



**HAL**  
open science

## Learning in Conjectural Stackelberg Games

Francesco Morri, H el ene Le Cadre, Luce Brotcorne

► **To cite this version:**

Francesco Morri, H el ene Le Cadre, Luce Brotcorne. Learning in Conjectural Stackelberg Games. 2025.  
hal-04908694

**HAL Id: hal-04908694**

**<https://hal.science/hal-04908694v1>**

Preprint submitted on 23 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

---

# LEARNING IN CONJECTURAL STACKELBERG GAMES

Francesco Morri<sup>1,\*</sup>, H el ene Le Cadre<sup>1</sup>, and Luce Brotcorne<sup>1</sup>

<sup>1</sup>Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

\*Corresponding author: francesco.morri@inria.fr

## Abstract

We extend the formalism of Conjectural Variations games to Stackelberg games involving multiple leaders and a single follower. To solve these nonconvex games, a common assumption is that the leaders compute their strategies having perfect knowledge of the follower’s best response. However, in practice, the leaders may have little to no knowledge about the other players’ reactions. To deal with this lack of knowledge, we assume that each leader can form conjectures about the other players’ best responses, and update its strategy relying on these conjectures. Our contributions are twofold: (i) On the theoretical side, we introduce the concept of Conjectural Stackelberg Equilibrium – keeping our formalism conjecture agnostic – with Stackelberg Equilibrium being a refinement of it. (ii) On the algorithmic side, we introduce a two-stage algorithm with guarantees of convergence, which allows the leaders to first learn conjectures on a training data set, and then update their strategies. Theoretical results are illustrated numerically.

## 1 Introduction

Game theory has been used in many fields to model and understand the outcomes of situations involving strategic agents in competition [Bagar & Olsder \(1998\)](#). In this paper, we focus on bridging the gap between Stackelberg games, involving hierarchical decision making, and Conjectural Variations (CVs) games by introducing Conjectural variations Stackelberg Equilibrium (CSE) as a broad solution concept that encompasses Stackelberg Equilibrium (SE) as a refinement.

**Conjectures in Games.** CV games are characterized by *conjectures*, which are functions used by players to model their adversaries’ reactions to their own action [Figui eres et al. \(2004\)](#). Conjectures are especially relevant in a context of bounded rationality [Simon \(1955\)](#), where each player chooses its action based on its subjective perception of the other players’ behaviors. CV games have been introduced in the context of duopolies [Bresnahan \(1981\)](#); [Friedman & Mezzetti \(2002\)](#); [Perry \(1982\)](#), considering classical Bertrand and Cournot equilibrium models. A new stream has appeared with the works of [Calderone et al. \(2023\)](#); [Chasnov et al. \(2019\)](#) who rely on variational analysis and dynamic system theory to provide generic interpretations of conjectural games. CV games have not been extensively applied in real-world problems, though some applications can be found in the context of electricity markets [D iaz et al. \(2010\)](#); [Chen et al. \(2021\)](#).

**Stackelberg Games.** We consider non-cooperative games involving a finite set of leaders at the upper level which compute their strategies based on their anticipations of the best response of the follower, which reacts rationally to the leaders’ strategies. Some cornerstone theoretical contributions have emerged on the analysis of multi-leader single follower (MLSF) Stackelberg games [Sherali \(1984\)](#); [Leyffer & Munson \(2010\)](#), but algorithmic contributions in this area remain scarce. Applications of MLSF Stackelberg games can be found in jamming problems [Zhang et al. \(2018\)](#), strategic bidding in deregulated markets [Morri et al. \(2024\)](#), security and privacy games [Gan et al. \(2018\)](#).

**Linking Conjectural and Stackelberg Games.** Most approaches in the literature consider leaders which have the ability to anticipate the best response of the follower. In reality, leaders

rely on conjectures to compute their adversaries' best responses. However, the use of conjectures in Stackelberg games has not been extensively studied. Close works are related to dynamic conjectural games for duopolies [Figuères et al. \(2004\)](#) and reverse Stackelberg games [Groot et al. \(2012\)](#) where the leaders announce their strategies as functions of the follower's action. CV Equilibrium (CVE) has been introduced in [Olsder \(2009\)](#) as a relevant solution concept to cope with incomplete information and implicit cooperation, as well as a shorthand for dynamic interactions. In [Rubinstein & Wolinsky \(1994\)](#), specific focus is put on the design of conjectures leading to a Nash Equilibrium. Conjectures can also be linked to models of bounded rationality, in place of prospect theory [Kahneman & Tversky \(1979\)](#); [Fochesato et al. \(2025\)](#), which requires to use complex subdifferential approaches to solve nonconvex nonsmooth games.

**Learning in Games with Hierarchical Structure.** There are two main approaches: either applying learning techniques to games, or using game theory to develop new learning algorithm. In our paper, we focus on the first class, leaving the possibility of future extensions in the second class. Learning in Stackelberg games is considered as a challenge. In this direction, [Fiez et al. \(2020\)](#) makes a significant contribution by developing a gradient-based learning rule for the leader, while the follower employs a best-response strategy. In the same vein, many algorithms involving a double-loop structure are proposed in the literature to compute stationary solutions [Grontas et al. \(2024\)](#). Single-loop stochastic algorithm has recently been proposed to tackle bilevel optimization problems [Hong et al. \(2023\)](#), allowing fast convergence rates. On the conjectural variations game side, [Wellman & Hu \(1998\)](#) propose learning dynamic rules for two player conjectural games. A similar idea is extended by [Chasnov et al. \(2019\)](#), considering gradient-based learning methods for anticipating the adversary's reaction in general conjectural games.

**Contributions.** We introduce CSE as a new solution concept to analyze Stackelberg games involving conjectures from the leaders on their adversaries' best responses. The distance between CV and Stackelberg game outcomes is upper bounded, and conditions on conjectures to reach a SE are identified. Then, we propose a two-stage algorithm to learn a CSE and provide convergence guarantees. Finally, the theoretical results are illustrated numerically.

**Structure.** In [Sec. 2](#) we introduce the main concepts of CV Games and MLSF Stackelberg Games, to then provide a definition for the novel class of Conjectural Stackelberg games and its equilibria. We then analyse CS Games in [Sec. 3](#), relating them to standard Stackelberg Games. In [Sec. 4](#) we introduce the **COSTAL** algorithm, with its proof of convergence, and in [Sec. 5](#) we provide numerical results obtained on multi player games.

## 1.1 Notations

We consider a set  $\mathcal{N} \stackrel{\text{def}}{=} \{1, \dots, N\}$  of  $N$  agents, each with strategy  $x_i \in \mathbb{R}^{m_i}$ ,  $m_i \in \mathbb{N}^*$ , with  $m \stackrel{\text{def}}{=} \sum_i m_i$ , and objective function  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ . We define  $\mathcal{X}_i \subseteq \mathbb{R}^{m_i}$  as the feasibility set of player  $i$ . We introduce  $\mathcal{X}_{-i} \stackrel{\text{def}}{=} \prod_{j \neq i} \mathcal{X}_j$  as the product of the feasibility sets of all the players in  $\mathcal{N}$  except  $i$ , and  $\mathcal{X} \stackrel{\text{def}}{=} \prod_{i \in \mathcal{N}} \mathcal{X}_i$  as the joint feasibility set. We denote  $x \stackrel{\text{def}}{=} (x_i)_{i \in \mathcal{N}}$  as the collective strategy of all the players. Throughout the paper  $\nabla$  will stand for the total derivative,  $\nabla_i$  the partial derivative w.r.t.  $x_i$  and  $\nabla_{-i}$  refers to all the terms  $(\nabla_j)_{j \in \mathcal{N} \setminus \{i\}}$ . For second-order derivatives we use the notation:  $\nabla_{i,j}$  or  $\nabla_i^2$ , while we let  $D$  be the gradient. Finally,  $\|\cdot\|$  stands for the  $L_2$  norm.

## 2 Conjectural Stackelberg Games

In CV games, each player builds conjectures about the other players' best responses. Two scenarios are considered in the literature. In the first scenario, each player  $i \in \mathcal{N}$  assumes that the other players react homogeneously to its strategy thus giving rise to the modified objective function  $f_i(x_i, \bar{x})$ ,  $\forall i \in \mathcal{N}$  where  $\bar{x} \stackrel{\text{def}}{=} \sum_{j \in \mathcal{N} \setminus \{i\}} x_j$ . In the second scenario, player  $i$  relies on its decision variable  $x_i$  to form conjectures

$$\gamma_i^j : \mathcal{X}_i \rightarrow \mathcal{X}_j, \forall j \in \mathcal{N} \setminus \{i\}.$$

The first scenario is the one discussed the most in the literature, since it can be simplified as a two-player game, and is easy to analyze. We instead focus on the second scenario which has more interest from a practical point of view. Before defining our game, we explicit our standing assumptions.

