Solving Stackelberg Equilibrium in Stochastic Games.

Víctor Bucarey López

vbucarey@ing.uchile.cl

FCEIA - UNR - Rosario

November 2nd, 2017

Stackelberg Game

Leader

π

Stackelberg Game

Leader
π

$\longrightarrow$

Follower
$\gamma(\pi)$

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

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

MIP formulation

$$\max\ v_A$$

$$v_A \leq 10x_1 + -8x_2 + M(1 - y_1)$$

$$v_A \leq -5x_1 + 6x_2 + M(1 - y_2)$$

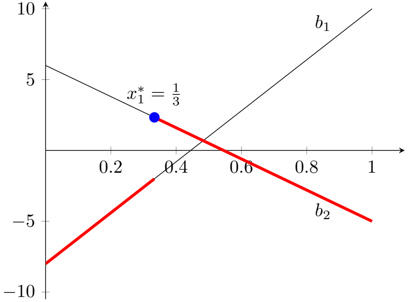
$$0 \leq v_B - (-10x_1 + 4x_2) \leq M(1 - y_1)$$

$$0 \leq v_B - (6x_1 + -4x_2) \leq M(1 - y_2)$$

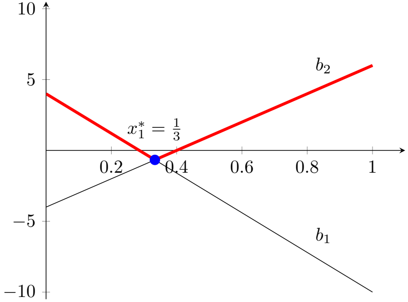
$$x_1 + x_2 = 1 \quad y_1 + y_2 = 1$$

$$x \geq 0,\ y \in \{0, 1\}$$

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



Leader



Follower

Multiple States

	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

Stochastic Games - Definition

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

$$s_0 \rightsquigarrow \begin{array}{c}\text{Player } A\\ \text{chooses } f_0\end{array} \rightsquigarrow \begin{array}{c}\text{Player } B\\ \text{observes } f_0\\ \text{and chooses } g_0\end{array} \underbrace{\rightsquigarrow}_{Q^{f_0 g_0}(s_1|s_0)}$$

$$\rightsquigarrow s_1 \rightsquigarrow \begin{array}{c}\text{Player } A\\ \text{chooses } f_1\end{array} \rightsquigarrow \begin{array}{c}\text{Player } B\\ \text{observes } f_1\\ \text{and chooses } g_1\end{array} \rightsquigarrow s_2 \cdots$$

Stochastic Games - Definition

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

$s_0 \rightsquigarrow$ Player A chooses f_0 $\rightsquigarrow$ Player B observes f_0 and chooses g_0 $\underbrace{\rightsquigarrow}_{Q^{f_0 g_0}(s_1|s_0)}$

$\rightsquigarrow s_1 \rightsquigarrow$ Player A chooses f_1 $\rightsquigarrow$ Player B observes f_1 and chooses g_1 $\rightsquigarrow s_2 \cdots$

Feedback Policies:

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

Stationary Policies:

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

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.

Framework

Contributions in Infinite horizon

- We define suitable Dynamic Programming operators.
- We used it to characterize value functions and to prove existence and unicity of stationary policies 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

(π, γ) $\longrightarrow$

Value Functions

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

Stackelberg equilibrium

$$(\pi, \gamma)$$

$$\longrightarrow$$

Value Functions

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

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

Stackelberg Equilibrium

$$(\pi^*, \gamma^*)$$

$$v_A^{\pi^*,\gamma^*}(s) = \max_{\pi,\gamma^*} v_A^{\pi,\gamma^*}(s)$$

$$\gamma^* \in \operatorname{argmax} v_B^{\pi,\gamma}(s)$$

Myopic Follower Strategies

Best response functional:

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

Myopic Follower Strategies

Best response functional:

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

Myopic follower strategies (MFS):

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

Myopic Follower Strategies

Best response functional:

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

Myopic follower strategies (MFS):

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

2 important cases:

