

# Non-Concave Games: A Challenge for Game Theory's Next 100 Years\*

Constantinos Daskalakis  
EECS and CSAIL, MIT

September 17, 2022

## Abstract

A pervasive assumption in Game Theory is that players' utilities are concave, or at least quasi-concave, with respect to their own strategies. While mathematically instrumental, enabling the existence of many kinds of equilibria in many kinds of settings, (quasi-)concavity of payoffs is too restrictive an assumption. For the same reasons that (quasi-)concave utilities can only go so far in capturing single-agent optimization problems, they can only go so far in modeling the considerations of an agent in a strategic interaction. Besides, the study of games with non-concave utilities is increasingly coming to the fore as Deep Learning ventures into multi-agent learning applications. In this article, we study what types of equilibria exist in such games, and whether they are computationally tractable, proposing paths for Game Theory and multi-agent learning in the next one hundred years.

## 1 A Century of (Quasi-)Concave Games

Convexity plays a central role in optimization. If your optimization problem can be fruitfully posed as a convex minimization — equivalently, a concave maximization — problem, over some convex set, this is a cause for celebration. A large gamut of optimization methods readily becomes available to you. These methods are versatile, accommodating different types of access to your objective and constraints and providing various tradeoffs between quality of optimization and use of computational resources. Additionally, convex programming duality provides a framework for gaining a deeper understanding of your problem.

Convexity has also played a key role in the development of equilibrium theory. Von Neumann's minimax theorem [vN28] establishes equilibrium existence in two-player zero-sum games under the assumption that each player's payoff is concave with respect to their own strategy — equivalently, that their cost is convex with respect to their own strategy. Under this assumption, computing equilibrium is in fact equivalent to convex programming [Dan51, Adl13, BR21], which makes the convex optimization toolkit readily available for equilibrium computation. It also enables a game-theoretic perspective to influence computation theory, and has motivated important developments at the interface of Game Theory, Machine Learning, and Optimization; see e.g. [CBL06, SS12, BC12, Haz16].

The assumption that a player's payoff is concave, or at least quasi-concave, with respect to their own strategy percolates through the foundations of Game Theory and Economic Theory. It is crucial in establishing generalizations of the minimax theorem (see e.g. [Fan53, Sio58]), the existence of Nash equilibrium [Nas50], and the existence of several other equilibrium notions in non-cooperative games; see

---

\*This article was written on the occasion of the Nobel Symposium "One Hundred Years of Game Theory: Future Applications and Challenges," held in Stockholm in December 2021.

e.g. [OR94]. For instance, the following is a general family of non-cooperative games, with coupled strategy constraints and concave utilities, for which a Nash equilibrium is guaranteed to exist.

**Definition 1** (Concave Games [Deb52, Ros65]). *A  $n$ -player continuous concave game has a finite number,  $n$ , of players, indexed by  $1, \dots, n$ . Each player,  $i$ , may choose some strategy  $x_i \in \mathcal{S}_i \subseteq \mathbb{R}^{d_i}$ , where  $d_i$  is the dimensionality of her strategy space, and the joint strategy profile  $(x_1, \dots, x_n)$  is constrained to lie in  $\mathcal{R} \subseteq \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ , where  $\mathcal{R}$  is convex and compact. The payoff to each player,  $i$ , is determined by some function  $u_i : \times_j \mathcal{S}_j \rightarrow [-1, 1]$  of the joint strategy profile  $x = (x_1, \dots, x_n)$ , which is continuous and is also concave in  $x_i$  for all  $x_{-i}$ .*

It is clear that the above definition captures normal form games by taking  $x_i$  to be the mixed strategy of player  $i$ , specializing  $\mathcal{R}$  to be a product of the players' mixed strategy simplices, and taking each player's payoff to be a multi-linear function of the mixed strategies. Thus the following theorem generalizes Nash's.

**Theorem 1** ([Deb52, Ros65]). *A continuous concave game has a Nash equilibrium, namely some strategy profile  $x^* = (x_1^*, \dots, x_n^*) \in \mathcal{R}$  such that, for all  $i$ ,  $u_i(x^*) = \max\{u_i(x) \mid x = (x_i, x_{-i}^*) \in \mathcal{R}\}$ .*

The afore-described family of concave games can, in fact, be expanded to accommodate quasi-concave utilities and even more general coupled constraints over player strategies and still maintain the existence of Nash equilibrium [Har91], but let us skip this generalization here to avoid overloading our notation.

Beyond the study of non-cooperative games, (quasi-)concavity of payoffs is a very common assumption in Economic Theory, and is crucial for many results including showing the existence of a competitive equilibrium in exchange economies; see [AD54, McK54] and the ensuing literature.

Ultimately, (quasi-)concavity of payoffs is key for setting up the fixed point formulations used to show the existence of various kinds of equilibria. It is also unavoidable, in general, as without this property equilibrium existence breaks, even in very simple games. Figure 1a shows the payoff to the column player (a.k.a. the cost to the row player) in the matching pennies game of Table 1, as a function of the Heads

	Heads	Tails
Heads	$(+1/2, -1/2)$	$(-1/2, +1/2)$
Tails	$(-1/2, +1/2)$	$(+1/2, -1/2)$

Table 1: The Matching Pennies Game

probabilities  $x$  and  $y$  of the row and column players respectively. The unique equilibrium of this game is, of course,  $(x^*, y^*) = \left(\frac{1}{2}, \frac{1}{2}\right)$ . In Figure 1b, we see the payoff to the column player in a modified zero-sum game where the row player is additionally rewarded from a high-entropy strategy, while the column player is penalized from a high-entropy strategy. In the resulting game, the column player's payoff is not concave/quasi-concave in  $y$ , and the game has no Nash equilibrium.

While mathematically instrumental, enabling the existence of many kinds of equilibria in many kinds of settings, concavity or quasi-concavity of payoffs is unfortunately too restrictive an assumption. For the same reasons that (quasi-)concave utilities can only go so far in modeling single-agent decision making, they can also go so far in modeling the considerations of an agent in a strategic interaction. Motivated by considerations such as concave production costs arising from economies of scale, the study of exchange economies with violations of (quasi-)concave utilities/(quasi-)convex costs has received some, but not much study, since the early days; see e.g. [Gra57, Bat57, Koo57, Far59, Rot60, Sta69, PW11]. Similarly, non-cooperative games with non-(quasi-)concave utilities have also received some, but not much study in recent works; see e.g. [RBS16]. Ultimately, the lack of progress on these fronts is not because of lack of interest in non-concave settings, but because these are incompatible with equilibrium existence. And without guaranteed existence of equilibria, the value of equilibrium analysis gets diminished.