**Assumptions 1.** *We consider the following assumptions for the game:*

1.  $f_i(\cdot) \in C^d, \forall i \in \mathcal{N}$ , with  $d \geq 2$ ;
2.  $f_i(\cdot), \forall i \in \mathcal{N}$  is  $M_1$ -Lipschitz and convex;
3. The set  $\mathcal{X}_i$  is compact and convex.

We now present some standard definitions from the CV game literature.

**Definition 1** (Conjectural Variations (CV) Game [Başar & Olsder \(1998\)](#)). *Given a set  $\mathcal{N}$  of players, each with decision variables  $x_i \in \mathcal{X}_i \subseteq \mathbb{R}^{m_i}$  and objective function  $f_i(x_i, x_{-i})$ , we define a conjecture as the function  $\gamma_i^j : \mathcal{X}_i \rightarrow \mathcal{X}_j, \forall j \in \mathcal{N} \setminus \{i\}$ . A conjecture variations (CV) game is defined as:*

$$\forall i \in \mathcal{N}, \quad \min_{x_i} \{f_i(x_i, \tilde{x}_{-i}) \mid \tilde{x}_{-i} = (\gamma_i^j(x_i))_{j \in \mathcal{N} \setminus \{i\}}\}. \quad (1)$$

The outcome of CV games can be analyzed relying on CVE. We let  $\{\gamma_i^j(x_i)\}_{j \in \mathcal{N} \setminus \{i\}, i \in \mathcal{N}}$  be the stack of all the players' conjectures.

**Definition 2** (Conjectural Variations Equilibrium [Figuères et al. \(2004\)](#)). *Given the conjectures  $\{\gamma_i^j(x_i)\}_{j \in \mathcal{N} \setminus \{i\}, i \in \mathcal{N}}$ , the collective strategy  $x^c \stackrel{\text{def}}{=} (x_i^c)_i$  is a Conjectural Variations equilibrium (CVE) if  $x^c$  is a solution of the optimization problem*

$$\forall i \in \mathcal{N}, \quad \min_{x_i} \{f_i(x_i, \tilde{x}_{-i}) \mid \tilde{x}_{-i} = (\gamma_i^j(x_i))_{j \neq i}\}. \quad (2)$$

Definition 2 may generate a multiplicity of solutions, since it is always possible to find a set of conjectures for any given point such that it becomes a CVE [Figuères et al. \(2004\)](#). Thus, the concept of consistency has been introduced as a refinement of a CVE.

**Definition 3** (Consistent Conjectural Variations Equilibrium [Bresnahan \(1981\)](#)). *Given the conjectures  $\{\gamma_i^j(x_i)\}_{j \in \mathcal{N} \setminus \{i\}, i \in \mathcal{N}}$ , the collective strategy  $x^c \stackrel{\text{def}}{=} (x_i^c)_i$  forms a Consistent Conjectural Variations Equilibrium (CCVE) if*

1.  $x^c$  is a CVE for the conjectures  $\gamma_i^j(x_i), \forall j \in \mathcal{N} \setminus \{i\}, \forall i \in \mathcal{N}$ ;
2. There exists  $\epsilon > 0$  such that  $\nabla \gamma_i^j(x_i) = \nabla_i x_j^c(x_{-j}), \forall x_i \in \mathcal{X}_i$  such that  $\|x_i^c - x_i\| < \epsilon, \forall i, j \in \mathcal{N}, i \neq j$ ;

where  $x_i(x_{-i})$  denotes player  $i$ 's best-response strategy.

The existence of CCVE is not easy to establish [Figuères et al. \(2004\)](#). For instance, duopolies with quadratic marginal costs may not allow for polynomial or symmetric analytic consistent conjectures. Similarly, [Calderone et al. \(2023\)](#) show that for quadratic games with affine conjectures, the existence of CCVE boils down to finding solutions to coupled asymmetric Riccati equations. Positive results for the existence require, in general, specific settings; such results exist, e.g., for specific cases of Cournot's duopoly and voluntary contributions to a public good [Figuères et al. \(2004\)](#), and in electricity markets where uniqueness is also considered [Liu et al. \(2007\)](#).

## 2.1 Multi-Leader Single-Follower Stackelberg Games

We consider Stackelberg games involving a set  $\mathcal{N}$  of leaders and a single follower. The  $N$  leaders act simultaneously, anticipating the reaction of the follower; then the follower replies after having observed the decision variables of all the leaders. We formally define a Stackelberg game as follows:

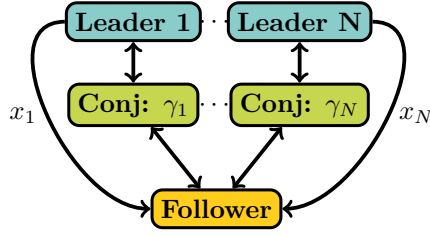


Figure 1: Schematic representation of a Conjectural Stackelberg Game.

**Definition 4** (Multi-Leader Single-Follower Stackelberg Game). *Consider a set  $\mathcal{N}$  of leaders and a follower, with respective decision variables  $x_i \in \mathcal{X}_i \subseteq \mathbb{R}^{m_i}, \forall i \in \mathcal{N}$ , and  $y \in \mathcal{Y} \subseteq \mathbb{R}^{m_y}$ , and objective functions:  $f_i(x_i, x_{-i}, y), \forall i \in \mathcal{N}$ , and  $g(x, y)$ . The Stackelberg game can be described as follows:*

$$\forall i \in \mathcal{N}, \min_{x_i \in \mathcal{X}_i, y \in \mathcal{Y}} f_i(x_i, x_{-i}, y), \quad (3a)$$

$$s.t. \quad y \in \underset{y \in \mathcal{Y}}{\operatorname{argmin}} g(x, y). \quad (3b)$$

**Assumptions 2.** *We need additional assumptions on the follower's problem:*

1. *The follower has a unique solution once  $x \in \mathcal{X}$  is fixed;*
2. *The follower's best response,  $y(x)$ , is  $M_2$ -Lipschitz;*
3.  *$g(\cdot) \in C^{d'}$  with  $d' \geq 2$ .*

We can see how the hierarchy introduced in Definition 4 already implicitly uses the concept of conjectures: the leaders need to act before the follower, meaning that they have to use an approximation, or a *conjecture*, in place of the actual strategy of the follower. If we extend the anticipation to all players, meaning that the leaders have conjectures about the follower's strategy and also about the other leaders, we obtain a new class of games, that we call Conjectural Stackelberg Games. Considering Definitions 1 and 4, a Conjectural Stackelberg Game can be formalized as follows:

$$\forall i \in \mathcal{N}, \min_{x_i \in \mathcal{X}_i} f_i \left( x_i, (\gamma_i^j(x_i))_{j \neq i}, \gamma_i^y(x_i) \right), \quad (4a)$$

$$\min_{y \in \mathcal{Y}} g(x, y). \quad (4b)$$

In Figure 1, we represent schematically the structure of these games. The  $N$  leaders receive feedback from their conjectures and update their strategies based on that information; then the follower optimizes its strategy after having observed the leaders' strategies. Following the literature on conjectural games, we introduce the novel concept of Conjectural Stackelberg Equilibrium (CSE):

**Definition 5** (Conjectural Stackelberg Equilibrium). *Given the conjectures  $\{ \{ \gamma_i^j(x_i) \}_{j \in \mathcal{N} \setminus \{i\}}, \gamma_i^y(x_i) \}_{i \in \mathcal{N}}$ ,  $(x^*, y^*)$  is a Conjectural Stackelberg Equilibrium (CSE) if it is a solution of Eqs. (4).*

Notice that Definition 5 does not consider the notion of consistency. Following the same steps as for standard conjectural games, we define requirements for CSE to be consistent:

**Proposition 1** (Consistency for CSE). *Given the conjectures  $\{ \{ \gamma_i^j(x_i) \}_{j \in \mathcal{N} \setminus \{i\}}, \gamma_i^y(x_i) \}_{i \in \mathcal{N}}$ , a CSE  $(x^*, y^*)$  is a Consistent CSE (CCSE) if*

1. *There exists  $\epsilon_L > 0$  such that  $\nabla \gamma_i^j(x_i) = \nabla_i x_j^c(x_{-j}), \forall x_i \in \mathcal{X}_i; \|x_i^* - x_i\| < \epsilon_L, \forall i, j \in \mathcal{N}, i \neq j$ ;*
2. *There exists  $\epsilon_F > 0$  such that  $\nabla \gamma_i^y(x_i) = \nabla_i y(x), \forall x_i \in \mathcal{X}_i, \|x_i^* - x_i\| < \epsilon_F, \forall i \in \mathcal{N}$  and  $\forall y \in \mathcal{Y}, \|y^* - y\| < \epsilon_F$ .*

The set of CCSEs is a superset of the set of Stackelberg Equilibria (SE), since if the conjectures coincide with the best responses, the game outcome coincides with a SE.

