Learning in Non-convex Games with an Optimization Oracle

Naman Agarwal NAMANAGARWAL@GOOGLE.COM
Alon Gonen AGONEN@CS.PRINCETON.EDU
Elad Hazan EHAZAN@CS.PRINCETON.EDU
Google AI Princeton & Princeton University

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

We consider online learning in an adversarial, non-convex setting under the assumption that the learner has an access to an offline optimization oracle. In the general setting of prediction with expert advice, (Hazan and Koren, 2016) established that in the optimization-oracle model, online learning requires exponentially more computation than statistical learning. In this paper we show that by slightly strengthening the oracle model, the online and the statistical learning models become computationally equivalent. Our result holds for any Lipschitz and bounded (but not necessarily convex) function. As an application we demonstrate how the offline oracle enables efficient computation of an equilibrium in non-convex games, that include GAN (generative adversarial networks) as a special case.

Keywords: Online Learning, Online Convex Optimization

1. Introduction

The setting of *online learning in games* is a fundamental paradigm which allows formulation of tasks such as spam detection, online routing, online recommendation systems, and more (Cesa-Bianchi and Lugosi, 2006; Hazan, 2016; Shalev-Shwartz, 2011). A key feature of this model is the ability of the environments to evolve over time, possibly in an adversarial manner. Consequently, this framework can be used to produce more robust learners compared to the classic stationary and statistical learning framework. A fundamental question investigated in recent literature is whether this robustness comes with a computational price. While it is well-known that any efficient online learner can be transformed into an efficient *statistical* (or *batch*) learner (Cesa-Bianchi et al., 2004), it is important to understand to what extent is the online model harder.

To enable a systematic comparison between the two models we must allow a reduction in the opposite direction. To this end we adopt the *offline optimization oracle* model suggested in (Hazan and Koren, 2016), where the online learner submits a sequence of loss functions and the oracle returns any minimizer of the cumulative loss. For the well-established setting of learning with expert advice, (Hazan and Koren, 2016) demonstrated an exponential gap between the oracle complexity in the online and the statistical settings.

In this paper we study the same question in the more general non-convex setting.[1] Deviating from (Hazan and Koren, 2016), we allow the learner to linearly perturb the objective submitted to the oracle. Arguably, adding a linear term to a non-convex function

1. We explain how to reduce the expert setting to the non-convex setting in Section 1.2.

should not increase the overall complexity of the oracle. Perhaps surprisingly, we show that this moderate modification renders the online adversarial setting computationally equivalent to the statistical setting. We show this by extending the powerful Follow-the-Perturbed-Leader (FTPL) meta-algorithm to the non-convex setting and derive a polynomial bound on its oracle complexity.

1.1. Setting and Main Result

1.1.1. Basic definitions and assumptions

Let $\mathcal{W} \subseteq \mathbb{R}^d$ be the decision set (a.k.a. hypothesis space in the statistical setting) with ℓ_∞-diameter at most D, and let $\mathcal{L} \subseteq \mathbb{R}^{\mathcal{W}}$ be the set of all G-Lipschitz functions w.r.t. the ℓ_1-norm. We assume that both G and D are polynomial in the ambient dimension d.[2]

Consider the setting of online learning, where an online algorithm predicts a point $w_t \in \mathcal{W}$ in iterative fashion and receives a feedback according to an adversarially chosen loss function $\ell_t \in \mathcal{L}$. The goal of the learner is to minimize the *average regret*, which is defined as the difference between the average loss of the learner and that of the best fixed point $w^* \in \mathcal{W}$ in hindsight. We define the *sample complexity* as the number of rounds required for attaining expected average regret at most ϵ.

The statistical setting differs from the online setting in two important aspects. a) We assume that the loss functions are drawn according to some unknown fixed distribution. b) The learner receives a sample of loss functions drawn according to the same distribution. Then it has to output a single predictor $\hat{w}$. The goal of the learner is minimize the expected *excess risk*, which is defined as $\mathbb{E}_\ell[\ell(\hat{w})] - \inf_{w\in\mathcal{W}} \mathbb{E}_\ell[\ell(w)]$. The sample complexity in this model is the size of a sample (of loss functions) that is required for attaining expected excess risk at most ϵ.

