Solving Stackelberg Equilibrium in Stochastic Games.

Víctor Bucarey López

vbcarey@ing.uchile.cl

FCEIA - UNR - Rosario

November 2nd, 2017
Stackelberg Game

Leader

\( \pi \)
Stackelberg Game

Leader
\( \pi \) \rightarrow \text{Follower} \gamma(\pi)
**Stackelberg Game**

- **Leader** \( \pi \) \( \rightarrow \) **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

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$(10,-10)$</td>
<td>$(-5,6)$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$(-8,4)$</td>
<td>$(6,-4)$</td>
</tr>
</tbody>
</table>

MIP formulation

\[
\begin{align*}
\text{max } v_A \\
v_A & \leq 10x_1 - 8x_2 + M(1 - y_1) \\
v_A & \leq -5x_1 + 6x_2 + M(1 - y_2) \\
0 & \leq v_B - (-10x_1 + 4x_2) \leq M(1 - y_1) \\
0 & \leq v_B - (6x_1 - 4x_2) \leq M(1 - y_2) \\
\end{align*}
\]

\[
\begin{align*}
x_1 + x_2 &= 1 \\
y_1 + y_2 &= 1 \\
x & \geq 0, \quad y \in \{0, 1\}
\end{align*}
\]
<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>(10, -10)</td>
<td>(-5, 6)</td>
</tr>
<tr>
<td>$a_2$</td>
<td>(-8, 4)</td>
<td>(6, -4)</td>
</tr>
</tbody>
</table>

Leader

Follower
## Multiple States

<table>
<thead>
<tr>
<th></th>
<th>( b_1 )</th>
<th>( b_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>( (\frac{1}{2}, \frac{1}{2}) )</td>
<td>( (0, 1) )</td>
</tr>
<tr>
<td></td>
<td>( (10, -10) )</td>
<td>( (-5, 6) )</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>( (\frac{1}{4}, \frac{3}{4}) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td></td>
<td>( (-8, 4) )</td>
<td>( (6, -4) )</td>
</tr>
</tbody>
</table>

### State \( s_1 \)

<table>
<thead>
<tr>
<th></th>
<th>( b_1 )</th>
<th>( b_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>( (\frac{1}{2}, \frac{1}{2}) )</td>
<td>( (0, 1) )</td>
</tr>
<tr>
<td></td>
<td>( (7, -5) )</td>
<td>( (-1, 6) )</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>( (\frac{1}{4}, \frac{3}{4}) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td></td>
<td>( (-3, 10) )</td>
<td>( (2, -10) )</td>
</tr>
</tbody>
</table>

### State \( s_2 \)
Stochastic Games - Definition

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

\[ s_0 \overset{\sim}{\Rightarrow} \text{Player A chooses } f_0 \overset{\sim}{\Rightarrow} \text{Player B observes } f_0 \text{ and chooses } g_0 \overset{\sim}{\Rightarrow} Q_{f_0g_0}(s_1|s_0) \]

\[ \overset{\sim}{\Rightarrow} s_1 \overset{\sim}{\Rightarrow} \text{Player A chooses } f_1 \overset{\sim}{\Rightarrow} \text{Player B observes } f_1 \text{ and chooses } g_1 \overset{\sim}{\Rightarrow} s_2 \cdots \]
Stochastic Games - Definition

\[ G = (S, A, B, Q, r_A, r_B, \beta_A, \beta_B, \tau) \]

\[ s_0 \sim \text{Player A chooses } f_0 \sim \text{Player B observes } f_0 \text{ and chooses } g_0 \sim Q^{f_0g_0}(s_1|s_0) \sim s_1 \sim \text{Player A chooses } f_1 \sim \text{Player B observes } f_1 \text{ and chooses } g_1 \sim s_2 \ldots \]

Feedback Policies:
\[ \pi = \pi(s, t) = \{ f_1, \ldots, f_\tau \} \]

Stationary Policies:
\[ \pi = \pi(s) = \{ f, \ldots, f \} \]
## Framework

<table>
<thead>
<tr>
<th>General Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existence and characterization of value functions.</td>
</tr>
<tr>
<td>Existence of equilibrium strategies.</td>
</tr>
<tr>
<td>Algorithms to compute them.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State of the Art</th>
</tr>
</thead>
<tbody>
<tr>
<td>For finite horizon, Stackelberg equilibrium in stochastic games via Dynamic programming.</td>
</tr>
<tr>
<td>Mathematical programming approach to compute stationary values.</td>
</tr>
</tbody>
</table>
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

\[(\pi, \gamma)\]

Value Functions

\[v_{\pi,\gamma}^A(s) = \mathbb{E}_{s,\tau}^{\pi,\gamma} \left[ \sum_{t=0}^{\tau} \beta_t r_{A_t}^{A_t, B_t}(S_t) \right] \]

\[v_{\pi,\gamma}^B(s) = \mathbb{E}_{s,\tau}^{\pi,\gamma} \left[ \sum_{t=0}^{\tau} \beta_t r_{B_t}^{A_t, B_t}(S_t) \right] \]
Stackelberg equilibrium

\[ (\pi, \gamma) \]