**Differential Equilibrium Definition** Given conjectures  $\{\{\gamma_i^j(x_i)\}_{j \in \mathcal{N} \setminus \{i\}}, \gamma_i^y(x_i)\}_{i \in \mathcal{N}}$ , we define first and second-order conditions for an equilibrium  $(x^*, y^*)$  to be a CSE

$$\begin{aligned} \nabla f_i \left( x_i^*, (\gamma_i^j(x_i^*))_{j \neq i}, \gamma_i^y(x_i^*) \right) &= 0, \\ \nabla^2 f_i \left( x_i^*, (\gamma_i^j(x_i^*))_{j \neq i}, \gamma_i^y(x_i^*) \right) &> 0, \forall i \in \mathcal{N}, \end{aligned}$$

and  $\nabla_y g(x^*, y^*) = 0, \nabla_y^2 g(x^*, y^*) > 0$ . Notice that under Assumptions 1 and 2, this definition is equivalent to Definition 5.

## 2.2 Form of the Conjectures

Our approach is *conjecture-agnostic* since we do not impose any priors on the form for the conjectures.

**Assumptions 3.** *The conjecture functions  $\gamma_i^j(\cdot), \forall i, j \in \mathcal{N}, i \neq j$  are twice differentiable.*

Some possible forms for these functions, that we will consider throughout the paper are the following:

- *Affine conjectures:*  $\gamma_i^j(x_i) = a_i x_i + b_i$ , for  $a_i, b_i \in \mathbb{R}$ ;
- *Polynomial conjectures:*  $\gamma_i^j(x_i) = \sum_{g=0}^G c_{i,g} x_i^g$ , for  $c_{i,g} \in \mathbb{R}$ ;
- *Neural networks:*  $\gamma_i^j(x_i) = \Phi_i^j(x_i)$ , where  $\Phi_i^j$  is the architecture for the network used.

## 3 Game Analysis

We can now proceed with a theoretical analysis of the game, comparing CV and Stackelberg games. We equip  $\mathcal{X}$  and  $\mathcal{Y}$  with distances  $d_{\mathcal{X}}$  and  $d_{\mathcal{Y}}$  respectively. For  $x^s, x^c \in \mathcal{X} \times \mathcal{Y}$  we define  $d(x^s, x^c) \stackrel{\text{def}}{=} \|(d_{\mathcal{X}}(x^s, x^\gamma), d_{\mathcal{Y}}(y^*(x^s), \gamma^y(x^\gamma)))\|$  as the distance of the projections of  $x^s, x^c$  on  $\mathcal{X}$  and  $\mathcal{Y}$  respectively.

### 3.1 Bounding the Objective Functions

**Proposition 2** (Objective Function Bound). *Suppose Assumptions 1-3 hold. Consider a SE  $x^s \stackrel{\text{def}}{=} (x^*, y^*(x^*))$  and a CSE  $x^c \stackrel{\text{def}}{=} (x^\gamma, \gamma^y(x^\gamma))$ , then the distance between the objective function of any leader  $i \in \mathcal{N}$  at a SE and at a CSE is upper bounded:*

$$\|f_i(x^s, y^*(x^*)) - f_i(x^\gamma, y^*(x^\gamma))\| \leq R d_{\mathcal{X}}(x^s, x^\gamma), \quad (5)$$

with  $R \stackrel{\text{def}}{=} M_1 \sqrt{1 + M_2^2}$ .

*Proof.* Let  $x^c$  correspond to the decision variable obtained at equilibrium by the leaders considering its conjecture, but evaluated with the *actual* follower's response  $y^*(x^\gamma)$ . Using the assumption of Lipschitz continuity of  $f_i$  from Assumption 1, we obtain:

$$\|f_i(x^s) - f_i(x^c)\| \leq M_1 d(x^s, x^c), \quad (6)$$

which is a first bound for the distance between the objective functions evaluated in  $x^s$  and  $x^c$ . If we explicit the right-hand side of Eq. (6), we obtain

$$d(x^s, x^c) = \sqrt{d_{\mathcal{X}}(x^s, x^\gamma)^2 + d_{\mathcal{Y}}(y^*(x^s), y^*(x^\gamma))^2}.$$

The second term, considering the distance between the follower's best responses, can be developed further using the Lipschitz continuity of the follower's best response from Assumption 2:

$$d_Y(y^*(x^*), y^*(x^\gamma)) \leq M_2 d_X(x^*, x^\gamma). \quad (7)$$

Combining Eq. (6) and Eq. (7) concludes the proof.  $\square$

### 3.2 Reaching a Stackelberg Equilibrium

The previous result linked a Stackelberg Equilibrium with a Conjectural Stackelberg Equilibrium, now instead we derive connections between a Stackelberg Equilibrium and a *Consistent* Conjectural Stackelberg Equilibrium. Using the differential equilibrium definition from Section 2, we write explicitly the full gradient of leader  $i$  considering either a SE or a CCSE. First, in  $x^s = (x^*, y^*)$ , we get:

$$\begin{aligned} \nabla f_i(x^s) &= \nabla_i f_i(x^s) + \nabla_{-i} f_i(x^s) \nabla_i x_{-i}^* \\ &\quad + \nabla_y f_i(x^s) \nabla_i y^*. \end{aligned} \quad (8)$$

Second, in a CCSE, in  $x^c = (x_i^\gamma, \gamma_i^{-i}(x_i^\gamma), \gamma_i^y(x_i^\gamma))$ , and using  $\gamma_i^{-i}(x_i) = (\gamma_i^j(x_i))_{j \neq i}$ , we get:

$$\begin{aligned} \nabla f_i(x^c) &= \nabla_i f_i(x^c) + \nabla_{-i} f_i(x^c) \nabla_i \gamma_i^{-i}(x_i^\gamma) \\ &\quad + \nabla_y f_i(x^c) \nabla_i \gamma_i^y(x_i^\gamma). \end{aligned} \quad (9)$$

We note that Eq. (8) and Eq. (9) are equal if

$$\begin{aligned} \nabla_i f_i(x^s) - \nabla_i f_i(x^c) &= 0, \\ \nabla_{-i} f_i(x^s) \nabla_i x_{-i}^* - \nabla_{-i} f_i(x^c) \nabla_i \gamma_i^{-i}(x_i^\gamma) &= 0, \\ \nabla_y f_i(x^s) \nabla_i y^* - \nabla_y f_i(x^c) \nabla_i \gamma_i^y(x_i^\gamma) &= 0. \end{aligned}$$

Since in a CCSE the variations of the conjectures must be equal to the real best-response variations, we can focus on the terms related to the derivatives of the objective function. Considering the derivative w.r.t. variable  $i$  we can rewrite the term  $\nabla_i f_i(x^s) - \nabla_i f_i(x^c)$  as:  $\nabla_i (f_i(x^s) - f_i(x^c))$ , and given Assumptions 1 for  $f_i$ , using the Mean Value Theorem we obtain:

$$\nabla_i [f_i(x^s) - f_i(x^c)] = \nabla_i [Df_i((1 - \tau)x^c + \tau x^s)(x^s - x^c)^\top] \quad (11)$$

with  $\tau \in [0, 1]$ . We set  $h \stackrel{\text{def}}{=} (1 - \tau)x^c + \tau x^s$ , then expanding the right-hand side of Eq. (11) we obtain the following term:

$$\begin{aligned} \nabla_i [f_i(x^s) - f_i(x^c)] &= \nabla_i^2 f_i(h)(x_i^\gamma - x_i^*) \\ &\quad + \nabla_{-i,i} f_i(h)(\gamma_i^{-i} - x_{-i}^*) \\ &\quad + \nabla_{y,i} f_i(h)(\gamma_i^y - y^*), \end{aligned} \quad (12)$$

and with the same procedure we equivalently obtain similar terms for the derivatives w.r.t.  $-i$  and  $y$ :

$$\begin{aligned} \nabla_{-i} [f_i(x^s) - f_i(x^c)] &= \nabla_{i,-i} f_i(h)(x_i^\gamma - x_i^*) \\ &\quad + \nabla_{-i}^2 f_i(h)(\gamma_i^{-i} - x_{-i}^*) \\ &\quad + \nabla_{y,-i} f_i(h)(\gamma_i^y - y^*), \end{aligned} \quad (13)$$

$$\begin{aligned} \nabla_y [f_i(x^s) - f_i(x^c)] &= \nabla_{i,y} f_i(h)(x_i^\gamma - x_i^*) \\ &\quad + \nabla_{-i,y} f_i(h)(\gamma_i^{-i} - x_{-i}^*) \\ &\quad + \nabla_y^2 f_i(h)(\gamma_i^y - y^*). \end{aligned} \quad (14)$$

