

GANGs: Generative Adversarial Network Games

Frans A. Oliehoek
University of Liverpool
frans.oliehoek@liverpool.ac.uk

Rahul Savani
University of Liverpool
rahul.savani@liverpool.ac.uk

Jose Gallego-Posada
University of Amsterdam
jose.gallegoposada@student.uva.nl

Elise van der Pol
University of Amsterdam
elisevanderpol@gmail.com

Edwin D. de Jong
IMC Financial Markets
jong.de.edwin@gmail.com

Roderich Groß
The University of Sheffield
r.gross@sheffield.ac.uk

ABSTRACT

Generative Adversarial Networks (GAN) have become one of the most successful frameworks for unsupervised generative modeling. As GANs are difficult to train much research has focused on this. However, very little of this research has directly exploited game-theoretic techniques. We introduce Generative Adversarial Network Games (GANGs), which explicitly model a finite zero-sum game between a generator (G) and classifier (C) that use *mixed* strategies. The size of these games precludes exact solution methods, therefore we define resource-bounded best responses (RBBRs), and a resource-bounded Nash Equilibrium (RB-NE) as a pair of mixed strategies such that neither G or C can find a better RBBR. The RB-NE solution concept is richer than the notion of ‘local Nash equilibria’ in that it captures not only failures of escaping local optima of gradient descent, but applies to any approximate best response computations, including methods with random restarts. To validate our approach, we solve GANGs with the Parallel Nash Memory algorithm, which provably monotonically converges to an RB-NE. We compare our results to standard GAN setups, and demonstrate that our method deals well with typical GAN problems such as mode collapse, partial mode coverage and forgetting.

KEYWORDS

GANs; adversarial learning; game theory

1 INTRODUCTION

Generative Adversarial Networks (GANs) [14] are a framework in which two neural networks compete with each other: the *generator* (G) tries to trick the *classifier* (C) into classifying its generated fake data as true. GANs hold great promise for the development of accurate generative models for complex distributions, such as the distribution of images of written digits or faces. Consequently, in just a few years, GANs have grown into a major topic of research in machine learning. A core appeal is that they do not need to rely on distance metrics [22]. However, GANs are difficult to train and much research has focused on this [1, 2, 38]. Typical problems are *mode collapse* in which the generator only outputs points from a single mode of the distribution and *partial mode coverage* in which the generator only learns to represent a (small) subset of the modes of the distribution. Moreover, while learning the players may *forget*: e.g., it is possible that a classifier correctly learns to classify part of the input space as ‘fake’ only to forget this later when the generator no longer generates examples in this part of space. Finally, when learning via gradient descent one can get stuck in *local* Nash equilibria [32].

We introduce a novel approach that does not suffer from local equilibria: *Generative Adversarial Network Games* (GANGs) formulate adversarial networks as *finite* zero-sum games, and the solutions that we try to find are saddle points in *mixed strategies*. This approach is motivated by the observation that, considering a GAN as a finite zero-sum game, in the space of mixed strategies, any local Nash equilibrium is a global one. Intuitively, the reason for this is that whenever there is a profitable pure strategy deviation one can move towards it in the space of mixed strategies.

However, as we cannot expect to find such exact best responses due to the extremely large number of pure strategies that result for sensible choices of neural network classes, we introduce Resource-Bounded Best-Responses (RBBRs), and the corresponding Resource-Bounded Nash equilibrium (RB-NE), which is a pair of mixed strategies in which no player can find a better RBBR. This is richer than the notion of local Nash equilibria in that it captures not only failures of escaping local optima of gradient descent, but applies to *any* approximate best response computations, including methods with random restarts, and allows us to provide convergence guarantees.

The key features of our approach are that:

- It is based on a framework of finite zero-sum games, and as such it enables the use of existing game-theoretic methods. In this paper we focus on one such method, Parallel Nash Memory (PNM).
- We show that PNM will provably and monotonically converge to an RB-NE.
- It enables us to understand existing GAN objectives (WGAN, Goodfellow’s training heuristic) in the context of zero-sum games.
- Moreover, it works for any network architectures (unlike previous approaches, see Related Work). In particular, future improvements in classifiers/generator networks can directly be exploited.

We investigate empirically the effectiveness of PNM and show that it can indeed deal well with typical GAN problems such as mode collapse, partial mode coverage and forgetting, especially in distributions with less symmetric structure to their modes. At the same time, a naive implementation of our method does have some disadvantages and we provide an intuitively appealing way to deal with these. Our promising results suggest several interesting directions for further work.

2 BACKGROUND

We defer a more detailed treatment of related work on GANs and recent game theoretic approaches until the end of the paper. Here, we start by introducing some basic game-theoretic notation.

Definition 1. A *strategic game* (also ‘normal-form game’), is a tuple $\langle \mathcal{D}, \{S_i\}_{i=1}^n, \{u_i\}_{i=1}^n \rangle$, where $\mathcal{D} = \{1, \dots, n\}$ is the set of

players, \mathcal{S}_i is the set of *pure strategies* for player i , and $u_i : \mathcal{S} \rightarrow \mathbb{R}$ is i 's payoff function defined on the set of pure strategy profiles $\mathcal{S} := \mathcal{S}_1 \times \dots \times \mathcal{S}_n$. When the set of players and their action sets are finite, the strategic game is *finite*.

A fundamental concept in game theory is the Nash equilibrium (NE), which is a strategy profile such that no player can unilaterally deviate and improve his payoff.

Definition 2 (Pure Nash equilibrium). A pure strategy profile $s = \langle s_1, \dots, s_n \rangle$ is an NE if and only if $u_i(s) \geq u_i(\langle s_1, \dots, s'_i, \dots, s_n \rangle)$ for all players i and $s'_i \in \mathcal{S}_i$.

A finite game may not possess a pure NE. A *mixed strategy* of player i is a probability distribution over i 's pure strategies \mathcal{S}_i . The payoff of a player under a profile of mixed strategies $\mu = \langle \mu_1, \dots, \mu_n \rangle$ is defined as the expectation:

$$u_i(\mu) := \sum_{s \in \mathcal{S}} \left[\prod_{j \in \mathcal{D}} \mu_j(s_j) \right] \cdot u_i(s).$$

Then an NE in mixed strategies is defined as follows.

Definition 3 (Mixed Nash equilibrium). A μ is an NE if and only if $u_i(\mu) \geq u_i(\langle \mu_1, \dots, s'_i, \dots, \mu_n \rangle)$ for all players i and $s'_i \in \mathcal{S}_i$.

Every finite game has at least one NE in mixed strategies [25]. In this paper we deal with two-player *zero-sum* games, where $u_1(s_1, s_2) = -u_2(s_1, s_2)$ for all $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$. The equilibria of zero-sum games, also called *saddle points*, have several important properties, as stated in the following theorem.

Theorem 4 (Minmax Theorem [39]). *In a zero-sum game, we have*

$$\min_{\mu_1} \max_{\mu_2} u_i(\mu) = \max_{\mu_2} \min_{\mu_1} u_i(\mu) = v.$$

We call v the *value of the game*. All equilibria have payoff v .

