Gradient Conjectures for Strategic Multi-Agent Learning

Benjamin Chasnov* BCHASNOV@UW.EDU
Department of Electrical and Computer Engineering
University of Washington

Tanner Fiez* FIEZT@UW.EDU
Department of Electrical and Computer Engineering
University of Washington

Lillian J. Ratliff RATLIFFL@UW.EDU
Department of Electrical and Computer Engineering
University of Washington

Abstract

We introduce a general framework for gradient-based learning that incorporates opponent behavior in continuous, general-sum games. In contrast to much of the work on learning in games, which primarily analyzes agents that myopically update a strategy under the belief opposing players repeat the last joint action, we study agents that model the dynamic behavior of opponents and optimize a surrogate cost. The surrogate cost functions embed *conjectures*, which anticipate the dynamic behavior of opponents. We show that agents with heterogeneous conjectures can result in a number of game-theoretic outcomes including Nash, Stackelberg, and general conjectural variations equilibrium. We review the suitability of each equilibrium concept for implicit and gradient conjectures and analyze the limiting outcomes of conjectural learning. Moreover, we demonstrate our framework generalizes a number of learning rules from recent years.

1. Introduction

The central topic of study in game theory is equilibrium analysis [2]. The famed Nash equilibrium solution concept attracts the bulk of attention in non-cooperative game theory. A Nash equilibrium is a solution that characterizes the outcome of rational agents optimizing individual cost functions that depend on the actions of competitors. At equilibrium, no player can benefit by unilaterally deviating from the joint strategy of the players. It is commonly adopted in games where each player selects an action simultaneously.

However, in a number of games, players select actions sequentially. A well-known sequential game is a Stackelberg game. In the simplest formulation, there is a leader and a follower who select actions respecting the hierarchical decision-making structure. In particular, the leader selects an action with knowledge that the follower will play the rational best response strategy to that action. The Stackelberg equilibrium is an outcome where the follower acts rationally given the leader strategy and the leader behaves optimally considering the reaction curve of the follower.

The study of equilibrium solution concepts is primarily motivated by the desire to efficiently compute equilibria of multiple competing costs and to characterize the process that guides players toward equilibria. The field of learning in games seeks to provide solutions to each objective by seeking equilibria with iterative learning algorithms and analyzing the convergence of learning algorithms reflecting the underlying game dynamics [13].

In simultaneous play games, gradient play defines the process in which each player descends a cost function they seek to optimize via an individual gradient update. The gradient play learning dynamics have been studied extensively since they are a simple analogue to gradient descent from optimization. As a result, the convergence characteristics are reasonably well-known for a broad class of games. Unfortunately, a number of results for gradient play show that the naive extension of classical optimization techniques to learning in games often fails to reach an equilibrium [23, 24]. In some sense, the undesirable behavior of the combined dynamics is unsurprising since each player is optimizing a cost function as if the joint strategy of the competitors is to remain stationary. Motivated by the shortcomings of the gradient play dynamics, learning rules have been proposed to mitigate the problems that come from implicitly neglecting the role of competing players [15, 25, 1, 22]. While useful for the purpose of computing equilibria, the emerging methods primarily arise from adjusting learning dynamics in a manner that cannot be explained by game-theoretic reasoning and may even directly conflict with competitive game play.

The convergence behavior of learning dynamics in hierarchical games has not been analyzed as broadly as in simultaneous play games. However, there is a surge of recent work on this topic [8, 6, 17, 20], which can primarily be attributed to a number of important applications naturally exhibiting a sequential order of play. As in simultaneous play games, gradient play dynamics do not always provide a suitable solution and can avoid equilibria even in zero-sum games. Conversely, dynamics reflecting the underlying decision-making structure in Stackelberg games come with strong convergence guarantees to Stackelberg equilibria [8]. The promising outcome can intuitively be attributed to the structure in the game from the order of play that is imposed. Explicitly, the ability of the leader to select an action while anticipating the follower response stabilizes the learning process in the game. This observation leads to the following questions:

> *What is the impact of players anticipating the behavior of competitors on the stability of learning dynamics in a broad class of games and is there a suitable solution concept to reconcile dynamic agent behavior in a rational manner?*

We study the preceding questions in a principled manner. The motivation for doing so stems from the lack of existing theoretically grounded work on this topic and the opportunity to adopt a relatively unexplored perspective on learning in games that can provide novel solutions to issues blocking stable, desirable convergence guarantees in a broad class of games. Importantly, gradient play dynamics are widely analyzed, yet the dynamics do not yield stable convergence guarantees to equilibria. The underlying problem with the dynamics is that each agent is updating a strategy without being cognizant of how competing agents may react. Moreover, it is implicitly assumed that each agent has a symmetric learning rule. Effectively, the learning process of each player disregards the highly non-stationary system embodied by opponent behavior. A number of dynamics analyzed in the literature beyond gradient play bear similar shortfalls. On the other hand, the analysis of learning dynamics in Stackelberg games demonstrates that the ability to model and predict the behavior of a competitor can stabilize the learning process and result in improved convergence guarantees [8]. Collectively, it is clear that no general framework for learning in games has fully harnessed the tools game theory provides to model strategic interactions in a sufficiently complex, yet elegant manner.

In this paper, we present strategic *implicit conjectures*, *gradient conjectures*, and *fast conjectures* learning dynamics. The dynamics are such that each agent has a conjecture about the anticipated behavior of the competitors and seeks to minimize a surrogate cost function that depends on the predicted behavior of opponents via gradient updates. In simple terms, we take a view that agents can look ahead one step and act as if the future behavior of competitors is known; consequently, agents explicitly consider the impacts of actions on competing agents and the role of the opponent learning process on the cost function being optimized. We restrict the class of conjectured opponent behavior to be gradient-based learning rules. The primary focus of this work is to analyze the convergence behavior of the general framework we introduce for gradient-based learning incorporating conjectures of opponent behavior in continuous, general-sum games.

In our analysis, we consider several solution concepts including Nash, Stackelberg, and general conjectural variations equilibrium. This work introduces the study of general conjectural variations equilibrium to the modern learning in games literature. In a general conjectural variations equilibrium, each agent plays the action minimizing the respective surrogate cost functions incorporating the conjectured play of competitors. Indeed, it is a natural solution concept to study when analyzing the gradient-based conjectured learning dynamics. We draw connections between each solution concept and review the suitability of them contingent on the conjectured behavior of opponents.

The conjectures learning framework in this paper is extremely general. As an artifact, we are able to analyze the role the conjectured opponent behavior has on the convergence of the dynamics. It is also interpretable, in the sense that we are able to look at the combined dynamics and tease out the role each component of the update plays in determining convergence. This is useful for obtaining a deep understanding of learning dynamics and for synthesizing novel learning dynamics for any given objective. Moreover, we are able to derive several prominent learning rules from our general framework. This demonstrates that our paper generalizes a number of works on learning in games.

Contributions. We formulate a learning framework for incorporating conjectures of opponent behavior in general-sum, continuous games. To study the dynamic learning rules, we define a differential general conjectural variations equilibrium concept that lends naturally to gradient-based learning. The primary contributions of this paper are summarized as follows:

- We provide asymptotic convergence guarantees for gradient conjectures.
- We give intuitive explanations for the role components of the gradient conjectures play in the convergence behavior.
- We show several dynamics from prominent works on learning in games can be derived from the general framework we present.

Related Work. The conjectural variations model and equilibrium concepts have a long history in a non-cooperative dynamic games and decision-theory in economics [2, 9]. The equilibrium concept was introduced in the 1920's [3], but since then the theory of conjectural variations has endured several controversies. In early descriptions, conjectures were considered exogenous; consequently, the outcome of a game could be determined by the choice of conjecture—providing an irrefutable theory of behavior. The concept was refined to require consistent conjectures, meaning that the conjectured opponent behavior should correspond to a best response [4]. However, in static games the requirement of consistency limits where the solution concept can be applied. Fortunately, the theory of conjectural variations is exempt from the primary objections when viewed as a dynamic concept. Indeed, it is now common theory in dynamic games and bounded rationality models [7, 16, 12].

In the machine learning literature, significant attention has been given to learning in games in the last few years. This is primarily owed to the applications of Generative Adversarial Networks (GANs) [14] and multi-agent learning. The gradient play learning dynamics have been analyzed extensively and the shortcomings are well documented [23, 24, 5]. A number of works develop learning rules to seek Nash equilibria in continuous games in sophisticated ways including lookahead [30], consensus optimization [25], symplectic gradient adjustment [1], and symplectic surgery [22]. Each learning rule either attempts to reach a stable point quickly by neutralizing rotational dynamics or only converge to an equilibria by avoiding non-equilibria attractors. Unfortunately, it is not clear how each learning rule derives from game-theoretic reasoning in comparison to mathematical reasoning. In our work, we clear this up by deriving each learning rule from our general framework.

There are a select number of learning dynamics that model opponent behavior including learning with opponent learning awareness (LOLA) [11], stable opponent shaping (SOS) [19], and learning in Stackelberg games [8]. The LOLA update was derived with the goal of designing an agent that could predict the future action of a naive gradient play learning agent and use the information to outwit the opponent. However, the paper only presented empirical results and came with no formal convergence guarantees. In the process of analyzing learning dynamics in Stackelberg games [8], it was discovered that the LOLA dynamics with a Newton step-size are equivalent to Stackelberg dynamics. The learning dynamics in Stackelberg games are guaranteed to converge to Stackelberg equilibria in zero-sum games, which contains the set of Nash equilibria. The SOS learning rule is a carefully designed set of dynamics building on the LOLA update. It is shown to converge to Nash equilibrium under certain conditions.

Organization. In Section 2, we present multi-agent learning with conjectures, including implicit conjectures and gradient conjectures learning dynamics. To formally study each learning rule, we define the differential general conjectural variations (DGCVE) equilibrium concept and give asymptotic convergence guarantees to a stable DGCVE under a suitable initialization and learning rate. In Section A, we present links between the limiting outcomes of the conjectures updates and learning dynamics in simultaneous and hierarchical play games. We present a unified framework for heterogeneous multi-agent learning in Section 3. We show numerical examples and conclude in Sections 4 and 5, respectively. The appendix of this paper includes a primer on dynamical systems theory definitions and learning dynamics in simultaneous and hierarchical play games.

2. Multi-agent Learning with Conjectures

Towards introducing the conjectures framework, we define the game-theoretic abstraction which describes players objectives. A continuous n-player general-sum game is a collection of costs $(f_1, \ldots, f_n)$ defined on $X = X_1 \times \cdots \times X_n$ where $f_i \in C^r(X, \mathbb{R})$ with $r \geq 2$ is player i's cost function, $X_i = \mathbb{R}^{d_i}$ is their action space, and $d = \sum_{i=1}^n d_i$. Each player $i \in \mathcal{I}$ aims to select an action $x_i \in X_i$ that minimizes their cost $f_i(x_i, x_{-i})$ given the actions of all other players, namely $x_{-i} \in X_{-i}$.

The learning rules we study are such that agents are adaptively updating their actions by repeatedly playing the game $(f_1, \ldots, f_n)$ and using the feedback they receive to make adjustments. In the conjectures framework, we consider learning rules employed by agents seeking a strategy that minimizes their objective given its dependence on other players' actions, initialization, and beliefs about how other players are updating. In particular, agents form *conjectures* about how their choice of action affects other agents' responses. The framework provides principled learning rules for agents to minimize their cost while modeling how their opponents might react.

The form of learning rule studied in this paper is given by

$$x_{k+1,i} = x_{k,i} - \gamma_{k,i} g_i(x_{k,i}, \xi_i(x_{k,i}, x_{k,-i})),\ i \in \{1, \ldots, n\}$$

where g_i is derived from gradient information of the player's cost function and ξ_i is player i's conjecture about the rest of the players given by

$$\xi_i(x_k) = (\xi_i^1(x_k), \ldots, \xi_i^{i-1}(x_k), \xi_i^{i+1}(x_k), \ldots, \xi_i^n(x_k))$$

in which $\xi_i^j(x_k)$ is player i's conjecture about player j. To provide some intuition, consider the following familiar learning framework. Simultaneous gradient play is defined by taking $g_i \equiv D_i f_i$ (or $g_i \equiv \mathbb{E}[\widehat{D_i f_i}]$ in the stochastic case) and $\xi_i(x_k) = x_k$. In words, this would mean player i conjectures that players $-i$ select a *static best response policy* in the sense that they conjecture opponents will play the same strategy as in the previous round.

While there are numerous classes of conjectures, we focus our study on conjectures motivated by multi-agent gradient-based learning defined as follows:

1. Implicit conjectures: $\xi_i(x)$ defined implicitly by $D_{-i}f_{-i}(x_i, x_{-i}) = 0$.
2. Gradient-based conjectures: $\xi_i(x)$ defined by the gradient update $x_{-i} - \widetilde{\gamma}_i D_{-i}f_{-i}(x_i, x_{-i})$.

The class of gradient-based conjecture learning rules we study can be encapsulated by the myopic update rule

$$x_{k+1,i} = x_{k,i} - \gamma_i D_i f_i(x_{k,i}, \xi_i(x_k))$$

for each player $i \in \{1, \ldots, n\}$. Using the multivariable chain rule, given the conjecture, the derivative[1] of the cost for player i with respect to x_i is

$$D_i f_i(x) + D_i \xi_i(x)^\top D_{-i} f_i(x).$$

The equilibrium notion we consider is known as a *general conjectural variations equilibrium* [9].