Comparing Eqs. (12) - (14) we obtain the following link between SE and CCSE:



---

**Proposition 3** (Existence of Conjectures). *Given a SE  $x^*$ , if each player’s best response is differentiable, then we can always find a set of conjectures to reach  $x^*$ .*

*Proof.* It suffices to choose the best responses as the conjectures. □

We can further analyze Eqs. (12) - (14) from the point of view of the objective functions, obtaining two interesting cases:

1. **Linear Objective Functions:** Given linear objectives functions for the leaders  $f_i(x, y) = a_i x_i + \sum_{j \neq i} b_{i,j} x_j + c_i y$ , we observe that all the second-order derivatives in Eqs. (12) - (14) vanish, meaning that SE and CCSE coincide. This means that the full derivative of the conjectural game for leader  $i$  becomes:

$$\nabla_i f_i(x, y) = a_i + \sum_{j \neq i} b_{i,j} \nabla_i \gamma_i^j(x_i^j) + c_i \nabla_i \gamma_i^y(x_i^y),$$

which coincides with the SE definition, given the conditions of Proposition 1 on the conjectures.

2. **Quadratic Objective Functions:** Consider quadratic objective functions such as:  $f_i(x, y) = z A^i z^\top$ , where  $z = (x, y)$  and  $A^i$  is the coefficients matrix for leader  $i$ , with dimensions  $(m + m_y) \times (m + m_y)$ . In Eqs. (12) - (14) all the second derivatives become constant terms, i.e., the term  $\nabla_i^2 f_i(m)$  becomes  $A_{i,i}^i$ , the term  $\nabla_{y,i} f_i(m)$  becomes  $A_{i,y}^i$ , where the subscripts  $i, j$  are the row and column indexes of the matrix  $A$ . This implies that the difference between the equilibria depends solely on how close the conjectures at a CCSE are to the other players’ best responses at a SE.

Note that the conjecture parameters are generally not known and need to be learned. To that purpose, we propose a two-stage algorithm in the next section.

## 4 The COSTAL Algorithm

In our two-stage algorithm, we first train the leaders’ conjecture models, and then let the players play the Stackelberg game relying on the learned conjectures.

### 4.1 Training Conjectures

Before the game is played, a training phase takes place during which the leaders collect data about the other players. This can be seen as a similar process to *cheap talk* Farrell (1995), or Generative Adversarial Networks Farnia & Ozdaglar (2020), where a pre-trained discriminator is used. At each step of the training phase we sample uniformly a different decision variable for each leaders, then the best response for each player is computed, considering the randomly sampled variables and adding noise:  $\forall i \in \mathcal{N}, \tilde{x}_i = x_i^*(x_{-i}, y) + \xi$  with  $\xi \sim \mathcal{N}(0, \sigma)$ . The tuples  $(x_i, \tilde{x}_j), \forall j \in \mathcal{N} \setminus \{i\}$ , and  $(x_i, \tilde{y}), \forall i \in \mathcal{N}$ , where  $\tilde{y}$  is the noisy follower’s reaction to the strategy of the leaders, are stored in datasets  $(\mathcal{D}_i^j)_{i,j}$  and  $(\mathcal{D}_i^y)_i$ . This process is repeated  $T$  times, creating for each player a dataset of size  $T$  for every other player in the game. The next step is then to update the conjectures to learn the other players’ behaviors over these data streams using stochastic gradient descent with batch size  $|\mathcal{B}|$ , and with a quadratic loss defined as follows:

$$L_i^j(\gamma_i^j) \stackrel{\text{def}}{=} \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B} \subset \mathcal{D}_i^j} \left( x_j^b - \gamma_i^j(x_i^b) \right)^2, \forall i, j \in \mathcal{N}, j \neq i.$$

The algorithm to learn the players’ conjectures is described in Algorithm 1. In the sampling part (lines 3-7), the operations scale as  $T \cdot N$ , thus it is linear in the number of agents. In the training



---

**Algorithm 1** Pseudo-code for training the conjectures.

---

```

1: Input:  $\sigma, T, |\mathcal{B}|$ 
2: for  $t \in [0, T]$  do
3:   Sample  $x^t \in \mathcal{X}, y^t \in \mathcal{Y}, \xi \sim \mathcal{N}(0, \sigma)$ 
4:   Compute  $\tilde{y}(x^t) = y^*(x^t) + \xi$ 
5:   for  $i \in \mathcal{N}$  do
6:     Sample  $\xi \sim \mathcal{N}(0, \sigma)$ 
7:     Compute  $\tilde{x}_i(x^t_{-i}, y^t) = x_i^*(x^t_{-i}, y^t) + \xi$ 
8:   end for
9:   for  $i \in \mathcal{N}$  do
10:    Save  $(x_i^t, \tilde{y})$  in  $\mathcal{D}_i^y$ 
11:    for  $j \in \mathcal{N}, j \neq i$  do
12:      Save  $(x_i^t, \tilde{x}_j)$  in  $\mathcal{D}_i^j$ 
13:    end for
14:  end for
15: end for
16: Train the conjectures on the data sets  $\mathcal{D}_i^j, \mathcal{D}_i^y, \forall i, j \in \mathcal{N}$ 
17: Output: trained conjectures  $\gamma_i^j, \gamma_i^y, \forall i, j \in \mathcal{N}$ 

```

---

**Algorithm 2** Pseudo-code for learning a Conjectural Stackelberg Equilibrium.

---

```

1: Input:  $T, \eta, \{\gamma_i^j, \gamma_i^y, \forall i, j \in \mathcal{N}\}, \{f_i(\cdot), \forall i \in \mathcal{N}\}$ 
2: Initialize  $x^0 \in \mathcal{X}$ 
3: for  $t \in [0, T]$  do
4:   for  $i \in \mathcal{N}$  do
5:     Compute  $D_i = \nabla_i f_i$ 
6:     Compute  $D_{-i} = \nabla_{-i} f_i \nabla_i \gamma_i^{-i}$ 
7:     Compute  $D_y = \nabla_y f_i \nabla_i \gamma_i^y$ 
8:      $x_i^{t+1} = x_i^t - \eta(D_i + D_{-i} + D_y)$ 
9:   end for
10: end for
11: Output: Final strategies  $\{x_i^T \forall i \in \mathcal{N}\}, y^*(x^T)$ 

```

---

part, we have to train  $N \cdot (N - 1)$  conjecture functions, thus it scales quadratically with the number of agents. Notice that this part could be run in parallel for each agent (or by groups of agents), if  $N$  becomes large.

## 4.2 Learning Conjectural Stackelberg Equilibria

The leaders participating in the Stackelberg game update their strategies through gradient descent. Considering the continuous counterpart of our system, the players update their strategies as follows:

$$\dot{x}_i = -\nabla f_i(x_i, x_{-i}, y), \quad \forall i \in \mathcal{N}, \quad (15a)$$

$$\dot{y} = -\nabla_y g(x, y). \quad (15b)$$

The main issue with this formulation is that, since all the leaders are coupled, computing the full gradient for each of them is not feasible. For the settings we consider, the presence of conjectures helps at rewriting this system in a simpler way. Substituting the conjectures in the objective function of leader  $i$ ,  $f_i$ , we obtain  $f_i(x_i, (\gamma_i^j(x_i))_{j \neq i}, \gamma_i^y(x_i))$ , which depends only on  $x_i$ . We define  $\tilde{f}_i(x_i) \stackrel{\text{def}}{=} f_i(x_i, (\gamma_i^j(x_i))_{j \neq i}, \gamma_i^y(x_i))$ , with  $\tilde{f}_i$  the conjectured objective function. Eqs. (15) can then be rewritten

as follows:

$$\dot{x}_i = -\nabla_i \tilde{f}_i(x_i), \forall i \in \mathcal{N}, \quad (16a)$$

$$\dot{y} = -\nabla_y g(x, y). \quad (16b)$$

We now explicit  $\nabla_i \tilde{f}_i$ :

$$\begin{aligned} \nabla_i \tilde{f}_i(x_i) &= \nabla_i f_i(x_i, \gamma_i^{-i}, \gamma_i^y) + \nabla_{-i} f_i(x_i, \gamma_i^{-i}, \gamma_i^y) \nabla_i \gamma_i^{-i} \\ &\quad + \nabla_y f_i(x_i, \gamma_{-i}, \gamma_i^y) \nabla_i \gamma_i^y. \end{aligned} \quad (17)$$