Value Functions

\[
\begin{align*}
\nu_{A}^{\pi, \gamma}(s) &= E_{s}^{\pi, \gamma} \left[ \sum_{t=0}^{\tau} \beta^{t} A(t) R_{A}(s_{t}) + B(t) R_{B}(s_{t}) \right] \\
\nu_{B}^{\pi, \gamma}(s) &= E_{s}^{\pi, \gamma} \left[ \sum_{t=0}^{\tau} \beta^{t} B(t) R_{B}(s_{t}) \right]
\end{align*}
\]

Stackelberg Equilibrium

\[ (\pi^{*}, \gamma^{*}) \]

\[
\begin{align*}
\nu_{A}^{\pi^{*}, \gamma^{*}}(s) &= \max_{\pi, \gamma^{*}} \nu_{A}^{\pi, \gamma^{*}}(s) \\
\gamma^{*} &\in \text{argmax} \nu_{B}^{\pi^{*}, \gamma}(s)
\end{align*}
\]
Myopic Follower Strategies

Best response functional:

\[
g(f, v_B) = \arg \max_{b \in B_s} \sum_{a \in A_s} f(a) \left[ r_{ab}^B(s) + \beta_B \sum_{z \in S} Q_{ab}^B(z|s)v_B(z) \right]
\]
Myopic Follower Strategies

Best response functional:

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

Myopic follower strategies (MFS):

\[
g(f, v_B) = g(f)
\]
Myopic Follower Strategies

Best response functional:

\[ g(f, v_B) = \arg \max_{b \in B_s} \sum_{a \in A_s} f(a) \left[ r_{B}^{ab}(s) + \beta_B \sum_{z \in S} Q^{ab}(z|s) v_B(z) \right] \]

Myopic follower strategies (MFS):

\[ g(f, v_B) = g(f) \]

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.
- $T^f_A : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|}$.

$$T^f_A(v_A)(s) = \sum_{a \in A_s} f(a) \left[ r^a_{ag}(f)(s) + \beta_A \sum_{z \in S} Q^a_{ag}(f)(z|s)v_A(z) \right]$$
Myopic Follower Strategies

- $f$ a stationary policy.
- $T^f_A : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|}$:

$$T^f_A(v_A)(s) = \sum_{a \in A_s} f(a) \left[ r^a_{ag}(f)(s) + \beta_A \sum_{z \in S} Q^a_{ag}(f)(z|s)v_A(z) \right]$$

Operator for the MFS case

$$T_A(v_A)(s) = \max_{f \in \mathbb{P}(A_s)} T^f_A(v_A)(s) \quad (1)$$
Myopic Follower Strategies

Theorem 1.

a) $T^f_A$, $T_A$ are monotone.

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

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

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: \( \varepsilon > 0 \)
1: Initialize with \( n = 1 \), \( v_A^0(s) = 0 \) for every \( s \in S \) and \( v_A^1 = T_A(v_A^0) \)
2: while \( ||v_A^n - v_A^{n-1}||_\infty > \varepsilon \) do
3: Compute \( v_A^{n+1} \) by
   \[
   v_A^{n+1}(s) = T_A(v_A^n)(s).
   \]
   Finding \( f^* \) and \( g^*(f) \) at stage \( n \).
4: \( n := n + 1 \)
5: end while
6: return Stationary Stackelberg policies \( \pi^* = \{f^*, \ldots\} \) and \( \gamma^* = \{g^*, \ldots\} \)
Myopic Follower Strategies

Theorem 3.

The sequence of value functions $v^n_A$ converges to $v^*_A$. Furthermore, $v^*_A$ is the fixed point of $T_A$ with the following bound

$$\|v^*_A - v^n_A\|_\infty \leq \frac{\|r_A\|_\infty \beta^n_A}{1 - \beta_A}.$$
Policy Iteration - MFS

- Begin with $f^0$ and $g(f^0)$ (e.g. $f^0 = \frac{1}{|A|}$).
- 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.

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

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

1: Choose a stationary Stackelberg pair \((f_0, g(f_0))\).
2: \textbf{while} \(\|u_{A,n} - u_{A,n+1}\| > \varepsilon\) \textbf{do}
3: \hspace{1em} Evaluation Phase: Find \(u_{A,n}\) fixed point of the operator \(T_{A}^{f_n}\).
4: \hspace{1em} Improvement Phase: Find a strategy \(f_{n+1}\) such that
\[
T_{A}^{f_{n+1}}(u_{A,n}) = T_{A}(u_{A,n}) .
\]
5: \hspace{1em} \textbf{n:= n+1}
6: \hspace{1em} \textbf{end while}
7: \textbf{return} \text{ Stationary Stackelberg policies } \pi^* = \{f^*, \ldots\} \text{ and } \gamma^* = \{g(f^*), \ldots\}
Computational Results - MFS

The diagram shows the solution time in seconds as a function of the set size $|S|$. There are two methods compared: Value Iteration (black crosses) and Policy Iteration (red line). As $|S|$ increases, the solution time for both methods increases, with Value Iteration generally taking longer than Policy Iteration.
Computational Results - MFS