Figure 1: On the left, we see the payoff to the column player (a.k.a. the cost to the row player) in the matching pennies game of Table 1 as a function of the Heads probabilities  $x$  and  $y$  of the row and column players respectively. The unique Nash equilibrium is  $(x^*, y^*) = (1/2, 1/2)$ . On the right, we see the payoff to the column player in a modified zero-sum game where the row player is additionally rewarded for a high entropy mixed strategy, while the column player is penalized for a high entropy mixed strategy, resulting in the column player receiving payoff  $(x + y - 2xy - \frac{1}{2}) - \frac{1}{2}H(x) - \frac{1}{2}H(y)$ , where  $H$  is the entropy function. This function is not concave/quasi-concave in  $y$  and the game has no Nash equilibrium.

## 2 The Dawn of Non-Concave Games

### 2.1 Embracing Non-Convexity in Machine Learning

Considerations of mathematical elegance, sharpness of prediction and computational tractability had steered models of single-agent decision-making in the safe harbor of convex programming formulations for decades. Yet, the avoidance of non-convexity has recently landed on its face, with the advent of Deep Learning, and the breakthroughs it has delivered in many heretofore impenetrable problems in the field of Artificial Intelligence.

Deep Learning has not shied away from formulating problems as non-convex optimization problems. It has also not shied away from targeting approximate locally optimal, as opposed to globally optimal, solutions for these problems. While the latter are intractable, the former are easily attainable via lightweight optimization methods such as gradient descent; see e.g. [GHJY15, LSJR16, AZL18]. Whether, why and under what conditions locally optimal solutions selected by gradient descent-based methods result in good models is still not fully understood despite extensive study. Yet, Deep Learning’s attitude of embracing non-convexity has led the field to groundbreaking advances in important learning challenges such as speech and image recognition, text generation, protein folding, automated translation, and more.

In recent years, Deep Learning has been rapidly expanding its scope to the domain of Game Theory, energizing the importance of studying *non-concave games*. Indeed, many outstanding challenges in this field, such as training deep neural networks that are robust to adversarial attacks, training Generative Adversarial Networks (GANs), and Multi-Agent Reinforcement Learning (MARL) are born as multi-player games with utility functions that are not concave (or even quasi-concave) in players’ strategies, as players in these games choose parameters in deep neural networks and their utilities are functions of these neural networks. It seems that applications of non-concave games will explode in a near future where complex agents using deep neural network models learn, make decisions, and receive rewards in a shared environment, whose state might also be affected by their decisions.

Unfortunately, embracing non-convexity hits a wall on the multi-agent front. It hits a wall as soon as we step away from single-agent settings, considering how to train two agents with non-concave payoffs in a zero-sum interaction. Indeed, as illustrated earlier with the simple example of Figure 1b, such games

may not have Nash equilibria. Without much clarity about the target solution concept, practitioners have still been trying to train agents in non-concave games by having them perform gradient ascent procedures in tandem to maximize their individual payoffs (equivalently gradient descent procedures to minimize their individual losses). Unlike single-agent settings, however, gradient descent is not effective in such settings. It commonly exhibits unstable, oscillatory or divergent behavior and the quality of the solutions encountered in the course of training can be poor; see e.g. [Goo16, MPPSD17, DISZ18, MGN18, DP18, MRS20, ADLH19]. For example, training GANs, which are formulated as two-player zero-sum non-concave games, is, at this point, more engineering than science, with rounds and rounds of fine-tuning before they can be reasonably trained, and, even when they appear to be reasonably trained, we lack formal guarantees about the quality of samples they produce — on the contrary, they commonly fail to pass basic statistical tests of quality; see e.g. [ARZ18]. The emerging importance of non-concave games and the practical frustration with solving them is the starting point of our investigation.

**Question 1.** *Is there a theory of non-concave games? What solution concepts are meaningful, universal, and tractable?*

## 2.2 Non-Concave Games and Local Nash Equilibria

Motivated by the emerging applications of non-concave games in Machine Learning, such as those discussed in the previous section, we propose a formal model of continuous games, dropping the concavity assumption from Definition 1 and adding differentiability, in accordance with the types of games arising in those applications.<sup>1</sup> We'll focus our attention to simultaneous games, although sequential games are also well-motivated and can be studied through a lens similar to the one used in our exposition; for some recent pursuits in this direction see e.g. [JNJ20, FCR20, MV21].

**Definition 2** (Differentiable/Smooth Games). *A  $n$ -player differentiable game is a simultaneous game with  $n$  players, indexed by  $1, \dots, n$ . Each player,  $i$ , may choose some strategy  $x_i \in \mathcal{S}_i \subset \mathbb{R}^{d_i}$ , where  $d_i$  is the dimensionality of her strategy space, and the joint strategy profile  $(x_1, \dots, x_n)$  is constrained to lie in  $\mathcal{R} \subset \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ , where  $\mathcal{R}$  is convex and compact. The payoff to each player,  $i$ , is determined by some function  $u_i : \times_j \mathcal{S}_j \rightarrow [-1, 1]$  of the joint strategy profile  $x = (x_1, \dots, x_n)$ , whose gradient with respect to  $x_i$  is continuous. We will refer to  $\sum_i d_i$  as the dimensionality of the game. Moreover, we will call a differentiable game smooth if the gradient of  $u_i$  with respect to  $x_i$  is  $L_i$ -Lipschitz with respect to the  $\ell_2$  norm. In this case, we will refer to  $\max_i L_i$  as the game's smoothness.*

As illustrated in Figure 1b, in the absence of concave utilities, a game may not have any Nash equilibrium. It can also be shown that checking if a game has one is NP-hard, even if the game is two-player zero-sum [DSZ21]. In accordance with the turn to local optimality driving recent advances in Deep Learning, we also advocate a turn to a notion of local Nash equilibrium, which we will argue exists in every differentiable game.