The key feature of Eq. (17) is that it is easily obtained once the conjectures are given. Furthermore, a convergence guarantee for the system described in Eqs. (16) can be obtained following [Fiez et al. \(2020\)](#) and [Borkar \(2008\)](#). Specifically, we can write the discrete version of this update rule for leader  $i$ :

$$x_i^{t+1} = x_i^t - \eta^t (\nabla_i \tilde{f}_i(x_i^t) + \zeta_i^t), \quad (18)$$

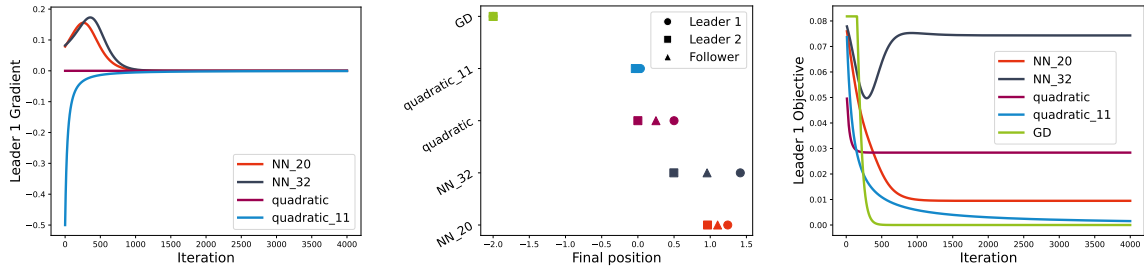
where we also consider the noise  $\zeta$  deriving from the fact that the leader has access to an unbiased estimate of the gradient. This allows us to use well-known results from stochastic approximation theory [Borkar \(2008\)](#), from which we also state the following assumptions:

**Assumptions 4** ( [Borkar, 2008](#), Sec. 2.1, Assumptions A1-A2-A3 ). *For all leader  $i \in \mathcal{N}$ ,  $\nabla_i \tilde{f}_i$  is Lipschitz continuous and  $\|\nabla_i \tilde{f}_i\| < +\infty$ ,  $\{\zeta_i^t\}$  is a martingale difference sequence, and the gradient step sizes are chosen such that  $\sum_t \eta^t = \infty$ ,  $\sum_t (\eta^t)^2 < +\infty$ .*

**Theorem 1.** *Suppose Assumptions 4 hold, then  $\{(x_i^t)_i\}_t$  converges almost surely to a local CSE  $(x_i^\gamma)_i$  of Eqs. (16).*

*Proof.* We consider the differential CSE  $(x^\gamma, y^\gamma)$ , given the set of conjectures  $\{(\gamma_i^j(x_i))_{j \neq i}, \gamma_i^y(x_i); i \in \mathcal{N}\}$ . From the differential version of Definition 5, we know that for any leader  $i$ ,  $\nabla_i f_i(x_i^\gamma) = 0$  and  $\nabla_i^2 \tilde{f}_i(x_i^\gamma) > 0$ , meaning that it is a stable point for the differential equation:  $\dot{x}_i = -\nabla_i f_i(x_i)$ . The discrete update  $x_i^{t+1} = x_i^t - \eta^t (\nabla_i \tilde{f}_i(x_i^t) + \zeta_i^t)$  is then a stochastic approximation of the continuous process, tracking the ODE asymptotically. Given the assumptions we described in the statement of the theorem and using [Borkar, 2008](#), Theorem 2, Corollary 4), the discrete sequence of each leader  $i$  converges almost surely to a compact internally chain transitive set of the ODE. This is a closed compact set  $\mathcal{A} \subset \mathcal{X}_i$  such that: for any trajectory  $x_i(t)$  with  $x_i(0) \in \mathcal{A}$ ,  $x_i(t) \in \mathcal{A} \forall t \geq 0$ ; and for any  $x^a, x^b \in \mathcal{A}$ ,  $\epsilon > 0, T > 0$ , there exists  $n \geq 1$  and  $x^a = x^0, x^1, \dots, x^n = x^b$  such that the trajectory of the ODE starting at  $x^l, 0 \leq l > n$ , reaches the  $\epsilon$ -neighbourhood of  $x^{l+1}$  after a time  $\geq T$ . Finally, since any stable attractor of the dynamic has to satisfy  $\nabla_i^2 \tilde{f}_i > 0$  for all leaders [Strogatz \(2015\)](#), and the followers always replies with the best response, by definition the only internally chain transitive sets are CSEs. We can then conclude that the algorithm using Eq. (18) converges a.s. to a CSE of Eqs. (16).  $\square$

In Algorithm 2 we report the structure of the algorithm in pseudo-code. Notice that the update of the strategy of player  $i$  (lines 5-8), obtained from Eq. (17), only depends on variables of player  $i$ . This means that the loop over all the players  $\mathcal{N}$  could be run in parallel instead than sequentially, speeding up COSTAL for large number of agents. Furthermore, this update does not even need to be run on the same machine, allowing full decentralization for the leaders. Note that COSTAL may never reach a global solution of Eqs. (15). There exist integrated approaches based on the reformulation of the Stackelberg games as an equivalent mathematical program with constraints [Dempe & Zemkoho \(2020\)](#) to compute global solutions, but they do not allow to account for the learning capabilities of the players. Further, reaching a global equilibrium can be expensive from a computational point of view, thus focusing on a local equilibrium might be preferable. In addition, from a system point of view, the adjustments of the players' strategies are usually small, meaning that it makes sense to study local solutions [Zheng et al. \(2021\)](#). From [Bubeck \(2015\)](#), we know that the convergence rate of the gradient descent given a smooth function, which we already require in Assumptions 4, is  $O(1/T)$ . Note that the loop over agents in Algorithm 2 (lines 4-9) includes an inner loop (line 6),



(a) Evolution of the gradient for leader 1. (b) Final position for all players. (c) Evolution of the objective of leader 1.

Figure 2: Results of the simulation for the game "Leader's Dilemma" with  $K = -1.5$ , we report only the results for one leader as the game is symmetric. The labels in the plots refer to which type of conjecture is used: NN stands for neural network, with the number referring to the size of the hidden layer; for the quadratic conjecture, just 'quadratic' refers to  $\gamma = x^2$ , while 'quadratic\_11' is  $\gamma = x^2 + x$ .

which runs over the  $N - 1$  remaining agents. This entails that the operations for each gradient step scale as  $O(N^2)$ . On the other hand, we mentioned that the gradient update can be run in parallel for all agents, bringing back the scaling to a linear behavior.

## 5 Experiments

We present numerical results for the **COSTAL** algorithm applied to two stylized games; for both games, we report additional data and a full script to reproduce the experiments in the Appendix. We focus on comparing the learned strategies with known equilibria, to check whether a CSE could benefit the players of the game, and if so, with which conjectures. In particular we compare our algorithm against a naive gradient descent update, with full access to the other players' variables, described by:

$$x_i^{t+1} = x_i^t - \eta \nabla_i f_i(x_i^t, x_{-i}^t, y^t), \forall t. \quad (19)$$

We will refer to this benchmark algorithm in the next sections as **GD**. Note that this method is not decentralized and cannot be run in parallel, as every update needs all the players' variables.

### 5.1 Leader's Dilemma

We consider here a Stackelberg game where two leaders should try to cooperate by getting close to a target set by the follower. The objective function for the leaders are:

$$f_i(x_i, x_j, y) = -(x_i - y)^2 - K \left(1 - e^{-(x_j - y)^2}\right), \forall i \in \{1, 2\}, \quad (20)$$

and for the follower:

$$g(x, y) = \left(\frac{x_1 + x_2}{2} - y\right)^2. \quad (21)$$

The leaders maximize their objective function, while the follower minimizes it, all variables are in the action space  $[-2, 2]$  and  $K < -1$ . It is easy to prove that the strategy  $x_1 = x_2$  for the leaders is a saddle point, thus they could obtain better payoffs by staying far from the target. In Fig. 2a we check convergence speed of the algorithm: the leaders' gradient converges to 0 rather quickly for each class of conjectures, as we expected from our theoretical analysis. Furthermore, in Fig. 2c we plot the leaders' objective function at equilibrium: we can see that the **GD** stays trapped in the saddle point, the two leaders stay in the same position, obtaining 0 as reward; the **COSTAL** algorithm with the different conjectures instead, finds stable solutions that guarantee a reward larger than 0,

beating the GD. Finally, in Fig. 2b, we plot the final solutions reached in each case. We infer that the conjectures allow the leaders to find an equilibrium which is at a distance from the target set by the follower, while the GD method cannot.