- Myopic follower: $\beta_B = 0$
- Leader-Controller Discounted Games: $Q^{ab}(z|s) = Q^a(z|s)$

Myopic Follower Strategies

- f a stationary policy.
- $T_A^f : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$:

$$T_A^f(v_A)(s) \; = \; \sum_{a \in \mathcal{A}_s} f(a) \left[r_A^{ag(f)}(s) + \beta_A \sum_{z \in \mathcal{S}} Q^{ag(f)}(z|s) v_A(z) \right]$$

Myopic Follower Strategies

- f a stationary policy.
- $T_A^f : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$:

$$T_A^f(v_A)(s) \;=\; \sum_{a \in \mathcal{A}_s} f(a) \left[r_A^{ag(f)}(s) + \beta_A \sum_{z \in \mathcal{S}} Q^{ag(f)}(z|s) v_A(z) \right]$$

Operator for the MFS case

$$T_A(v_A)(s) \;=\; \max_{f \in \mathbb{P}(\mathcal{A}_s)} T_A^f(v_A)(s) \tag{1}$$

Myopic Follower Strategies

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 $(\mathbb{R}^{|\mathcal{S}|}, ||\cdot||_\infty)$ of modulus β_A.

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

Myopic Follower Strategies

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 $\gamma^* = \{g^*, \ldots\}$

Myopic Follower Strategies

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

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

Policy Iteration - MFS

- Begin with f^0 and $g(f^0)$ (e.g. $f^0 = \frac{1}{|\mathcal{A}|}$).
- Compute: $u_{A,0} = T_A^{f_0}(u_{A,0})$
- Find f_1:

$$T_A^{f_1}(u_{A,0}) = T_A(u_{A,0})$$

- Compute: $u_{A,1} = T_A^{f_1}(u_{A,1})$
- ...
- Repeat until convergence.

Policy Iteration - MFS

- Begin with f^0 and $g(f^0)$ (e.g. $f^0 = \frac{1}{|\mathcal{A}|}$).
- Compute: $u_{A,0} = T_A^{f_0}(u_{A,0})$
- Find f_1:

$$T_A^{f_1}(u_{A,0}) = T_A(u_{A,0})$$

- Compute: $u_{A,1} = T_A^{f_1}(u_{A,1})$
- $\ldots$
- Repeat until convergence.

Theorem 4.

The sequence of functions $u_{A,n}$ verifies $u_{A,n} \uparrow v_A^*$. Even more, 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 - MFS

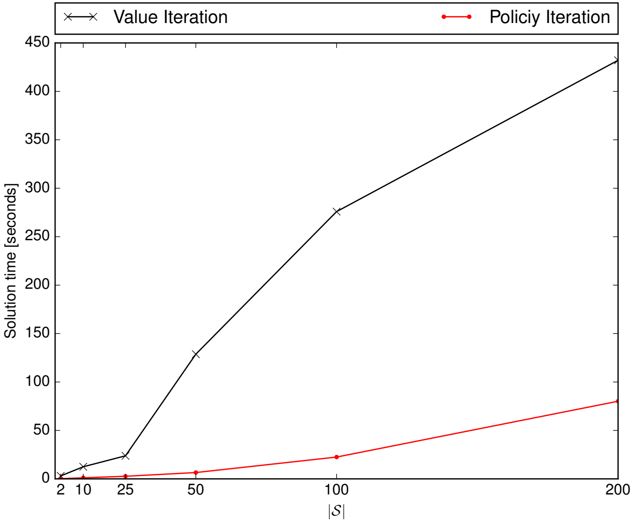
Algorithm 2 Policy Iteration (PI)