Definition 1 (General Conjectural Varations Equilibrium)**.** *For a game $\mathcal{G} = (f_1, \ldots, f_n)$ with each $f_i \in C^r(X, \mathbb{R})$ and conjectures $\xi_i : X \to X_{-i}$, a point $x^* \in X$ constitutes a* general conjectural variations equilibrium *if*

$$D_i f_i(x^*) + D_i \xi_i(x^*)^\top D_{-i} f_i(x^*) = 0, \;\; i \in \{1, \ldots, n\}.$$

The general conjectural variations equilibrium concept generalizes the conjectural variations equilibrium, which is defined equivalently but restricted to a limited class of conjectures of the form $\xi_i : X_i \to X_{-i}$ where the conjecture of each player can only depend on its own action. Further, unlike a Nash equilibrium, which is defined with respect to the cost evaluation of each player at a candidate point, the conjectural variations equilibrium is defined with respect to first order conditions on the cost function of each player. Consequently, players may have an incentive to deviate at a general conjectural variations equilibrium. Towards mitigating the limitations of the solution concept, we define the notion of a differential general conjectural variations equilibrium. In particular, a differential general conjectural variations equilibrium includes a second order condition so that locally no player has a direction in which they can incrementally adjust their action and benefit.

Definition 2 (Differential General Conjectural Variations Equilibrium)**.** *For a game $\mathcal{G} = (f_1, \ldots, f_n)$ with each $f_i \in C^r(X, \mathbb{R})$ and conjectures $\xi_i : X \to X_{-i}$, a general conjectural variations equilibrium $x^* \in X$ is a differential general conjectural variations equilibrium if, for each $i \in \{1, \ldots, n\}$, $D_i^2 f_i(x_i, \xi_i(x^*)) > 0$.*

For simplicity and to reduce notational overhead that needs to be maintained by the reader, let us consider two player settings in the remainder. The extension to n–player games is straightforward.

For the purpose of analyzing the convergence properties of the gradient-based conjecture learning rules, we introduce the limiting continuous time dynamics $\dot{x} = -g(x)$ where $g(x) = (D_i f_i(x_i, \xi_i(x)))_{i=1}^n$. With respect to these dynamics, we define the notion of a stable general conjectural variations equilibrium.

Definition 3 (Stable General Conjectural Variations Equilibrium)**.** *A general conjectural variations equilibrium x^* is stable when the Jacobian of $-g$ has its spectrum in the open left-half complex plane, meaning* $\mathrm{spec}(-Dg(x^*)) \subset \mathbb{C}_-^\circ$.

1. We use the following convention for the dimensions of derivatives: $D_i f_i \in \mathbb{R}^{d_i \times 1}$, $D_i^2 f_i \in \mathbb{R}^{d_i \times d_i}$, $D_{ij} f_i = D_j(D_i f_i) \in \mathbb{R}^{d_i \times d_j}$, and $D_i \xi_i = -(D_j^2 f_j)^{-1} D_{ji} f_j \in \mathbb{R}^{d_j \times d_i}$.

We note that unlike single player optimization, Dg is in general not symmetric. Dependent on the choice of the learning rates γ_i and the conjectured learning rates $\widetilde{\gamma}_i$, we provide convergence guarantees for the gradient-based conjectured learning rules to stable general conjectural variations equilibrium.

We also remark that *consistency* of conjectures can call into question the realistic nature of the general conjectural variations equilibrium—that is, if a player observes the other player not playing consistently with their conjecture, then is it reasonable that a player should maintain this conjecture? To address this, we make a few comments. First, many multi-agent learning paradigms are already utilizing conjectures of some form [1, 22, 30, 10, 19]; we formalize this in Section 3 by showing rules that can be derived from a conjectures-based perspective. Second, we introduce a notion of stability of general conjectural variations equilibrium with respect to the dynamical system that the *myopic* players are utilizing. This provides some notion of robustness to the fallacy of conjecture. The players are assumed to be myopic in many approaches to multi-agent learning in the sense that they blindly follow the update rule provided to them at the beginning of the learning process. We adopt that viewpoint in this paper. That having been said, a notion of consistent conjectural variations equilibrium exists in the conjectural variations literature [21, 4, 12, 9]. However, by and large, the multi-agent learning paradigms being studied lately will not lead to such outcomes. Nevertheless, an interesting avenue for future research in multi-agent learning is the consideration of learning and adapting conjectures.

2.1 Implicitly Defined Conjectures

For simplicity, we derive the update for two players, however the extension to n players is straightforward. A natural conjecture to consider is one in which each player conjectures that the other player is selecting a *best response*. In this conjecture, each player conjectures that the opponent selects an action by minimizing its own cost while holding the actions of the remaining players to be constant. For example, the conjecture formed by player 1 about how player 2 reacts to action $x_1 \in X_1$ is

$$\xi_1(x_1) \in \arg\min_w f_2(x_1, w)$$

and, vice versa, player 2 conjectures the reaction of player 1 to its action $x_2 \in X_2$ is

$$\xi_2(x_2) \in \arg\min_z f_1(z, x_2).$$

Players may not know the opponents cost function. Moreover, in general best responses define sets. Given the gradient-based learning framework we consider, it is reasonable to consider optimality conditions of the above best response maps. Indeed, the implicit conjectures framework is derived from consideration of conjectures defined implicitly via first-order optimality conditions. For example, at iteration k, $\xi_i : X \to X_{-i}$ is defined implicitly by

$$D_{-i} f_{-i}(x_k) = 0.$$

Hence, the component of player i's update derived from the dependence on the conjecture ξ_i can be constructed via the implicit function theorem [18]. Indeed,

$$D_i \xi_i(x_1, x_2) = -(D_j^2 f_j(x))^{-1} D_{ji} f_j(x)$$

so that the gradient-based conjecture learning rule given implicitly defined conjectures is

$$x_{k+1,i} = x_{k,i} - \gamma_i (D_i f_i(x_k) - (D_{ji} f_j(x_k))^\top (D_j^2 f_j(x_k))^{-1} D_j f_i(x_k))$$

for each player $i = 1, 2$.

We can write the update in a more concise form by defining the standard simultaneous gradient-play objects $g(x) = (D_1 f_1(x), D_2 f_2(x))$ and its Jacobian

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}$$

Using this notation, we can define the following analogous object to $g(x)$ in the implicit conjectures case:

$$\begin{aligned} g_{\mathtt{i}}(x) &= g(x) - \begin{bmatrix} 0 & D_{12} f_2(x) \\ D_{21} f_1(x) & 0 \end{bmatrix} \begin{bmatrix} (D_1^2 f_1(x))^{-1} & 0 \\ 0 & (D_2^2 f_2(x))^{-1} \end{bmatrix} \begin{bmatrix} D_1 f_2(x) \\ D_2 f_1(x) \end{bmatrix} \\ &= g(x) - J_{\mathrm{o}}^\top(x) J_{\mathrm{d}}^{-1}(x) \begin{bmatrix} D_1 f_2(x) \\ D_2 f_1(x) \end{bmatrix}, \end{aligned}$$

where J_{d} and J_{o} are the block diagonal and block off-diagonal of J such that $J \equiv J_{\mathrm{o}} + J_{\mathrm{d}}$. Equivalently, the implicit update can be written using

$$g_{\mathtt{i}}(x) = g(x) - \mathrm{diag}(J_{\mathrm{o}}^\top(x) J_{\mathrm{d}}^{-1}(x) \nabla f(x))$$

where

$$\nabla f(x) = \begin{bmatrix} D_1 f_1(x) & D_1 f_2(x) \\ D_2 f_1(x) & D_2 f_2(x) \end{bmatrix}$$

with $f \equiv (f_1, f_2)$, a vector-valued function defined by players' costs. In the special case of zero-sum games, the form of update simplifies to

$$g_{\mathtt{i}}(x) = (I + J_{\mathrm{o}}^\top(x) J_{\mathrm{d}}^{-1}(x)) g(x)$$

since $(D_1 f_2(x), D_2 f_1(x)) = -g(x)$. Hence, the gradient-based conjecture learning rule with implicit conjectures can be written as

$$x_{k+1} = x_k - \Gamma_k g_{\mathtt{i}}(x_k)$$

where $\Gamma_k = \mathrm{diag}(\gamma_{k,1}, \gamma_{k,2})$. The update rule for n–player games is equivalent using the analogous game and operator definitions.

Consider that agents have oracle access to their individual gradients given their conjectures so the components $g_{\mathtt{i},i}$ of $g_{\mathtt{i}}$ corresponding to each player are known. Let $J_{\mathtt{i}}$ denote the Jacobian of $g_{\mathtt{i}}$ and $\gamma_{k,i} = \gamma_i$ to indicate players have constant learning rates.

Proposition 1. *Suppose agents use learning rates such that $\rho(I - \Gamma J_{\mathtt{i}}(x)) < 1$ where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_n)$, then if x_0 is initialized in the region of attraction of a stable general conjectural variations equilibrium x^*, then $x_k \to x^*$ exponentially as $k \to \infty$.*

If players have unbiased estimators of their updates, as opposed to oracle access in the deterministic setting, the update rule can be viewed as a stochastic approximation update. Under the assumptions of zero-mean, finite variance noise and uniform square summable but not summable step-size sequences γ_k for all players, the standard stochastic approximation convergence results apply. Similarly, if the learning rates are non-uniform but satisfy an 'ordered' relationship of the form $\gamma_{i,k} = o(\gamma_{j,k})$ for all $i < j$, $i, j \in \{1, \ldots, n\}$ such that $j = i+1$, then a multi-timescale approach can be applied to obtain convergence guarantees and even concentration bounds.

The following zero-sum, quadratic game example provides some intuition for how the implicit conjectures change the Jacobian in the neighborhood of a critical point of the gradient play dynamics.

Example 1. *Consider the two-player game defined by*

$$f_i(x_1, x_2) = \frac{1}{2}x_i^\top A_i x_i + x_i^\top B_{ij} x_j + \frac{1}{2}x_j^\top C_{ij} x_j,\ \ i, j = 1, 2,\ \ j \neq i.$$

Then, the Jacobian of the implicit conjecture dynamics is given by

$$J_{\mathtt{i}}(x_1, x_2) = \begin{bmatrix} A_1 - B_{21}^\top A_2^{-1} B_{12}^\top & B_{12} - B_{21}^\top A_2^{-1} C_{12} \\ B_{21} - B_{12}^\top A_1^{-1} C_{21} & A_2 - B_{12}^\top A_1^{-1} B_{21}^\top \end{bmatrix}.$$

In the special case of zero-sum games $\mathcal{G} = (f, -f)$ where

$$f(x, y) = \frac{1}{2}x^\top A_1 x + x^\top B y + \frac{1}{2}y^\top A_2 y,$$

$A_1 > 0$ and $-A_2 > 0$, the Jacobian has the block diagonal form given by

$$J_{\mathtt{i}}(x, y) = \begin{bmatrix} A_1 - B A_2^{-1} B^\top & 0 \\ 0 & -A_2 + B^\top A_1^{-1} B \end{bmatrix}.$$

With implicit conjectures, the block off-diagonal terms are zero and the diagonals are the Schur complements of the original Jacobian. Under this lens, we can interpret this result as agent i choosing $(D_j^2 f_j(x))^{-1}$ as their conjectured learning rate.

2.2 Gradient-Based Conjectures

The implicitly defined conjecture assumes that players conjecture that their opponents player a local best response. A natural relaxation of this conjecture is the class of conjectures defined by individual gradient-based updates—that is, players assume their opponents are simply playing simultaneous gradient play with response map

$$x_{k+1,-i} = x_{k,-i} - \widetilde{\Gamma}_{k,i} D_{-i} f_{-i}(x_k)$$

where $\widetilde{\Gamma}_{k,i} = \mathrm{diag}(\gamma_{k,1}^i, \ldots, \gamma_{k,i-1}^i, \gamma_{k,i+1}^i, \ldots, \gamma_{k,n}^i)$ with $\gamma_{k,j}^i$ being the learning rate player i conjectures of player j. In a two-player game with constant conjectured learning rates $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$, the conjectures are defined as

$$\xi_1(x_1, x_2) = x_2 - \widetilde{\gamma}_1 D_2 f_2(x_1, x_2) \quad \text{and} \quad \xi_2(x_1, x_2) = x_1 - \widetilde{\gamma}_2 D_1 f_1(x_1, x_2).$$

Since $D\xi_i(x) = -\widetilde{\gamma}_i D_{ji} f_j(x)$, the update rules become

$$x_{k+1,i} = x_{k,i} - \gamma_{k,i}(D_i f_i(x_k) - \widetilde{\gamma}_i D_{ij} f_j(x_k) D_j f_i(x_k)).$$

As in the preceding subsection, we can similarly write the gradient-based conjectures in a more compact form given by

$$\begin{aligned} g_{\mathtt{g}}(x) &= g(x) - \begin{bmatrix} 0 & \widetilde{\gamma}_1 D_{12} f_2(x) \\ \widetilde{\gamma}_2 D_{21} f_1(x) & 0 \end{bmatrix} \begin{bmatrix} D_1 f_2(x) \\ D_2 f_1(x) \end{bmatrix} \\ &= g(x) - \widetilde{\Gamma} J_{\mathrm{o}}^\top \begin{bmatrix} D_1 f_2(x) \\ D_2 f_1(x) \end{bmatrix} \end{aligned}$$

or, equivalently,

$$g_{\mathtt{g}}(x) = g(x) - \mathrm{diag}(\widetilde{\Gamma} J_{\mathrm{o}}^\top(x) \nabla f(x)),$$

where $g(x)$ is the original game form for gradient play and $\widetilde{\Gamma}$ is a diagonal matrix of the conjectured learning rates. For zero-sum games, the update reduces to $(I + \widetilde{\Gamma} J_{\mathrm{o}}^\top)g(x)$, since the vector $(D_1 f_2(x), D_2 f_1(x)) = -g(x)$. The compact form does not extend from two-player games to n–player games unless the conjectured learning rates are consistent between players in which event there would be a shared conjectured learning rate $\widetilde{\gamma}$ or conjectured learning rate matrix $\widetilde{\Gamma}_k$ for all players.