Moreover, these equilibria μ can be expressed in terms of so-called maxmin strategies. A *maxmin strategy* of player 1 is a $\hat{\mu}_1$ that solves $\min_{\mu_2} u_1(\hat{\mu}_1, \mu_2) = v$, and a maxmin strategy of player 2 is a $\hat{\mu}_2$ that solves $\min_{\mu_1} u_2(\mu_1, \hat{\mu}_2) = v$. Any pair of maxmin strategies of the players is an equilibrium. This directly implies that equilibrium strategies are interchangeable: if $\langle \mu_1, \mu_2 \rangle$ and $\langle \mu'_1, \mu'_2 \rangle$ are equilibria, then so are $\langle \mu'_1, \mu_2 \rangle$ and $\langle \mu_1, \mu'_2 \rangle$ [28]. Moreover, the convex combination of two equilibria is an equilibrium, meaning that the game has either one or infinitely many equilibria.

We will not always be able to compute an exact NE, and so we employ the standard, additive notion of *approximate equilibrium*:

Definition 5. Let μ_{-i} denote a strategy of the opponent of player i . A pair of (possibly pure) strategies (μ_i, μ_{-i}) is an ϵ -NE if

$$\forall i \quad u_i(\mu_i, \mu_{-i}) \geq \max_{\mu'_i} u_i(\mu'_i, \mu_{-i}) - \epsilon. \quad (1)$$

In other words, no player can gain more than ϵ by deviating.

In the literature, GANs have not typically been considered as finite games. The natural interpretation of the standard setup of GANs is of an infinite game where payoffs are defined over all possible weight parameters for the respective neural networks. With this view we do not obtain existence of saddle points, nor the desirable properties of Theorem 4. Some results on the existence of saddle points in infinite action games are known, but they require properties like convexity and concavity of utility functions [5], which we

cannot apply as they would need to hold w.r.t. the neural network parameters. This is why the notion of *local Nash equilibrium (LNE)* has arisen in the literature [32, 38]. Roughly, an LNE is a strategy profile where neither player can improve in a small neighborhood of the profile. In finite games every LNE is an NE, as, whenever there is a global deviation, one can always deviate locally in the space of mixed strategies towards a pure best response.

3 GANGS

Intuitively, it is clear that GANs are closely related to games, and the original paper by [14] already points out that G and C essentially play a minimax game. The exact relation has not been explicitly described, leading to confusion on whether the sought solution is a saddle point or not [8]. We set out to address this confusion by introducing the *Generative Adversarial Network Game (GANs)* formalism, which explicitly phrases adversarial networks as zero-sum strategic-form games. It builds on the building blocks of regular GANs, but it emphasizes the fact that G and C play a zero-sum game, and formalizes the action space available to these players.

In this section, we provide a ‘fully rational’ formulation of GANGs, i.e., one where we make no or limited assumptions on the bounded resources that we face in practice. Bounded resources will be introduced in the next section.

Definition 6. A **GANG** is a tuple $\mathcal{M} = \langle p_d, \langle G, p_z \rangle, C, \phi \rangle$ with

- $p_d(x)$ is the distribution over (‘true’ or ‘real’) data points $x \in \mathbb{R}^d$.
- G is a neural network class with parameter vector $\theta_G \in \Theta_G$ and d outputs, such that $G(z; \theta_G) \in \mathbb{R}^d$ denotes the (‘fake’ or ‘generated’) output of G on a random vector z drawn from some distribution $z \sim p_z$. We will typically leave dependence on the parameter vector θ_G implicit and just write $G(z)$.
- C is a neural network class with parameter vector $\theta_C \in \Theta_C$ and a single output, such that the output $C(x; \theta_C) \in [0, 1]$ indicates the ‘realness’ of x according to C . We will interpret it as the probability that C assigns to a point x being real, even though it does not need to be a strict probability.
- ϕ is a *measuring function* [3]: $\phi : [0, 1] \rightarrow \mathbb{R}$ —typically log, for GANs, or the identity mapping, for WGANs—that is used to specify the payoffs of the agents, explained next.

A GANG induces a zero-sum game in an intuitive way:

Definition 7. The induced zero-sum strategic-form game of a GANG is $\langle \mathcal{D} = \{G, C\}, \{\mathcal{S}_G, \mathcal{S}_C\}, \{u_G, u_C\} \rangle$ with:

- $\mathcal{S}_G = \{G(\cdot; \theta_G) \mid \theta_G \in \Theta_G\}$, elements of which we denote by s_G ;
- $\mathcal{S}_C = \{C(\cdot; \theta_C) \mid \theta_C \in \Theta_C\}$, elements of which we denote by s_C ;
- The payoff of G , for all $(\theta_G \in \Theta_G, \theta_C \in \Theta_C)$ and their induced strategies s_G, s_C is $u_G(s_G, s_C) = -u_C(s_G, s_C)$;
- The payoff of the classifier is given by:

$$u_C(s_G, s_C) = \mathbb{E}_{x \sim p_d} [\phi(s_C(x))] - \mathbb{E}_{z \sim p_z} [\phi(s_C(G(z)))].$$

That is, the score of the classifier is the expected correct classification on the real data minus the expected incorrect classification on the fake data.

In practice, GANs are represented using floating point numbers, of which, for a given setup, there is only a finite (albeit large) number. In GANGs, we formalize this:

Definition 8. Any GANG where G, C are finite classes—i.e., classes of networks constructed from a finite set of node types (e.g., {Sigmoid, ReLU, Linear})—and with architectures of bounded size, is called a *finite network class GANG*. A finite network class GANG in which the sets Θ_G, Θ_C are finite is called a *finite GANG*.

From now on, we deal with finite GANGs. These correspond to finite (zero-sum) strategic games and, as explained earlier, they will possess one or infinitely many mixed NEs with the same payoff. Note that we only point out that finiteness of floating point systems leads to finite GANGs, but we do not impose any additional constraints or discretization. Therefore, our finite GANGs have the same representational capacity as normal GANs that are implemented using floating point arithmetic.

Zero-sum vs Non-zero-sum. In contrast to much of the GAN literature, we explicitly formulate GANGs as being zero-sum games. GANs [14] formulate the payoff of the generator as a function of the fake data only: $u_G = \mathbb{E}_{z \sim p_z} [\phi(s_C(s_G(z)))]$. However, it turns out that this difference typically has no implications for the sought solutions. We clarify this with the following theorem, and investigate particular instantiations below. In game theory, two games are called *strategically equivalent* if they possess exactly the same set of Nash equilibria; this (standard) definition is concerned only about the mixed strategies played in equilibrium and not the resulting payoffs. The following is a well-known game transformation (folklore, see [23]) that creates a new strategically equivalent game:

Fact 9. Consider a game $\Gamma = (\{1, 2\}, \{S_1, S_2\}, \{u_1, u_2\})$. Fix a pure strategy $s_2 \in S_2$. Define \bar{u}_1 as identical to u_1 except that $\bar{u}_1(s_1, s_2) = u_1(s_1, s_2) + c$ for all $s_1 \in S_1$ and some constant c . We have that Γ and $\bar{\Gamma} = (\{1, 2\}, \{S_1, S_2\}, \{\bar{u}_1, u_2\})$ are strategically equivalent.