1.1.2. The offline oracle model

In order to compare between the online and the statistical models, we assume an access to two types of oracles:

1. **Value oracle** whose input is a pair $(w, \ell) \in \mathcal{W} \times \mathcal{L}$ and its output is $\ell(w)$.
2. **Offline optimization oracle** whose input consists of a sequence of loss functions $(\ell_1, \ldots, \ell_k) \in \mathcal{L}^k$ and a d-dimensional vector σ, and its output is output has the form

$$\hat{w} \in \operatorname{argmin}\{\sum_{i=1}^{k} \ell_i(w) - \sigma^\top w : \ w \in \mathcal{W}\} \ .$$

We define the oracle complexity as

$$\text{sample complexity} + \#\text{ of calls to value oracle} + \#\text{ of calls to offline oracle}$$

2. The choice of norm in our setting is inconsequential as norms are equivalent up to $poly(d)$.

1.1.3. Main result

Our online Algorithm 1 applies the offline oracle with a random linear perturbation σ whose coordinates are i.i.d. exponential random variables with parameter η. Our main result can be stated as follows.

Theorem 1 *The oracle complexity of Algorithm 1 is poly$(d, 1/\epsilon)$.*

Notably, both the loss functions and the domain $\mathcal{W}$ are not assumed to be convex. The oracle complexity in the statistical setting (under the same assumptions) is also poly$(d, 1/\epsilon)$.[3] We thus conclude that both statistical and the online oracle complexities for non-convex learning setting are polynomially equivalent. We deduce the following game theoretic result:

Corollary 2 *(informal) Convergence to equilibrium in two player zero-sum non-convex games is as hard as the corresponding offline best-response optimization problem.*

We elaborate on this implication and specify it to GANs in Section 4.

1.2. Related Work

Follow-the-perturbed-leader. The ubiquitous Follow-the-Perturbed-Leader (FTPL) algorithm (Hannan, 1957; Kalai and Vempala, 2004) is the canonical example of using an optimization oracle: the algorithm returns the result of a single optimization oracle call per iteration. Since its introduction, an extensive study of FTPL has yielded new insights and efficient variants in various different settings (e.g. (Hazan and Kale, 2012; Devroye et al., 2013; Van Erven et al., 2014; Cohen and Hazan, 2015)).

Online Convex Optimization. If the problem admits a convex structure, then the oracle complexity is polynomial in the dimension via bandit convex optimization (Cesa-Bianchi and Lugosi, 2006; Hazan, 2016; Bubeck et al., 2012). If one considers the number of oracle calls to the optimization oracle only, and does not have access to a value oracle, then it is still possible to obtain a polynomial bound on the oracle complexity. This is due to the fact that online convex optimization reduces to online linear optimization (Zinkevich, 2003), and this enables extension of FTPL to the convex case. However, this extension requires access to the gradient, which does not fall into our oracle model. We are not aware of any analysis of direct application of FTPL to a convex loss (i.e., without access to the gradients). In a sense, our treatment of the non-convex case gives the first direct analysis for FTPL to the convex case.

The experts setting: Overcoming the lower bound It is instructive to revisit the experts setting and understand why our result does not contradict the exponential lower bound of (Hazan and Koren, 2016). After all, one can easily embed the general experts problem in the d-dimensional hypercube for $d = \lceil \log N \rceil$ using the following standard technique:

1. Associate each vertex $z \in \{0,1\}^d$ with some expert $i(z)$.
2. Associate each $x \in [0,1]^d$ with a random expert according to $p(z) = \prod_{i=1}^{d}(z_i x_i + (1 - z_i)(1 - x_i))$.

3. This follows from standard covering argument

3. Perform optimization over $[0,1]^d$, where the loss of each $x \in [0,1]^d$ is $\sum_{z\in\{0,1\}^d} p(z)\ell(i(z))$.

It can be verified that the parameters G and D are polynomial in d, as required. Consequently, our main result applies to this setting as well.