In two-player games and n–player games with consistent conjectured learning rates, the combined update rule of gradient-based conjectures is

$$x_{k+1} = x_k - \Gamma_k g_{\mathtt{g}}(x_k).$$

Consider that agents have oracle access to their individual gradients given their conjectures so the components $g_{\mathtt{g},i}$ of $g_{\mathtt{g}}$ corresponding to each player are known. Define $J_{\mathtt{g}}$ to be the Jacobian of $g_{\mathtt{g}}$. Furthermore, let the learning rates be constant and given by $\gamma_{k,i} = \gamma_i$ for each player.

Proposition 2. *Suppose agents use learning rates such that $\rho(I - \Gamma J_{\mathtt{g}}(x)) < 1$ where $\Gamma = (\gamma_1, \ldots, \gamma_n)$. If x_0 is initialized in the region of attraction of a stable GCVE x^*, then $x_k \to x^*$ exponentially as $k \to \infty$.*

As discussed for implicit conjectures learning, we can also provide convergence guarantees in the stochastic learning regime in which players have unbiased estimators of their individual conjectured gradients instead of oracle access.

Quadratic games can provide intuition for the behavior of learning dynamics in a neighborhood of a critical point.

Example 2. *Consider the two-player game defined by*

$$f_i(x_1, x_2) = \frac{1}{2} x_i^\top A_i x_i + x_i^\top B_{ij} x_j + \frac{1}{2} x_j^\top C_{ij} x_j, \ \ i, j = 1, 2, \ \ j \neq i.$$

Then, the Jacobian of the gradient-based conjectures dynamics is given by

$$J_{\mathtt{g}}(x_1, x_2) = \begin{bmatrix} A_1 - \widetilde{\gamma}_1 B_{21}^\top B_{12}^\top & B_{12} - \widetilde{\gamma}_1 B_{21}^\top C_{12} \\ B_{21} - \widetilde{\gamma}_2 B_{12}^\top C_{21} & A_2 - \widetilde{\gamma}_2 B_{12}^\top B_{21}^\top \end{bmatrix}.$$

In the special case of zero-sum games $\mathcal{G} = (f, -f)$ where

$$f(x, y) = \frac{1}{2} x^\top A_1 x + x^\top B y + \frac{1}{2} y^\top A_2 y,$$

the Jacobian has the block diagonal form given by

$$J_{\mathtt{g}}(x, y) = \begin{bmatrix} A_1 + \widetilde{\gamma}_1 B B^\top & B(I - \widetilde{\gamma}_1 A_2) \\ -B^\top (I - \widetilde{\gamma}_2 A_1) & -A_2 + \widetilde{\gamma}_2 B^\top B \end{bmatrix}.$$

Compared to the Jacobian of simultaneous gradient play, the diagonal terms are modified by positive definite terms and the off diagonal terms are modified by $-\widetilde{\gamma}_1 B A_2$ and $\widetilde{\gamma}_2 B^\top A_1$, respectively. Intuitively, the former serves to increase the real part of the eigenvalues while with a careful choice of the conjectured learning rates, the latter can be used to control the imaginary component. In particular, if the imaginary component decreases and the real part increases, then the effect is to make the dynamics exhibit behavior more like a potential flow.

In Section 3.2, we investigate exactly how the conjectured learning rates impact the dynamics relative to simultaneous gradient play, which can be interpreted as players conjecturing opponents perform a static best response strategy.

3. A Unified Framework for Multi-agent Learning with Conjectures

Several existing multi-agent learning algorithms can be derived from the conjectures framework. Towards this end, we consider Taylor expansions of a cost function for a player in order to explain and provide intuition for how players envision the reactions of opponents impacting the function they are optimizing. For simplicity of notation, we often focus on two-player games.

Consider a two-player game (f_1, f_2) with conjectures (ξ_1, ξ_2). Let us consider conjectures of the form

$$\xi_i(x_i, x_j) = x_j + v$$

for some vector v and each player $i = 1, 2$ such that $i \neq j$. Now, consider the k-th iteration of the learning process and suppose that player 1 wants to compute a best response by selecting a minimizer in the set

$$\arg\min_{x_1} f_1(x_1, x_2 + v_k),$$

given that player 1 conjectures that player 2 is using the simple update rule $x_2 + v_k$ and an oracle provides the update direction v_k. This problem in general may be complicated to solve, so instead player 1 may take a gradient step in a direction of steepest descent of the cost function. Toward that end, they could first approximate their function with a local conjectured cost. That is, given $x_{2,k}$, the cost function f_1 can be approximated locally by its Taylor expansion

$$f_1(x_1, x_{k,2} + v_k) \simeq f_1(x_1, x_{k,2}) + D_2 f_1(x_1, x_{k,2}) v_k + O(v_k^2).$$

Dropping higher order terms, let us define the left-hand side as $\widetilde{f}_{k,1}(x_1)$. Now, player 1 faces the optimization problem

$$\arg\min_{x_1} \widetilde{f}_{k,1}(x_1) = \arg\min_{x_1} \{ f_1(x_1, x_{k,2}) + D_2 f_1(x_1, x_{k,2}) v_k \}$$

Looking at the Taylor expansion of $\widetilde{f}_{k,1}(x_1)$ we get some intuition for how the direction of steepest descent is computed. Indeed,

$$\widetilde{f}_{k,1}(x_1 + w_k) \simeq f_1(x_1, x_{k,2}) + D_2 f_1(x_1, x_{k,2}) v_k + (D_1 f_1(x_1, x_{2,k}) + D_{12} f_1(x_1, x_{2,k}) v_k) w_k + O(w_k^2)$$

so that the gradient-based learning rule is of the form

$$x_{k+1,1} = x_{k,1} - \gamma_1 (D_1 f_1(x_{k,1}, x_{k,1}) + D_{12} f_1(x_{k,1}, x_{k,2}) v_{k,1}) \quad (1)$$

$$x_{k+1,2} = x_{k,2} - \gamma_2 (D_2 f_2(x_{k,1}, x_{k,1}) + D_{21} f_2(x_{k,1}, x_{k,2}) v_{k,2}), \quad (2)$$

or equivalently

$$x_{k+1} = x_k - \Gamma g(x_k) - \Gamma J_{\mathrm{o}}(x_k) v_k,$$

where $v_{k,i}$ is player i's conjectured direction for player j, provided by an oracle[2]. We call the dynamics in (1)–(2) faster conjecture learning. The update form gives rise to a vast number of learning schemes depending on how $v_{k,i}$ and γ_i are defined. We discuss several choices in the sections that follow. Essentially, update rules of the form overhead give players the ability to adjust for the effect an opponent has on the descent direction. That is, if player j is updating in direction v, then that can have the effect of perturbing player i away from its optimal steepest descent direction. The ability to adjust for perturbations from opposing agents leads to an update scheme that converges much faster to attractors; however, the set of attractors is not equivalent to the set of simultaneous gradient descent attractors. This is an important point and we discuss the connections between the critical points of learning dynamics later on.

2. We leave developing estimation schemes for estimating such directions from samples of actions or costs from opponents for future work.

Proposition 3. *For zero-sum settings, if $v_i \equiv \alpha D_j f_j$ for any scalar α, then all critical points of simultaneous gradient descent are critical points of faster conjecture learning.*

Proof. Consider a zero-sum game such that $(f_1, f_2) \equiv (f, -f)$. Let x be a critical point of simultaneous gradient descent so that $D_i f(x) = 0$, $i = 1, 2$. Then, $D_i f(x) + \alpha D_{ij} f(x) D_j f_i(x) = 0$. $\square$

Since the Jacobian has potentially a different eigenstructure at these critical points for each set of dynamics, the set of attractors may not coincide except in special cases which we characterize in later sections.

Suppose now that player 1 no longer believes player 2 myopically updates its choice in the direction v_k, but rather envisions them as a strategic entity that is reacting to x_1 via some *reaction curve*, say $x_2 + v_1(x_1, x_2)$. As an example, consider the reaction curve to be defined by player 2's best response given the objective f_2, which is

$$v_{k,1} = -\widetilde{\gamma}_1 D_2 f_2(x_1, x_{2,k}).$$

Indeed, the reaction curve is gradient conjectures of the form

$$\xi_i(x_i, x_j) = x_j - \widetilde{\gamma}_i D_j f_j(x_i, x_j),\ \ i = 1, 2,\ \ j \neq i$$

where $\widetilde{\gamma}_i$ is player i's conjecture about player j's learning rate. Then, following the same procedure as before, locally player 1 faces the optimization problem

$$\arg\min_{x_1} \{f_1(x_1, x_{2,k}) + D_2 f_1(x_1, x_{k,2}) v_1(x_1, x_{k,2})\}.$$

To get intuition for the direction of steepest descent we can consider

$$\begin{aligned}
f_1(x_1 + w, x_{k,2}) &+ D_2 f_1(x_1 + w, x_{k,2}) v_1(x_1 + w, x_{k,2}) \\
&= f_1(x_1, x_{2,k}) + D_2 f_1(x_1, x_{k,2}) v_1(x_1, x_{k,2}) + D_1 f_1(x_1, x_{2,k}) w \\
&\quad + D_{12} f_1(x_1, x_{k,2}) v_1(x_1, x_{2,k}) w + D_2 f_1(x_1, x_{k,2}) D_1 v_1(x_1, x_{2,k}) w + O(w^2)
\end{aligned}$$

For gradient conjectures, this is exactly

$$\begin{aligned}
f_1(x_1 + w, x_{2,k}) &- \widetilde{\gamma}_1 D_2 f_1(x_1 + w, x_{k,2}) D_2 f_2(x_1 + w, x_{2,k}) \\
&= f_1(x_1, x_{2,k}) - \widetilde{\gamma}_1 D_2 f_1(x_1, x_{k,2}) D_2 f_2(x_1, x_{2,k}) + D_1 f_1(x_1, x_{2,k}) w \\
&\quad - \widetilde{\gamma}_1 D_{12} f_1(x_1, x_{k,2}) D_2 f_2(x_1, x_{2,k}) w - \widetilde{\gamma}_1 D_{12} f_2(x_1, x_{2,k}) D_2 f_1(x_1, x_{k,2}) w + O(w^2).
\end{aligned}$$

The resulting learning rules are given by

$$\begin{aligned}
x_{k+1,1} &= x_{k,1} - \gamma_1\big(D_1 f_1(x_k) - \widetilde{\gamma}_1\big(D_{12} f_1(x_k) D_2 f_2(x_k) + D_{12} f_2(x_k) D_2 f_1(x_k)\big)\big) \\
x_{k+1,2} &= x_{k,2} - \gamma_2\big(D_2 f_2(x_k) - \widetilde{\gamma}_2\big(D_{21} f_2(x_k) D_1 f_1(x_k) + D_{21} f_1(x_k) D_1 f_2(x_k)\big)\big),
\end{aligned}$$

or equivalently

$$x_{k+1} = x_k - \Gamma(I - J_{\mathrm{o}}(x_k)\widetilde{\Gamma}) g(x_k) - \Gamma \mathrm{diag}(J_{\mathrm{o}}^\top(x_k) \nabla f(x_k) \widetilde{\Gamma}).$$

The update component $-\widetilde{\gamma}_i D_{ij} f_i(x_k) D_j f_j(x_k)$ adjusts for the distortions player j causes on player i's component of the vector field and the component $-\widetilde{\gamma}_i D_{ij} f_j(x_k) D_j f_i(x_k)$ accounts for the strategic effects that player j has on player i. We refer to this set of dynamics as fast strategic conjectures.