## 5.2 Revisiting Olsder’s Paradox

For our second experiment we take a game described in Olsder (2009) (players are maximizing). The game considers two players with action space  $[0, +\infty]$  and with objective functions:

$$f_1(x_1, x_2) = (x_1 - 84)(-12.5x_1 + 21x_2 + 756),$$

$$f_2(x_1, x_2) = (x_2 - 50)(25x_1 - 50x_2 + 560).$$

This game can be interpreted as a paradox because it leads to a CCE with affine conjectures, which is more efficient than the NE and the SE (with player 1 as leader, and player 2 as follower) solutions of the game. This result is counter-intuitive, because it means that being bounded rational can be more profitable for the players and more efficient for a system point of view than being fully rational. The numerical results are reported in Tab. 1, where we also highlight the social welfare optimum for the game, which is reached by maximizing the sum of the objective functions. For the numerical simulations, we consider two settings: simultaneous play (N at the front of the name in the graphs) and Stackelberg play (S), with player 1 using the algorithm as the leader, while player 2 replies with the best response. In Fig. 3 we report the values reached by the objective function of player 1 and 2 respectively. We can see how using the COSTAL algorithm produces results that always beat the NE, and for player 1 in the simultaneous case we even achieve a better result than the SE. The CCE is far more profitable for the players and more efficient than the NE and the SE. However, this result needs to be mitigated in practice, because a large amount of information is needed to compute a CCE: relying on Proposition 1, player 1 would need to have access to player 2’s objective function, together with its first and second-order derivatives. However, it is worth noticing that COSTAL enables to reach CSE without any information exchange among the players, achieving better performance (players’ objective values, efficiency) than the NE and the SE. Furthermore, under the hierarchical play setting, we can observe that Proposition 2 holds: both players’ objective functions are upper bounded by the SE.

Solution	$x_1$	$x_2$	$f_1$	$f_2$
<b>CCE</b>	164.4	81	32320.8	19220
<b>SE</b>	138.04	65.11	21411.6	11415.8
<b>NE</b>	123.98	61.6	19979.8	6722.13
<b>SWO</b>	300.04	150.98	38141.2	56548.7

Table 1: Values at CCE, SE, NE and SWO for the Olsder’s game.

## 6 Conclusions

In this work we presented a new class of games, called Conjectural Stackelberg Games, aimed at bridging the gap between the use of conjectures and Stackelberg games. In particular we formalized how leaders may model opposite players’ behavior. To characterize the game we also propose an equilibrium definition, using the notion of consistency, first introduced in Conjectural Variations Games. As a first effort to study these new games, we compare the Conjectural Stackelberg Equilibrium with the standard Stackelberg Equilibrium for multiple settings. We also want to stress the fact that we kept our whole formalism *conjecture-agnostic*, meaning that we do not assume a specific form for the conjecture functions, which is in stark contrast with the existing literature. Finally, we develop a multi agent learning algorithm (COSTAL) that can be used by the players of Conjectural Games to update their conjectures and their strategy. We also provide a proof of convergence for

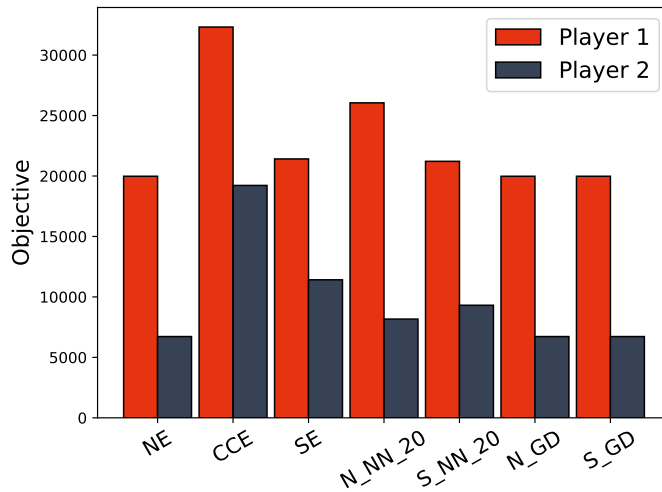


Figure 3: Final value of the objective function of both players of the Olsder’s Paradox.

said algorithm. We believe Conjectural Stackelberg Games could be a useful and interesting class of games to study further, both on the theoretical side, obtaining more results regarding their relation with Nash and Stackelberg equilibria; but also on the algorithmic side, using their structure to develop game-theory informed learning algorithm.

## References

- Tamer Başar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory, 2nd Edition*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, January 1998. ISBN 978-0-89871-429-6. doi: 10.1137/1.9781611971132. URL <https://epubs.siam.org/doi/book/10.1137/1.9781611971132>.
- Vivek S. Borkar. *Stochastic Approximation*, volume 48 of *Texts and Readings in Mathematics*. Hindustan Book Agency, Gurgaon, 2008. ISBN 978-81-85931-85-2 978-93-86279-38-5. doi: 10.1007/978-93-86279-38-5. URL <http://link.springer.com/10.1007/978-93-86279-38-5>.
- Timothy F. Bresnahan. Duopoly Models with Consistent Conjectures. *The American Economic Review*, 71(5):934–945, 1981. ISSN 0002-8282. URL <https://www.jstor.org/stable/1803475>. Publisher: American Economic Association.
- Sébastien Bubeck. Convex Optimization: Algorithms and Complexity, November 2015. URL <http://arxiv.org/abs/1405.4980>. arXiv:1405.4980 [math].
- Daniel J. Calderone, Benjamin J. Chasnov, Samuel A. Burden, and Lillian J. Ratliff. Consistent Conjectural Variations Equilibrium: Characterization & Stability for a Class of Continuous Games, June 2023. URL <http://arxiv.org/abs/2305.11866>. arXiv:2305.11866 [cs].
- Benjamin Chasnov, Tanner Fiez, and Lillian Ratliff. Opponent Anticipation via Conjectural Variations. 2019.
- Sheng Chen, Antonio J. Conejo, and Zhinong Wei. Conjectural-Variations Equilibria in Electricity, Natural-Gas, and Carbon-Emission Markets. *IEEE Transactions on Power Systems*, 36(5):4161–4171, September 2021. ISSN 1558-0679. doi: 10.1109/TPWRS.2021.3066459. URL <https://ieeexplore.ieee.org/abstract/document/9380720>. Conference Name: IEEE Transactions on Power Systems.

- 
- Stephan Dempe and Alain Zemkoho (eds.). *Bilevel Optimization: Advances and Next Challenges*, volume 161 of *Springer Optimization and Its Applications*. Springer International Publishing, Cham, 2020. ISBN 978-3-030-52118-9 978-3-030-52119-6. doi: 10.1007/978-3-030-52119-6. URL <http://link.springer.com/10.1007/978-3-030-52119-6>.
- Cristian A. Díaz, José Villar, Fco. Alberto Campos, and Javier Reneses. Electricity market equilibrium based on conjectural variations. *Electric Power Systems Research*, 80(12):1572–1579, December 2010. ISSN 0378-7796. doi: 10.1016/j.epsr.2010.07.012. URL <https://www.sciencedirect.com/science/article/pii/S0378779610001756>.
- Farzan Farnia and Asuman Ozdaglar. Do GANs always have Nash equilibria? In *Proceedings of the 37th International Conference on Machine Learning*, pp. 3029–3039. PMLR, November 2020. URL <https://proceedings.mlr.press/v119/farnia20a.html>. ISSN: 2640-3498.
- Joseph Farrell. Talk is Cheap. *The American Economic Review*, 85(2):186–190, 1995. ISSN 0002-8282. URL <https://www.jstor.org/stable/2117916>. Publisher: American Economic Association.
- Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit Learning Dynamics in Stackelberg Games: Equilibria Characterization, Convergence Analysis, and Empirical Study. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 3133–3144. PMLR, November 2020. URL <https://proceedings.mlr.press/v119/fiez20a.html>. ISSN: 2640-3498.
- Charles Figuières, Alain Jean-Marie, Nicolas Quérou, and Mabel Tidball. *Theory of Conjectural Variations*. WORLD SCIENTIFIC, February 2004. ISBN 978-981-238-736-3 978-981-279-488-8. doi: 10.1142/5453. URL <http://www.worldscientific.com/worldscibooks/10.1142/5453>.
- Marta Fochesato, Fredy Pokou, Helene Le Cadre, and John Lygeros. Noncooperative games with prospect theoretic preferences. *Preprint HAL* <https://hal.science/hal-04866629>, 2025.
- James W. Friedman and Claudio Mezzetti. Bounded rationality, dynamic oligopoly, and conjectural variations. *Journal of Economic Behavior & Organization*, 49(3):287–306, November 2002. ISSN 0167-2681. doi: 10.1016/S0167-2681(02)00005-7. URL <https://www.sciencedirect.com/science/article/pii/S0167268102000057>.
- Jiarui Gan, Edith Elkind, and Michael Wooldridge. Stackelberg Security Games with Multiple Uncoordinated Defenders. *Game Theory*, 2018.
- P. D. Grontas, G. Belgioioso, C. Cenedese, M. Fochesato, J. Lygeros, and F. Dörfler. Bighype: Best intervention in games via distributed hypergradient descent. *IEEE Transactions on Automatic Control*, 69(12):8338–8353, 2024.
- Noortje Groot, Bart De Schutter, and Hans Hellendoorn. Reverse Stackelberg games, Part I: Basic framework. In *2012 IEEE International Conference on Control Applications*, pp. 421–426, Dubrovnik, Croatia, October 2012. IEEE. ISBN 978-1-4673-4505-7 978-1-4673-4503-3 978-1-4673-4504-0. doi: 10.1109/CCA.2012.6402334. URL <http://ieeexplore.ieee.org/document/6402334/>.
- Mingyi Hong, Hoi-To Wai, Zhaoran Wang, and Zhuoran Yang. A two-timescale stochastic algorithm framework for bilevel optimization: Complexity analysis and application to actor-critic. *SIAM Journal on Optimization*, 33(1), 2023.
- Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291, 1979. ISSN 0012-9682. doi: 10.2307/1914185. URL <https://www.jstor.org/stable/1914185>. Publisher: [Wiley, Econometric Society].
- Sven Leyffer and Todd Munson. Solving multi-leader–common-follower games. *Optimization Methods and Software*, 25(4):601–623, August 2010. ISSN 1055-6788. doi: 10.1080/10556780903448052. URL <https://doi.org/10.1080/10556780903448052>. Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/10556780903448052>.



