Stationary Strong Stackelberg Equilibrium in Discounted Stochastic Games

Alain Jean-Marie

Inria – Université de Montpellier, France

a joint work with:

Victor Bucarey López

Universidad de O'Higgins, Chile

Eugenio Della Vecchia

Universidad Nacional de Rosario, Argentina

Fernando Ordóñez

Universidad de Chile, Santiago, Chile

realized in part during the SticAmSud project DyGaMe

Dynamic Games and Applications Seminar

GERAD, 13 October 2022

Ideas of the paper

The Stackelberg solution to (stochastic) games is an appealing concept for Operations Research because of its predictive potential.
In this paper:

- we investigate the question of existence and computation of such equilibria in stochastic games
- we introduce the dynamic programming operator associated with the game
- we realize that
  - this operator does not necessarily have fixed points (FPE)
  - when it does, FPE are not necessarily equilibria for the game
  - and actually, there may be no equilibria at all...
- we provide sufficient conditions for everything to work well

Outline

Origin of the problem: security games

A very recent review on security games:

Trends and Applications in Stackelberg Security Games,
D. Kar, T.H. Nguyen, F. Fang, M. Brown, A. Sinha, M. Tambe, A.X. Jiang,
Chapter 28 in in Handbook of Dynamic Game Theory, T. Başar and G. Zaccour, eds.
Springer, 2018.

This reference and others on Security Games explains that the relevant solution concept is the Strong Stackelberg Equilibrium.

The Stackelberg solution concept

Consider a game with two players A and B.

- action sets A for Player A/Leader/Defender, B for Player B/Follower/Attacker
- set of strategies: W_A and W_B (typically $W_A \subset \mathbb{P}(A)$)
- payoffs r_A, r_B: $A \times B \to \mathbb{R}$.

The steps of the (sequential) game are:

- Player A plays some action $a \in A$
- Player B observes the action a
- Player B chooses optimally her action b
- payoffs $r_i(a, b)$ are obtained.

Goal of Player A: optimize her (expected) payoff over W_A

Stackelberg, ctd.

If B's reaction to A's action a is a unique strategy $\gamma(a) \in W_B$, then A can predict what B will do. She just choses the strategy that maximizes her own payoff:

$$\max_{f \in W_A} \sum_a \sum_b f(a) \times [\gamma(f)](b) \times r_A(a, b).$$

But if $|\arg\max_g \{r_B(a, g)\}| > 1$... bummer.

$\implies$ some more elaborate solution concept is needed.

Strong Stackelberg

Formal definition credited to:

Strong Stackelberg Equilibrium (Breton, Alj and Haurie, Def. 2.1)

Define the response/reaction set:

$$R_B(a) \;=\; \{b \in B \mid r_B(a,b) \geq \sup_{c \in B} r_B(a,c)\} \;.$$

A SSE is a pair (a^*, b^*) such that:

$$b^* \in R_B(a^*)$$
$$r_A(a^*, b^*) \geq \sup_{a \in A} \Big\{ \sup_{b \in R_B(a)} \{r_A(a,b)\} \Big\} \;.$$

They themselves refer to bilevel programming.

Leadership Games

Leadership Games are variants of Stackelberg games (von Stengel & Zamir, *GEB*, 2010).

The steps of the game are:

- Player A announces a strategy in W_A
- Player B reacts optimally to this known strategy

Main difference: Player B does not observe the action but does observe the strategy.
No difference if $W_A = A$ (pure strategies).

$\Longrightarrow$ concept credible if there is some sort of "commitment" on the part of Player A.
$\Longrightarrow$ If the game is repeated and Player B makes statistics, she can

- test the commitment
- react to the observed strategy instead

SSE for dynamic games

What about *dynamic* games?

In dynamic games, there is a state space $\mathcal{S}$;

- rewards depend on $s \in \mathcal{S}$: $r_A(s, a, b)$, $r_B(s, a, b)$
- there is a probability transition function $Q(z|s, a, b)$
- players optimize the total expected discounted gain

$$V_i(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta_i^t r_i(X_t, A_t, B_t)\right] \qquad i = A, B, X_0 = s \ .$$

  This goal is multi-objective: maximize $V_i(s)$ for all s
- the set of strategies is... what? the observation/information is... what?

SSE for dynamic games, ctd.