Conjecture learning	$x_{k+1,i} = x_{k,i} - \gamma_{k,i}(D_i f_i(x_k) + D_i \xi_i^\top(x_k) D_j f_i(x_k)) \quad \forall i$
Strategic implicit conjectures	$x_{k+1} = x_k - \Gamma_k(g(x_k) - \mathrm{diag}(J_{\mathrm{o}}^\top(x_k) J_{\mathrm{d}}^{-1}(x_k) \nabla f(x_k)))$
Strategic gradient conjectures	$x_{k+1} = x_k - \Gamma_k(g(x_k) - \mathrm{diag}(\widetilde{\Gamma} J_{\mathrm{o}}^\top(x_k) \nabla f(x_k)))$
Fast conjectures	$x_{k+1} = x_k - \Gamma_k g(x_k) - \Gamma_k J_{\mathrm{o}}(x_k) v_k$, v_k is oracle direction
Fast strategic conjectures	$x_{k+1} = x_k - \Gamma_k(I - J_{\mathrm{o}}(x_k)\widetilde{\Gamma}) g(x_k) - \Gamma_k \mathrm{diag}(J_{\mathrm{o}}^\top(x_k) \nabla f(x_k) \widetilde{\Gamma})$
Individual Learning	
Simultaneous gradient	$x_{k+1} = x_k - \Gamma_k g(x_k)$
Faster dynamics	
Lookahead [30]	$x_{k+1} = x_k - \gamma(I - \alpha J_o(x_k)) g(x_k)$
Symplectic gradient adjustment [1]	$x_{k+1} = x_k - \gamma(I - \eta(x_k) A^\top(x_k)) g(x_k)$, where $A \equiv \frac{1}{2}(J - J^\top)$ and $\eta(x_k) \in \{-1, 1\}$
Strategic updates	
Learning with opponent learning awareness [11]	$x_{k+1,1} = x_{k,1} - \gamma_1(D_1 f_1(x_k) - \gamma_2 (D_{21} f_2(x_k))^\top D_2 f_1(x_k))$ $x_{k+1,2} = x_{k,2} - \gamma_2 D_2 f_2(x_k)$
Stackelberg learning [8]	$x_{k+1,1} = x_{k,1} - \gamma_{k,1}(D_1 f_1(x_k) - (D_{21} f_2(x_k))^\top (D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k)$ $x_{k+1,2} = x_{k,2} - \gamma_{k,2} D_2 f_2(x_k)$, where $\gamma_{k,1} = o(\gamma_{k,2})$
Stable opponent shaping [19]	$x_{k+1} = x_k - \gamma(I - \alpha J_o(x_k)) g(x_k) - p(x_k) \alpha \mathrm{diag}(J_{\mathrm{o}}^\top(x_k) \nabla f(x_k))$, where $0 < p(x) < 1$ and $p(x) \to 0$ as $\lVert g(x) \rVert \to 0$
Regularization	
Consensus optimization [25]	$x_{k+1} = x_k - \gamma(I - \alpha J^\top(x_k)) g(x_k)$
Competitive gradient descent [28]	$x_{k+1} = x_k - \gamma(I - \gamma^2 J_{\mathrm{o}}^2(x_k))^{-1}(I - \gamma J_{\mathrm{o}}(g_k)) g(x_k)$
Estimated conjectures	
Approximate Fictitious play	$x_{k+1,i} = x_k - \gamma_i g(x_i, \bar{x}_{-i}), \quad \forall i$, where $\bar{x}_{-i}$ is the empirical mean of agent $-i$'s actions
Least squares conjectures	$x_{k+1,i} = x_{k,i} - \gamma_i \big(D_i f_i(x_k) + \hat{P}_{i\mid k}^\top D_{-i} f_i(x_k)\big), \quad \forall i$

3.1 Deriving Related Updates Using the Conjectures Framework

The conjectures framework allows one to derive suites of game learning dynamics. By choosing appropriate strategic conjectures ξ_i for each player and an oracle-provided update direction $v_{k,i}$ or regularization term, one can design different learning rules for heterogeneous strategic or accelerated agents.

Individual Learning

The individual learning dynamics can be derived from each player conjecturing that opponents repeat the last joint strategy that was selected.

Example 3 (Simultaneous Gradient Descent). *If each player has the conjecture* $\xi_i(x_k) = x_{k,-i}$*, then we derive the simultaneous gradient descent update rule*

$$x_{k+1} = x_k - \Gamma g(x_k).$$

Fast Learning Dynamics

We now show a number of fast learning dynamics can be derived from the conjectures framework.

Example 4 (LookAhead [30]). *If the oracle provides the conjectured descent directions*

$$v_{k,1} = -\alpha D_1 f_1(x_k),\ v_{k,2} = -\alpha D_2 f_2(x_k),$$

where $\alpha > 0$ *and the learning rate* $\gamma > 0$ *is common between players, then we derive the update rule*

$$x_{k+1} = x_k - \gamma(I - \alpha J_o(x_k))g(x_k).$$

Example 5 (Symplectic gradient adjustment [1]). *Consider that the oracle provides the conjectured descent direction*

$$v_k = \eta(x_k)\begin{bmatrix} D_1 f(x_k) \\ D_2 f_2(x_k) \end{bmatrix},$$

where $\eta(x_k) \in \{-1, 1\}$ *is negative except when* x_k *is in a neighborhood of an unstable equilibrium and* $J_o(x_k)g(x_k)$ *points away from it or when* x_k *is in a neighborhood of an stable equilibrium and* $J_o(x_k)g(x_k)$ *points towards it. Then,*

$$x_{k+1} = x_k - \gamma(I + \eta(x_k)J_o(x_k))g(x_k).$$

In zero-sum games, $J_o(x_k) = J(x_k) - J^\top(x_k) =: -A^\top(x_k)$*, which results in the update*

$$x_{k+1} = x_k - \gamma(I - \eta(x_k)A^\top(x_k))g(x_k).$$

Strategic Learning Dynamics

Example 6 (Learning with opponent-learning awareness [11]). *Consider a two-player game in which player 1 has a gradient conjecture* $\xi_1(x_k) = x_{k,2} - \gamma_2 D_2 f_2(x_k)$ *and player 2 has the conjecture* $\xi_2(x_k) = x_{k,1}$*. Then, we obtain the combined LOLA dynamics given by*

$$\begin{aligned} x_{k+1,1} &= x_{k,1} - \gamma_1(D_1 f_1(x_k) - \gamma_2(D_{21} f_2(x_k))^\top D_2 f_1(x_k)) \\ x_{k+1,2} &= x_{k,2} - \gamma_2 D_2 f_2(x_k). \end{aligned}$$

Example 7 (Stackelberg Learning [8]). *Consider a two-player game in which player 1 has an implicit conjecture* $\xi_1(x_k)$ *defined by* $D_2 f_2(x_k) = 0$ *and* $D_1\xi_1(x_k) = -(D_2^2 f_2(x_k))^{-1}(D_{21} f_2(x_k))$ *and player 2 has the conjecture* $\xi_2(x_k) = x_{k,1}$*. Then, given learning rates satisfying* $\gamma_{k,1} = o(\gamma_{k,2})$*, we obtain the Stackelberg learning update given by*

$$\begin{aligned} x_{k+1,1} &= x_{k,1} - \gamma_{k,1}(D_1 f_1(x_k) - (D_{21} f_2(x_k))^\top (D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k) \\ x_{k+1,2} &= x_{k,2} - \gamma_{k,2} D_2 f_2(x_k). \end{aligned}$$

Example 8 (Stable opponent shaping [19]). *Suppose agents form conjectures* $\xi_i(x) = p(x_k)\alpha D_{-i} f_{-i}(x)$*, where* $p(x_k) > 0$ *and* $p(x_k) \to 0$ *as* $x \to x^*$*, and an oracle provides conjectured direction* $v_k = \alpha g(x_k)$*. Then we derive*

$$x_{k+1} = x_k - \gamma(I - J_o(x_k))g(x_k) - p(x_k)\alpha \mathrm{diag}(J_o^\top(x_k)\nabla f(x_k)).$$

In equilibrium, $p(x_k) = 0$*, hence the dynamics are equivalent to faster conjecture learning dynamics.*

Regularized Learning Dynamics

Example 9 (Consensus optimization [25])**.** *Suppose agents adopt regularized conjectured costs* $\tilde{f}_i(x) = f_i(x) + \alpha\|g(x_k)\|^2$*. Then we derive*

$$x_{k+1} = x_k - \gamma(I - \alpha J^\top(x_k))g(x_k).$$

Estimated Conjectures

Example 10 (Approximate Fictitious play)**.** *Suppose agents form the conjectures* $\xi_i(x_k) = \bar{x}_{-i} = \frac{1}{k}\sum_{j=1}^{k} x_{j,-i}$ *representing the empirical mean of the observed history of actions from opponents. Then, we obtain the approximate fictitious play learning rule given by*

$$x_{k+1,i} = x_{k,i} - \gamma_i g(x_i, \bar{x}_{-i}), \quad \forall i.$$

Example 11 (Least squares estimated conjectures)**.** *Suppose each agent stores a history of its own actions* $x_{i|k} := (x_{1,i}, \ \dots \ , x_{k,i})$ *and the observed opponent actions* $x_{-i|k} := (x_{1,-i}, \ \dots \ , x_{k,-i})$ *and forms a least squares estimator that fits a regression conjecture* $\xi_i(x_k) = \hat{P}_i x_{k,i} + \hat{q}_i$ *from data, where* $P_i \in \mathbb{R}^{d_{-i}} \times \mathbb{R}^{d_i}$ *and* $q_i \in \mathbb{R}^{d_{-i}}$*. The estimator minimizes least squares loss*

$$(\hat{P}_i, \hat{q}_i) = \arg\min_{P_i, q_i} \|P_i x_{i|k} + q_i - x_{-i|k}\|^2.$$

Since we are ultimately interested in $D\xi_i(x_i) = \hat{P}_{i|k}$*, which can be computed via*

$$\hat{P}_{i|k} = (x_{i|k}^\top x_{i|k})^{-1} x_{i|k}^\top x_{-i|k}$$

at each iteration k*. This yields dynamics*

$$x_{k+1,i} = x_{k,i} - \gamma_i\big(D_i f_i(x_k) + \hat{P}_{i|k}^\top D_{-i} f_i(x_k)\big), \quad \forall i.$$

3.2 Spectral Properties of Fast Conjecture Learning Dynamics

Empirically, many of the updates that fall under a variant of the fast conjecture learning dynamics exhibit rapid convergence to Nash equilibria. To show the behavior of the dynamics is expected theoretically, we can compare the spectral properties of the adjusted or conjectured learning rules to those for simultaneous gradient play. In comparison to the gradient play dynamics, if the eigenvalues of the Jacobian near an equilibrium for an adjusted or conjectured learning rule contain larger real components and smaller imaginary components, then the learning rule will converge faster to an equilibrium than the gradient play dynamics.

We provide results characterizing the spectrum of the fast conjecture learning dynamics for the purpose of improving the understanding of how the real and imaginary parts of the Jacobian can be controlled or how they depend on the conjectured learning rates. To do so, we leverage the theory of indefinite linear operators to construct bounds on the spectrum of the Jacobian of the different proposed learning dynamics nearby equilibria. The significance is two-fold: first, we can utilize the bounds to analyze convergence properties of the fast conjecture learning dynamics (and its derivatives described in Section 3), and second, we can utilize the bounds to synthesize algorithms and heuristics. Indeed, the spectral properties along the learning path can be obtained via computation of the *quadratic numerical range* and then utilized to generate well-informed heuristics to guide agents toward a desired type of equilibria such as cooperative (lower social cost or Pareto) or competitive (Nash equilibria) outcomes.

Towards this end, consider the fast conjecture update:

$$x_{k+1} = x_k - \Gamma_k(g(x_k) + J_o(x_k)v_k).$$

For two-player games, we can demonstrate through bounds on the spectrum of the Jacobian in a neighborhood of a critical point that the fast conjectures update converges faster since the imaginary component is smaller relative to simultaneous gradient play. Consider a two-player game $\mathcal{G} = (f_1, f_2)$ so that

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \end{bmatrix} = \begin{bmatrix} x_{k,1} \\ x_{k,2} \end{bmatrix} - \begin{bmatrix} \gamma_{k,1} & 0 \\ 0 & \gamma_{k,2} \end{bmatrix} \left(\begin{bmatrix} D_1 f_1(x_k) \\ D_2 f_2(x_k) \end{bmatrix} + \begin{bmatrix} 0 & D_{12} f_1(x_k) \\ D_{21} f_2(x_k) & 0 \end{bmatrix} \begin{bmatrix} v_{1,k} \\ v_{2,k} \end{bmatrix} \right).$$

Assuming for simplicity that players have constant, uniform learning rates[3] γ and the conjectures are $v_{i,k} = -\widetilde{\gamma}_i D_i f_i(x_k)$, the limiting dynamics are

$$\dot{x} = -\left(I - \begin{bmatrix} \widetilde{\gamma}_1 & 0 \\ 0 & \widetilde{\gamma}_2 \end{bmatrix} \begin{bmatrix} 0 & D_{12} f_1(x_k) \\ D_{21} f_2(x_k) & 0 \end{bmatrix} \right) \begin{bmatrix} D_1 f_1(x_k) \\ D_2 f_2(x_k) \end{bmatrix}.$$

Let us consider all critical points $\mathcal{C}(\mathcal{G}) = \{x |\ g(x) = 0\}$ where recall that $g(x) = (D_1 f_1(x), D_2 f_2(x))$. The set of critical points contains all local Nash equilibria and GCVE. Our goal is to analyze the spectral properties of the update in a neighborhood of these points. At critical points $x \in \mathcal{C}(\mathcal{G})$, the Jacobian has the structure

$$\begin{aligned} J_{\mathtt{f}}(x) &= \begin{bmatrix} D_1^2 f(x) - \widetilde{\gamma}_1 D_{12} f_1(x) D_{21} f_2(x) & D_{12} f_1(x)(I - \widetilde{\gamma}_1 D_2^2 f_2(x)) \\ D_{21} f_2(x)(I - \widetilde{\gamma}_2 D_1^2 f_1(x)) & D_2^2 f_2(x) - \widetilde{\gamma}_2 D_{21} f_2(x) D_{12} f_1(x) \end{bmatrix} \\ &= J(x) - \widetilde{\Gamma} J_{\mathrm{o}}(x) J(x). \end{aligned}$$

Similarly, recall the Jacobian of the simultaneous gradient play update:

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}.$$

Each Jacobian is a block operator matrix. We can find bounds on the *point spectrum*[4] of $J_{\mathtt{f}}(x)$, denoted $\sigma_p(J_{\mathtt{f}}(x))$, by bounding elements of the *quadratic numerical range*, denoted $\mathcal{W}^2(J_{\mathtt{f}}(x))$, since $\sigma_p(J_{\mathtt{f}}(x)) \subset \mathcal{W}^2(J_{\mathtt{f}}(x))$ [29, Theorem 4.1]. In zero-sum games, $J(x)$ has a very special block structure resembling a J-frame on a Krein space; indeed, it has the form

$$\begin{bmatrix} A & B \\ -B^T & -D \end{bmatrix}.$$

Given a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$, for a block operator

$$\mathcal{A} = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

the numerical range is defined by $\mathcal{W}(\mathcal{A}) = \{\langle \mathcal{A}x, x \rangle |\ x \in \mathcal{H}, \|x\| = 1\}$, and the quadratic numerical range is defined by

$$\mathcal{W}^2(\mathcal{A}) = \bigcup_{\substack{(v,w)^\top \in \mathcal{D}(\mathcal{A}) \\ \|v\| = \|w\| = 1}} \sigma_p(\mathcal{A}_{v,w})$$

3. If player's have different constant learning rates γ_1 and γ_2, then a scaling factor can be introduced such that the limiting dynamics are on the slowest player's timescale.
4. The point spectrum of an operator is the set of eigenvalues.

where $\mathcal{D}(\cdot)$ is the domain of its argument and $\mathcal{A}_{v,w} \in \mathbb{C}^{2\times 2}$ is defined by

$$\mathcal{A}_{v,w} = \begin{bmatrix} \langle Av, v\rangle & \langle Bw, v\rangle \\ \langle Cv, w\rangle & \langle Dw, w\rangle \end{bmatrix}.$$

Note that $\mathcal{D}(J_{\mathfrak{f}}(x))$ is simply $\mathbb{C}^{d_1+d_2}$ where d_i is the dimensions of player i's action space X_i.

