On the Value Iteration method for dynamic Strong Stackelberg Equilibria

Víctor Bucarey, Alain Jean-Marie, Eugenio Della Vecchia, Fernando Ordóñez

On the Value Iteration method for dynamic Strong Stackelberg Equilibria

Víctor Bucarey[1], Alain Jean-Marie[2], Eugenio Della Vecchia[3], Fernando Ordóñez[4]

[1] Département d'Informatique, Université Libre de Bruxelles, vbucarey@gmail.com
[2] Inria, Université Côte d'Azur, alain.jean-marie@inria.fr
[3] Departamento de Matemática, Universidad Nacional de Rosario, eugenio@fceia.unr.edu.ar
[4] Departamento de Ingeniería Industrial, Universidad de Chile, fordon@dii.uchile.cl

Mots-clés : *Game Theory, Stackelberg Equilibria, Value Iteration*

1 Introduction

The concept of Strong Stackelberg Equilibrium has received a renewed attention in the recent years, due notably to its application to security questions [2]. In Stackelberg games, it is assumed that one of the players, called the Leader, commits to a strategy. The other player, called the Follower, learns or observes this strategy, and reacts rationally to it. The game being with complete information, the Leader is able to predict the reaction of the follower, and is then able to optimize her strategy to maximize her own rewards. It may be that the follower's response is not unique, in which case this prediction of the leader is not possibly anymore. A *strong* Stackelberg Equilibrium (SSE) occurs when the Follower breaks such ties in favor of the Leader : she chooses the best option for the Leader among the best ones for her.

Initially defined in static games, this concept has been extended to inifinite-horizon, discounted dynamic games. Calculating SSE in such games within the set of general strategies appears to be very difficult in general [3]. With the hope of reducing this complexity, as well as memory storage for the strategies, many authors concentrate on *stationary feedback* policies, although they are known to be suboptimal [4]. Even in this restricted class, the problem is still NP-Complete to compute in general [3] and several algorithms based on Mathematical Programming exist in the literature for solving or approximating this problem [4].

In this paper, we study the opportunity to use instead the Value Iteration algorithm, the well-known dynamic programming (DP) method used to solve, in particular, Markov Decision Processes. In this context, it proves useful to introduce a DP (or Bellman) operator acting on value functions. We therefore seek to adapt the technique to dynamic Stackelberg games. Our contribution is as follows : we first define a DP operator T corresponding to one-step strong Stackelberg games. Equilibria are defined as fixed points of T, and the Value Iteration algorithm is defined as the iterations of T until convergence. We discuss through examples (non)existence of equilibria and (non)convergence of the algorithm. [1]

2 Definitions

The system has a finite state space $\mathcal{S}$. In each state s, the Leader and the Follower, respectively denoted with A and B, have strategy sets $\mathcal{A}_s$ and $\mathcal{B}_s$ and play a one-shot game where they gain rewards $r_i(s, a, b)$, $i = A, B$, and cause the system to jump in state s' with probability $Q(s, a, b, s')$ where a scrap value $v_i(s')$ is obtained, discounted by a factor β_i. The sets of mixed stationary feedback strategies W_A and W_B are defined as probability distributions over $\mathcal{A}_s$ and $\mathcal{B}_s$ in each state s. For any mixed strategies f and g, the expected gain of each

1. Part of this research was realized during the SticAmSud project 16-STIC-10 DyGaMe.

player is given by fixed point of linear operator : $(T_i^{fg}v)(s) = \sum_{a\in\mathcal{A}_s, b\in\mathcal{B}_s} f(s,a)g(s,b)[r_i(s,a,b)+\beta_i \sum_{z\in\mathcal{S}} Q(s,a,b,z)v(z)]$ acting on functions from $\mathcal{S}$ to $\mathbb{R}$. The SSE is defined using reaction sets. A reaction set is the set of actions that are optimal for one player, given her knowledge of the opponent's strategy. Here, they are defined as follows :

$$R_B(s,f,v_B) := \{g \in W_B \mid (T_B^{fg}v_B)(s) \geq (T_B^{fh}v_B)(s), \forall h \in W_B\} \quad (1)$$

$$\gamma_B(s,f,v) := \max_{\prec_B}\{g \in R_B(s,f,v_B) \mid (T_A^{fg}v_A)(s) \geq (T_A^{fh}v_A)(s), \forall h \in R_B(s,f,v_B)\} \quad (2)$$

$$R_A(s,v) := \max_{\prec_A}\{f \in W_A \mid (T_A^{f\gamma_B(s,f,v)}v_A)(s) \geq (T_A^{h\gamma_B(s,h,v)}v_A)(s), \forall h \in W_A\}\ . \quad (3)$$

These definitions depend on some total orderings $\prec_A$ and $\prec_B$ on action sets. Finally, the dynamic programming operator T is defined on value functions (v_A, v_B) as, for $i = A, B$:

$$[(Tv)_i](s) = [T_i^{R_A(s,v),\gamma_B(s,R_A(s,v),v)}v_i](s).$$

3 Examples and discussion

The first situation we will present is cases where the operator T has a fixed point to which iterations converge. This extends the class of Myopic Follower Strategies, introduced in [1], where the operator is actually globally contractive. This property (summarized as : $|Tv - Tw| \leq \beta|v-w|$, $\beta < 1$), classically implies the existence and uniqueness of a fixed point $Tv = v$.

The second example we present has the feature that there exists a fixed point to operator T. However, as we will argue, the operator is not continuous at this fixed point, and iterations of the operator do not converge, unless the starting point is very finely chosen.

Finally, we present an example where no fixed point exist. It is opportune to note here that no such example has been presented yet. The question of the existence dynamic SSE is not even discussed in the current literature, where all examples have structures where the existence of a solution is almost evident. This example has two states (one being absorbing), three actions per player, and the structure of security games where players have rewards of opposite signs.

When convergence does not occur, we observe that the successive values draw, asymptotically, a periodic pattern, as displayed in Figure 1.

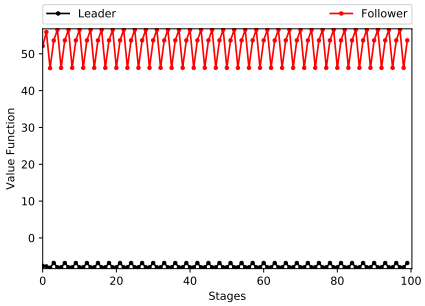

FIG. 1 – Iterates of the value function of both players in state 1 for the third example.

Références

[1] V. Bucarey. *Addressing Problem Size in Stackelberg Security Games.* PhD thesis, Universidad de Chile, 2017.

[2] D. Kar, T. Nguyen, F. Fang, M. Brown, A. Sinha, M. Tambe, and A. Jiang. Trends and applications in Stackelberg security games. In T. Başar and G. Zaccour, editors, *Handbook of Dynamic Game Theory*, chapter 28, pages 1223–1269. Springer, 2018.

[3] J. Letchford, L. MacDermed, V. Conitzer, R. Parr, and C. L. Isbell. Computing optimal strategies to commit to in stochastic games. In *26th AAAI*, 2012.

[4] Y. Vorobeychik and S. Singh. Computing Stackelberg equilibria in discounted stochastic games (corrected version). Retreived online on Oct. 19, 2018, 2012.