- 
- Youfei Liu, Y. X. Ni, Felix F. Wu, and Bin Cai. Existence and uniqueness of consistent conjectural variation equilibrium in electricity markets. *International Journal of Electrical Power & Energy Systems*, 29(6):455–461, July 2007. ISSN 0142-0615. doi: 10.1016/j.ijepes.2006.11.006. URL <https://www.sciencedirect.com/science/article/pii/S0142061506002043>.
- Francesco Morri, H el ene Le Cadre, Luce Brotcorne, and Pierre Gruet. Learning in Stackelberg Games with Application to Strategic Bidding in the Electricity Market. In *2024 20th International Conference on the European Energy Market (EEM)*, pp. 1–7, June 2024. doi: 10.1109/EEM60825.2024.10608880. URL <https://ieeexplore.ieee.org/document/10608880>. ISSN: 2165-4093.
- G.J. Olsder. Phenomena in Inverse Stackelberg Games, Part 1: Static Problems. *Journal of Optimization Theory and Applications*, 143 (3), 143(3), December 2009. doi: 10.1007/s10957-009-9573-9.
- Martin K. Perry. Oligopoly and Consistent Conjectural Variations. *The Bell Journal of Economics*, 13(1):197–205, 1982. ISSN 0361-915X. doi: 10.2307/3003440. URL <https://www.jstor.org/stable/3003440>. Publisher: [RAND Corporation, Wiley].
- Ariel Rubinstein and Asher Wolinsky. Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability. *Games and Economic Behavior*, 6(2):299–311, March 1994. ISSN 08998256. doi: 10.1006/game.1994.1016. URL <https://linkinghub.elsevier.com/retrieve/pii/S0899825684710165>.
- Hanif D. Sherali. A Multiple Leader Stackelberg Model and Analysis. *Operations Research*, April 1984. doi: 10.1287/opre.32.2.390. URL <https://pubsonline.informs.org/doi/abs/10.1287/opre.32.2.390>. Publisher: INFORMS.
- Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, February 1955. ISSN 0033-5533. doi: 10.2307/1884852. URL <https://doi.org/10.2307/1884852>.
- Steven H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview Press, a member of the Perseus Books Group, Boulder, CO, second edition edition, 2015. ISBN 978-0-8133-4910-7. OCLC: ocn842877119.
- Michael P. Wellman and Junling Hu. Conjectural Equilibrium in Multiagent Learning. *Machine Learning*, 33(2):179–200, November 1998. ISSN 1573-0565. doi: 10.1023/A:1007514623589. URL <https://doi.org/10.1023/A:1007514623589>.
- Yuli Zhang, Yuhua Xu, Yitao Xu, Yang Yang, Yunpeng Luo, Qihui Wu, and Xin Liu. A Multi-Leader One-Follower Stackelberg Game Approach for Cooperative Anti-Jamming: No Pains, No Gains. *IEEE Communications Letters*, 22(8):1680–1683, August 2018. ISSN 1558-2558. doi: 10.1109/LCOMM.2018.2843374. URL <https://ieeexplore.ieee.org/abstract/document/8371277>. Conference Name: IEEE Communications Letters.
- Liyuan Zheng, Tanner Fiez, Zane Alumbaugh, Benjamin Chasnov, and Lillian J. Ratliff. Stackelberg Actor-Critic: Game-Theoretic Reinforcement Learning Algorithms, September 2021. URL <http://arxiv.org/abs/2109.12286>. arXiv:2109.12286 [cs].

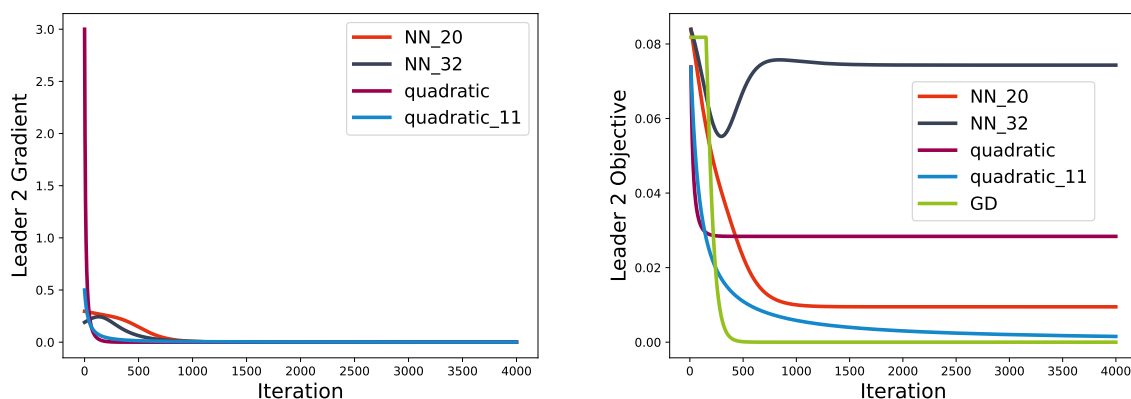
## Appendix

We report here further data regarding the simulation discussed in the main paper, focusing on the trajectories of the players during the learning of their strategy and the convergence speed, which we can show empirically looking at the gradient at each step of the algorithm.

A full script to run all the experiments can be found at this [link](#).

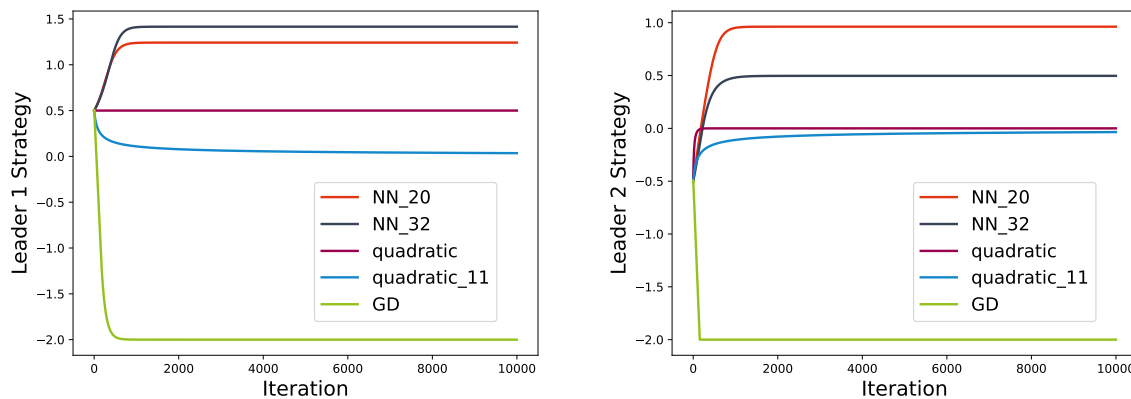


**Leader’s Dilemma** For this game we already reported the gradient evolution and objective evolution for leader 1 in the main paper, so we show here in Fig. 4a and Fig. 4b the same plots but for leader 2, and is clearly visible the gradient quickly goes to 0 and the the results obtained by the COSTAL algorithm are superior to that of the basic GD approach. Furthermore, in Fig. 5 we report the evolution of the strategies of the two leaders, where the different results between our algorithm and the gradient descent method are even more distinct. Finally, in Fig. 6 we report a comparison between the co-evolution of the two leaders and the Stackelberg equilibria of the game. In particular this game has a continuum of equilibria, between which the line  $x_1 = x_2$  is not stable and it also achieves a lower reward than  $x_2 = x_1 \pm 2\sqrt{\log |K|}$ . We can see that the GD method converge to the unstable point, while the equilibria reached by the conjectures are not on the Stackelberg Equilibria line, even though some are getting close. Once more this highlights the fact that the COSTAL algorithm converges to a stable equilibrium that is different from the standard notions of Nash/Stackelberg.