**Definition 3** (Local Nash Equilibrium). *Consider a differentiable game as in Definition 2. A strategy profile  $x^* = (x_1^*, \dots, x_n^*) \in \mathcal{R}$  is a (first-order) local Nash equilibrium iff for all  $i$ :*

$$x_i^* = \Pi_{\mathcal{R}(x_{-i}^*)}(x_i^* + \nabla_{x_i} u_i(x_i^*; x_{-i}^*)), \quad (1)$$

where  $\mathcal{R}(x_{-i}^*) = \{x_i \mid (x_i; x_{-i}^*) \in \mathcal{R}\}$ , and  $\Pi_{\mathcal{R}(x_{-i}^*)}(\cdot)$  is the  $\ell_2$  projection onto the set  $\mathcal{R}(x_{-i}^*)$ . A strategy profile  $x^* = (x_1^*, \dots, x_n^*) \in \mathcal{R}$  is an  $\varepsilon$ -approximate local Nash equilibrium iff for all  $i$ :

$$\left\| x_i^* - \Pi_{\mathcal{R}(x_{-i}^*)}(x_i^* + \nabla_{x_i} u_i(x_i^*; x_{-i}^*)) \right\|_2 \leq \varepsilon. \quad (2)$$

<sup>1</sup>Some families of Deep Neural Networks are non-differentiable at a measure zero set of points. We could extend our exposition here to accommodate this by using sub-differentials. To avoid the unnecessary complexity that this would entail, we assume differentiability everywhere. Moreover, as noted earlier, we may consider even more general coupled constraints over player strategies. Our results can be extended to the types of coupled strategies considered in [Har91] but we skip such generalizations to avoid overloading our notation.

Intuitively, a first-order local Nash equilibrium  $x^*$  comprises strategies that satisfy a weak form of local optimality, assuming that each player has a myopic view of their utility as a function of their own strategy, consistent with a first-order Taylor approximation. This assumption fits well with the type of function access (through value and gradient value queries) that is practically available to agents in complex settings such as those motivating our discussion. When the players have such a myopic view of their payoffs, Condition (1) captures that player  $i$  is playing a best response to her opponents' strategies. Accordingly, Condition (2) captures that the player is playing an approximate best response. While the players would not be able to tell from first-order approximations to their utilities, their local Nash equilibrium strategies may, of course, be suboptimal for their true utilities. The following proposition characterizes how suboptimal players' utilities might be at an (approximate) local Nash equilibrium when considering deviations within a small ball around the local Nash equilibrium strategies. It says that the amount by which the utility might increase under small deviations is second-order, namely deviations at small distance  $\delta$  from a local Nash equilibrium can only give  $\frac{L}{2}\delta^2$  payoff improvements, where  $L$  is the smoothness of the game, i.e. a bound on the Lipschitz constant of the gradients of the payoffs. Deviations around an  $\varepsilon$ -approximate local Nash equilibrium can benefit by an additional  $\varepsilon\delta$ .

**Proposition 1** (Local Payoff Near-Optimality). *Consider a smooth game such that, for all  $i$ ,  $\nabla_{x_i} u_i(x_i; x_{-i})$  is  $L$ -Lipschitz continuous with respect to the  $\ell_2$  norm. If  $x^* = (x_1^*, \dots, x_n^*)$  is a local Nash equilibrium of this game, then for all  $i$ :*

$$u_i(x^*) \geq u_i(x_i; x_{-i}^*) - \frac{L}{2} \|x_i - x_i^*\|_2^2, \text{ for all } x_i \in \mathcal{R}(x_{-i}^*).$$

If  $x^*$  is an  $\varepsilon$ -approximate local Nash equilibrium of the game, then for all  $i$ :

$$u_i(x^*) \geq u_i(x_i; x_{-i}^*) - \varepsilon \|x_i - x_i^*\|_2 - \frac{L}{2} \|x_i - x_i^*\|_2^2, \text{ for all } x_i \in \mathcal{R}(x_{-i}^*).$$

Importantly, local Nash equilibria exist, under the assumption that the gradients of the payoff functions are continuous, as per the following proposition.<sup>2</sup>

**Proposition 2** (Existence of Local Nash). *Every differentiable game has at least one local Nash equilibrium.*

*Proof of Proposition 2:* Consider the mapping  $F : \mathcal{R} \rightarrow \mathcal{R}$  defined as follows:

$$x \mapsto \Pi_{\mathcal{R}} \left( \begin{bmatrix} x_1 + \nabla_{x_1} u_1(x) \\ \vdots \\ x_n + \nabla_{x_n} u_n(x) \end{bmatrix} \right).$$

The mapping is continuous because, for all  $i$ , it is assumed that  $\nabla_{x_i} u_i(x)$  is continuous, and because the  $\ell_2$  projection to a convex and compact set is continuous. Given the continuity of  $F$  and the convexity and compactness of  $\mathcal{R}$ , it follows from Brouwer's fixed point theorem that  $F$  has a fixed point  $x^* = F(x^*)$ .

We will argue that, for all  $i$ , Eq. (1) holds for the fixed point  $x^*$  of  $F$ . Indeed, denote by

$$z^* = \begin{bmatrix} x_1^* + \nabla_{x_1} u_1(x^*) \\ \vdots \\ x_n^* + \nabla_{x_n} u_n(x^*) \end{bmatrix}.$$

<sup>2</sup>We note that the local Nash equilibrium condition (1) can be written as a quasi-variational inequality problem [BJL73] so a proof of Proposition 2 can alternatively be pursued by appealing to existence theorems for quasi-variational inequalities; see e.g. [CP82, Har91]. We stick to our proof here using Brouwer's fixed point theorem as it is simple and self-contained.