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

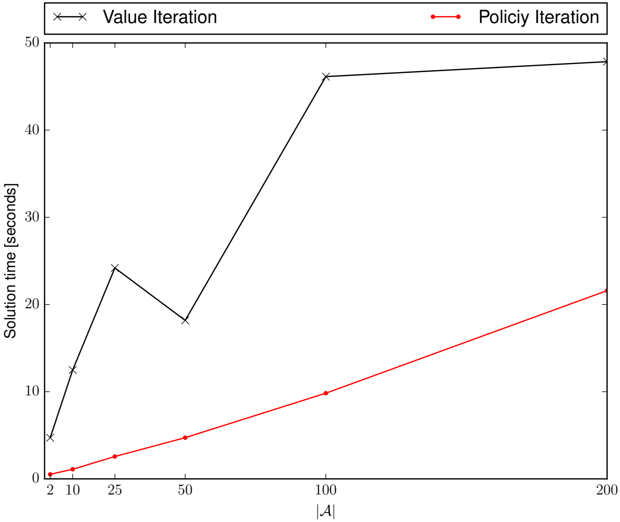
$$T_A^{f_{n+1}}(u_{A,n}) = T_A(u_{A,n}) \ .$$

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

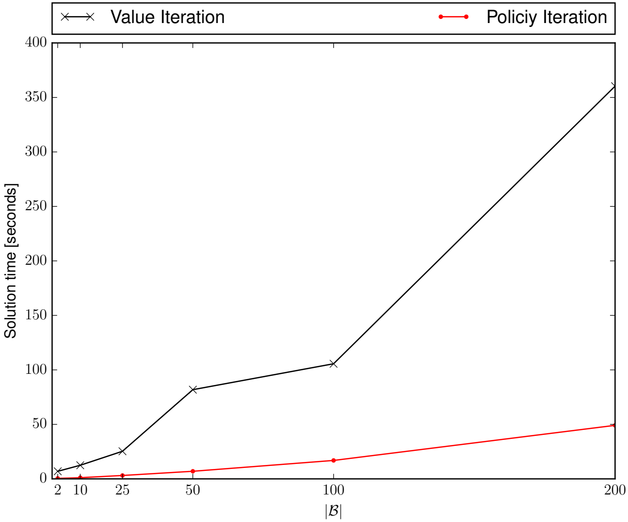
Computational Results - MFS



Computational Results - MFS



Computational Results - MFS



General Case

- f and g fixed stationary policies
- $T_i^{f,g} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$, $i \in \{A, B\}$

$$T_i^{f,g}(v_i)(s) = \sum_{a \in \mathcal{A}_s} f(a) \sum_{b \in \mathcal{B}_s} g(b) \left[r_i^{ab}(s) + \beta_i \sum_{z \in \mathcal{S}} Q^{ab}(z|s) v_i(z) \right]$$

Operator for the General case

$T : \mathbb{R}^{|\mathcal{S}|} \times \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|} \times \mathbb{R}^{|\mathcal{S}|}$