Theorem 10. Any finite (non-zero-sum) two-player game between G and C with payoffs of the following form:

$$\begin{aligned} u_G &= \text{Fake}_G(s_G, s_C) = -\text{Fake}_C(s_G, s_C), \\ u_C &= \text{Real}_C(s_C) + \text{Fake}_C(s_G, s_C), \end{aligned}$$

is strategically equivalent to a zero-sum game where G has payoff $\bar{u}_G \triangleq -\text{Real}_C(s_C) - \text{Fake}_C(s_G, s_C)$.

PROOF. By adding $-\text{Real}_C(s_C)$ to G 's utility function, for each pure strategy s_C of C we add a different constant to all utilities of G against s_C . Thus, by applying Fact 9 iteratively for all $s_C \in S_C$ we see that we produce a strategically equivalent game. \square

Next, we formally specify the conversion of existing GAN models to GANGs. We consider the general measure function that covers GANs and WGANs. In these models, the payoffs are specified as

$$u_G(s_G, s_C) \triangleq -\mathbb{E}_{z \sim p_z} [\phi(1 - s_C(s_G(z)))],$$

$$u_C(s_G, s_C) \triangleq \mathbb{E}_{x \sim p_d} [\phi(s_C(x))] + \mathbb{E}_{z \sim p_z} [\phi(1 - s_C(s_G(z)))].$$

These can be written using $\text{Fake}_G(s_G, s_C) = -\text{Fake}_C(s_G, s_C) = -\mathbb{E}_{z \sim p_z} [\phi(1 - s_C(s_G(z)))]$ and $\text{Real}_C(s_C) = \mathbb{E}_{x \sim p_d} [\phi(s_C(x))]$. This means that we can employ Theorem 10 and equivalently define a GANG with zero-sum payoffs that preserves the NEs.

In practice, most work on GANs uses a different objective, introduced by [14]. They say that [formulas altered]: ‘‘Rather than training G to minimize $\log(1 - s_C(s_G(z)))$ we can train G to maximize $\log s_C(s_G(z))$. This objective function results in the same fixed

point of the dynamics of G and C but provides much stronger gradients early in learning.’’ This means that they redefine $u_G(s_G, s_C) \triangleq \mathbb{E}_{z \sim p_z} [\phi(s_C(s_G(z)))]$, which still can be written as $\text{Fake}_G(s_G, s_C) \triangleq \mathbb{E}_{z \sim p_z} [\phi(s_C(s_G(z)))]$. Now, as long as the classifier’s payoff is also adapted we can still write the payoff functions in the form of Theorem 10. That is, the trick is compatible with a zero-sum formulation, as long as it is also applied to the classifier. This then yields the formulation of the payoffs as used in GANGs (in Def. 7).

4 RESOURCE-BOUNDED GANGS

While GANGs clarify the relation of existing adversarial network models to zero-sum games, we will not be able to solve them exactly. Even though they are finite, the number of pure strategies will be huge when we use reasonable neural network classes and parameter sets. This means that finding an NE, or even an ϵ -NE will be typically beyond our capabilities. Thus we need to deal with players with bounded computational resources. In this section, we propose a formal notion of such ‘bounded rationality’.

Resource-Bounded Best-Responses (RBBR). A Nash equilibrium is defined by the absence of better responses for any of the players. As such, best response computation is a critical tool to verify whether a strategy profile is an NE, and is also common subroutine in algorithms that compute an NE. However, like the computation of an (ϵ -)NE, computing an (ϵ -)best response will generally be intractable for GANGs. Therefore, to clarify the type of solutions that we actually can expect to compute with our bounded computational power, we formalize the notion of *resource-bounded best response* and show how it naturally leads to a solution concept that we call *resource-bounded NE (RB-NE)*.

Definition 11. We say that $S_i^{RB} \subseteq S_i$ is the subset of strategies of player i , that i can compute as a best response, given its bounded computational resources.

This computable set is an abstract formulation to capture phenomena like gradient descent being stuck in local optima, but also more general reasons for not computing a best response, such as not even reaching a local optimum in the available time.

Definition 12. A strategy $s_i \in S_i^{RB}$ of player i is a *resource-bounded best-response (RBBR)* against a (possibly mixed) strategy s_j , if $\forall s'_i \in S_i^{RB} \quad u_i(s_i, s_j) \geq u_i(s'_i, s_j)$.

That is, s_i only needs to be amongst the best strategies that player i can compute in response to s_j . We denote the set of such RBBRs to s_j by $S_i^{RBBR(s_j)} \subseteq S_i^{RB}$.

Definition 13. A *resource-bounded best-response function* $f_i^{RBBR} : S_j \rightarrow S_i^{RB}$ is a function that maps from the set of possible strategies of player j to an RBBR for i , s.t. $\forall s_j \quad f_i^{RBBR}(s_j) \in S_i^{RBBR(s_j)}$.

Now we can define a class of GANGs which are a better model for games played by neural networks:

Definition 14. A *resource-bounded GANG* is a tuple $\langle \mathcal{M}, \{f_g^{RBBR}, f_c^{RBBR}\} \rangle$ with \mathcal{M} a finite GANG as above, and f_g^{RBBR}, f_c^{RBBR} the RBBR functions for both players.

For these games, we define an intuitive specialization of NE:

Definition 15. A strategy profile $s = \langle s_i, s_j \rangle$ is a *Resource-Bounded Nash Equilibrium (RB-NE)* iff $\forall i \quad u_i(s_i, s_j) \geq u_i(f_i^{RBBR}(s_j), s_j)$.

That is, an RB-NE can be thought of as follows: we present s to each player i and it gets the chance to switch to another strategy, for which it can apply its bounded resources (i.e., use f_i^{RBBR}) exactly once. After this application, the player’s resources are exhausted and if the found $f_i^{RBBR}(s_j)$ does not lead to a higher payoff it will not have an incentive to deviate.

Clearly, an RB-NE can be linked to the familiar notion of ϵ -NE by making assumptions on the power of the best response computation.

Theorem 16. *If the players are powerful enough to compute an ϵ -best response function, then an RB-NE for their game is an ϵ -NE.*

PROOF. Starting from the RB-NE (s_i, s_j) , assume an arbitrary i . By definition of RB-NE $u_i(s_i, s_j) \geq u_i(f_i^{RBBR}(s_j), s_j) \geq \max_{s'_i} u_i(s'_i, s_j) - \epsilon$. \square

Non-deterministic Best Responses. The above definitions assumed deterministic RBBR functions f_i^{RBBR} . However, in many cases the RBBR function can be non-deterministic (e.g., due to random restarts), which means that the sets \mathcal{S}_i^{RB} are non-deterministic. This is not a fundamental problem, however, and the same approach can be adapted to allow for such non-determinism. In particular, now let f_i^{RBBR} be a non-deterministic function, and define \mathcal{S}_i^{RB} as the (non-deterministic) range of this function. That is, we define \mathcal{S}_i^{RB} as that set of strategies that our non-deterministic RBBR function delivers. Given this modification the definition of the RB-NE remains unchanged: a strategy profile $s = \langle s_i, s_j \rangle$ is a *non-deterministic RB-NE* if each player i uses all its computational resources by calling $f_i^{RBBR}(s_j)$ once, and no player finds a better strategy to switch to.