Because  $x^* = \Pi_{\mathcal{R}}(z^*)$ , we have by the properties of the  $\ell_2$  projection that

$$\langle z^* - x^*, x^* - y \rangle \geq 0, \forall y \in \mathcal{R}.$$

For an arbitrary  $x_i \in \mathcal{R}(x_{-i}^*)$ , plugging  $y = (x_i; x_{-i}^*)$  into the above inequality we get that:

$$\langle \nabla_{x_i} u_i(x^*), x_i^* - x_i \rangle \geq 0.$$

Thus, we have

$$\langle \nabla_{x_i} u_i(x^*), x_i^* - x_i \rangle \geq 0, \forall x_i \in \mathcal{R}(x_{-i}^*),$$

which is tantamount to Eq. (1).  $\square$

We remark that the development of this section assumes that players are myopic, having a first-order understanding of their payoffs around the equilibrium, which is motivated by the value and gradient value access to the payoffs that is practically available in the Deep Learning applications motivating our discussion. The local Nash equilibrium concept of Definition 3 reflects that such myopic players are best-responding. If the players have a richer local understanding of their utilities, e.g. if they have higher-order access to their payoffs, it would be appropriate to incorporate that richer player knowledge into refined local Nash equilibrium concepts. We note, however, that improperly doing so might break equilibrium existence even at the second order. This can be seen by considering a two-player zero-sum game, where the row chooses  $x \in [0, 1]$ , the column player chooses  $y \in [0, 1]$  and the column's payoff is:  $(x + y - 2xy - \frac{1}{2}) - \frac{1}{2}x^2 - \frac{1}{2}y^2$ . In this game, the second-order view of the utilities incurs no information loss. So second-order local Nash equilibria correspond to Nash equilibria, but this game has no Nash equilibria!

### 3 The Complexity of Local Nash Equilibria

As discussed in the previous section, first-order local Nash equilibria are guaranteed to exist in all differential games, making them a reasonable first cut at a solution concept. We say "first cut" as one might be dissatisfied with the first-order best-response nature of the solution concept. As discussed at the end of last section, one might want to incorporate richer understanding that players might have about their utilities to remove some first-order local Nash equilibria from consideration, ideally not eliminating them all. Regardless of what equilibrium refinement one might want to pursue, it is reasonable to ask at this point whether the weak solution concept of first-order local Nash equilibrium, which is guaranteed to exist, is also tractable, as any refinement would be at least as hard to compute.

**Question 2.** *Are local Nash equilibria tractable?*

As noted in Section 2.1, practical experience with non-concave games considered in Deep Learning applications reveals that gradient descent-based methods have a real challenge converging, necessitating hyper-optimized training frameworks to get any reasonable performance. It is justified to wonder if there is some lingering intractability explaining this phenomenon. It turns out that this is indeed the case.

**Theorem 2** ([DSZ21]). *Any method, which accesses players' payoffs in a smooth differentiable game through value queries and gradient value queries, needs a number of queries that is exponential in at least one of  $1/\epsilon$ , the smoothness of the game, or the dimensionality of the game to compute an  $\epsilon$ -approximate first-order local Nash equilibrium. This is true even if the game is a two-player zero-sum game.*

This result is in sharp contrast to single-agent non-concave maximization (a.k.a. the restriction of Definition 3 to  $n = 1$ ) for which a polynomial number, in  $1/\epsilon$  and the smoothness, queries to the objective are sufficient to compute first-order approximate local maxima. In fact, a stronger intractability result actually

holds subject to a widely held complexity-theoretic assumption. Unless the complexity class PPAD<sup>3</sup> collapses to the class P of polynomial-time solvable problems, no method can compute local Nash equilibria in polynomial time, even if it has exact knowledge of the players' payoffs (rather than just first-order query access to them).

**Theorem 3** ([DSZ21]). *Unless  $P = \text{PPAD}$ , no method can compute an  $\varepsilon$ -approximate first-order local Nash equilibrium of a smooth differentiable game in time polynomial in  $1/\varepsilon$  and the game's smoothness and dimensionality. In particular, the problem is PPAD-complete. This is true even if the game is a two-player zero-sum game.*

In other words, computing local Nash equilibria in smooth differentiable games, even in two-player zero-sum ones, is exactly as hard as computing Brouwer fixed points of Lipschitz functions and computing mixed Nash equilibria in general-sum normal-form games [DGP09, CDT09], and is at least as hard as any other problem in the class PPAD. For a discussion of PPAD, its relationship to complexity theory and the complexity of topological problems see e.g. [Das18].

**Intuition for Intractability.** It is worth providing some intuition about the intractability results of this section, and compare to non-convex minimization/non-concave maximization where similar intractability does not arise.

**Question 3.** *Why is it that local optima in optimization problems are tractable, but local Nash equilibria in multi-player non-concave games, even in two-player zero-sum ones, are intractable?*

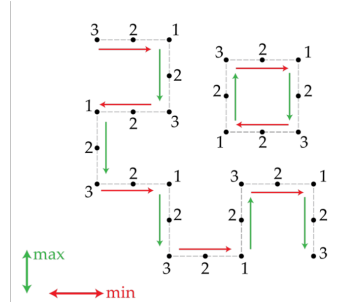
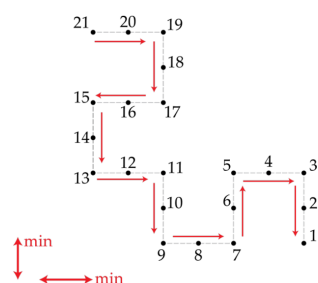
To answer this question, it is illustrative to compare differentiable two-player *common-interest* games, in which two players share an objective that they both want to minimize, to differentiable two-player *zero-sum* games, in which one player wants to minimize an objective that the other player wants to maximize. Approximate local Nash equilibria in a common-interest game are the same as approximate local minima of the shared cost function, and these are tractable as noted above. On the other hand, approximate local Nash equilibria in a zero-sum game are intractable as per the above theorems. Why are local Nash equilibria in these two families of games so much different in complexity?

To shed some light on this, in Figure 2a we show how a better-response sequence of moves may look like in a two-player common-interest game, where one player controls moves along the horizontal axis, the other player controls moves along the vertical axis, and both players want to minimize a shared cost function. Because the game is common-interest, it is clear that the shared cost must decrease along a better response sequence. Indeed, a horizontal move only makes sense for the horizontal player if it decreases the shared cost, and a vertical move only makes sense if it decreases the shared cost. Thus, a better response sequence makes progress towards a local minimum. In particular, if we normalize the shared cost function to lie in  $[-1, 1]$  as in Definition 2, and only take steps that improve the shared cost by at least  $\varepsilon$ , then the number of steps until we find an  $\varepsilon$ -approximate local minimum is  $O(1/\varepsilon)$ .