Crucially, unlike our oracle model, (Hazan and Koren, 2016) does not allow a linear perturbation of the cumulative loss in this low-dimensional presentation. As it seems, this arguably moderate modification of the model rendered the offline-to-online reduction tractable.

Experts with low-dimensional structure. In the context of contextual bandits, (Dudik et al., 2017) formulate abstract conditions under which the randomness can be shared between the experts, and allow efficient regret minimization in the oracle complexity model.

(Gonen and Shalev-Shwartz, 2017) study stability in non-convex settings, and bound the stability rate of ERM for strict saddle problems. In this paper we derive stable algorithms under much more moderate assumptions.

Generative adversarial networks. Several works have studied GANs in the regret minimization framework (e.g. (Schuurmans and Zinkevich, 2016; Kodali et al., 2017; Hazan et al., 2017)). We provide the first evidence that achieving equilibrium in GANs can be reduced to the offline problems associated with the players.

1.3. Overview and Techniques

1.3.1. Why standard approaches do not work?

A common approach which works well in the convex setting is to apply the Follow-the-Regularized-Leader (FTRL) with ℓ_2-regularization:

$$w_t \in \operatorname{argmin}\left\{\sum_{i<t}\ell_i(w) + \eta\|w\|^2\right\}, \quad \eta \approx T^\alpha,\ \alpha \in (0,1)\ .$$

In the convex case ℓ_2-regularization stabilizes the solution by pushing it towards zero. However, we argue that in the non-convex setting, this approach does not help. To demonstrate this claim, consider a 1-dimensional setting, where the loss functions have the form $w \mapsto (\sigma(wx) - y)^2$, where $\sigma(x) = \max\{x, 0\}$ is the ReLU function and $x \in [-1,1], y \in [0,1]$. Due to the ReLU term, the magnitude of the loss incurred by classifying x negatively is not important (i.e., there is no difference between $wx = -10^{-6}$ and $wx = -1$). Informally, if all x's are bounded away from zero, we mostly care for the ratio between positive and negative examples. Therefore, adding ℓ_2-regularization does not make solutions near zero more appealing. It is not hard to formalize this argument and show that FTRL with ℓ_2 (or ℓ_1) regularization can not yield sublinear regret.

1.3.2. Extending FTPL to the non-convex case

Our result is proved by extending the Follow-The-Perturbed-Leader algorithm to the non-convex setting. As we detail in the preliminaries section, online learnability requires algorithmic stability between consecutive rounds. For linear loss functions, (Kalai and Vempala,

2004) proved that linear perturbation of the loss stabilized the loss function itself, and consequently the minimizer is stable as well. The proof relies heavily on the fact that the perturbation and the loss function are of the same type.

In the non-convex case, we can not hope to stabilize the loss itself using a linear perturbation. Nevertheless, our main contribution is to establish that the randomness injected by FTPL does stabilize the predictions of the learner. We prove this result by investigating how the outputs of FTPL change as we vary the the noise vector $\sigma \in \mathbb{R}^d_{\geq 0}$. In the 1-dimensional case, this investigation yields a useful monotonicity property which helps us bounding the expected distance between consecutive minimizers. While the general d-dimensional introduces some challenges, we are able to effectively reduce the analysis to the 1-dimensional setting by varying each coordinate of the noise separately.

2. Preliminaries

2.1. Online to batch conversion

The following well-known result due to (Cesa-Bianchi et al., 2004) tells us that the online sample complexity dominates the batch sample complexity. The intuition that online learning is at least as hard as batch learning is formalized by the following online-to-batch theorem.