Lemma 1 (Lemma 5.1 [29])**.** *Let $a, b, c, d \in \mathbb{C}$ and*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The eigenvalues λ_1, λ_2 of A have the following properties.

(i) If $\mathrm{Re}(d) < 0 < \mathrm{Re}(a)$ and $bc \geq 0$, then
   - *(a) $\mathrm{Re}(\lambda_2) \leq \mathrm{Re}(d) < 0 < \mathrm{Re}(a) \leq \mathrm{Re}(\lambda_1)$*
   - *(b) $\min\{\mathrm{Im}(a), \mathrm{Im}(d)\} \leq \mathrm{Im}(\lambda_1)$, $\mathrm{Im}(\lambda_2) \leq \max\{\mathrm{Im}(a), \mathrm{Im}(d)\}$*
   - *(c) $\lambda_1, -\lambda_2 \in \{z \in \mathbb{C}|\ |\arg(z)| \leq \max\{|\arg(a)|, \pi - |\arg(d)|\}$*

(ii) If $\mathrm{Re}(d) < \mathrm{Re}(a)$ and $bc \leq 0$, then
   - *(a) $\mathrm{Re}(d) \leq \mathrm{Re}(\lambda_2) \leq \mathrm{Re}(\lambda_1) \leq \mathrm{Re}(a)$,*
   - *(b) $\mathrm{Re}(\lambda_2) \leq \mathrm{Re}(d) + \sqrt{|bc|} < \mathrm{Re}(a) - \sqrt{|bc|} \leq \mathrm{Re}(\lambda_1)$ if $\sqrt{|bc|} < (\mathrm{Re}(a) - \mathrm{Re}(d))/2$, and $\lambda_1, \lambda_2 \in \mathbb{R}$ if, in addition, $a, d \in \mathbb{R}$,*
   - *(c) $\mathrm{Re}(\lambda_1) = \mathrm{Re}(\lambda_2) = (a+d)/2$, $|\mathrm{Im}(\lambda_1)| = |\mathrm{Im}(\lambda_2)| = \sqrt{|bc| - (a-d)^2/4}$ if $a, d \in \mathbb{R}$, and $\sqrt{|bc|} \geq (a-d)/2$.*

We can use the lemma and the fact that the point spectrum is contained in the quadratic numerical range to provide bounds on the eigenvalues of the Jacobian for update rules in various game structures.

As before, let $\dot{x} = -g(x)$ be the limiting continuous time dynamics of simultaneous gradient play where $g \equiv (D_1 f_1, D_2 f_2)$. For a game $\mathcal{G} = (f_1, f_2)$, recall that the individual player Hessians $D_i^2 f_i$ are real, symmetric matrices. Hence, it can be shown that the numerical range is bounded by the minimum and maximum eigenvalues. With this in mind, define the following quantities:

$$\lambda_i^- = \min \sigma_p(D_i^2 f_i(x)),\ \ i = 1, 2 \tag{3}$$

$$\lambda_i^+ = \max \sigma_p(D_i^2 f_i(x)),\ \ i = 1, 2 \tag{4}$$

We first characterize the spectrum of the simultaneous gradient play update.

Proposition 4. *Consider a zero-sum game $\mathcal{G} = (f, -f)$. The Jacobian $J(x)$ of the dynamics $\dot{x} = -g(x)$ at critical points $x \in \mathcal{C}(\mathcal{G})$ is such that*

$$\sigma_p(J(x)) \cap \mathbb{R} \subset [\min\{\lambda_1^-, \lambda_2^-\}, \max\{\lambda_1^+, \lambda_2^+\}] \tag{5}$$

and

$$\sigma_p(J(x)) \backslash \mathbb{R} \subset \left\{ z \in \mathbb{C} \,\middle|\, \frac{\lambda_1^- + \lambda_2^-}{2} \leq \mathrm{Re}(z) \leq \frac{\lambda_1^+ + \lambda_2^+}{2},\ |\mathrm{Im}(z)| \leq \|D_{12} f(x)\| \right\}. \tag{6}$$

Furthermore, if $\lambda_1^- - \lambda_2^+ > 0$ or $\lambda_2^- - \lambda_1^+ > 0$ then the following two implications hold for $\delta = \lambda_1^- - \lambda_2^+$ or $\delta = \lambda_2^- - \lambda_1^+$, respectively:

(i) $\|D_{12} f(x)\| \leq \delta/2 \implies \sigma_p(J(x)) \subset \mathbb{R}$;

(ii) $\|D_{12} f(x)\| > \delta/2 \implies \sigma_p(J(x)) \backslash \mathbb{R} \subset \{z \in \mathbb{C}|\ |\mathrm{Im}(z)| \leq \sqrt{\|D_{12} f(x)\|^2 - \delta^2/4}\}$.

Proof. The proof follows from the application of Lemma 1-(ii), since $-w^*D_{12}f(x)^\top vv^*D_{12}f(x)w \le 0$, and the fact that $\overline{\det(J_{v,w}(x)-\lambda I)} = \det(J_{v,w}(x)-\bar{\lambda}I)$ for $v,w\in\mathbb{C}^{d_1+d_2}$ such that $\|v\|=\|w\|=1$, since $D_1^2f(x)$ and $-D_2^2f(x)$ are symmetric, which in turn implies that $\mathcal{W}^2(J(x)) = \mathcal{W}^2(J(x))^*$. □

Now, we provide a similar result on the spectral properties of the fast conjecture learning dynamics given above at critical points $x\in\mathcal{C}(\mathcal{G})$ for a given two-player game $\mathcal{G}=(f_1,f_2)$. We note that in the simultaneous gradient play case, for a zero-sum game $\mathcal{G}=(f,-f)$, $J(x)$ has $D_{12}f_1(x)$ and $-D_{12}f^\top(x)$ as off-diagonal elements so that the condition $bc\le 0$ is always satisfied in Lemma 1-(ii). This is not the case, however, for the fast conjecture learning dynamics, even in the zero-sum case, since the off-diagonal terms are modified by $(I-\tilde{\gamma}D_2^2f_2(x))$ and $(I-\tilde{\gamma}D_1^2f_1(x))$, respectively. We note, however, both these matrices have norm less than one for appropriately chosen $\tilde{\gamma}_i$'s and they are both symmetric.

For a given two-player game $\mathcal{G}=(f_1,f_2)$, define the following quantities:

$$a_- = \inf \mathrm{Re}(\mathcal{W}(D_1^2f_1(x) - \tilde{\gamma}_1 D_{12}f_1(x)D_{21}f_2(x))) \tag{7}$$
$$a_+ = \sup \mathrm{Re}(\mathcal{W}(D_1^2f_1(x) - \tilde{\gamma}_1 D_{12}f_1(x)D_{21}f_2(x))) \tag{8}$$
$$d_- = \inf \mathrm{Re}(\mathcal{W}(D_2^2f_2(x) - \tilde{\gamma}_2 D_{21}f_2(x)D_{12}f_1(x))) \tag{9}$$
$$d_+ = \sup \mathrm{Re}(\mathcal{W}(D_2^2f_2(x) - \tilde{\gamma}_2 D_{21}f_2(x)D_{12}f_1(x))) \tag{10}$$

Note that in the case of a zero-sum game, $D_i^2f_i(x)+\tilde{\gamma}_iD_{ij}f_i(x)D_{ji}f_j(x)$ is real, symmetric so that the above quantities are simply the minimum and maximum eigenvalues of their arguments as above. In the general-sum case, however, these matrices may not be symmetric and hence, the quadratic range is used in place of the minimum and maximum eigenvalues. Define also the quantities

$$\alpha(x) = \|D_{12}f_1(x)(I-\tilde{\gamma}_1D_2^2f_2(x))\|\|D_{21}f_2(x)(I-\tilde{\gamma}_2D_1^2f_1(x))\|$$

and

$$\beta(x) = \langle D_{12}f(x)(I+\tilde{\gamma}_1D_2^2f(x))w, v\rangle\langle -D_{12}f(x)^\top(I-\tilde{\gamma}_2D_1^2f(x))v, w\rangle$$

Unlike in the case of simultaneous gradient play where we need to only use Lemma 1-(ii), since $\beta(x)$ may be positive or negative, we need to consider both the cases of Lemma 1-(i) and Lemma 1-(ii).

Proposition 5. *Consider a game $\mathcal{G}=(f_1,f_2)$. Suppose that $\max_{(v,w)\in\mathcal{B}}\beta(x)\le 0$ where $\mathcal{B}=\{v\in\mathbb{C}^{d_1},\ w\in\mathbb{C}^{d_2}|\ \|w\|=\|v\|=1\}$. The spectrum of the Jacobian $J_{\mathfrak{f}}(x)$ at critical points $x\in\mathcal{C}(\mathcal{G})$ is such that*

$$\min\{a_-,d_-\}\le \mathrm{Re}(\sigma_p(J_{\mathfrak{f}}(x)))\le\max\{a_+,d_+\},$$

and

$$|\mathrm{Im}(\sigma(J_{\mathfrak{f}}(x)))|\le\sqrt{\alpha(x)}\le(\|D_{12}f_1(x)\|\|D_{21}f_2(x)\|)^{1/2}.$$

Furthermore, $J_{\mathfrak{f}}(x)$ has the following spectral properties:

(i) If $d_+<a_-$, then $\sqrt{|\beta(x)|}<\frac{1}{2}(a_--d_+)$ implies that $\sigma_p(J_{\mathfrak{f}}(x))\subset\mathcal{W}^2(J_{\mathfrak{f}}(x))\subset\mathbb{R}$ and, on the other hand, $\sqrt{\alpha(x)}\ge\sqrt{|\beta(x)|}\ge\frac{1}{2}(a_--d_+)$ implies that

$$|\mathrm{Im}(\sigma_p(J_{\mathfrak{f}}(x)))|\le|\mathrm{Im}(\mathcal{W}^2(J_{\mathfrak{f}}(x)))|\le\Big(\alpha(x)-\frac{(a_--d_+)^2}{4}\Big)^{1/2}.$$

(ii) If $a_+<d_-$, then $\sqrt{|\beta(x)|}<\frac{1}{2}(d_--a_+)$ implies $\sigma_p(J_{\mathfrak{f}}(x))\subset\mathcal{W}^2(J_{\mathfrak{f}}(x))\subset\mathbb{R}$ and, on the other hand, $\sqrt{\alpha(x)}\ge\sqrt{|\beta(x)|}>\frac{1}{2}(d_--a_+)$ implies that

$$|\mathrm{Im}(\sigma_p(J_{\mathfrak{f}}(x)))|\le|\mathrm{Im}(\mathcal{W}^2(J_{\mathfrak{f}}(x)))|\le\Big(\alpha(x)-\frac{(d_--a_+)^2}{4}\Big)^{1/2}.$$

Moreover, if the game is zero-sum (i.e., $f_1 \equiv f$ and $f_2 \equiv -f$), then $|\mathrm{Im}(\sigma(J_{\mathtt{f}}(x)))| \leq \sqrt{\alpha(x)} \leq \|D_{12}f(x)\|$.

Proof. The proof follows from Lemma 1-(ii) where we note that

$$\begin{aligned}|bc| &= |\langle D_{12}f_1(x)(I-\tilde{\gamma}_1 D_2^2 f_2(x))w, v\rangle\langle D_{21}f_2(x)(I-\tilde{\gamma}_2 D_1^2 f_1(x))v, w\rangle| = |\beta(x)|\\ &\leq \|D_{12}f_1(x)(I-\tilde{\gamma}_1 D_2^2 f_2(x))\|\|D_{21}f_2(x)(I-\tilde{\gamma}_2 D_1^2 f_1(x))\| = \alpha(x)\\ &\leq \|D_{12}f_1(x)\|\|D_{21}f_2(x)\|\end{aligned}$$

for $(v,w)\in\mathcal{B}$ and where the last inequality holds since by assumption, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are such that the eigenvalues of $I-\tilde{\gamma}_i D_j^2 f_j(x)$ are in the unit circle for $i,j=1,2$ with $j\neq i$. □

Let us now consider the case such $\min_{(v,w)\in\mathcal{B}}\beta(x)\geq 0$. That is, $bc\geq 0$ for all 2×2 matrices $(J_{\mathtt{f}}(x))_{v,w}$.

Proposition 6. *Consider a game $\mathcal{G}=(f_1,f_2)$. Suppose that $\min_{(v,w)\in\mathcal{B}}\beta(x)\geq 0$ and that $d_+ < 0 < a_-$. The Jacobian $J_{\mathtt{f}}$ has the following spectral properties at critical points $x\in\mathcal{C}(\mathcal{G})$:*

(i) The real-part of the spectrum $\mathrm{Re}(\sigma_p(J_{\mathtt{f}}(x)))$ splits into two disjoint intervals $(-\infty, d_+]$ and $[a_-,\infty)$.