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

Algorithm

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

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

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

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

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)	$(1, 0)$ / $(6, -4)$

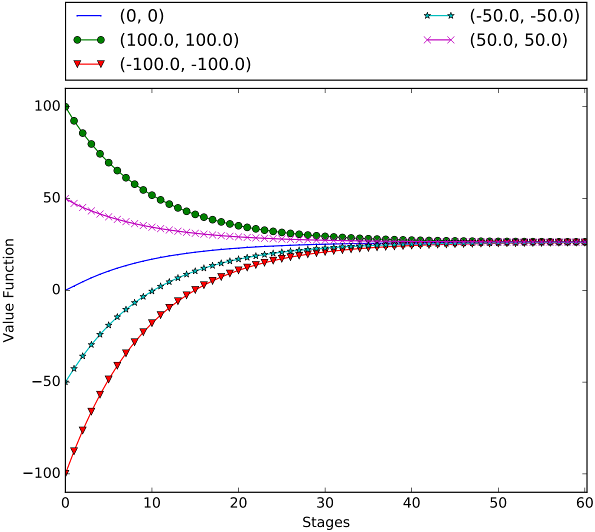
State s_1

	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

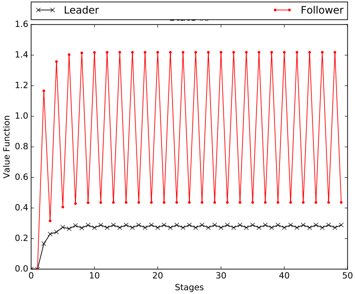
$$\beta_A = \beta_B = 0.9$$

Example

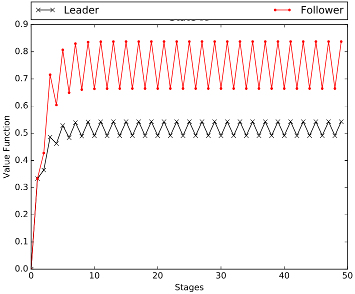


(0, 0)
(100.0, 100.0)
(-100.0, -100.0)
(-50.0, -50.0)
(50.0, 50.0)
Value Function
Stages
100
50
0
−50
−100
0
10
20
30
40
50
60

Counterexample

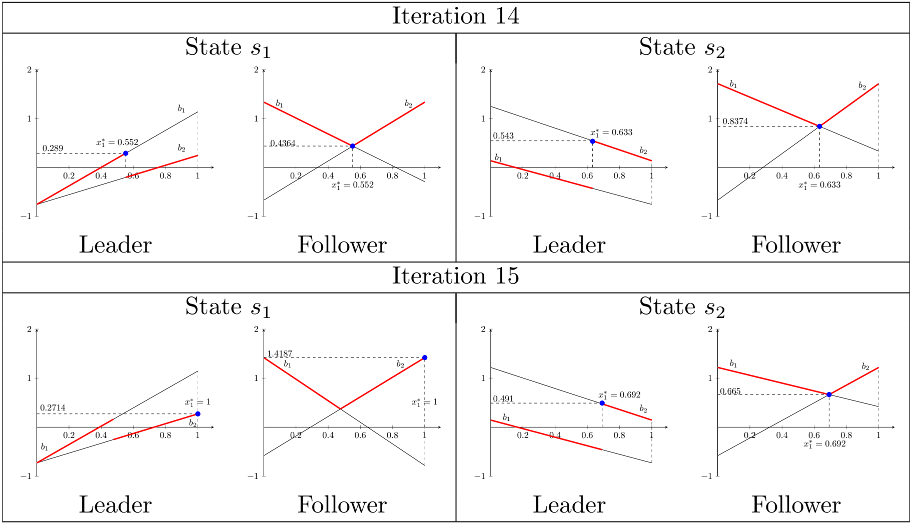


State s_1



State s_2

Counterexample



Computational Results - General Instances

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.

Security Games

$$r_A^{ab}(s) = \begin{cases} R_A(b) > 0 & \text{if } b = a \\ P_A(b) < 0 & \text{otherwise} \end{cases}$$

$$r_B^{ab}(s) = \begin{cases} P_B(b) < 0 & \text{if } b = a \\ R_B(b) > 0 & \text{otherwise} \end{cases}$$

- Non pure strategies seems to be optimal for the leader.
- Computationally all instances in Security games VI converges with the geometric bound.

Security Games

$$r_A^{ab}(s) = \begin{cases} R_A(b) > 0 & \text{if } b = a \\ P_A(b) < 0 & \text{otherwise} \end{cases}$$

$$r_B^{ab}(s) = \begin{cases} P_B(b) < 0 & \text{if } b = a \\ R_B(b) > 0 & \text{otherwise} \end{cases}$$

- 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 β contractive, with $\beta = \max\{\beta_A, \beta_B\}$.

Computational Results - General Instances

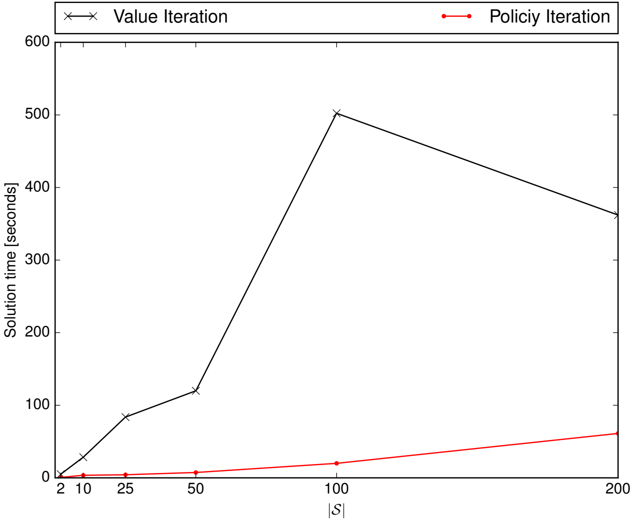


Figure: Performance of VI and PI in general random instances generated.