Theorem 3 *(Cesa-Bianchi et al., 2004) Suppose that $\mathcal{A}$ is an online learner with $\frac{\mathbb{E}[\mathrm{Regret}_T]}{T} \leq \epsilon(T)$ for any T. Consider the following algorithm for the batch setting: given a sample $(\ell_1, \ldots, \ell_n) \sim \mathcal{P}^n$, the algorithm applies $\mathcal{A}$ to the sample in an online manner. Thereafter, it draws a random round $j \in [n]$ uniformly at random and returns $\hat{w} = w_j$. Then the expected excess risk of the algorithm, $\mathbb{E}[L(\hat{w})] - L(w^\star)$, is at most $\epsilon(T)$.*

2.2. Online learning via stability

The main challenge in online learning stems from the fact that the learner has to make a decision before observing the adversarial action. Intuitively, we expect that the performance after shifting the actions of the learner by one step (i.e. considering the loss $\ell_t(w_{t+1})$ rather than $\ell_t(w_t)$) to be optimal. This view suggests that online learning is all about balancing between optimal performance w.r.t. previous rounds and ensuring stability between consecutive rounds. Similarly to the statistical setting, the most common algorithmic tool for achieving stability is regularization. In particular, the well-established Follow-the-Regularized-Leader is a meta-algorithm whose instances are determined by choosing a concrete regularization function. Precisely, given a regularizer $R : \mathbb{R}^d \to \mathbb{R}$, the t-iterate of the algorithm is

$$w_t = \operatorname{argmin} \left\{ \sum_{i<t} \ell_i(w) + R(w) \right\} .$$

The next well-known lemma provides a systematic approach for analyzing *Follow-the-Regularized-Leader*-type algorithms.

Lemma 4 *(FTL-BTL (Kalai and Vempala, 2004)) The regret of Follow-the-Regularized-Leader is at most*

$$\mathbb{E}[\text{Regret}_T] \leq \mathbb{E}[R(w^\star) - R(w_1)] + \sum_{i=1}^{T} \mathbb{E}[\ell_t(w_t) - \ell_t(w_{t+1})] \ ,$$

where $w^\star = \text{argmin}\{\sum_{t=1}^{T} \ell_i(w) : \ w \in \mathcal{W}\}$.

2.3. The exponential distribution

We use the following properties of the exponential distribution.

Lemma 5 *Let* X *be an exponential random variable with parameter* η.[4] *The following properties hold: a) for any* $s \in \mathbb{R}$, $P(X \geq s) = \exp(-\eta s)$. *b) Memorylessness: for any* $s, q \in \mathbb{R}$, $P(X \geq q + s | X \geq q) = P(X \geq s)$. *c) if* $X_1, \ldots, X_d$ *are i.i.d. with* $X_i \sim Exp(\eta)$, *then* $\mathbb{E}[\|(X_1, \ldots, X_d)\|_\infty] \leq \eta^{-1}(\log(d) + 1)$.

3. Non-convex FTPL

In this section we present and analyze the non-convex FTPL method presented in Algorithm 1. Our analysis completes the proof of our main theorem (Theorem 1). Along the proof we distinguish between the one-dimensional and the general d-dimensional case. For the former case we obtain better regret bound in terms of the dependence on the horizon parameter T. Omitted proofs are provided in the Appendix.

Algorithm 1 Non-convex FTPL

Parameter: $\eta > 0$

for $t = 1$ **to** T **do**

Draw i.i.d. random vector $\sigma_t \sim (Exp(\eta))^d$

Prediction at time t:

$$w_t \in \text{argmin} \left\{ \sum_{i<t} \ell_i(w) - \sigma_t^\top w : \ w \in \mathcal{W} \right\} \ , \tag{1}$$

end for

3.1. Reduction to oblivious setting

To simplify the presentation we make the following standard modification:

1. The adversary is oblivious in the sense that the sequence $(\ell_t)_{t=1}^T$ is chosen in advance.
2. This allows us to analyze a slightly different algorithm which draws only a single noise vector $\sigma \sim Exp(\eta)^d$ rather than drawing a fresh noise vector on every round.

It follows from (Cesa-Bianchi and Lugosi, 2006)[Lemma 4.1] that proving regret bounds for this variant translates into asymptotically equivalent (expected) regret bounds for non-oblivious adversaries using Algorithm 1.