5 SOLVING GANGS

The treatment of GANGs as finite games in mixed strategies opens up the possibility of extending many of the existing tools and algorithms for these classes of games [12, 13, 30]. In this section, we consider the use of the Parallel Nash Memory, which is particularly tailored to finding approximate NEs with small support, and which monotonically converges to such an equilibrium [27].

Here we give a concise description of a slightly simplified form of Parallel Nash Memory (PNM) and how we apply it to GANGs.¹ For ease of explanation, we focus on the setting with deterministic best responses.² The algorithm is shown in Algorithm 1. Intuitively, PNM incrementally grows a strategic game SG , over a number of iterations, using the AUGMENTGAME function. It also maintains a mixed strategy NE $\langle \mu_G, \mu_C \rangle$ of this smaller game at all times. In each of the iterations it uses a ‘search’ heuristic to deliver new promising strategies. In our GANG setting, we use the resource-bounded best-response (RBBR) functions of the players for this purpose. After having found new strategies, the game is augmented with these and solved again to find a new NE of the sub-game SG .

¹This ignores some aspects of PNM that we do not use, such as ways to ‘discard’ old strategies [11] that have not been used for a long time.

²In our experiments, we use random initializations for the best responses. To deal with this non-determinism we simply ignore any tests that are not able to find a positive payoff over the current mixed strategy NE $\langle \mu_G, \mu_C \rangle$, but we do not terminate. Instead, we run for a pre-specified number of iterations.

Algorithm 1 PARALLEL NASH MEMORY FOR GANGS

```

1:  $\langle s_G, s_C \rangle \leftarrow \text{INITIALSTRATEGIES}()$ 
2:  $\langle \mu_G, \mu_C \rangle \leftarrow \langle \{s_G\}, \{s_C\} \rangle$  ▷ set initial mixtures
3: //Main loop:
4: while not done do
5:    $s_G = \text{RBBR}(\mu_C)$  ▷ get new bounded best resp.
6:    $s_C = \text{RBBR}(\mu_G)$ 
7:   //expected payoffs of these ‘tests’ against mixture:
8:    $u_{BRs} = u_G(s_G, \mu_C) + u_C(\mu_G, s_C)$ 
9:   if  $u_{BRs} \leq 0$  then
10:     done  $\leftarrow$  True
11:   else
12:      $SG \leftarrow \text{AUGMENTGAME}(SG, s_G, s_C)$ 
13:      $\langle \mu_G, \mu_C \rangle \leftarrow \text{SOLVEGAME}(SG)$ 
14:   end if
15: end while
16: return  $\langle \mu_G, \mu_C \rangle$  ▷ found an RB-NE

```

In order to augment the game, PNM evaluates (by simulation) each newly found strategy for each player against all of the existing strategies of the other player, thus constructing a new row and column for the maintained payoff matrix.

In order to implement the best response functions, any existing neural network architectures (e.g., for GANs) can be used. However, we need to compute RBBRs against *mixtures* of networks of the other player. For C this is trivial: we can just generate a batch of fake data from the mixture μ_G . Implementing a RBBR for G against μ_C is slightly more involved, as we need to back-propagate the gradient from all the different $s_C \in \mu_C$ to G . In our implementation we subsample a number of such pure s_C and construct a larger network for that purpose. Better ways of doing this are an interesting line of future work.

Intuitively, it is clear that PNM converges to an RB-NE, which we now prove formally.

Theorem 17. *If PNM terminates, then it has found an RB-NE.*

PROOF. We show that $u_{BRs} \leq 0$ implies we found an RB-NE:

$$\begin{aligned}
 u_{BRs} &= u_G(f_G^{RBBR}(\mu_C), \mu_C) + u_C(\mu_G, f_C^{RBBR}(\mu_G)) \\
 &\leq 0 = u_G(\mu_G, \mu_C) + u_C(\mu_G, \mu_C)
 \end{aligned} \tag{2}$$

Note that, per Def. 12, $u_G(f_G^{RBBR}(\mu_C), \mu_C) \geq u_G(s'_G, \mu_C)$ for all computable $s'_G \in \mathcal{S}_G^{RB}$ (and similar for C). Therefore, the only way that $u_G(f_G^{RBBR}(\mu_C), \mu_C) \geq u_G(\mu_G, \mu_C)$ could fail to hold, is if μ_G would include some strategies that are not computable (not in \mathcal{S}_G^{RB}) that provide higher payoff. However, as the support of μ_G is composed of computed (i.e., computable) strategies in previous iterations, this cannot be the case. As such we conclude $u_G(f_G^{RBBR}(\mu_C), \mu_C) \geq u_G(\mu_G, \mu_C)$ and similarly $u_C(\mu_G, f_C^{RBBR}(\mu_G)) \geq u_C(\mu_G, \mu_C)$. Together with (2) this directly implies $u_G(\mu_G, \mu_C) = u_G(f_G^{RBBR}(\mu_C), \mu_C)$ and $u_C(\mu_G, \mu_C) = u_C(\mu_G, f_C^{RBBR}(\mu_G))$, indicating we found an RB-NE. \square

When there are only finitely many pure best responses that we can compute, as for finite GANGs, the algorithm will terminate. Moreover, using the same finiteness argument, one can show that this method *monotonically* converges to an equilibrium [27].

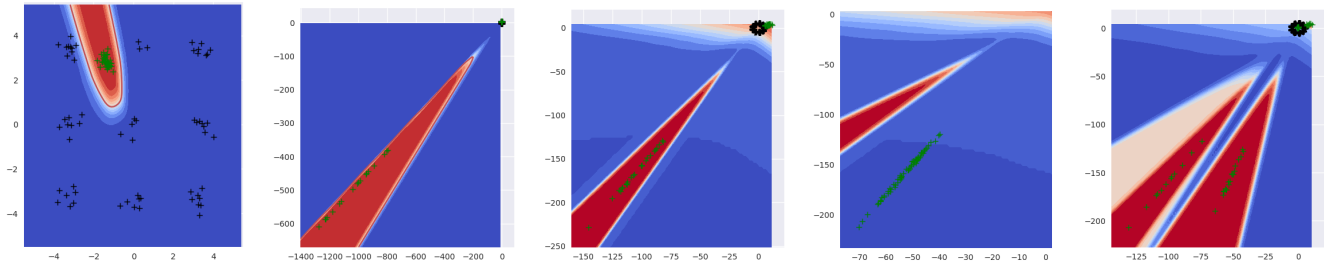


Figure 1: Illustration of exploitation of ‘overfitted’ classifier best-response.