Same basic idea, relevant in particular to Security Games:

- A announces a strategy $f = (f_0, f_1, f_2, \ldots)$
  (like in Leadership Games)
- B reacts to it by $g = (g_0, g_1, \ldots) = \gamma(f_0, f_1, f_2, \ldots)$
- A maximizes $r_A(s)$ with respect to $f_0, f_1, f_2, \ldots$

Problem:

- A's optimum is not a stationary strategy in general
  (Vorobeychik & Singh, counterexample attributed to Conitzer)
- computing the optimum is hard (Letchford & al, 2012)

SSE in stationary feedback

Despite their suboptimality, many authors recommend to focus on stationary feedback strategies:

$$f : \mathcal{S} \to \mathbb{P}(A)$$
$$g : \mathcal{S} \to \mathbb{P}(B).$$

Then the proximity to Markov Decision Processes (MDPs) is striking:

- B's optimal response to a stationary policy of A is indeed solving a MDP
  $\to$ existence of a solution in pure feedback strategies
  $\to$ *strong* reaction set $R_B(f)$

Definition of dynamic SSE

$V_i^{fg}(s)$: value of state s for Player i when stationary policies f and g are played.

Strong Stationary Stackelberg Equilibrium

A strategy pair (f^*, g^*) is a Strong Stackelberg Equilibrium in Stationary Strategies (SSSE) if

$$g^* \in R_B(f^*)$$
$$V_A^{f^*,g^*}(s) = \sup\{V^{f,g}(s); f \text{ stationary}, g \in R_B(f)\}$$

for all $s \in S$.

A diverse litterature...

Litterature on Stackelberg Equilibria and their Strong form comes from several sources with imperfect communications...

- Mathematics, Mathematical Economics: Game Theory
  Simaan & Cruz, Breton, Alj & Haurie, Başar & Olsder, Osborne & Rubinstein, ...
- Artificial Intelligence: Complexity, Algorithmic Game Theory
  Conitzer *et al*, Letchford *et al*, Vorobeychik & Singh, ...
- Operations Research: Mathematical Programming, Bilevel Programming
  Kar *et al*, Tambe *et al*, ...

A diverse litterature... but incomplete?

In all this literature, the question of the existence of SSSE is hardly touched.

Question

Does there always exist a Strong Stackelberg Equilibrium in Stationary Strategies (SSSE) in finite-state discounted stochastic games?

We try to tackle the question with the operator approach.

Related approaches

Why operators?
The approach is successful for:

- MDPs
- Competitive MDPs/Nash Stochastic Games (Vilar & Vrieze)
- Sequential Stackelberg Games (Breton, Alj & Haurie, 1988)

each time with an existence result, at least in mixed strategies.

Progress

1. Introduction
2. The operator approach
   - Classical Operators
   - Operators for SSE
3. Negative results
4. Positive Results

Standard Operator Approach in MDP

One-slide reminder of basic MDP theory:

- In a discounted MDP, the optimal value exists and satisfies a Bellman equation:

$$V(x) \;=\; \max_a \left\{ r(x,a) + \beta \sum_z Q(z|s,a) V(z) \right\}$$

- The right-hand side defines an operator T on value functions
- The optimal value is a fixed point: $TV^* = V^*$
- Two major uses of the operator:
  - Existence: T is contractive $\to$ existence & uniqueness of V^*
  - Computation: V^* approximated by *value iteration*: $V_{n+1} = TV_n$.

Standard Operator Approach in Stochastic Games

A reminder of "competitive MDP" theory: Shapley's stochastic zero-sum game with the Nash solution.

$$(Uv)(s) \;=\; \mathrm{val}\left[r(s,a,b) + \beta \sum_{z} Q(z|s,a,b)v(z)\right] \; .$$

Existence (Filar&Vrieze)

U is contractive, and there exists an equilibrium point.

Also, for general-sum games:

Existence (Filar&Vrieze, Theorem 4.6.4)

Every non-zero sum stochastic game has an equilibrium point in stationary strategies.

Operators for Stackelberg Games

We wish to reproduce this scheme in Stackelberg games.
One-step Dynamic programming operator on functions v: $S \times \{A, B\} \to \mathbb{R}$:

$$(T^{fg}v)_i(s) = \mathbb{E}^{fg}\left(r_i(s,a,b) + \beta_i \sum_z Q(z|s,a,b)v_i(z)\right)$$

$$= \underbrace{\sum_a \sum_b f(s,a)g(s,b)\left[r_i(s,a,b) + \beta_i \sum_z Q(z|s,a,b)v_i(z)\right]}_{:= h_i(s,f,g,v)}$$

Reaction sets

Strong Reaction set of follower with "scrap value":

$$R_B(s,f,v) \;=\; \left\{ \beta \in \mathbb{P}(B) \mid h_B(s,f,\beta,v) = \sup_{g\in\mathbb{P}(B)} h_B(s,f,g,v) \right\}$$

+ ties broken in favor of A + ordering on W_B.