4. That is, X has density $p(x) = \eta \exp(-\eta x)$.

3.2. Main Lemma

Throughout this section we use the notation $w_t(\sigma)$ to emphasize that w_t as defined in (1), is determined by the noise vector σ. Following Lemma 4 we would like to to establish a bound on the expected instability at time t, i.e. $\mathbb{E}[\ell_t(w_t(\sigma)) - \ell_t(w_{t+1}(\sigma))]$. This is bounded above by $G \cdot \mathbb{E}\|w_t(\sigma) - w_{t+1}(\sigma)\|_1$. Note that the distance between w_t and w_{t+1} is ill-defined since both w_t and w_{t+1} are not unique. However, as we show below, we will be able to derive a uniform bound on the distance between any consecutive minimizers for **every** choice of minimizers. Note that this not really needed. As we are primarily interested in stability with respect to the function value, we can make any assumptions on the tie-breaking mechanism. However, we found it both interesting and surprisingly easier to prove the stronger result.

Lemma 6 *Fix an iteration t and let $\delta > 0$ be a* margin *parameter. There is a tie-breaking rule for choosing minimizers such that $\mathbb{E}[\|w_t(\sigma) - w_{t+1}(\sigma)\|_1] = O\left(\frac{\mathrm{poly}(d)\eta}{\delta} + d\delta\right)$. In the one-dimensional case we obtain the improved bound $\mathbb{E}[|w_t - w_{t+1}|] = O(\eta)$.*

Proof (of Theorem 1) We start with the multidimensional case. Applying the FTL-BTL lemma (Lemma 4) with the regularizer $R(w) = -\sigma^\top w$ and using Hölder inequality, we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{Regret}_T] &\le \mathbb{E}\left[\|\sigma\|_\infty \cdot \|w^\star - w_1\|_1\right] + G\sum_{t=1}^{T}\mathbb{E}[\|w_t(\sigma) - w_{t+1}(\sigma)\|_1] \\ &\le \mathbb{E}\left[\|\sigma\|_\infty\right] D + G\sum_{t=1}^{T}\mathbb{E}[\|w_t(\sigma) - w_{t+1}(\sigma)\|_1]\ .\end{aligned}$$

For the multidimensional case, we use that $\mathbb{E}[\|\sigma\|_\infty] \le \eta^{-1}(\log d + 1)$, $D, G \in \mathrm{poly}(d)$, and apply Lemma 6 to obtain

$$\mathbb{E}[\mathrm{Regret}_T] \le \mathrm{poly}(d)\left((\eta^{-1}(\log d + 1) + T(\eta\delta^{-1} + \delta))\right)\ .$$

By setting $\eta = T^{-2/3}$ and $\delta = T^{-1/3}$, we obtain the regret bound $\mathbb{E}[\mathrm{Regret}_T] \le O(T^{2/3}\mathrm{poly}(d))$. Online-to-batch conversion yields a sample complexity bound of $O\left(\frac{\mathrm{poly}(d)}{\epsilon^3}\right)$.

In the 1-dimensional case we simply set $\eta = T^{-1/2}$ to obtain $\mathbb{E}[\mathrm{Regret}_T] = O(T^{1/2})$. This translates into a sample complexity bound of $O\left(\frac{\mathrm{poly}(d)}{\epsilon^2}\right)$. ■

3.3. Proof of Lemma 6

We begin with the following lemma which provides a bound on the gap between minimizers with respect to the change in noise parameter σ.

Lemma 7 *For any two functions $f_1, f_2 : \mathcal{W} \to \mathbb{R}$ and vectors $\sigma_1, \sigma_2 \in \mathbb{R}^d$, let*

$$w_i(\sigma_i) \in \mathrm{argmin}\left\{f_i(w) - \sigma_i^\top w\right\}, \qquad i = 1, 2.$$

Letting $f = f_1 - f_2$ and $\sigma = \sigma_1 - \sigma_2$, we have that

$$f(w_1(\sigma_1)) - f(w_2(\sigma_2)) \le \sigma^\top(w_1(\sigma_1) - w_2(\sigma_2)) \tag{2}$$