Computational Results - General Instances

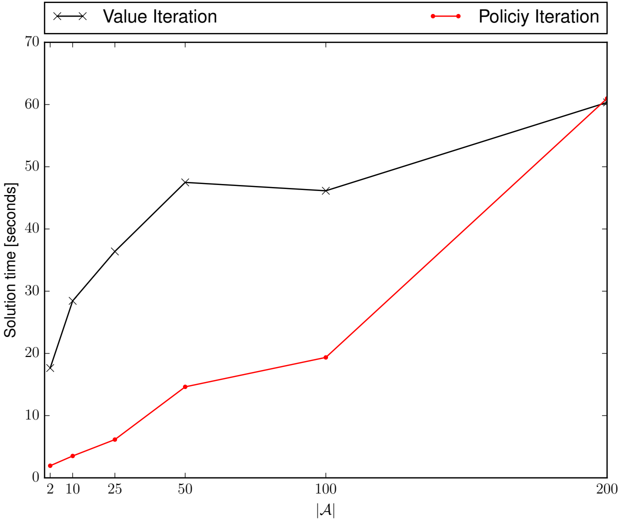


Figure: Performance of VI and PI in general random instances generated.

Computational Results - General Instances

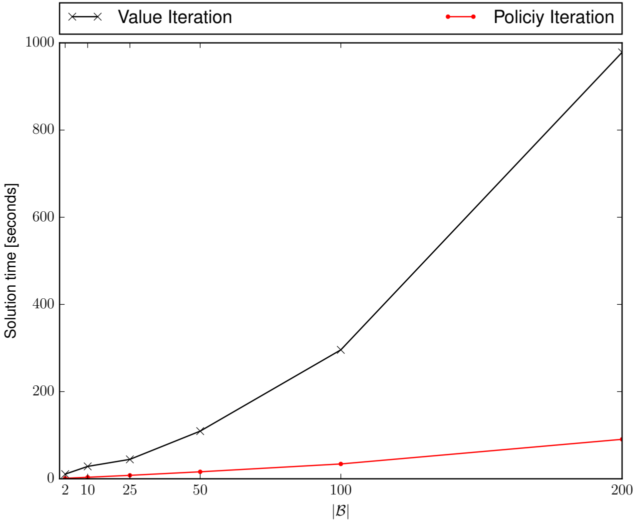


Figure: Performance of VI and PI in general random instances generated.

Computational Results - % UNDEFINED.

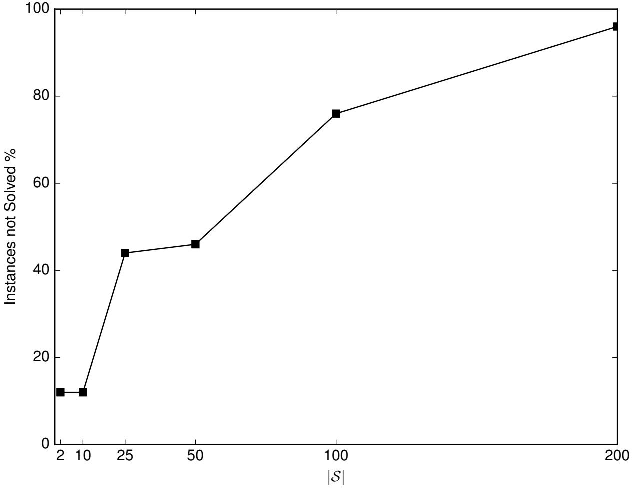


Figure: Percentage of instances where VI returns UNDEFINED.

Computational Results - % UNDEFINED.