Reaction set of the leader with "scrap value": for each state s,

$$R_A(s,v) \;=\; \left\{ f(s) \mid (T_A^{fR_B(s,f,v)} v_A)(s) \geq (T_A^{hR_B(s,h,v)} v_A)(s), \forall h \right\}.$$

Operators T^{fg} contractive $\rightarrow$ unique fixed point V^{fg}.

Operator definition

Let T be the operator on pairs of functions v:

$$(Tv)_i(s) \; = \; T_i^{R_A(s,v),R_B(s,R_A(s,v),v)} v_i(s).$$

Fixed-Point Equilibrium

A strategy pair (f^*, g^*) is a Fixed-Point Equilibrium if value $v^* \equiv V^{f^*,g^*}$ is such that, equivalently,

- $Tv^* = v^*$
- for all $s \in S$,

$$\begin{aligned} g^* &\in R_B(s, f^*, v^*) \\ v_A^*(s) = &\sup_{\alpha \in \mathbb{P}(A)} \left\{ \sup\{\mathbb{E}^{\alpha,\gamma} h_A(s,\alpha,\gamma,v^*), \gamma \in R_B(s,f^*,v^*)\} \right\} \end{aligned}$$

Progress

A two-state counterexample

Let $\varepsilon > 0$ and $M > 0$. It is assumed that

$$M\beta_B - \varepsilon > 0.$$

Data: (transition distribution/costs)

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

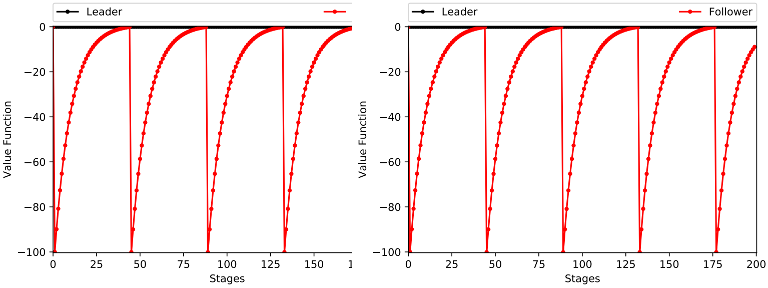
State s_1

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

State s_2

Application of VI

We run Value Iteration. Indeed, it does not converge!



state s_1 state s_2

There seems to be a cycle of large period.

Fail

The scheme of proof based on the operator approach fails!

- there is no Fixed Point
- there is not even a Strong Stackelberg Equilibrium in stationary strategies!

Principle

Features:

- gains do not depend on the state

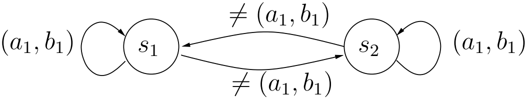
r_A:

	b_1	b_2
a_1	1	0
a_2	0	0

r_B:

	b_1	b_2
a_1	0	ε
a_2	$-M$	$-M$

- state changes if not (a_1, b_1)



Story:

- Player A has interest to stay in the same state and win 1 every turn $\rightarrow$ play a_1
- But Player B's response to a_1 is b_2, not b_1!

Principle (ctd.)

- So Player A needs to *menace* B with playing a_2 in the other state; B will anticipate she loses $-M$ and the state will come back to s_1
- The menace is effective if B loses less by playing b_1:

$$\underbrace{\varepsilon - \beta_B \times M}_{\text{B plays } b_2} < \underbrace{0 + \beta_B \times 0}_{\text{B plays } b_1} .$$

- Player A's optimum in state s_1 is to announce: $s_1 \rightarrow a_1; s_2 \rightarrow a_2$
- By symmetry, in state s_2 she must announce: $s_1 \rightarrow a_2; s_2 \rightarrow a_1$
- $\rightarrow$ no SSSE.

Findings

Conclusion of this example:

- There does not exist a SSSE in general
- Value Iteration does not necessarily converge

Findings on other examples

- When Value Iteration converges, the FPE is not necessarily a stationary SSE
- There may be cases where a FPE does exist, but VI does not converge to it from any initial solution
  $\rightarrow$ the operator is not contractive, actually not even continuous.