Proof Using optimality conditions for $w_i(\sigma_i)$, we have that

$$f_1(w_1(\sigma_1)) - \sigma_1^T w_1(\sigma_1) \leq f_1(w_2(\sigma_2)) - \sigma_1^T w_2(\sigma_2)$$

$$f_2(w_2(\sigma_2)) - \sigma_2^T w_2(\sigma_2) \leq f_2(w_1(\sigma_1)) - \sigma_2^T w_1(\sigma_1)$$

Adding the above two inequalities and rearranging finishes the proof. ■

We now provide the proof of Lemma 6 in the two considered cases.

Proof (of Lemma 6: one-dimensional case) We wish to use Lemma 7. For any time t, consider substituting $f_1(w) = \sum_{i<t} l_i(w)$, $f_2(w) = \sum_{i<t+1} l_i(w)$, $\sigma_2 = \sigma$ and $\sigma_1 = \sigma' \triangleq \sigma + 2G$. We immediately get that

$$l_t(w_t(\sigma')) - l_t(w_{t+1}(\sigma)) \leq 2G(w_t(\sigma') - w_{t+1}(\sigma))$$

Using the fact that l_t is G-lipschitz, we get that

$$-G|w_t(\sigma') - w_{t+1}(\sigma)| \leq l_t(w_t(\sigma')) - l_t(w_{t+1}(\sigma)) \leq 2G(w_t(\sigma') - w_{t+1}(\sigma))$$

which immediately implies that $w_t(\sigma') \geq w_{t+1}(\sigma)$. Similar calculations show that $w_{t+1}(\sigma') \geq w_t(\sigma)$ and $w_t(\sigma') \geq w_t(\sigma)$.

For the rest of the proof we will omit the dependence on t as it will be clear from context. We denote by $w_{\min}(\sigma) = \min\{w_t(\sigma), w_{t+1}(\sigma\}\}$, $w_{\max}(\sigma) = \max\{w_t(\sigma), w_{t+1}(\sigma)\}$. First we observe that

$$\mathbb{E}[|w_t(\sigma) - w_{t+1}(\sigma)|] = \mathbb{E}[w_{\max}(\sigma)] - \mathbb{E}[w_{\min}(\sigma)] \ .$$

Secondly the computation above implies that

$$w_{\min}(\sigma') \geq w_{\max}(\sigma) \ . \tag{3}$$

This powerful monotonicity property (see Figure 1) is now used to lower bound $\mathbb{E}[w_{\min}(\sigma)]$ in terms $\mathbb{E}[w_{\max}(\sigma)]$. Letting $\sigma' = \sigma + 2G$, we have

$$\begin{aligned}
\mathbb{E}[w_{\min}(\sigma)] &= \int_{\sigma=0}^{2G} \eta \exp(-\eta\sigma) w_{\min}(\sigma)\, d\sigma + \int_{\sigma>2G} \eta \exp(-\eta\sigma) w_{\min}(\sigma)\, d\sigma \\
&\geq (1 - \exp(-2\eta G))(\mathbb{E}[w_{\max}(\sigma)] - D) + \int_{\sigma>0} \eta \exp(-\eta(\sigma')) w_{\min}(\sigma')\, d\sigma \\
&\geq (1 - \exp(-2\eta G))(\mathbb{E}[w_{\max}(\sigma)] - D) + \int_{\sigma>0} \eta \exp(-\eta(\sigma')) w_{\max}(\sigma)\, d\sigma \\
&= (1 - \exp(-2\eta G))(\mathbb{E}[w_{max}(\sigma)] - D) + \exp(-2\eta G)\mathbb{E}[w_{\max}(\sigma)] \\
&= \mathbb{E}[w_{\max}(\sigma)] - D(1 - \exp(-2\eta G)) \geq \mathbb{E}[w_{\max}(\sigma)] - 2\eta DG \ ,
\end{aligned}$$

where the second inequality uses Equation 3 and the last inequality uses the inequality $\exp(x) \geq 1 + x$. ■