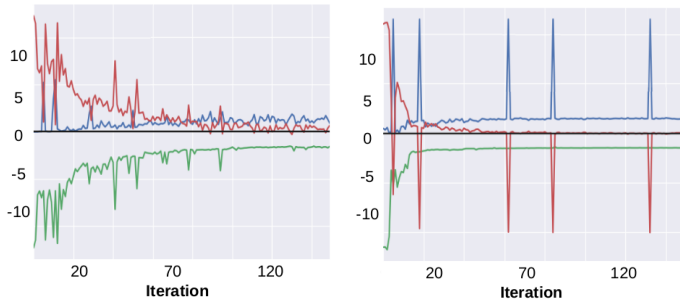


Figure 2: Convergence without (left) and with (right) adding uniform fake data. Shown is payoff as a function of the number of iterations of PNM: generator (blue), classifier (green), tests (red). The tests that do not generate positive payoff (red line < 0) are not added to the mixture.

6 EXPERIMENTS

Here we report on experiments that aim to test if using existing tools such as PNM can help in reducing problems with training GANs, such as missed modes. This is very difficult to assess on complex data like images; in fact, there is debate about whether GANs are overfitting (memorizing the data). Assessing this from samples is very difficult; only crude methods have been proposed e.g., [4]. Therefore, we restrict our proof-of-concept results to synthetic mixtures of Gaussians for which the distributions can readily be visualized.

Experimental setup. We compare to a GAN implementation by Shih [36]. For PNM, we use the same architectures for G and C as the GAN implementation. The settings for GAN and PNM training are summarized in Table 1. The mixture components comprise

	GAN	RBBR
Iterations	2500	1000
Learning Rate	$2 \cdot 10^{-4}$	10^{-3}
Batch Size	64	128
Measuring Function	log	10^{-5} -bounded log

Table 1: Settings used to train GANs and RBBRs.

grids and annuli with equal-variance components, as well as non-symmetric cases with randomly located modes and with a random covariance matrix for each mode. For each domain we create test cases with 9 and 16 components. In our plots, black points are real

data, green points are generated data. Blue indicates areas that are classified as ‘realistic’ while red indicates a ‘fake’ classification by C . **Plain application of PNM.** In GANGs, G informs C about what good strategies are and vice versa. However, as we will make clear here, this G has limited incentive to provide the best possible training signal to C . This is illustrated in Figure 1. The leftmost two plots show the *same* best response by C : zoomed in on the data and zoomed out to cover some fake outliers. Clearly, C needs to really find creative solutions to try and get rid of the faraway points, and also do good near the data. As a result, it ends up with the shown narrow beams in an effort to give high score to the true data points (a very broad beam would lower their scores), but this exposes C to being exploited in later iterations: G needs to merely shift the samples to some other part of the vast empty space around the data. This phenomenon is nicely illustrated by the remaining three plots (that are from a different training run, but illustrate it well): the middle plot shows an NE that targets one beam, this is exploited by G in its next best response (fourth image, note the different scales on the axes, the ‘beam’ is the same). The process continues, and C will need to find mixtures of all these type of complex counter measures (rightmost plot). This process can take a long time.

PNM with Added Uniform Fake Data. However, the GANG formalism allows us to incorporate a simple way to resolve this issue and make training more effective. In each iteration, we look at the total span (i.e., bounding box) of the real and fake data, and we add some uniformly sampled fake data in this bounded box (we used the same amount as fake data produced by G). In that way, we further guide C in order to better guide the generator (by directly making clear that all the area beyond the true data is fake). The impact of this procedure is illustrated by Figure 2, which shows the payoffs that the maintained mixtures μ_G, μ_C achieve against the RBBRs computed against them (so this is a measure of security), as well as the ‘payoff for tests’ (u_{BR}). Clearly, adding uniform fake data leads to much faster convergence.

As such, we perform our main comparison to GANs with this uniform fake data component added in. These results are shown in Figure 3 and 4, and clearly convey three main points: first, the PNM mixed classifier has a much flatter surface than the classifier found by the GAN, which is in line with the theoretical predictions about the equilibrium [14]. More importantly, however, we see that this flatter surface is not coming at the cost of inaccurate samples. In contrast: nearly all samples shown are hitting one of the modes and thus are highly accurate, much more so than the GAN’s samples.

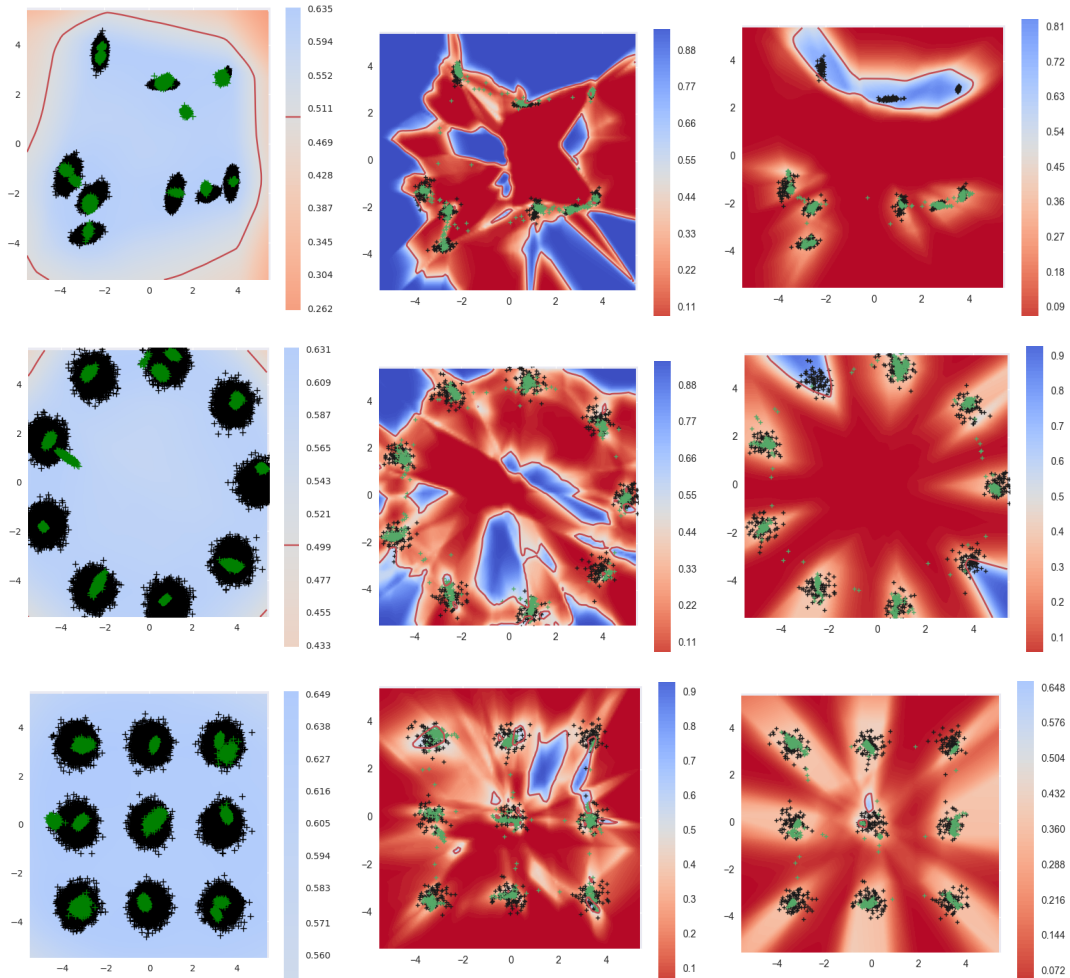


Figure 3: Results on mixtures with 9 components. Left: PNM, Center: GAN with BN, Right: GAN without BN. True data is shown in black, while fake data is green.