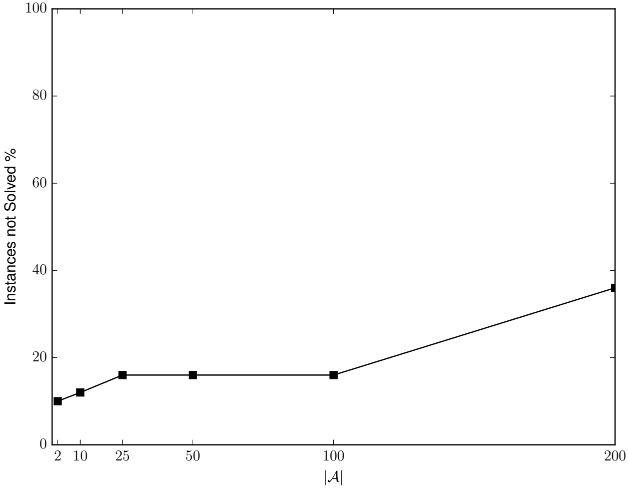


Figure: Percentage of instances where VI returns UNDEFINED.

Computational Results - % UNDEFINED.

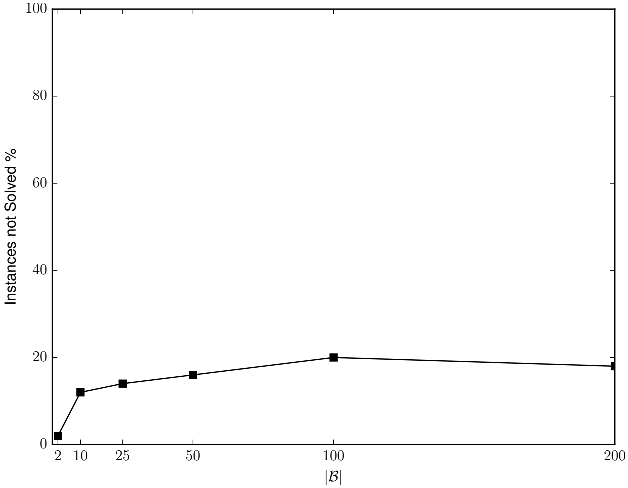


Figure: Percentage of instances where VI returns UNDEFINED.

Conclusions

- We define suitable Dynamic Programming operators.
- We used it to characterize value functions and to prove existence and unicity of stationary policies 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.
- We study security games and we conjecture that operators this type of games are contractive.

Future Work

- We aim to prove the convergence of VI procedure for security games.
- Rolling horizon techniques.
- Applicability Approximate Dynamic Programming techniques.
- To formalize and understand the behavior of Cyclic policies forming strong Stackelberg equilibrium.

Thank you!

Víctor Bucarey López
vbucarey@ing.uchile.cl

FCEIA - UNR - Rosario
November 2nd, 2017

References

1. Tansu - Alpcan and Tamer Basar. Stochastic security games, page 74-97. Cambridge. University Press, 2010.
2. Tamer Basar, Geert Jan Olsder. Dynamic noncooperative game theory, volume 200. SIAM, 1995.
3. Francesco Maria Delle Fave, Albert Xin Jiang, Zhengyu Yin, Chao Zhang, Milind Tambe, Sarit Kraus, and John P Sullivan. Game-theoretic patrolling with dynamic execution uncertainty and a case study on a real transit system. Journal of Artificial Intelligence Research, 2014.
4. Jerzy Filar and Koos Vrieze. Competitive Markov decision processes. Springer Science & Business Media, 2012.
5. Yevgeniy Vorobeychik, Bo An, Milind Tambe, and Satinder Singh. Computing solutions in infinite-horizon discounted adversarial patrolling games. In Proc. 24th International Conference on Automated Planning and Scheduling (ICAPS 2014)(June 2014), 2014.
6. Yevgeniy Vorobeychik and Satinder Singh. Computing Stackelberg equilibria in discounted stochastic games (corrected version). 2012

Counterexample

	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

Table: Transition matrix and payoffs for each player in the numerical example 2.

Back-up slides: Stochastic games