Proof (of Lemma 6:multiple dimensions) Once again we wish to use Lemma 7. For any time t and any coordinate k, consider substituting $f_1(w) = \sum_{i<t} l_i(w)$, $f_2(w) =$

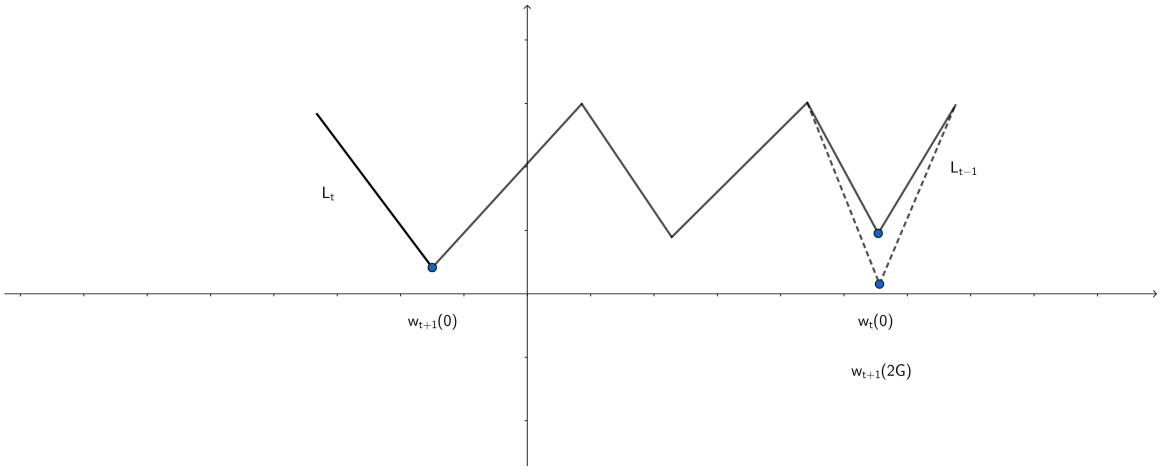

Figure 1: Illustration of the monotonicity property used in the Proof of Lemma 6: The unperturbed minimizer of L_t (solid line), denoted $w_{t+1}(0)$, can be significantly smaller than the unperturbed minimizer of L_{t-1} (dashed line), $w_t(0)$. This can be balanced by increasing the noise parameter corresponding to w_{t+1}.

$\sum_{i<t+1} l_i(w)$, $\sigma_2 = \sigma$ and $\sigma_1 = \sigma' \triangleq \sigma + 3B\delta^{-1} \cdot e_k$ (where e_k is the k^{th} vector in the canonical basis). We immediately get that

$$l_t(w_t(\sigma')) - l_t(w_{t+1}(\sigma)) \leq 3B\delta^{-1}(w_{t,k}(\sigma') - w_{t+1,k}(\sigma)) \ ,$$

where $w_{t,k}$ is the k-th coordinate of w_t. Using the fact that the range of l_t is $[-B, B]$, we get that

$$-2B \leq l_t(w_t(\sigma')) - l_t(w_{t+1}(\sigma)) \leq 3B\delta^{-1}(w_{t,k}(\sigma') - w_{t+1,k}(\sigma))$$

which immediately implies that $w_{t,k}(\sigma') \geq w_{t+1,k}(\sigma) - \delta$. A similar calculation also derives that $w_{t+1,k}(\sigma') \geq w_{t,k}(\sigma) - \delta$.