<table>
<thead>
<tr>
<th></th>
<th>Solution time [seconds]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value Iteration</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph showing solution time in seconds for Value Iteration and Policy Iteration with respect to |A|. The solution time increases significantly with increasing |A| for both methods, but the graph indicates that Policy Iteration has a more linear increase compared to Value Iteration.
Computational Results - MFS

![Graph showing solution time vs. |\(B|\) with two distinct lines for Value Iteration and Policy Iteration.](image)
General Case

- $f$ and $g$ fixed stationary policies
- $T^{f,g}_i : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|}$, $i \in \{A, B\}$

$$T^{f,g}_i(v_i)(s) = \sum_{a \in A_s} f(a) \sum_{b \in B_s} g(b) \left[ r_{i}^{ab}(s) + \beta_i \sum_{z \in S} Q^{ab}(z|s)v_i(z) \right]$$

Operator for the General case

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

$$(T(v_A, v_B))(s) = \left( \max_{f \in \mathcal{P}(A_s)} T^f_{A, g(f, v_B)}(v_A)(s), T^f_{B, g(f^*, v_B)}(v_B)(s) \right)$$
Algorithm 3 Value Iteration (VI): Finite horizon for the general case

1: Initialize with $v_A^{τ+1}(s) = v_B^{τ+1}(s) = 0$ for every $s \in S$
2: for $t = τ, \ldots, 0$, and for every $s \in S$ do
3: Solve
   \[(v_A^t(s), v_B^t(s)) = T(v_A^{t+1}, v_B^{t+1})(s) \quad \forall s \in S\]
   Finding $f_t^*$ and $g_t^*$ SSE strategies at stage $t$.
4: end for
5: return Stackelberg policies $π^* = \{f_0^*, \ldots, f_τ^*\}$ and $γ^* = \{g_0^*, \ldots, g_τ^*\}$
Example

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$(\frac{1}{2}, \frac{1}{2})$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>(10, -10)</td>
<td>$(-5, 6)$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$(\frac{1}{4}, \frac{3}{4})$</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td>(-8, 4)</td>
<td>$(6, -4)$</td>
</tr>
</tbody>
</table>

State $s_1$

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$(\frac{1}{2}, \frac{1}{2})$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>(7, -5)</td>
<td>$(-1, 6)$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$(\frac{1}{4}, \frac{3}{4})$</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td>(-3, 10)</td>
<td>$(2, -10)$</td>
</tr>
</tbody>
</table>

State $s_2$

$$\beta_A = \beta_B = 0.9$$
Example

![Graph showing value function over stages]

- (0, 0)
- (100.0, 100.0)
- (-100.0, -100.0)
- (-50.0, -50.0)
- (50.0, 50.0)
Counterexample

State $s_1$

State $s_2$
### Iteration 14

<table>
<thead>
<tr>
<th>State $s_1$</th>
<th>State $s_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Leader**

**Follower**

### Iteration 15

<table>
<thead>
<tr>
<th>State $s_1$</th>
<th>State $s_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Leader**

**Follower**
Computational Results - General Instances

**Algorithm 4** VI modified: Infinite horizon for the general case

1: Initialize with $n = 0$, $v_A^0(s) = v_B^0(s) = 0$ for every $s \in S$.
2: for $n = 1, \cdots, \text{MAX\_IT}$ do
3:   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 $n - 1$.
4:   if $(v_A^n, v_B^n) = (v_A^{n-1}, v_B^{n-1})$ then
5:     return $(v_A^n, v_B^n)$ fixed point of $T$.
6:   end if
7:   if $\|(v_A^n, v_B^n) - (v_A^{n-1}, v_B^{n-1})\| > 2^{\frac{n-1}{1-\beta}} \|(r_A, r_B)\|$ then
8:     return UNDEFINED 1.
9:   end if
10: end for
11: return UNDEFINED 2.
Non pure strategies seems to be optimal for the leader.

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

\[ r_{ab}^A(s) = \begin{cases} R_A(b) > 0 & \text{if } b = a \\ P_A(b) < 0 & \text{otherwise} \end{cases} \]

\[ r_{ab}^B(s) = \begin{cases} P_B(b) < 0 & \text{if } b = a \\ R_B(b) > 0 & \text{otherwise} \end{cases} \]

- 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 $\beta$ contractive, with $\beta = \max\{\beta_A, \beta_B\}$. 
Computational Results - General Instances

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


## Counterexample

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$(1, 0)$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>$(1, -1)$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$(0, 1)$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>$(-1, 1)$</td>
<td>$(-1, -1)$</td>
</tr>
</tbody>
</table>

**State $s_1$**

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$(0, 1)$</td>
<td>$(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$(-1, 0)$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$(1, 0)$</td>
<td>$(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>$(0, 1)$</td>
<td>$(1, -1)$</td>
</tr>
</tbody>
</table>

**State $s_2$**

**Table:** Transition matrix and payoffs for each player in the numerical example 2.
Back-up slides: Stochastic games