In zero-sum games, however, the situation is much trickier. First, a sequence of better-response moves may actually be cyclic, as shown at the top right of Figure 2b. Second, the payoff values along a better response sequence may repeat. As a result, payoff values along a better-response sequence do not reveal whether or not the sequence is cyclic. And, even if it is non-cyclic, we gain no information, by querying the payoff function along the sequence, about how far from the end of the sequence the query was made.

This apparent lack of much information gain about the location of local Nash equilibria when querying the payoff function of a zero-sum game lies at the heart of what makes local Nash equilibria so much harder than local minima/local maxima. To turn this intuition into an intractability proof what we would like to do is to hide an exponentially long, better-response path with recycled payoff values, like the one

<sup>3</sup>The complexity class PPAD was introduced in [Pap94] and used in [DGP09, CDT09] to characterize the complexity of computing a mixed Nash equilibrium in normal-form games. See also [Das18] for a recent survey of PPAD, its uses and relationship to problems in Topology, Combinatorics, Optimization and Game Theory.



(a) A better-response path in a common-interest game. (b) A better-response path and a better-response cycle in a zero-sum game.

Figure 2: An illustration of better-response move sequences in a two-player common-interest game (on the left) and a zero-sum game (on the right), where one player is choosing strategies on the horizontal axis and the other player is choosing strategies on the vertical axis. In both cases, the values shown represent the cost (i.e. minus the payoff) of the horizontal player. The vertical player shares the same cost on the left, and has the opposite cost on the right.

shown in Figure 2b, within some ambient space so that querying the payoff function along the path or in the ambient space provides very little information about the location of local Nash equilibria. However, we do not see how to implement such a construction directly, but use the complexity-theoretic machinery of PPAD, establishing Theorem 3 first, and exploiting special properties of that construction to obtain Theorem 2 as a corollary. See [DSZ21] for the details.

## 4 Philosophical Corollary and Ways Forward

Motivated by emerging, multi-agent applications in Deep Learning, we discussed the importance of non-concave games, and specifically those of the differentiable kind, captured by Definition 2. As we noted, these games may have no Nash equilibria, and this prompted us to study what might be plausible solution concepts to target in these games. Motivated by the practical consideration that agents have restricted access to their payoff functions, typically via value and gradient value queries, in the complex settings motivating our study, we proposed the notion of first-order local Nash equilibrium of Definition 3, which are points where all players are best-responding to their opponents' strategies, as far as the first-order Taylor approximation to their payoff functions can tell.

Local Nash equilibria are a natural generalization to the game-theoretic setting of local optima, whose computation in single-agent problems has been driving the spectacular progress that Deep Learning has made in the past decade. Importantly, local Nash equilibria are guaranteed to exist, and their universality is appealing to theorists and practitioners alike.

Unfortunately, Theorems 2 and 3 of the previous section, preclude the existence of some variant of gradient descent — or subject to the widely-held complexity-theoretic assumption that  $P \neq PPAD$  the existence of any algorithm — computing local Nash equilibria efficiently in general differentiable games.

In the balance, our results establish the existence of a methodological roadblock in extending the Deep Learning to multi-agent settings. While postulating a complex parametric model for a learning agent and training this model using gradient descent has been a successful framework in single-agent settings, choosing complex models for a bunch of different agents and having them train their models via competing gradient descent-based procedures is just not going to work, even in two-agent zero-sum settings, and even if one were to compromise with the solution concept of first-order local Nash equilibrium.



Our work advocates that progress in multi-agent learning will be unlocked by modeling and methodological insights, as well as better clarity about the types of solutions that are useful and attainable, obtained by combining the perspectives of Game Theory and Economics with those of Machine Learning and Optimization. We conclude with some musings on where to go from here.

#### 4.1 Asymptotic Local Nash Equilibrium Convergence in General Non-Concave Games

While local Nash equilibria are intractable in general games, it is still worth identifying iterative procedures whose steps are efficiently computable, and which exhibit guaranteed asymptotic convergence to local Nash equilibrium. Since local Nash equilibria are fixed points of a continuous map — see proof of Proposition 2, one may try to obtain such procedures by translating to the setting of differentiable games, and optimizing the steps of iterative fixed point computation algorithms. We have recently pursued this avenue of research, proposing second-order methods which exhibit guaranteed asymptotic convergence to local Nash equilibria in smooth differentiable games [DGSZ22]. Our work advances a recent line of work proposing methods that attain local convergence [JNJ20, FCR20, FR20, WZB20]. It is important to identify beyond-worst-case assumptions under which asymptotically convergent methods become computationally efficient, a direction that is underexplored.

#### 4.2 Non-Concave Games with Tractable Equilibria, Stochastic Games and MARL

The intractability results that we presented are worst-case. It is plausible that vast families of non-concave games have tractable local Nash equilibria, or even global Nash equilibria. If those families are expressive enough to capture broad types of multi-agent interactions, they would go a long way to alleviate the impact of our intractability results.

On this front, a prominent family of games to study are *stochastic games*. These are games where two or more agents interact and receive payoffs over multiple stages, or indefinitely, in some environment whose state gets updated at every stage depending on the actions chosen by the agents, whose payoff functions are also state-dependent. These games, defined in Shapley’s foundational work [Sha53], have received broad study in Game Theory literature — see e.g. [SV15] for a recent review, and have found many applications due to their generality and versatility — see e.g. [Ne03]. Among those, of prominent importance is becoming their application to *multi-agent reinforcement learning* (MARL) — see e.g. [Lit94, HW03, BBDS08, ZYB21]. In this application, the stochastic game’s state transitions and payoffs are not known to the players a priori so the equilibrium has to be *learned* through interaction.

Since Shapley’s work, a compelling notion of equilibrium considered in stochastic games are stationary Markov Nash equilibria, i.e. Nash equilibria in strategies (a.k.a. policies) that depend on the current state of the game but not on the stage count or the history of play so far — hence both *stationary* and *Markovian*. In terms of their stationary Markov strategies, player utilities in a stochastic game are non-concave. Yet, under broad conditions, e.g. future payoff discounting and a finite number of states, stationary Markov Nash equilibria exist; see e.g. [Sha53, Tak62, Fin64] and an overview of such results in [SV15]. In fact, in the two-player zero-sum case, it is even known that the minimax theorem holds with respect to stationary Markov strategies [Sha53].