(ii) The imaginary part satisfies $\min\{a_-, d_-\}\leq \mathrm{Im}(\sigma_p(J_{\mathtt{f}}(x)))\leq \max\{a_+, d_+\}$.

Hence, in this case, the imaginary component of the spectrum can be controlled by the choice of $\tilde{\gamma}$ and how it impacts spectrum of the diagonal components of the Jacobian.

The previous two propositions apply to general-sum settings, which of course subsumes zero-sum settings. However, in the zero-sum setting we can provide sufficient conditions on the conjectured learning rates under which convergence is guaranteed to be faster. Towards this end, the next set of results demonstrate that fast conjecture learning provably speeds up convergence in zero-sum games.

Recall that the fast conjecture learning dynamics with uniform conjectured learning rates are defined by

$$g(x)-\tilde{\gamma}J_{\mathrm{o}}(x)g(x)$$

so that the Jacobian is

$$J_{\mathtt{f}}(x)=J(x)-\tilde{\gamma}J_{\mathrm{o}}(x)J(x)$$

in a neighborhood of stable critical points $\mathcal{C}_s(g)=\{x|\ g(x)=0,\ \sigma_p(J(x))\in\mathbb{C}_+^\circ\}$.

Now, let us examine the imaginary component of the eigenvalues of the Jacobian with the goal of showing that for certain choices of $\tilde{\gamma}$ the imaginary part decreases relative to the imaginary component of the Jacobian of simultaneous gradient play.

Proposition 7. *Consider a zero-sum game $\mathcal{G}=(f,-f)$. Suppose that x is a stable critical point of $g(x)$. Let $c=\max_i|\mathrm{Re}(\lambda_i(J(x)))|$ where $\lambda_i(J(x))$ is an eigenvalue of $J(x)$. For $\tilde{\gamma}<2/c$, the magnitude of imaginary components of the eigenvalues of $J_{\mathtt{f}}(x)$ are smaller than the magnitude of the imaginary components of the eigenvalues of $J(x)$.*

Proof. The imaginary component can be computed by taking into consideration the antisymmetric part of the Jacobian, which is given by

$$\begin{aligned}J_{\mathtt{f}}(x)-J_{\mathtt{f}}^\top(x) &= J(x)-\tilde{\gamma}J_{\mathrm{o}}(x)J(x)-(J(x)-\tilde{\gamma}J_{\mathrm{o}}(x)J(x))^\top\\ &= J(x)-\frac{\tilde{\gamma}}{2}J(x)J(x)-(J(x)-\frac{\tilde{\gamma}}{2}J(x)J(x))^\top\end{aligned}$$

using the fact that $J_{\text{o}}(x) = J(x) - J(x)^\top$ for zero sum games. Note also that since $J(x)$ commutes with it self, given an eigenpair (λ, v) of $J(x)$, $(\lambda - \frac{\tilde{\gamma}}{2}\lambda^2, v)$ is an eigenpair of $J(x) - \frac{\tilde{\gamma}}{2}J(x)J(x)$. Hence, we can just examine the imaginary parts of the eigenvalues $\lambda - \frac{\tilde{\gamma}}{2}\lambda^2$ and compare it to λ where $\lambda \in \sigma_p(J(x))$. Let $\lambda = a + ib$ so that

$$\operatorname{Im}\left(\lambda - \frac{\tilde{\gamma}}{2}\lambda^2\right) = (1 - a\tilde{\gamma})b.$$

Since x is a stable attractor, $a > 0$. Hence, since $0 < \tilde{\gamma} < 2/c < 2/a$ for all $a = \operatorname{Re}(\lambda)$ such that $\lambda \in \sigma_p(J(x))$,

$$\left|\operatorname{Im}\left(\lambda - \frac{\tilde{\gamma}}{2}\lambda^2\right)\right| = |(1 - a\tilde{\gamma})b| < |b|$$

which proves the desired result. $\square$

The above proposition shows that the imaginary part of the eigenvalues of $J_{\mathtt{f}}$ relative to those of J for a range of conjectured learning rates. We can also use $\tilde{\gamma}$ to control the ratio of the imaginary part to the real part by varying $\tilde{\gamma}$. The technique we use combines properties of the fast conjecture learning dynamics with an approach to bounding this ratio introduced in [25]. Define $\tilde{J}(x) = J(x) - \frac{\tilde{\gamma}}{2}J(x)J(x)$. For any matrix $A \in \mathbb{C}^{d\times d}$, recall that the numerical range is defined by $\mathcal{W}(A) = \{\langle Av, v\rangle |\ v \in \mathbb{C}^d,\ \|v\| = 1\}$. It is known that $\mathcal{W}(A)$ lies in a disc of radius $\|A\|$.

Proposition 8. *Consider a zero-sum game $\mathcal{G} = (f, -f)$ and a stable critical point x of $g(x)$ such that $D_i^2 f_i(x) > 0$—i.e., x is a differential Nash equilibrium. Suppose $\tilde{\gamma}$ is chosen such that $0 < \tilde{\gamma} < 2/\max\{\lambda_{\min}(D_1^2 f(x)), \lambda_{\min}(-D_2^2 f(x))\}$ and*

$$\frac{1}{\alpha + \tilde{\gamma}\frac{\sigma_{\min}(J(x))}{\rho}} \le \min_{v:\ \|v\|=1} \frac{|v^*(J(x) - J^\top(x))v|}{|v(J(x) + J^\top(x))v|}$$

where $\sigma_{\min}(J(x))$ is the minimum singular value of $J(x)$, $\rho = \|\tilde{J}(x) - \tilde{J}^\top(x)\|$, and

$$\alpha = \min_{v:\ \|v\|=1} \frac{|v^*(\tilde{J}(x) + \tilde{J}(x)^\top)v|}{|v^*(\tilde{J}(x) - \tilde{J}(x)^\top)v|}$$

Then, the maximum ratio of imaginary to real parts of the eigenvalues of the Jacobian of the fast conjecture learning rate is less than the minimum ratio of the imaginary to real parts of the eigenvalues of the Jacobian of simultaneous gradient play.

Proof. For any eigenvector v of $J_{\mathtt{f}}(x)$ with $\|v\| = 1$, consider the ratio of the imaginary to the real part of eigenvalue which is given by

$$\frac{|v^*(J_{\mathtt{f}}(x) - J_{\mathtt{f}}^\top(x))v|}{|v^*(J_{\mathtt{f}}(x) + J_{\mathtt{f}}^\top(x))v|} \tag{11}$$

Further,

$$J_{\mathtt{f}}(x) - J_{\mathtt{f}}^\top(x) = \tilde{J}(x) - \tilde{J}^\top(x)$$

Note that since x is a stable critical point, $\|J(x)v\|^2 \ge \sigma_{\min}(J(x))\|v\| > 0$ for any non-trivial vector v. Hence, the expression in (11) can be upper bounded. The first step towards constructing such a bound is to show that

$$|v^*(\tilde{J}(x) + \tilde{J}(x)^\top)v + \tilde{\gamma}v^*J^\top(x)J(x)v| = |v^*(\tilde{J}(x) + \tilde{J}(x)^\top)v| + \tilde{\gamma}\|J(x)v\|^2 \tag{12}$$

Let v be a norm one eigenvector of $J_{\mathtt{f}}(x)$. We note that $\tilde{\gamma}v^* J^\top(x)J(x)v$ is always positive. In addition, using the notion of a numerical range, $v^*(\tilde{J}(x)+\tilde{J}^\top(x))v$ is lower bounded by the minimum eigenvalue of $\tilde{J}(x)+\tilde{J}^\top(x)$. Note that $\tilde{J}(x)+\tilde{J}^\top(x)$ is a block-diagonal matrix of the form

$$\begin{bmatrix} 2D_1^2f(x)+\tilde{\gamma}D_{12}f(x)D_{12}f^\top(x)-\tilde{\gamma}(D_1^2f(x))^2 & 0 \\ 0 & -2D_2^2f(x)+\tilde{\gamma}D_{12}f^\top(x)D_{12}f(x)+\tilde{\gamma}(-D_2^2f(x))^2 \end{bmatrix}$$

so that its eigenvalues are the union of the eigenvalues of the two block-diagonal entries which are both symmetric matrices. Since for two symmetric matrices A and B, $\lambda_{\min}(A+B)\geq\lambda_{\min}(A)+\lambda_{\min}(B)$, we have that

$$\begin{aligned} &\lambda_{\min}(2D_1^2f(x)+\tilde{\gamma}D_{12}f(x)D_{12}f^\top(x)-\tilde{\gamma}(D_1^2f(x))^2) \\ &\qquad\geq 2\lambda_{\min}(D_1^2f(x))+\tilde{\gamma}(\lambda_{\min}(D_{12}f(x)D_{12}f^\top(x))-\lambda_{\min}((D_1^2f(x))^2)), \end{aligned}$$

and an analogous inequality holds for the other block-diagonal entry. Now, if $\lambda_{\min}(D_{12}f(x)D_{12}f^\top(x))-\lambda_{\min}((D_1^2f(x))^2)\geq 0$ and $\lambda_{\min}(D_{12}f^\top(x)D_{12}f(x))+\lambda_{\min}((D_2^2f(x))^2)\geq 0$, then any $\tilde{\gamma}>0$—and in particular the range of $\tilde{\gamma}$'s assumed—ensures that that the minimum eigenvalue of $\tilde{J}(x)+\tilde{J}^\top(x)$ is positive. On the other hand, if either of the values is negative, then choosing $\tilde{\gamma}<2/\max\{\lambda_{\min}(D_1^2f(x)),\lambda_{\min}(-D_2^2f(x))\}$ ensures that $\lambda_{\min}(\tilde{J}(x)+\tilde{J}^\top(x))>0$. Hence, the equality in (12) holds since both quadratic terms are positive.

Now, using this fact, we have that

$$\begin{aligned} \frac{|v^*(\tilde{J}(x)-\tilde{J}(x)^\top)v|}{|v^*(\tilde{J}(x)+\tilde{J}(x)^\top)v+\tilde{\gamma}v^*J^\top(x)J(x)v|} &= \frac{|v^*(\tilde{J}(x)-\tilde{J}(x)^\top)v|}{|v^*(\tilde{J}(x)+\tilde{J}(x)^\top)v|+\tilde{\gamma}\|J(x)v\|^2} \\ &\leq \frac{1}{\alpha+\tilde{\gamma}\frac{\|J(x)v\|^2}{|v^*(\tilde{J}(x)-\tilde{J}(x)^\top)v|}} \\ &\leq \frac{1}{\alpha+\tilde{\gamma}\frac{\sigma_{\min}(J(x))}{\rho}} \end{aligned}$$

since

$$\alpha=\min_{v:\ \|v\|=1}\frac{|v^*(\tilde{J}(x)+\tilde{J}(x)^\top)v|}{|v^*(\tilde{J}(x)-\tilde{J}(x)^\top)v|}\leq\min_{v:\ \|v\|=1, J_{\mathtt{f}}(x)v=\lambda v}\frac{|v^*(\tilde{J}(x)+\tilde{J}(x)^\top)v|}{|v^*(\tilde{J}(x)-\tilde{J}(x)^\top)v|}$$

This gives the designed result as the lefthand side is simply the ratio of the imaginary to real components of the eigenvalues of $J_{\mathtt{f}}(x)$. $\square$

4. Numerical Examples

Cost landscapes are locally quadratic near fixed points. Thus, it is illuminating to investigate both the resulting stationary points (steady-state behavior) and dynamics (transient behavior) of quadratic games.

A key aspect of strategic conjectures is the agent's ability to modify the equilibria of the coupled interaction without modifying the value of agents' costs at a given joint action. Conjectures allow agents to impose their internal models of others to guide the learning process: the vector field of the underlying game warps and its equilibria shifts. The steady-state solution of the game is a point in the vector field where agents have no incentive to deviate from their choice of actions, with respect to their conjectured costs. The ability to form different conjectures provides agents with a 'tunable' parameter to relocate the equilibrium point to different strategies, and may be an effective way for agents to adaptively adjust the equilibria based off some individual or external preference that is not encoded by the cost itself.

The conjectures also allow agents to modify the transient dynamics of the interaction. How fast one might converge to the equilibria, the agents' incurred regret, the magnitude of oscillations and whether the equilibria is stable at all, are all properties of the transient learning dynamics in games that may be of interest. We investigate both these phenomena in general-sum and zero-sum two-player quadratic games below.

4.1 Shifting Equilibria in General-Sum Games

The general conjectural variations equilibria of a game, in general, do not coincide with its Nash equilibria. By forming different conjectures, agents are able to adjust the location of the equilibria of the learning dynamics, and converge to points that may be desirable as the outcome of the game.

We illustrate this concept by forming a quadratic game which has Nash, Pareto and Stackelberg equilibria all obtained through forming different conjectures. We compare the costs at equilibrium of these examples to motivate why agents might choose to adopt these conjectures to arrive at these strategies.

