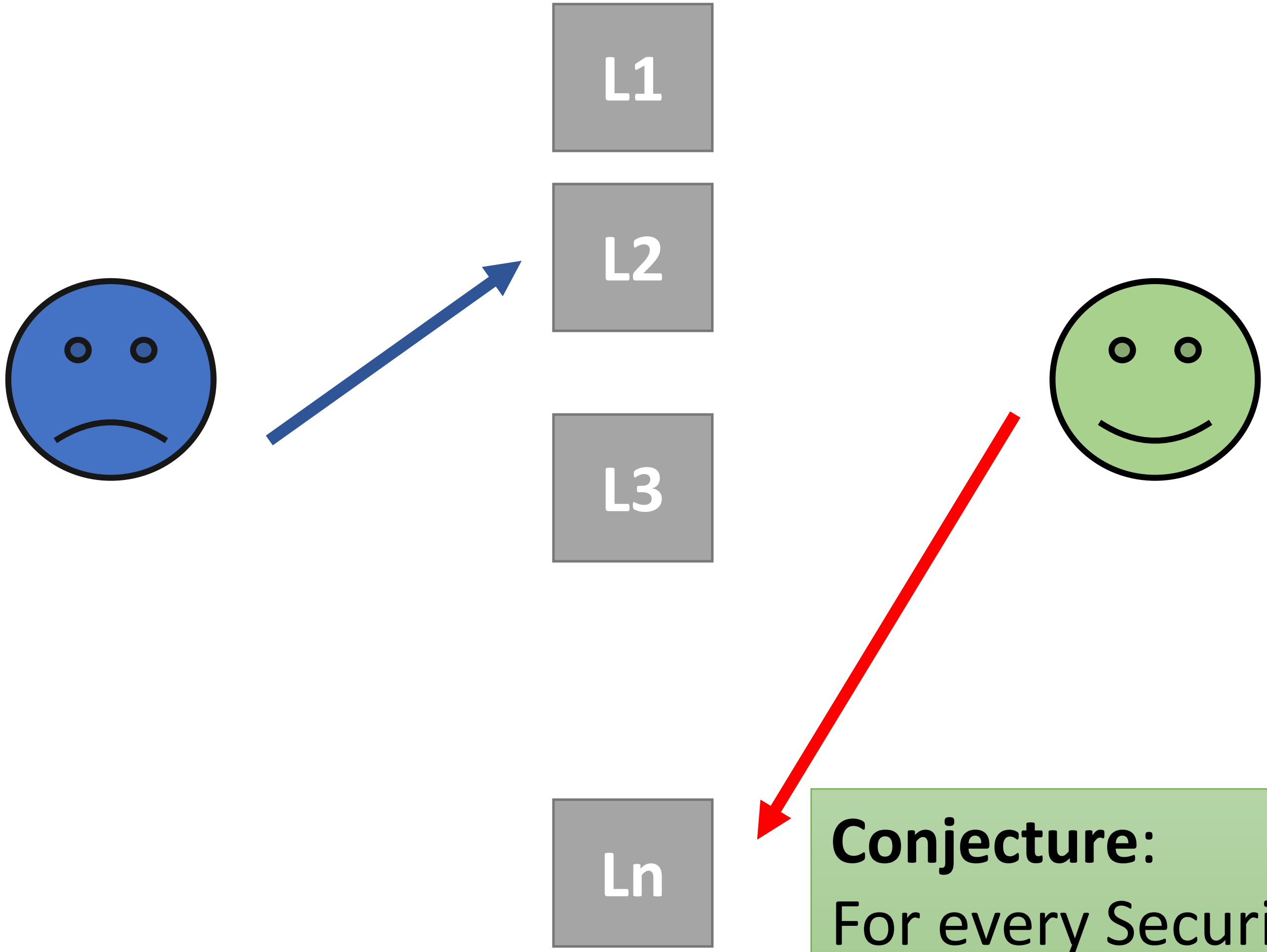
Payoffs only depends on whether a location is protected or not.

Security Games



$$\mathrm{Reward_D} > 0$$
$$\mathrm{Penalty_A} < 0$$

Security Games



$$\text{Penalty}_D < 0$$
$$\text{Reward}_A > 0$$

Non pure strategies seems to be optimal for the leader.

Computationally all instances in Security games VI converges with the geometric bound.

Conjecture:
For every Security game with this payoff structure, the operator T is contractive of modulus $\beta = \max\{\beta_A, \beta_B\}$

Conclusions

- We define suitable dynamic programming operators and we use it to prove unicity of values of Strong Stackelberg games in stationary policies for a family of problems.
- We define Value Iteration and Policy Iteration algorithms for finding Stackelberg stationary equilibrium.
- We prove via counterexample that this methodology is not always applicable for the general case.
- We study security games and we conjecture that operators for this type of games are contractive.

Thank You!

Víctor Bucarey López

vbucarey@ing.uchile.cl

INRIA – LILLE

December 2017