On the computation front, many works have studied the challenge of computing or learning Nash equilibria in stochastic games. Most of the focus has been on efficient computation and learning of (approximate) Nash equilibria in two-player zero-sum stochastic games; see e.g. [BT02, WHL17, XCWY20, ZKBY20, SWYY20, BJ20, DFG20, BJY20, LYBJ21, TWYS21] and their references. Some of these works provide decentralized learning procedures via which the players of a zero-sum stochastic game efficiently converge to approximate stationary Markov minimax equilibria. Those learning procedures are decentralized in that the players are not allowed to communicate but only interact in the game by each employing some strategy and observing their own sequence of rewards and the game’s state transitions.

Beyond these results, however, our understanding is lagging. First, there is still mystery surrounding whether exact stationary Nash equilibria of two-player zero-sum stochastic games can be computed efficiently when the discount factor is very close to 1 [Con92, DP11]. Beyond the two-player zero-sum case, stochastic games have intractable equilibria, since they are more expressive than normal-form games [DGP09, CDT09]. Can the weaker solution concept of (coarse) correlated equilibrium be computed/learned efficiently, and what kinds of correlation among agents’ strategies allow for efficient computation?

Some recent works have studied the computation and learning of (coarse) correlated equilibria in multi-player stochastic games. These works target *non-stationary* (coarse) correlated equilibria, which are easy to compute using backwards induction if the game is known. So they provide decentralized algorithms for *learning* approximate non-stationary (coarse) correlated equilibria. However, these works either require exponential time, in the number of players, to converge, or they compute equilibria that are *non-Markovian* [LYBJ21, SMB22, JLWY21, MB22].

In very recent work [DGZ22], we address some of the outstanding questions about (coarse) correlated equilibrium computation and learning in these games. We show that focusing on non-stationary equilibria is in general unavoidable, as stationary Markov coarse correlated equilibria are intractable, namely PPAD-hard, even when the game has two players and the discount factor is  $1/2$ , i.e. the game is expected to last for two rounds.<sup>4</sup> This is quite surprising as coarse correlated equilibrium computation in normal-form games is very tractable. Complementing our intractability results, we also provide decentralized procedures for efficiently learning non-stationary Markov coarse correlated equilibria, whose running time is polynomial in the number of players.

On the front of (coarse) correlated equilibrium learning in stochastic games, there is a lot more work to be pursued. First, it is important to provide simpler alternatives to the methods obtained in recent work. Moreover, in many games, the state/action spaces are infinite but we are interested in strategies that are expressible compactly by a policy in some class of functions, which map states to (distributions over) actions taken at those states. What types of equilibria exist and how is their computational and learning complexity affected by the choice of the policy class? We will touch upon this kind of questions in the next section but specialized work should be done for this important class of games.

Finally, it is important to study the afore-described problems of learning and computation of equilibria in stochastic games with incomplete information. Their special case of extensive-form games have played a central role in recent practical developments in human-level Poker-playing algorithms [BS18, BS19]. In these games, it is known how to compute Nash equilibria in the two-player zero-sum case [KMVS96, vS07]; Nash equilibria are intractable in the general case [DGP09, CDT09]; and there is an active line of investigation on different notions of correlated equilibrium [vSF08] and its computation and learning; see e.g. [HvS08, CMFG20] and their references. It is worth pursuing these directions further, and generalizing our existing understanding to general stochastic games with incomplete information.

### 4.3 Convexification and Global Equilibria

In many settings, the strategies available to players in a differentiable game correspond to a vector of parameters indexing a function from some function class. In Deep Learning applications, these could be the parameters of some Deep Neural Network model. In multi-agent reinforcement learning, discussed in the previous section, these could be indexing a parametric policy in a stochastic game with an infinite

---

<sup>4</sup>See also [JMS22] for a weaker version of this claim. We also note that both [DGZ22] and [JMS22] show intractability results for a type of stationary Markov coarse correlated equilibrium wherein the correlation of players’ actions is independent across stages and states of the game, and moreover the recommendations from the correlation device are revealed to the players step-by-step, as they reach a new state at each stage of the game, similar to the notion of extensive-form correlated equilibrium proposed in [vSF08] for extensive-form games. The complexity of the alternative solution concept of strategic-form coarse correlated equilibrium in stationary Markov strategies still remains open.

number of states and actions.

Making this perspective explicit, we might consider the following generalization of differentiable games. Each player,  $i$ , may choose a function from some (potentially infinite and potentially non-parametric) function class  $\mathcal{F}_i$ , or a distribution over elements of  $\mathcal{F}_i$ , and the payoff to each player for a given pure strategy profile is determined by some function  $u_i : \times_j \mathcal{F}_j \rightarrow \mathbb{R}$ . Since these games generalize differentiable games, pure Nash equilibria are not guaranteed to exist, motivating our interest in local pure Nash equilibria, as discussed earlier in this paper. If we do not want to give in to local optimality, we might alternatively turn to convexification approaches, considering mixed Nash or correlated equilibria in these games. Under what conditions are mixed Nash or correlated equilibria guaranteed to exist in infinite games? And under what conditions are they efficiently learnable?<sup>5</sup>

A recent line of work [RST15, HLM21, DG22] makes progress on these questions by developing equilibrium existence results and no-regret online learning algorithms converging to equilibrium, under necessary conditions about the complexity of the game’s strategy sets and utility functions.<sup>6</sup> Under these assumptions, it has been shown that the minimax theorem holds in the two-player zero-sum setting, and approximate coarse correlated equilibria (of arbitrary approximation) exist in the multi-player/general-sum setting. Moreover, decoupled online learning procedures have been obtained that the players can employ to converge to these kinds of equilibria, or to guarantee small regret against worst-case adversaries. Naturally, the rates of convergence depend on the complexity of the function classes, but their time-dependence in the context of games can match the fast rates known in finite games [DG22].

Going forward, it is also important to obtain conditions under which correlated equilibria and Nash equilibria exist. Especially for the latter it seems that entirely different techniques might be needed. Finally, it is important to identify settings where the convergence rates to equilibrium, in terms of the complexity of the game’s strategy sets and utility functions, can be improved.