Consider a general-sum game (f_1, f_2) where player 1 and 2's actions are $x_1 \in \mathbb{R}^{d_1}$ and $x_2 \in \mathbb{R}^{d_2}$ and their respective costs are given by

$$f_i(x_1, x_2) = \frac{1}{2} x_i^\top A_i x_i + x_i^\top B_{ij} x_j + \frac{1}{2} x_j^\top C_{ij} x_j + e_i^\top x_i + h_i^\top x_j,\ i, j = 1, 2,\ j \neq i$$

for positive definite A_1 and A_2. The Nash equilibrium of this game can be obtained by solving the linear equation of the sufficient first-order conditions of Nash,

$$\begin{bmatrix} x_1^{\mathtt{NE}} \\ x_2^{\mathtt{NE}} \end{bmatrix} = -\begin{bmatrix} A_1 & B_{12} \\ B_{21} & A_2 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ h_2 \end{bmatrix} = -J^{-1} \begin{bmatrix} e_1 \\ h_2 \end{bmatrix},$$

assuming the Jacobian matrix is invertible. The second-order conditions of Nash are satisfied since $A_1, A_2 > 0$. If agents employ gradient conjectures with conjectured learning rates $\widetilde{\gamma}_1, \widetilde{\gamma}_2$, then the Jacobian of the first and second player's conjecture are $D\xi_1(x) = \widetilde{\gamma}_1 B_{21}$ and $D\xi_2(y) = \widetilde{\gamma}_2 B_{12}$ respectively. The general conjectural variations equilibrium of the game can be obtained by solving for its first-order conditions

$$\left(\begin{bmatrix} A_1 & B_{12} \\ B_{21} & A_2 \end{bmatrix} - \begin{bmatrix} \widetilde{\gamma}_1 B_{21}^\top B_{12}^\top & \widetilde{\gamma}_1 B_{21}^\top C_{12} \\ \widetilde{\gamma}_2 B_{12}^\top C_{21} & \widetilde{\gamma}_2 B_{12}^\top B_{21}^\top \end{bmatrix} \right) \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} + \begin{bmatrix} e_1 \\ h_2 \end{bmatrix} - \begin{bmatrix} \widetilde{\gamma}_1 B_{21}^\top h_1 \\ \widetilde{\gamma}_2 B_{12}^\top e_2 \end{bmatrix} = 0.$$

Assuming the matrix is invertible, we apply the matrix inversion identity $(J - U\Gamma V)^{-1} = J^{-1} + J^{-1} U M^{-1} V J^{-1}$ where $M = \Gamma^{-1} - V J^{-1} U$ to compute the general conjectural variations equilibrium in terms of the Nash equilibrium,

$$\begin{aligned} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} &= -\left(J - \begin{bmatrix} B_{21}^\top & 0 \\ 0 & B_{12}^\top \end{bmatrix} \begin{bmatrix} \widetilde{\gamma}_1 I_{d_1} & 0 \\ 0 & \widetilde{\gamma}_2 I_{d_2} \end{bmatrix} \begin{bmatrix} B_{12}^\top & C_{12} \\ C_{21} & B_{21}^\top \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} e_1 \\ h_2 \end{bmatrix} - \begin{bmatrix} \widetilde{\gamma}_1 B_{21}^\top h_1 \\ \widetilde{\gamma}_2 B_{12}^\top e_2 \end{bmatrix} \right) \\ &= \begin{bmatrix} x_1^{\mathtt{NE}} \\ x_2^{\mathtt{NE}} \end{bmatrix} + \left(J^{-1} \begin{bmatrix} B_{21}^\top & 0 \\ 0 & B_{12}^\top \end{bmatrix} M^{-1} \begin{bmatrix} B_{12}^\top & C_{12} \\ C_{21} & B_{21}^\top \end{bmatrix} J^{-1} \right) \left(\begin{bmatrix} e_1 \\ h_2 \end{bmatrix} - \begin{bmatrix} \widetilde{\gamma}_1 B_{21}^\top h_1 \\ \widetilde{\gamma}_2 B_{12}^\top e_2 \end{bmatrix} \right) - J^{-1} \begin{bmatrix} \widetilde{\gamma}_1 B_{21}^\top h_1 \\ \widetilde{\gamma}_2 B_{12}^\top e_2 \end{bmatrix} \end{aligned}$$

where

$$M = \begin{bmatrix} (1/\widetilde{\gamma}_1) I_{d_1} & 0 \\ 0 & (1/\widetilde{\gamma}_2) I_{d_2} \end{bmatrix} - \begin{bmatrix} B_{12}^\top & C_{12} \\ C_{21} & B_{21}^\top \end{bmatrix} J^{-1} \begin{bmatrix} B_{21}^\top & 0 \\ 0 & B_{12}^\top \end{bmatrix},$$

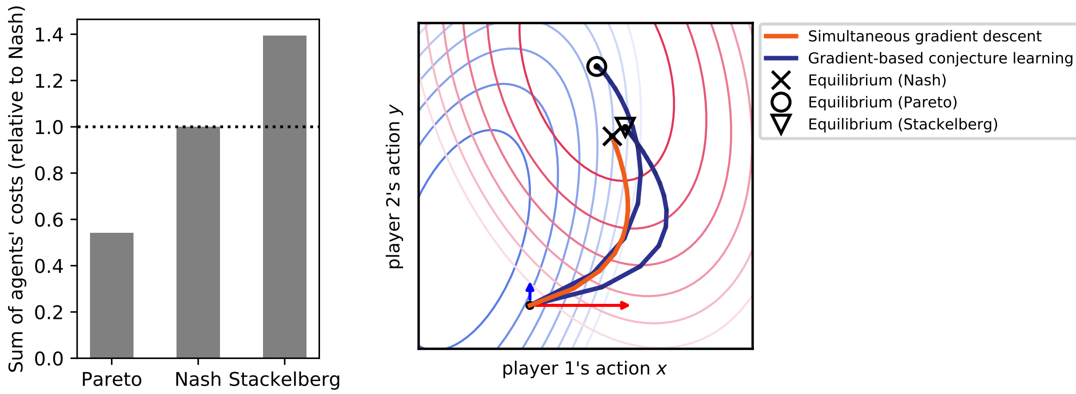

Figure 1: Agents employing different conjectures arrive at different general conjectural variations equilibria. When both players have the conjecture $\lambda_i(x) = x_{-i}$, the general conjectural variations equilibrium coincides with the Nash equilibrium. When player 1 formulates a gradient-based conjecture with $\widetilde{\gamma}_1 = A_2^{-1}$, and player 2 has no conjecture, the general conjectural variations equilibrium is a Stackelberg equilibrium with player 1 being the leader. When the joint conjectured learning rates $(.5, .325)$, the general conjectural variations equilibrium is on the Pareto front. The sum of the agents' costs differ at the three equilibria. The conjecture framework provides agents with tunable parameters to adjust properties at stable equilibria.

and I_{d_i} is the identity matrix of size $d_i \times d_i$. Writing the general conjectural variations equilibrium in terms of the Nash equilibrium allows us to quantify the difference between their solutions, $x^* - x^{\texttt{NE}}$. We demonstrate this through a scalar example.

We construct a general-sum quadratic game with $A_1 = 4.6$, $B_{12} = 1.0$, $C_{12} = 2$, $e_1 = -19.8$, $h_1 = -15.0$ and $A_2 = 1.4$, $B_{21} = -1.0$, $C_{21} = 4.0$, $e_2 = 10$, $h_2 = -0.2$. The level sets of the game are shown in

Figure 1 as red and blue contour lines for player 1 and player 2, respectively. Gradient play converges to the Nash equilibrium.

Depending on the choice of conjectured learning rates, the conjecture dynamics can converge to a Stackelberg equilibrium or a Pareto strategy. If player 1 is the leader and player 2 the follower, by choosing $\widetilde{\gamma}_1 = A_2^{-1} = 0.71$ and $\widetilde{\gamma}_2 = 0$, the dynamics converge to the Stackelberg equilibrium where the leaders' cost is lower but sum cost is higher. On the other hand, with $\widetilde{\gamma}_1 = .5, \widetilde{\gamma}_2 = .325$, the dynamics converges to the Pareto front, where player 1's cost is minimized with respect to both actions. By adjusting one's conjecture about the other, agents can manipulate the outcome of a game to their advantage or to satisfy some external criteria.

4.2 Dampening Oscillations in Zero-Sum Games

Zero-sum games have large rotational vector fields when the asymmetric component of the Jacobian is large. Indeed, the asymmetric components of a zero-sum game $\mathcal{G} = (f, -f)$ under simultaneous gradient descent are $D_{12}f(x)$ and $-D_{21}f(x)$. For large $\|D_{12}f(x)\|$ relative to the individuals' Hessians, we expect to have large rotational components in the game's learning dynamics, as indicated by the results in Section 3.2. We investigate the use of gradient-based conjectures to dampening these oscillations in zero-sum games.

In Figure 2, we show the learning trajectory of two players in a zero-sum quadratic game from Example 2. We randomly generate the individual Hessians A_1, A_2 and coupling term B by sampling each element from a normal distribution. The first player's action $x_1 \in \mathbb{R}^{25}$ and the second player's action $x_2 \in \mathbb{R}^{75}$ and we plot the first components each player's action in Figure 2. To ensure that

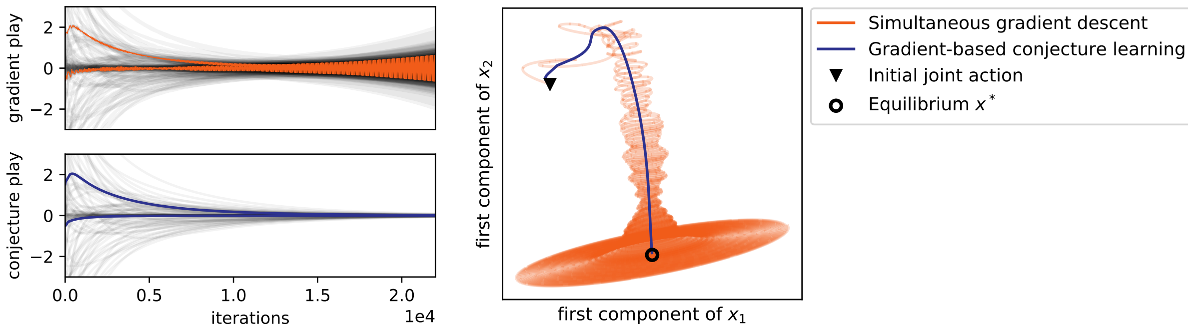

Figure 2: We observe the learning path of simultaneous gradient descent and gradient-based conjecture learning in a zero-sum quadratic game where $x_1 \in \mathbb{R}^{25}$ and $x_2 \in \mathbb{R}^{75}$. We plot the first components of each player's action. Gradient play fails to converge to the Nash equilibrium due to the large imaginary eigenvalues of the dynamics, whereas conjecture play dampens the oscillations around the equilibrium point.

the Hessians are positive definite and small relative to B, we sample a square matrix $P \sim \mathcal{N}(0, 1)$ and set $A_1 = P^\top P/100 + 10^{-6}I$, and similarly for A_2.

We run simultaneous gradient descent and gradient-based conjecture learning with uniform learning rates $\gamma_1 = \gamma_2 = 6.6 \times 10^{-3}$ initialized at $(0, 0)$ and observe their learning paths. Simultaneous gradient descent exhibits oscillation throughout its learning path and fails to converge to the stationary point. The gradient-based conjecture learning dampens these oscillations and converges directly to the general conjectural variations equilibrium.

5. Discussion and Conclusion

The need to optimize multiple objectives in large-scale machine learning has motivated much of the recent work in this area. For instance, a prominent application area is the training of generative adversarial networks, which are a zero-sum game between neural-network agents. As such, some of these updates are well-suited for two player zero-sum games, but provide little flexibility to extend to multiple, heterogeneous, non-zero-sum agents.

Our work develops a first principled framework to synthesize learning rules for multiple interacting objectives where agents may form models of others. We present three types of conjectures: gradient-based conjecture learning, which requires only gradient and Jacobian-gradient computations; implicit conjecture learning, which is a 'Newton-like' gradient conjecture where second order information corrects for poor conditioning and speeds up convergence; and faster conjecture learning, wherein an oracle provides a descent direction for agents to correct against.

The conjecture framework will enable the study and design of novel game dynamics. It will also encourage the study of new equilibria of games, ones that arise from a different choice of conjecture. What other conjectures agents can adopt and what the outcomes of those games are, amongst other questions, will be the topic of future work.

References

[1] David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pages 354–363, 2018.

[2] Tamer Basar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory.* Society for Industrial and Applied Mathematics, 2nd edition, 1998.

[3] Arthur Lyon Bowley. *Mathematical groundwork of economics.* California Press, Oxford, 1924.

[4] Timothy F Bresnahan. Duopoly models with consistent conjectures. *The American Economic Review*, 71(5):934–945, 1981.

[5] Benjamin Chasnov, Lillian J Ratliff, Eric Mazumdar, and Samuel A Burden. Convergence analysis of gradient-based learning with non-uniform learning rates in non-cooperative multi-agent settings. *arXiv preprint arXiv:1906.00731*, 2019.

[6] Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. *arXiv preprint arxiv:1807.03907*, 2018.

[7] Engelbert J Dockner. A dynamic theory of conjectural variations. *The Journal of Industrial Economics*, pages 377–395, 1992.

[8] Tanner Fiez, Benjamin Chasnov, and Lillian J Ratliff. Convergence of learning dynamics in stackelberg games. *arXiv preprint arXiv:1906.01217*, 2019.

[9] Charles Figuières. *Theory of conjectural variations*, volume 2. World Scientific, 2004.

[10] Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[11] Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.

[12] James W Friedman and Claudio Mezzetti. Bounded rationality, dynamic oligopoly, and conjectural variations. *Journal of Economic Behavior & Organization*, 49(3):287–306, 2002.

[13] Drew Fudenberg, Fudenberg Drew, David K Levine, and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

[16] Jun-ichi Itaya and Koji Shimomura. A dynamic conjectural variations model in the private provision of public goods: a differential game approach. *Journal of Public Economics*, 81(1): 153–172, 2001.

[17] Chi Jin, Praneeth Netrapalli, and Michael I Jordan. Minmax optimization: Stable limit points of gradient descent ascent are locally optimal. *arXiv preprint arXiv:1902.00618*, 2019.

[18] John M Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pages 1–31. Springer, 2013.