Finally, except in the 16-component grid plot, we see that the approach does not suffer from partial mode coverage as it leaves out no modes. We point out that these PNM results are without batch normalization, while the GAN results without batch normalization suffered from many inaccurate samples and severe partial mode coverage in many cases, as shown in Figures 3 and 4.

Impact of Generator Learning Rate. The above results show that PNM can accurately cover multiple modes, however, not all modes are *fully* covered; some amount of mode collapse still takes place. As also pointed out by, e.g., [2], the best response of G against μ_C is a single point with the highest ‘realness’, and therefore the WGAN they introduced uses fewer iterations for G than for C . Inspired by this, we investigate if we can reduce the mode collapse by reducing the learning rate of G (to 10^{-5}). The results in Figure 5 clearly show that more area of the modes are covered confirming this hypothesis. However, it also makes clear that by doing so, we are now generating some data outside of the true data. We point out that also with mode collapse, the PNM mechanism theoretically could still converge, by adding in more and more delta peaks covering parts of the modes. In fact, this process in work is already

illustrated in the plots in Figure 3: each plot in the left column contains at least one true data (black) mode which is covered by multiple fake data (green) modes. However, if these fake data modes would be true delta peaks, the number of mixture components required clearly would be huge, making this infeasible in practice. As such, this leads to parameter tweaking; investigation of better solutions is deferred to future work.

7 RELATED WORK

Progress in zero-sum games. [6] devise a double-oracle algorithm for computing exact equilibria in extensive-form games with imperfect information. Their algorithm uses *best response oracles*; PNM does so too, though in this paper using resource-bounded rather than exact best responses. Inspired by GANs, [19] deal with general zero-sum settings with non-convex loss functions. They introduce a weakening of local equilibria known as *smoothed local equilibria* and provide algorithms with guarantees on the smoothed local regret. In contrast, we work with a generalization of local equilibrium (RB-NE) that allows for stronger notions of equilibrium, not only weaker ones, depending on the power of one’s RBRR

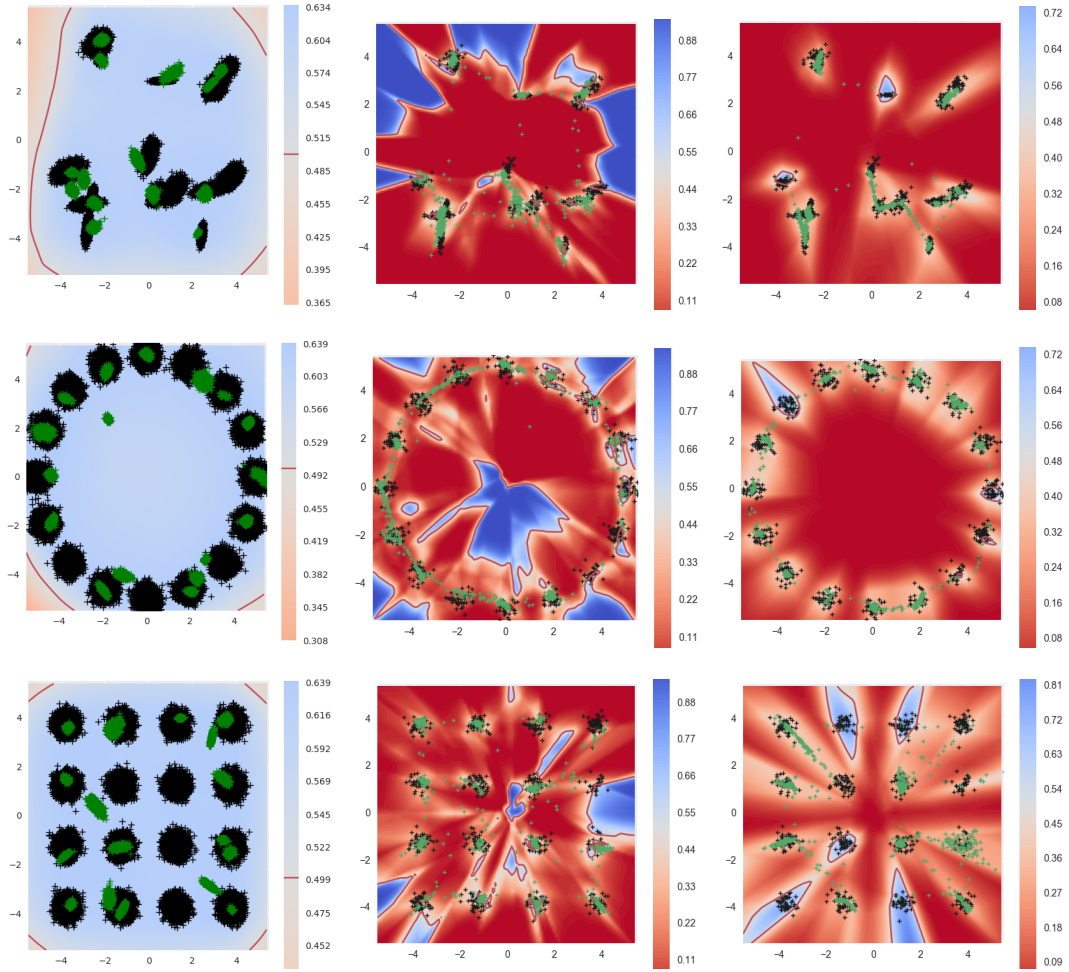


Figure 4: Results on mixtures with 16 components. Left: PNM, Center: GAN with BN, Right: GAN without BN. True data is shown in black, while fake data is green.