## Acknowledgements

This work was supported by NSF Awards CCF-1901292, DMS-2022448 and DMS-2134108, a Simons Investigator Award, the Simons Collaboration on the Theory of Algorithmic Fairness, a DSTA grant, and the DOE PhILMs project (DE-AC05-76RL01830). We thank Noah Golowich and Manolis Zampetakis for providing feedback on drafts of this paper.

## References

- [AD54] Kenneth J. Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, 1954.
- [Ad13] Ilan Adler. The equivalence of linear programs and zero-sum games. *International Journal of Game Theory*, 42(1):165–177, 2013.
- [ADLH19] Leonard Adolphs, Hadi Daneshmand, Aurelien Lucchi, and Thomas Hofmann. Local saddle point optimization: A curvature exploitation approach. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

---

<sup>5</sup>We note that in many applications the strategies available to players are not necessarily from a compact set, thus we cannot apply known results to show that a mixed Nash equilibrium exists [Gli52], and even show that a finitely supported Nash equilibrium exists if the compactly supported game has extra structure [DKS50, Par06, SOP08].

<sup>6</sup>Generalizing [RST15, HLM21], in [DG22], we assume, roughly speaking, that, for each player  $i$ , the function class  $\mathcal{F}_i^{u_i} = \{\times_{j \neq i} \mathcal{F}_j \ni f_{-i} \mapsto u_i(f_i; f_{-i}) \mid f_i \in \mathcal{F}_i\}$  has bounded sequential fat-shattering dimension at all scales. This is a necessary assumption to avoid the two-player zero-sum game wherein whoever says the largest number receives one dollar from the other player, which has no  $\varepsilon$ -approximate mixed Nash equilibria for any  $\varepsilon < 1$  [HLM21].

- [ARZ18] Sanjeev Arora, Andrej Risteski, and Yi Zhang. Do GANs learn the distribution? some theory and empirics. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, 2018.
- [AZL18] Zeyuan Allen-Zhu and Yuanzhi Li. Neon2: Finding local minima via first-order oracles. *Proceedings of the 32nd Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [Bat57] Francis M Bator. The simple analytics of welfare maximization. *The American Economic Review*, 47(1):22–59, 1957.
- [BBDS08] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008.
- [BC12] Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Found. Trends Mach. Learn.*, 5(1):1–122, 2012.
- [BJ20] Yu Bai and Chi Jin. Provable self-play algorithms for competitive reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.
- [BJL73] Alain Bensoussan and Lions Jacques-Louis. Nouvelle formulation de problèmes de contrôle impulsif et applications. *Comptes Rendus des Séances de l’Académie des Sciences*, 276:1189–1192, 1973.
- [BJY20] Yu Bai, Chi Jin, and Tiancheng Yu. Near-Optimal Reinforcement Learning with Self-Play. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [BR21] Benjamin Brooks and Philip J Reny. A Canonical Game – Nearly 75 Years in the Making – Showing the Equivalence of Matrix Games and Linear Programming. *Available at SSRN 3851583*, 2021.
- [BS18] Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.
- [BS19] Noam Brown and Tuomas Sandholm. Superhuman AI for multiplayer poker. *Science*, 365(6456):885–890, 2019.
- [BT02] Ronen I. Brafman and Moshe Tennenholtz. R-MAX – A General Polynomial-Time Algorithm for Near-Optimal Reinforcement Learning. *Journal of Machine Learning Research*, 3(Oct):213–231, 2002.
- [CBL06] Nicolò Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [CDT09] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM (JACM)*, 56(3):1–57, 2009.
- [CMFG20] Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [Con92] Anne Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

- [CP82] D. Chan and J. S. Pang. The generalized quasi-variational inequality problem. *Mathematics of Operations Research*, 7(2):211–222, 1982.
- [Dan51] George B. Dantzig. A proof of the equivalence of the programming problem and the game problem. *Koopmans, T. C., editor(s), Activity Analysis of Production and Allocation*, 1951.
- [Das18] Constantinos Daskalakis. Equilibria, Fixed Points, and Computational Complexity - Nevanlinna Prize Lecture. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 147–209. World Scientific, 2018.
- [Deb52] Gerard Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences (PNAS)*, 38(10):886–893, 1952.
- [DFG20] Constantinos Daskalakis, Dylan Foster, and Noah Golowich. Independent policy gradient methods for competitive reinforcement learning. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [DG22] Constantinos Daskalakis and Noah Golowich. Fast Rates for Nonparametric Online Learning: From Realizability to Learning in Games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2022.
- [DGP09] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
- [DGSZ22] Constantinos Daskalakis, Noah Golowich, Stratis Skoulakis, and Manolis Zampetakis. Sperner Continuous Dynamics: Guaranteed Convergence to Local Minimax Equilibrium in Nonconvex-Nonconcave Games. *arXiv preprint*, 2022.
- [DGZ22] Constantinos Daskalakis, Noah Golowich, and Kaiqing Zhang. The Complexity of Markov Equilibrium in Stochastic Games. *arXiv preprint arXiv:2204.03991*, 2022.
- [DISZ18] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with Optimism. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, 2018.
- [DKS50] M. Dresher, S. Karlin, and L. S. Shapley. Polynomial games. In *Contributions to the Theory of Games I*, volume 24, pages 161–180. Princeton University Press Princeton, NJ, 1950.
- [DP11] Constantinos Daskalakis and Christos Papadimitriou. Continuous local search. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.
- [DP18] Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In *Proceedings of the 32nd Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [DSZ21] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2021.
- [Fan53] Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42, 1953.
- [Far59] Michael J. Farrell. The convexity assumption in the theory of competitive markets. *Journal of Political Economy*, 67(4):377–391, 1959.