[19] Alistair Letcher, Jakob Foerster, David Balduzzi, Tim Rocktäschel, and Shimon Whiteson. Stable opponent shaping in differentiable games. In *International Conference on Learning Representations*, 2019.

[20] Tianyi Lin, Chi Jin, and Michael I Jordan. On gradient descent ascent for nonconvex-concave minimax problems. *arXiv preprint arXiv:1906.00331*, 2019.

[21] Thomas Lindh. The inconsistency of consistent conjectures: Coming back to cournot. *Journal of Economic Behavior & Organization*, 18(1):69–90, 1992.

[22] E. Mazumdar, M. Jordan, and S. S. Sastry. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arxiv:1901.00838*, 2019.

[23] Eric Mazumdar and Lillian J Ratliff. Fundamental Issues with Gradient-Based Learning in Continuous Games. *arXiv preprint arXiv:1804.05464*, 2018.

[24] Eric Mazumdar, Lillian J Ratliff, Michael I Jordan, and S Shankar Sastry. Policy-gradient algorithms have no guarantees of convergence in continuous action and state multi-agent settings. *arXiv preprint arXiv:1907.03712*, 2019.

[25] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. The numerics of gans. In *Advances in Neural Information Processing Systems*, pages 1825–1835, 2017.

[26] L. J. Ratliff, S. A. Burden, and S. S. Sastry. On the Characterization of Local Nash Equilibria in Continuous Games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.

[27] Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. Genericity and structural stability of non-degenerate differential Nash equilibria. In *Proceedings of the 2014 American Control Conference*, 2014.

[28] Florian Schäfer and Anima Anandkumar. Competitive gradient descent. *arXiv preprint arXiv:1905.12103*, 2019.

[29] C. Tretter. Spectral inclusion for unbounded block operator matrices. *Journal of Functional Analysis*, 256:3806–3829, 2009.

[30] Chongjie Zhang and Victor Lesser. Multi-agent learning with policy prediction. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

Appendix A. Linking to Common Game-Theoretic Equilibrium Notions

In this section, we draw connections between the Nash and Stackelberg equilibrium concepts in simultaneous and hierarchical play games, respectively.

The solution concept usually adopted in non-cooperative simultaneous play games is the Nash equilibrium. In a Nash equilibrium, the strategy of each player is a best response to the joint strategy of the competitors.

Definition 4 (Nash Equilibrium). *The joint strategy $x^* \in X$ is a Nash equilibrium if for each $i \in \mathcal{I}$,*

$$f_i(x^*) \leq f_i(x_i, x_{-i}^*), \quad \forall\ x_i \in X_i.$$

The strategy is a local Nash equilibrium on $W \subset X$ if for each $i \in \mathcal{I}$,

$$f_i(x^*) \leq f_i(x_i, x_{-i}^*), \quad \forall\ x_i \in W_i \subset X_i.$$

The local Nash equilibrium concept can be stated in terms of sufficient conditions.

Definition 5 (Differential Nash Equilibrium [26]). *The joint strategy $x^* \in X$ is a differential Nash equilibrium if $g(x^*) = 0$ and $D_i^2 f_i(x^*) > 0$ for each $i \in \mathcal{I}$.*

A differential Nash equilibrium is said to be non-degenerate if the determinant of the Jacobian at the equilibrium is non-zero.

On the other hand, the solution concept most commonly adopted in hierarchical play settings is a Stackelberg equilibrium. In a Stackelberg game, the leader aims to solve the optimization problem given by

$$\min_{x_1 \in X_1} \left\{ f_1(x_1, x_2) \middle|\ x_2 \in \arg\min_{y \in X_2} f_2(x_1, y) \right\}$$

and the follower aims to solve the optimization problem $\min_{x_2 \in X_2} f_2(x_1, x_2)$. The equilibrium concept is known as a Stackelberg equilibrium. The solution can be characterized as the intersection points of the reaction curves of the players [2].

Definition 6 (Stackelberg Equilibrium). *In a two-player game with player 1 as the leader, a strategy $x_1^* \in X_1$ is called a Stackelberg equilibrium strategy for the leader if*

$$\sup_{x_2 \in \mathcal{R}(x_1^*)} f_1(x_1^*, x_2) \leq \sup_{x_2 \in \mathcal{R}(x_1)} f_1(x_1, x_2), \quad \forall x_1 \in X_1,$$

where $\mathcal{R}(x_1) = \{y \in X_2 |\ f_2(x_1, y) \leq f_2(x_1, x_2), \forall x_2 \in X_2\}$ is the rational reaction set of x_2.

Sufficient conditions can be stated to characterize a *local* Stackelberg equilibrium strategy for the leader.

Definition 7 (Differential Stackelberg Equilibrium). *The pair $(x_1^*, x_2^*) \in X$ with $x_2^* = \xi(x_1^*)$, where ξ is implicitly defined by $D_2 f_2(x_1^*, x_2^*) = 0$, is a differential Stackelberg equilibrium for the game (f_1, f_2) with player 1 as the leader if $Df_1(x_1^*, \xi(x_1^*)) = 0$, and $D^2 f_1(x_1^*, \xi(x_1^*))$ is positive definite..*

For a game $\mathcal{G}$, we denote the set of differential Nash equilibria as $\texttt{DNE}(\mathcal{G})$, the set of stable GCVE as $\texttt{SGCVE}(\mathcal{G})$, and the set of local Nash equilibria as $\texttt{LNE}(\mathcal{G})$. For a two-player game $\mathcal{G}$, we denote the set of differential Stackelberg equilibria as $\texttt{DSE}(\mathcal{G})$. Also, for a dynamical system $\dot{x} = -g(x)$, we denote the set of locally asymptotically stable equilibria by $\texttt{LASE}(g)$.

A.1 Connections Between Local Nash and General Conjectural Variations Equilibria

Let us start by connecting the notion of a differential Nash equilibrium, a refinement of local Nash, to a stable general conjectural variations equilibrium.

Proposition 9 (Zero-sum $\mathcal{G}$, $\texttt{DNE}(\mathcal{G})\subset \texttt{SGCVE}(\mathcal{G})$)**.** *Consider a zero-sum game $\mathcal{G}=(f,-f)$. All differential Nash equilibria of $\mathcal{G}$ are stable general conjectural variations equilibrium of the gradient-based conjecture learning update with implicitly defined conjectures (ξ_1,ξ_2).*

Proof. Suppose x^* is a differential Nash equilibrium of $\mathcal{G}=(f,-f)$. Then, by definition, $D_1^2f(x^*)>0$, $-D_2f(x^*)>0$, and $D_if(x^*)=0$ for $i=1,2$. Hence,

$$D_if_i(x^*)+D_i\xi_i(x^*)^\top D_jf_i(x^*)=0,\ \ i=1,2,j\neq i$$

where $f_1\equiv f$ and $f_2\equiv -f$ so that x^* is a general conjectural variations equilibrium. To see that it is stable, we compute the Jacobian $J_{\mathtt{i}}(x^*)$ of the implicit conjecture dynamics at x^*:

$$\begin{bmatrix} D_1^2f(x^*)-D_{12}f(x^*)(D_2^2f)^{-1}(x^*)D_{21}f(x^*) & 0\\ 0 & -D_2^2f(x^*)+D_{21}f(x^*)(D_1^2f)^{-1}(x^*)D_{12}f(x^*)\end{bmatrix}$$

where the off-diagonal terms are zero since they reduce to $D_{ij}f(x^*)-D_{ij}f(x^*)$ using the fact that $D_if(x^*)=0$, $i=1,2$ which also implies that the higher-order terms evaluate to zero. Since $D_1^2f(x^*)>0$ and $-D_2^2f(x^*)>0$, both diagonal blocks of $J_{\mathtt{i}}(x^*)$ are positive definite so that the eigenvalues are in the open right-half complex plane. Thus, $x^*\in\texttt{SGCVE}(\mathcal{G})$. $\square$

This result implies that for zero-sum games, stable attractors of the simultaneous gradient-play update which are differential Nash equilibria The result also implies that for *almost all*, in a formal mathematical sense, zero-sum games $\mathcal{G}$, $\texttt{LNE}(\mathcal{G})\subset\texttt{SGCVE}(\mathcal{G})$.

Proposition 10. *Consider any generic zero-sum game $\mathcal{G}=(f,-f)$—i.e., f is a generic function. All local Nash equilibria are stable general conjectural variations equilibrium of the gradient-based conjecture learning update with implicitly defined conjectures (ξ_1,ξ_2).*

The proof follows from the fact that non-degenerate differential Nash equilibria—i.e., differential Nash equilibria x such that the Jacobian $J(x)$ of $g(x)$ has non-zero determinant—are generic amongst local Nash equilibria [27].

Proposition 11 (Zero-sum $\mathcal{G}$, $\texttt{SGCVE}(\mathcal{G})\subset\mathcal{C}(g)$)**.** *For any zero-sum game $\mathcal{G}=(f,-f)$, all stable general conjectural variations equilibria are critical points of g.*

Proof. Suppose not. That is, suppose x is a stable general conjectural variations equilibrium but $D_if(x)\neq 0$, $i=1,2$. Then,

$$\begin{aligned} D_1f(x)-D_{21}f^\top(x)(D_2^2f)^{-1}(x)D_2f(x)&=0\\ -D_2f(x)+D_{12}f^\top(x)(D_1^2f)^{-1}(x)D_1f(x)&=0\end{aligned}$$

implies that

$$-D_2f(x)+D_{12}f^\top(x)(D_1^2f)^{-1}(x)D_{21}f^\top(x)(D_2^2f)^{-1}(x)D_2f(x)=0$$

Since $D_2f(x)\neq 0$,

$$D_{12}f^\top(x)(D_1^2f)^{-1}(x)D_{21}f^\top(x)(D_2^2f)^{-1}(x)=I$$

so that

$$D_2^2f(x)-D_{12}f^\top(x)(D_1^2f)^{-1}(x)D_{21}f^\top(x)=0$$

which contradicts the stability condition for x. $\square$

Furthermore, not all locally asymptotically stable attractors of simultaneous gradient-play are stable general conjectural variations equilibria.

Proposition 12 ($\texttt{LASE}(\mathcal{G}) \not\subset \texttt{SGCVE}(\mathcal{G})$.)**.** *Not all locally asymptotically stable attractors of simultaneous gradient descent are stable general conjectural variations equilibria.*

Proof. To show this result, it is sufficient to consider a zero-sum game $\mathcal{G} = (f, -f)$ defined by cost

$$f(x, y) = \frac{a}{2}x^2 + \frac{1}{2}y^2 + 2xy.$$

Then, the eigenvalues of

$$J(x) = \begin{bmatrix} a & 2 \\ -2 & -1 \end{bmatrix}$$

have positive real parts for any $a \in (1, 4)$. Moreover,

$$D_1^2 f(x) - D_{12} f(x)(D_2^2 f)^{-1}(x) D_{21} f(x) = a - 4 < 0$$

and

$$-D_2^2 f(x) - D_{21} f(x)(D_1^2 f)^{-1}(x) D_{12} f(x) = -1 + \frac{4}{a} > 0$$

Hence, $(x, y) = 0$ is a locally asymptotically stable attractor of simultaneous gradient descent, yet it is not a stable general conjectural variations equilibrium for implicit conjecture learning. $\square$

We note the purpose of using the parameter a in the above is to show that the example is not a degenerate case.

For zero-sum quadratic games, we can show that all attractors of the implicit conjectures update are stable general conjectural variations equilibrium.

Proposition 13. *All attractors of the implicit conjectures update in zero-sum quadratic games $\mathcal{G} = (f, -f)$ are stable general conjectural variations equilibrium.*

Proof. This is trivial to see by simply examining the structure of the Jacobian which is block diagonal. Hence, if x is an attractor of the implicit conjectures update, it has to be a stable general conjectural variations equilibria. $\square$

The dynamics also correspond to a gradient flow and do not possess limit cycles due to the block diagonal structure; this is unlike general sum games where even in the case of simultaneous gradient play for quadratic zero-sum games there may exists non-Nash attractors and limit cycles [23].

For non-quadratic games, this above proposition does not hold since there are higher order terms; however, ignoring terms of order three, the Jacobian is block diagonal. In the case of general-sum games, however, the off-diagonal terms in do not evaluate to zero, even if higher order terms are ignored.

A.2 Connections Between Stackelberg and General Conjectural Variations Equilibria

If one player forms an implicit conjecture about the other while the other assumes a static best-response conjecture—i.e., they assume their opponent repeats its last play—the result is gradient-based learning for Stackelberg games.

Indeed, consider a game $\mathcal{G} = (f_1, f_2)$. Suppose player 1 adopts the implicit conjecture $\xi_1^2(x_1) = \arg\min_{x_2} f_2(x_1, x_2)$ and player 2 adopts $\xi_2^1(x_1, x_2) = x_1$. Then, the players will follow the vector field

$$\begin{bmatrix} D_1 f_1(x) - (D_{21} f_2(x))^\top (D_2^2 f_2(x))^{-1} D_2 f_1(x)) \\ D_2 f_2(x) \end{bmatrix}$$

which is indeed the Stackelberg update as shown in our recent work [8].

Proposition 14. *Differential Stackelberg equilibria are equivalent to stable general conjectural variations equilibria in games where one player adopts an implicit conjecture and the other a static best response conjecture.*

From our previous work [8], we already know several inclusions hold including the following. In zero-sum games $\mathcal{G}$, the differential Nash equilibria are differential Stackelberg equilibria. Proposition 9 extends this result to the case where both players have non-trivial conjectures; said another way, both players act as if they are the leader.

The connections between Stackelberg, Nash and stable general conjectural variations equilibria give us insight into how we might use conjectures to synthesize game dynamics with heterogeneous agents of different hierarchy or internal models of each other.