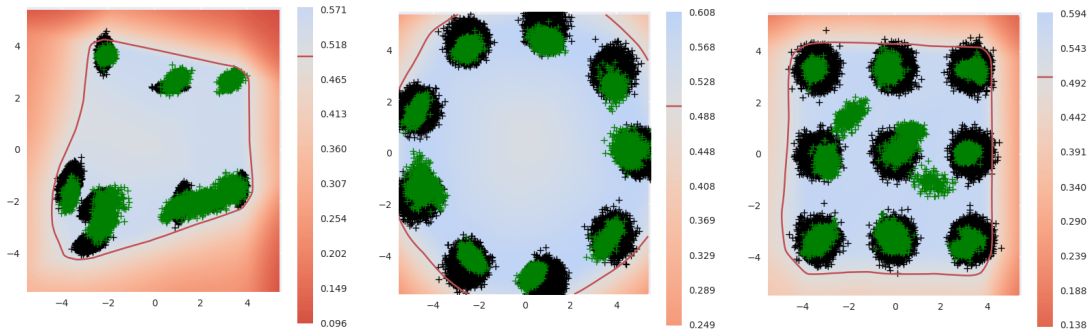


Figure 5: Results for PNM with a learning rate of 10^{-5} for the generator.

functions. For the more restricted class of convex-concave zero-sum games, it was recently shown that Optimistic Mirror Descent (a variant of gradient descent) and its generalization Optimistic Follow-the-Regularized-Leader achieve faster convergence rates than gradient descent [30, 31]. These algorithms have been explored

in the context of GANs by [10]. However, the convergence results do not apply as GANs are not convex-concave.

GANs. The literature on GANs has been growing at an incredible rate, and due to space constraints, we cannot give a full overview of all the related works, such as [1, 2, 4, 9, 16, 20, 26, 29, 35, 40].

Instead we refer to [38] for a comprehensive recent overview. That same paper also introduces Coulomb GANs [38]. As we do here, the authors show convergence, but for them only under the strong assumption that the “generator samples can move freely” (which is not the case when training via gradient descent; samples can only move small steps). Moreover, their approach essentially performs non-parametric density estimation, which is based on the (Euclidean) distance between data points, which we believe undermines one of the most attractive features of GANs (not needing a distance metric). Furthermore, it is widely known that using metrics in high-dimensional spaces is problematic (i.e., the “curse of dimensionality”, see, e.g., [18, Sect. 2.5]). Perhaps not surprisingly Unterthiner et al. report a higher frequency of “generated samples that are non-sensical interpolations of existing data modes”.

Explicit representations of mixtures of strategies. Recently, more researchers have investigated the idea of (more or less) explicitly representing a set or mixture of strategies for the players. For instance, [21] retains sets of networks that are trained by randomly pairing up with a network for the other player thus forming a GAN. This, like PNM, can be interpreted as a coevolutionary approach, but unlike PNM, it does not have any convergence guarantees.

Generally, explicit mixtures can bring advantages in two ways: (1) *Representation*: intuitively, a mixture of k neural networks could better represent a complex distribution than a single NN of the same size, and would be roughly on par with a single network that is k times as big. Arora et al. [3] show how to create such a bigger network using a ‘multi-way selector’. In preliminary experiments [not reported] we observed mixtures of simpler networks leading to better performance than a single larger network. (2) *Training*: Arora et al. use an architecture that is tailored to representing a mixture of components and train a single such network. We, in contrast, explicitly represent the mixture; given the observation that good solutions will take the form of a mixture, this is a form of domain knowledge that facilitates learning and convergence guarantees.

The most closely related paper that we have come across is by Grnarova et al. [15], which also builds upon game theoretic tools to give certain convergence guarantees. The main differences with our paper are as follows: 1) We clarify how zero-sum games relate to original GAN formulation, Wasserstein GAN objectives, and Goodfellow’s ‘heuristic trick’. 2) We provide a much more general form of convergence (to an RB-NE) that is applicable to *all* architectures, that only depends on the power to compute best responses, and show that PNM converges in this sense. We also show that if agents can compute an ϵ -best response, then the procedure converges to an ϵ -NE. 3) Grnarova et al. show that for a very specific GAN architecture their Alg. 1 converges to an ϵ -NE. This result is an instantiation of our more general theory: they assume they can compute exact (for G) and ϵ -approximate (for C) best responses; for such powerful players our Theorem 16 provides that guarantee. 4) Their Algorithm 2 does not provide guarantees.

Bounded rationality. The proposed notion of RB-NE is one of bounded rationality [37]. Over the years a number of different such notions have been proposed, e.g., see [34, 41]. Some of these also target agents in games. Perhaps the most well-known such a concept is the quantal response equilibrium [24]. Other concepts take into account an explicit cost of computation [17, 33], or explicitly limit the allowed strategy, for instance by limiting the size of finite-state

machines that might be employed [17]. However, these notions are motivated to explain *why* people might show certain behaviors or *how* a decision maker should use its limited resources. We on the other hand, take the why and how of bounded rationality as a given, and merely model the outcome of a resource-bounded computation (as the computable set $\mathcal{S}_i^{RB} \subseteq \mathcal{S}_i$). In other words, we make a minimal assumption on the nature of the resource-boundedness, and aim to show that even under such general assumptions we can still reach a form of equilibrium, an RB-NE, of which the quality can be directly linked (via Theorem 16) to the computational power of the agents.

8 CONCLUSIONS

We introduced GANGs—Generative Adversarial Network Games—a novel framework for representing adversarial generative models by formulating them as finite zero-sum games. The framework provides strong links to the rich literature in game theory, and makes available the rich arsenal of game theory solution techniques. It also clarifies the solutions that we try to find as saddle points in *mixed strategies* resolving the possibility of getting stuck in a local NE. As finite GANGs have extremely large action spaces we cannot expect to exactly (or ϵ -approximately) solve them. Therefore, we introduced a concept of bounded rationality, Resource-Bounded Nash Equilibrium (RB-NE). This notion is richer than the ‘local Nash Equilibria’ in that it captures not only failures of escaping local optima of gradient descent, but applies to any approximate best response computations, including methods with random restarts.

While in GANGs with mixed strategies gradient descent has no inherent problems (any local NE is a global NE), there are no known deep learning (i.e., gradient-based) methods to optimize mixed strategies with *very* large support sizes. However, in our RB-NE formulation, we can draw on a richer set of methods for solving zero-sum games [12, 13, 27, 31]. In this paper, we focus on PNM, which monotonically converges to an RB-NE, and we empirically investigated its effectiveness in search for good strategies in a setting where the real data generating process is a mixture of Gaussians. Our proof-of-concept results demonstrate this method indeed can deal well with typical GAN problems such as mode collapse, partial mode coverage and forgetting.

Future work. We presented a framework that can have many instantiations and modifications. For example, one direction is to employ different learning algorithms. Another direction could focus on modifications of PNM, such as to allow discarding “stale” pure strategies, which would allow the process to run for longer without being inhibited by the size of the resulting zero-sum “subgame” that must be maintained and repeatedly solved. The addition of fake uniform data as a guiding component suggests that there might be benefit of considering “deep-interactive learning” where there is deepness in the number of players that interact in order to give each other guidance in adversarial training. This could potentially be modelled by zero-sum polymatrix games [7].

ACKNOWLEDGMENTS

This research made use of a GPU donated by Nvidia.