Progress

Contribution: myopic followers

Myopic Follower Strategies
Consider the best response functional:

$$h_B(s,f,v) = \arg\max_{b\in B} \mathbb{E}^f[r_B(s,a,b) + \beta_B \sum_z Q^{ab}(z|s)v_B(z)]$$

It is a real-valued function, not set-valued, thanks to the tie-breaking rule of follower + additional tie-breaking rule for leader.

MFS

A game is with Myopic Follower Strategies (MFS) if:

$$R_B(s,f,v_B) = R_B(s,f), \qquad \forall f \in \mathbb{P}(A), s\in\mathcal{S}, v\in\mathcal{F}(\mathcal{S}).$$

Stackelberg with Myopic Follower Strategies

Existence theorem

If a finite-state, finite-action game is with Myopic Follower Strategies, then it has a unique FPE which is also a SSSE. Value Iteration converges geometrically to it.

Idea of Proof: If the game has MFS, the operator is such that $(Tv)_A$ depends only on v_A.
This operator on "v_A" functions is shown to be contractive $\Longrightarrow$ unique fixed point & geometric convergence.

Characterization of MFS

Theorem

MFS is equivalent to either

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

$\Longrightarrow$ no particular structure on the instantaneous reward $r(\cdot)$.

Multi-stage games

A particular case of Leader-Controlled Games, quite common in (counter-)examples from AI, are:

Multi-stage games

In a multi-stage game, the state evolves sequentially and deterministically through $s_1, s_2, \ldots, s_K$ and stops.

The evolution is actually not controlled at all!

Particular case of the particular case: single state.

Other existence results

The existence of SSSE or/and FPE can be proved for other classes of games:

- zero-sum games
- acyclic games
- team/common-goal games.

Proof: the operator is contractive on some specific subset of value functions.

Acyclic Games

Acyclic Games

The game is an Acyclic Game if the state space $\mathcal{S}$ admits the partition $\mathcal{S} = \mathcal{S}_\perp \cup \mathcal{S}_1$, with:

- for all $s \in \mathcal{S}_\perp$, $a \in \mathcal{A}_s$, $b \in \mathcal{B}_s$, $Q^{ab}(s|s) = 1$;
- for every pair $(s, s') \in \mathcal{S}_1 \times \mathcal{S}_1$, if s' is reachable from s, then s is not reachable from s'.

$\Longrightarrow$ no particular structure on the rewards.

Theorem

If the stochastic game $\mathcal{G}$ is an Acyclic Game, then it admits an FPE.

However, existence of SSSE is not guaranteed.

Team Games

Team Game (generalization)

The game is a Team Game (or Identical Goal Game) if $\beta_A = \beta_B$ and there exists real constants μ and $\nu > 0$ such that: $r_B^{ab}(s) = \mu + \nu r_A^{ab}(s)$.

More common definition: with $\mu = 0$ and $\nu = 1$.

$\Longrightarrow$ no particular structure on transitions.

Team Games (ctd.)

Steps for the solution:

- Construction of the cooperative MDP
- Existence of a set $\mathcal{H}$ of deterministic optimal stationary policies h: $s \to (a, b)$
- Optimal value: $\widetilde{V}^* = \widetilde{V}^h$ for each $h \in \mathcal{H}$
- For any $h \in \mathcal{H}$, define $f^h \in W_A$ and $g^h \in W_B$ as:

$$f^h(s, a) = 1 \quad \text{iff } h(s, (a, b)) = 1 \text{ for some } b$$
$$g^h(s, b) = \sum_{a \in \mathcal{A}_s} f^h(s, a) h(s, (a, b))$$

so that

$$h(s, (a, b)) = f(s, a) g(s, b)$$

Team Games (end)

Final step:

- Define h^*:

$$h^* = \arg\max_{\prec_B}\{g^h : h \in \mathcal{H}\} .$$

Theorem

The pair (f^{h^*}, g^{h^*}) forms an SSSE and an FPE with value $v_A^* = \widetilde{V}^*$ for the leader and

$$v_B^* = \frac{\mu}{1-\beta} + \nu \widetilde{V}^*$$

for the follower.

Conclusions and issues

Conclusions:

- FPE may or may not exist
  - find more sufficient conditions for existence
  - find ways (algorithms?) to test for existence or not in practice
- When FPE exist, how to compute it/them?
  - Value Iteration may or may not converge
  - They may or may not be Stationary SSE
- Stationary SSE are not optimal for the leader anyway
  - FPE as a way to get better policies?

More details in Inria Research Report #9271:

https://hal.inria.fr/hal-02144095