- [FCR20] Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.
- [Fin64] Arlington M. Fink. Equilibrium in a stochastic  $n$ -person game. *Journal of Science of the Hiroshima University, series A-I (mathematics)*, 28(1):89–93, 1964.
- [FR20] Tanner Fiez and Lillian J Ratliff. Local convergence analysis of gradient descent ascent with finite timescale separation. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.
- [GHJY15] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, 2015.
- [Gli52] Irving L. Glicksberg. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.
- [Goo16] Ian Goodfellow. NIPS 2016 tutorial: Generative Adversarial Networks. *arXiv preprint arXiv:1701.00160*, 2016.
- [Gra57] J de V. Graaff. *Theoretical welfare economics*. Cambridge University Press, 1957.
- [Har91] Patrick T. Harker. Generalized Nash games and quasi-variational inequalities. *European Journal of Operational research*, 54(1):81–94, 1991.
- [Haz16] Elad Hazan. Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.
- [HLM21] Steve Hanneke, Roi Livni, and Shay Moran. Online learning with simple predictors and a combinatorial characterization of minimax in 0/1 games. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, 2021.
- [HvS08] Wan Huang and Bernhard von Stengel. Computing an extensive-form correlated equilibrium in polynomial time. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE)*, 2008.
- [HW03] Junling Hu and Michael P. Wellman. Nash Q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4(Nov):1039–1069, 2003.
- [JLWY21] Chi Jin, Qinghua Liu, Yuanhao Wang, and Tiancheng Yu. V-Learning – A Simple, Efficient, Decentralized Algorithm for Multiagent RL. *arXiv preprint arXiv:2110.14555*, 2021.
- [JMS22] Yujia Jin, Vidya Muthukumar, and Aaron Sidford. The complexity of infinite-horizon general-sum stochastic games. *arXiv preprint arXiv:2204.04186*, 2022.
- [JNJ20] Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. What is Local Optimality in Nonconvex-Nonconcave Minimax Optimization? In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.
- [KMVS96] Daphne Koller, Nimrod Megiddo, and Bernhard Von Stengel. Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 14(2):247–259, 1996.

- [Koo57] Tjalling C. Koopmans. *Three Essays in the State of Economic Science*. New York: McGraw-Hill Book Co., 1957.
- [Lit94] Michael L. Littman. Markov Games as a Framework for Multi-Agent Reinforcement Learning. In *Proceedings of the 11th International Conference on Machine Learning (ICML)*, 1994.
- [LSJR16] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. In *Proceedings of the 29th Annual Conference on Learning Theory (COLT)*, 2016.
- [LYBJ21] Qinghua Liu, Tiancheng Yu, Yu Bai, and Chi Jin. A sharp analysis of model-based reinforcement learning with self-play. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.
- [MB22] Weichao Mao and Tamer Başar. Provably Efficient Reinforcement Learning in Decentralized General-Sum Markov Games. *Dynamic Games and Applications*, 2022.
- [McK54] Lionel McKenzie. On equilibrium in Graham’s model of world trade and other competitive systems. *Econometrica*, 22(2):147–161, 1954.
- [MGN18] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which Training Methods for GANs do actually Converge? In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.
- [MPPSD17] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled Generative Adversarial Networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017.
- [MRS20] Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On Gradient-Based Learning in Continuous Games. *SIAM J. Math. Data Sci.*, 2(1):103–131, 2020.
- [MV21] Oren Mangoubi and Nisheeth K. Vishnoi. Greedy adversarial equilibrium: an efficient alternative to nonconvex-nonconcave min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2021.
- [Nas50] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.
- [Ne03] Abraham Neyman and Sylvain Sorin (editors). *Stochastic games and applications*, volume 570. Springer Science & Business Media, 2003.
- [OR94] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.
- [Pap94] Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. Syst. Sci.*, 48(3):498–532, 1994.
- [Par06] Pablo A Parrilo. Polynomial games and sum of squares optimization. In *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006.
- [PW11] Christos H. Papadimitriou and Christopher A. Wilkens. Economies with non-convex production and complexity equilibria. In *Proceedings of the 12th ACM Conference on Electronic Commerce (EC)*, 2011.
- [RBS16] Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. On the characterization of local Nash equilibria in continuous games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.

- [Ros65] J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965.
- [Rot60] Jerome Rothenberg. Non-convexity, aggregation, and Pareto optimality. *Journal of Political Economy*, 68(5):435–468, 1960.
- [RST15] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *J. Mach. Learn. Res.*, 16:155–186, 2015.
- [Sha53] Lloyd S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.
- [Sio58] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.
- [SMB22] Ziang Song, Song Mei, and Yu Bai. When Can We Learn General-Sum Markov Games with a Large Number of Players Sample-Efficiently? In *Proceedings of the 10th International Conference on Learning Representations (ICLR)*, 2022.
- [SOP08] Noah D. Stein, Asuman Ozdaglar, and Pablo A. Parrilo. Separable and low-rank continuous games. *International Journal of Game Theory*, 37(4):475–504, 2008.
- [SS12] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.
- [Sta69] Ross M. Starr. Quasi-equilibria in markets with non-convex preferences. *Econometrica*, 37(1):25–38, 1969.
- [SV15] Eilon Solan and Nicolas Vieille. Stochastic games. *Proceedings of the National Academy of Sciences*, 112(45):13743–13746, 2015.
- [SWYY20] Aaron Sidford, Mengdi Wang, Lin Yang, and Yinyu Ye. Solving discounted stochastic two-player games with near-optimal time and sample complexity. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.
- [Tak62] Masayuki Takahashi. Stochastic games with infinitely many strategies. *Journal of Science of the Hiroshima University, Series A-I (mathematics)*, 26(2):123–134, 1962.
- [TWYS21] Yi Tian, Yuanhao Wang, Tiancheng Yu, and Suvrit Sra. Online Learning in Unknown Markov Games. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.
- [vN28] John von Neumann. Zur Theorie der Gesellschaftsspiele. In *Math. Ann.*, pages 295–320, 1928.
- [vS07] Bernhard von Stengel. Equilibrium computation for two-player games in strategic and extensive form. *Algorithmic Game Theory*, pages 53–78, 2007.
- [vSF08] Bernhard von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.
- [WHL17] Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. Online Reinforcement Learning in Stochastic Games. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.
- [WZB20] Yuanhao Wang, Guodong Zhang, and Jimmy Ba. On Solving Minimax Optimization Locally: A Follow-the-Ridge Approach. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.



- [XCWY20] Qiaomin Xie, Yudong Chen, Zhaoran Wang, and Zhuoran Yang. Learning zero-sum simultaneous-move markov games using function approximation and correlated equilibrium. In *Proceedings of the the 33rd Annual Conference on Learning Theory (COLT)*, 2020.
- [ZKBY20] Kaiqing Zhang, Sham M. Kakade, Tamer Basar, and Lin F. Yang. Model-based multi-agent RL in zero-sum markov games with near-optimal sample complexity. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [ZYB21] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, pages 321–384, 2021.