(a) Evolution of the gradient of leader 2. (b) Evolution of the objective function of leader 2.

Figure 4: Gradient and objective evolution of leader 2 in the Leader’s Dilemma game.



(a) Evolution of the strategy of leader 1. (b) Evolution of the strategy of leader 2.

Figure 5: Plots of the evolution of the strategies of the players of the Leader’s Dilemma game.

**Revisiting Olsder’s Paradox.** We report a similar analysis also for the second game. In Fig. 7 we can see once again how the strategies of the players converge very quickly, in this case also for the GD update. As before we show also the convergence of the gradients in Fig. 8, where we stress that in the two plots we are only showing the first 500 iterations, to highlight the fact that the gradient does go to 0 extremely quickly in this game. Furthermore, we also report in Fig. 9, the evolution of the

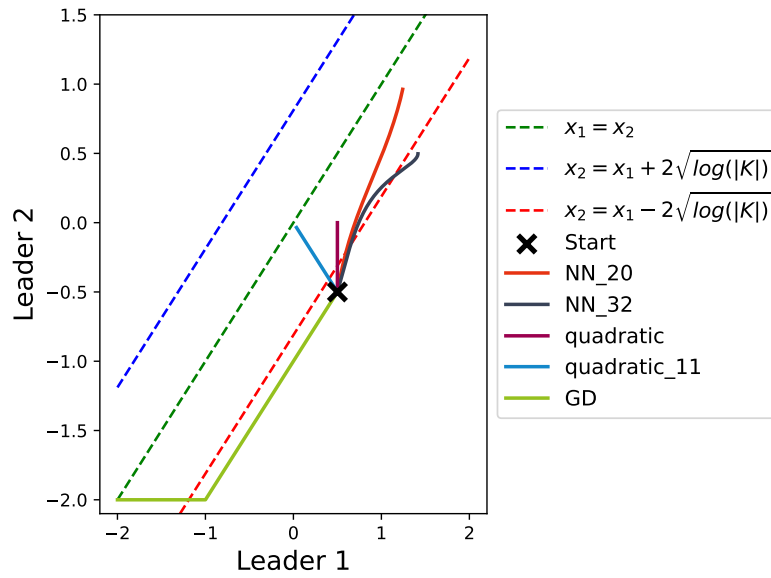
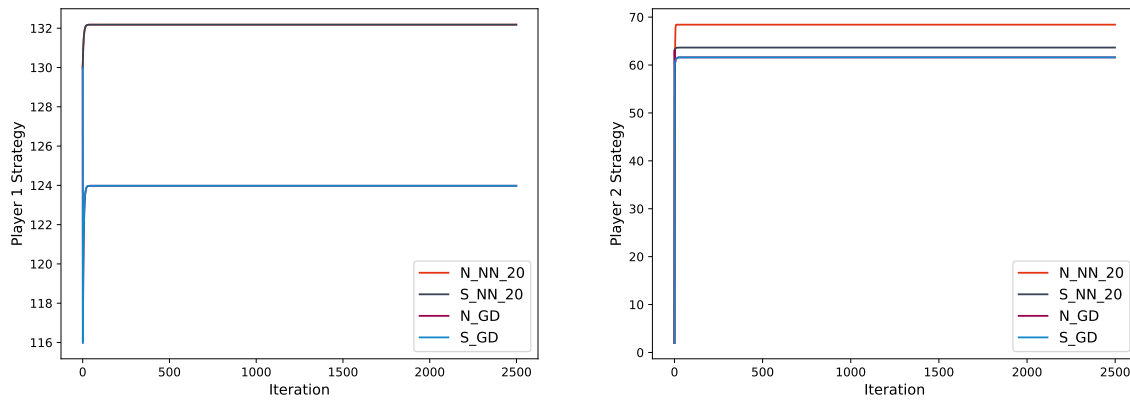


Figure 6: Comparison between the evolution of the strategies of the two leaders and the Stackelberg Equilibria of the Leader's Dilemma game

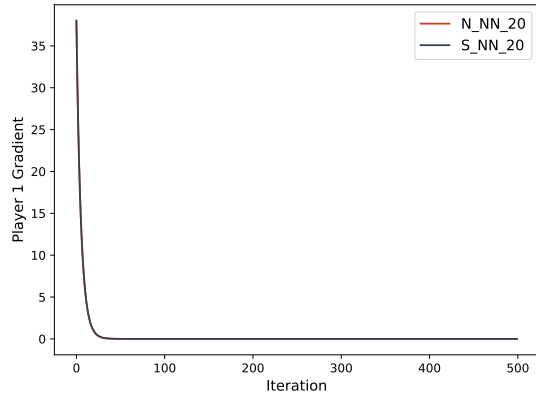
objective functions, comparing them with the value assumed in the different equilibria. Once again we can observe how quickly these functions converge to a stable point, without changing throughout the rest of the simulation.



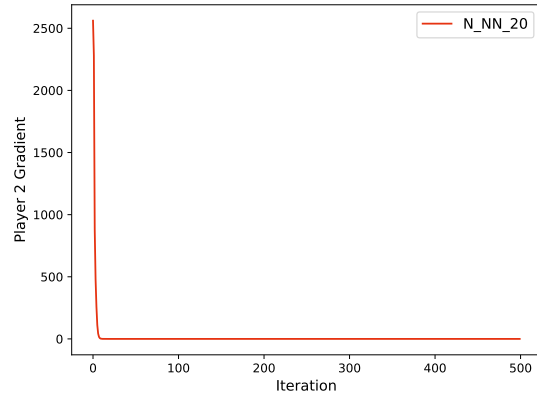
(a) Evolution of the strategy of player 1.

(b) Evolution of the strategy of player 2.

Figure 7: Strategies of the two players throughout the simulation.

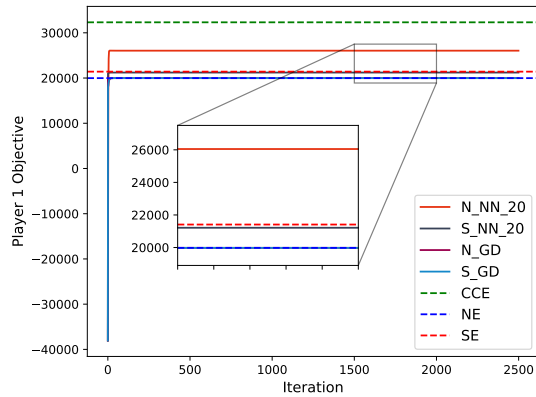


(a) Evolution of the gradient of player 1.

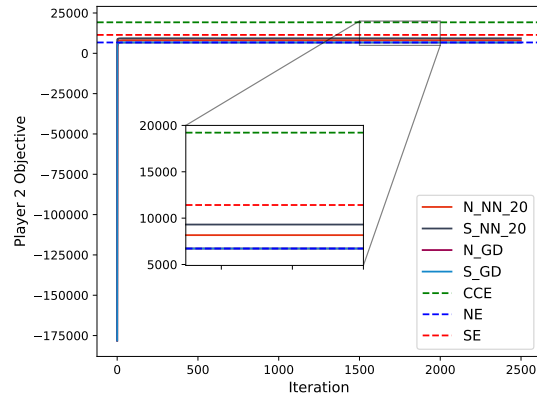


(b) Evolution of the gradient of player 2.

Figure 8: Gradients of the two players throughout the simulation.



(a) Evolution of the objective function of player 1.



(b) Evolution of the objective function of player 2.

Figure 9: Objective functions of the two players throughout the simulation.