REFERENCES

- [1] Martin Arjovsky and Léon Bottou. 2017. Towards Principled Methods for Training Generative Adversarial Networks. In *Proc. of the International Conference on Learning Representations (ICLR)*.
- [2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein Generative Adversarial Networks. In *Proc. of the International Conference on Machine Learning (ICML)*. 214–223.
- [3] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. 2017. Generalization and Equilibrium in Generative Adversarial Nets (GANs). In *Proc. of the International Conference on Machine Learning (ICML)*. 224–232.
- [4] S. Arora and Y. Zhang. 2017. Do GANs actually learn the distribution? An empirical study. *ArXiv e-prints* (June 2017). arXiv:cs.LG/1706.08224
- [5] Jean-Pierre Aubin. 1998. *Optima and equilibria: an introduction to nonlinear analysis*. Vol. 140. Springer Science & Business Media.
- [6] Branislav Bosanský, Christopher Kiekintveld, Viliam Lisý, and Michal Pechoucek. 2014. An Exact Double-Oracle Algorithm for Zero-Sum Extensive-Form Games with Imperfect Information. *J. Artif. Intell. Res.* 51 (2014), 829–866.
- [7] Yang Cai, Ozan Candogan, Constantinos Daskalakis, and Christos H. Papadimitriou. 2016. Zero-Sum Polymatrix Games: A Generalization of Minmax. *Math. Oper. Res.* 41, 2 (2016), 648–655.
- [8] Soumith Chintala, Aaron Courville, Emily Denton, Ian Goodfellow, Arthur Gretton, Yann LeCun, and Sebastian Nowozin. 2016. NIPS 2016 Workshop on Adversarial Training - Panel Discussion. <https://www.youtube.com/watch?v=WUoGbwGKE>. (2016).
- [9] Zihang Dai, Amjad Almahairi, Philip Bachman, Eduard H. Hovy, and Aaron C. Courville. 2017. Calibrating Energy-based Generative Adversarial Networks. In *Proc. of the International Conference on Learning Representations (ICLR)*.
- [10] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. 2017. Training GANs with Optimism. *ArXiv e-prints* (Oct. 2017). arXiv:cs.LG/1711.00141
- [11] Sevan G. Ficici and Jordan B. Pollack. 2003. A Game-Theoretic Memory Mechanism for Coevolution. In *Proc. of Genetic and Evolutionary Computation (GECCO)*. 286–297.
- [12] Dylan J Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Eva Tardos. 2016. Learning in games: Robustness of fast convergence. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*. 4734–4742.
- [13] Drew Fudenberg and David K Levine. 1998. *The theory of learning in games*. MIT press.
- [14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*. 2672–2680.
- [15] P. Grnarova, K. Y. Levy, A. Lucchi, T. Hofmann, and A. Krause. 2017. An Online Learning Approach to Generative Adversarial Networks. *ArXiv e-prints* (June 2017). arXiv:cs.LG/1706.03269
- [16] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C. Courville. 2017. Improved Training of Wasserstein GANs. *CoRR* abs/1704.00028 (2017). <http://arxiv.org/abs/1704.00028>
- [17] Joseph Y. Halpern, Rafael Pass, and Lior Seeman. 2014. Decision Theory with Resource-Bounded Agents. *Topics in Cognitive Science* 6, 2 (2014), 245–257.
- [18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2001. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.
- [19] Elad Hazan, Karan Singh, and Cyril Zhang. 2017. Efficient Regret Minimization in Non-Convex Games. In *International Conference on Machine Learning (ICML)*. 1433–1441.
- [20] F. Huszár. 2015. How (not) to Train your Generative Model: Scheduled Sampling, Likelihood, Adversary? *ArXiv e-prints* (Nov. 2015). arXiv:stat.ML/1511.05101
- [21] D. Jiwoong Im, H. Ma, C. Dongjoo Kim, and G. Taylor. 2016. Generative Adversarial Parallelization. *ArXiv e-prints* (Dec. 2016). arXiv:cs.LG/1612.04021
- [22] Wei Li, Melvin Gauci, and Roderich Groß. 2016. Turing learning: a metric-free approach to inferring behavior and its application to swarms. *Swarm Intelligence* 10, 3 (2016), 211–243.
- [23] Luchan Liu. 1996. The Invariance of Best Reply Correspondences in Two-Player Games. (1996). <http://www.cb.cityu.edu.hk/ef/getFileWorkingPaper.cfm?id=193>.
- [24] Richard D. McKelvey and Thomas R. Palfrey. 1995. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10, 1 (1995), 6 – 38.
- [25] John F. Nash. 1950. Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences of the United States of America* 36 (1950), 48–49.
- [26] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. 2016. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*. 271–279.
- [27] Frans A. Oliehoek, Edwin D. de Jong, and Nikos Vlassis. 2006. The Parallel Nash Memory for Asymmetric Games. In *Proc. Genetic and Evolutionary Computation (GECCO)*. 337–344.
- [28] Martin J. Osborne and Ariel Rubinstein. 1994. *A Course in Game Theory*. The MIT Press, Chapter Nash Equilibrium, 21 – 27.
- [29] A. Radford, L. Metz, and S. Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv e-prints* (Nov. 2015). arXiv:cs.LG/1511.06434
- [30] Alexander Rakhlin and Karthik Sridharan. 2013. Online Learning with Predictable Sequences. In *Proc. of the Annual Conference on Learning Theory (COLT)*. 993–1019.
- [31] Alexander Rakhlin and Karthik Sridharan. 2013. Optimization, Learning, and Games with Predictable Sequences. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*. 3066–3074.
- [32] Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. 2013. Characterization and computation of local Nash equilibria in continuous games. In *Communication, Control, and Computing, 2013 51st Annual Allerton Conference on*. IEEE, 917–924.
- [33] Ariel Rubinstein. 1986. Finite automata play the repeated prisoner’s dilemma. *Journal of Economic Theory* 39, 1 (1986), 83–96.
- [34] Stuart J Russell. 1997. Rationality and intelligence. *Artificial Intelligence* 94, 1-2 (1997), 57–77.
- [35] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved Techniques for Training GANs. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*. 2226–2234.
- [36] Yong-Siang Shih. 2017. GAN Implementations. <https://github.com/shaform/DeepNetworks>. (2017).
- [37] Herbert A Simon. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 69, 1 (1955), 99–118.
- [38] T. Unterthiner, B. Nessler, G. Klambauer, M. Heusel, H. Ramsauer, and S. Hochreiter. 2017. Coulomb GANs: Provably Optimal Nash Equilibria via Potential Fields. *ArXiv e-prints* (Aug. 2017). arXiv:1708.08819
- [39] J von Neumann. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100, 1 (1928), 295–320.
- [40] Junbo Jake Zhao, Michaël Mathieu, and Yann LeCun. 2017. Energy-based Generative Adversarial Network. In *Proc. of the International Conference on Learning Representations (ICLR)*.
- [41] Shlomo Zilberstein. 2011. Metareasoning and bounded rationality. In *Metareasoning: Thinking about Thinking*, M. Cox and A. Raja (Eds.). MIT Press.