Now for any $k \in [d]$, let $w_{k,\min}(\sigma) = \min\{w_{t,k}(\sigma), w_{t+1,k}(\sigma)\}$, $w_{\max}(\sigma) = \max\{w_{t,k}(\sigma), w_{t+1,k}(\sigma)\}$. First we observe that

$$\mathbb{E}[\|w_t(\sigma) - w_{t+1}(\sigma)\|_1] = \sum_{k=1}^{d} (\mathbb{E}[w_{k,\max}(\sigma)] - \mathbb{E}[w_{k,\min}(\sigma)]$$

Secondly the calculation above implies that for all k

$$w_{k,\min}(\sigma + 3B\delta^{-1}e_k) \geq w_{k,\max}(\sigma) - \delta \tag{4}$$

Now fix a coordinate $k \in [d]$ along with all noise coordinates σ_j for $j \neq k$. Denote by $\mathbb{E}_{-k}$ the corresponding conditional expectation. Up to the additional margin term δ, lower bounding $\mathbb{E}_{-k}[w_{k,\min}]$ in terms of $\mathbb{E}_{-k}[w_{k,\max}]$ reduces to the one-dimensional case; letting

$q = 3B\delta^{-1}$ and $\mu(x) = \eta \exp(-\eta x)$, we have

$$\begin{aligned}
\mathbb{E}_{-k}[w_{k,\min}(\sigma_k)] &= \int_{\sigma_k=0}^{q} \mu(\sigma_k) w_{k,\min}(\sigma_k)\, d\sigma_k + \int_{\sigma_k>q} \mu(\sigma_k) w_{k,\min}(\sigma_k)\, d\sigma_k \\
&\geq (1-\exp(-q\eta))(\mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - D) + \int_{\sigma_k>0} \mu(\sigma_k+q) w_{k,\min}(\sigma_k+q)\, d\sigma_k \\
&\geq (1-\exp(-q\eta))(\mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - D) + \int_{\sigma_k>0} \mu(\sigma_k+q)(w_{k,\max}(\sigma_k)-\delta)\, d\sigma_k \\
&= (1-\exp(-q\eta))(\mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - D) + \exp(-q\eta)(\mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - \delta) \\
&\geq \mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - D(1-\exp(-q\eta)) - \delta \geq \mathbb{E}_{-k}[w_{k,\max}(\sigma_k)] - 3B\eta\delta^{-1}D - \delta\ .
\end{aligned}$$

The second inequality uses Equation 4 and the last inequality follows by substituting $q = 3B\delta^{-1}$ and using the inequality $\exp(x) \geq 1 + x$. Since the above holds for any fixed $\sigma_{-k} = (\sigma_j)_{j\neq k}$, the unconditioned expectations also satisfy

$$\mathbb{E}[w_{k,\min}(\sigma)] \geq \mathbb{E}[w_{k,\max}(\sigma)] - \frac{\mathrm{poly}(d)\eta}{\delta} - \delta\ .$$

Summing over all coordinates we conclude the bound. ■

4. Implications to Non-convex Games

Consider the following formulation of a non-convex zero-sum game. Let $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, where $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ are compact with diameter at most D. The x-th player wishes to minimize F and whereas the y-th player wishes to maximize F. We assume that for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, both $F(\cdot, y)$ and $-F(x, \cdot)$ are G-Lipschitz and B-bounded. A known approach for achieving equilibrium is to apply (for each of the players) an online method with vanishing average regret. Precisely, on each round t both players choose a pair (x_t, y_t) which induces the losses $F(x_t, y_t)$ and $-F(x_t, y_t)$, respectively. Finally, we draw a random index $[j] \in [T]$ and output the pair $(\hat{x}, \hat{y}) \triangleq (x_j, y_j)$. By endowing the players with access to an offline oracle and playing according to non-convex FTPL we can reach approximate equilibrium.

Theorem 8 *Suppose that both the x-player and the y-player have an access to an offline oracle and play according to non-convex FTPL (Algorithm 1). Given $\epsilon > 0$, let $T \in \mathrm{poly}(d)/\epsilon^3$ such that the expected average regret of non-convex FTPL is at most ϵ. Then, $(\hat{x}, \hat{y})$ forms an ϵ-approximated equilibrium, i.e., for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,*

$$\mathbb{E}[F(\hat{x}, \hat{y})] \leq \mathbb{E}[F(x, \hat{y})] + \epsilon, \quad \mathbb{E}[F(\hat{x}, \hat{y})] \geq \mathbb{E}[F(\hat{x}, y)] - \epsilon\ .$$

Note that the players can use their offline oracle to amplify their confidence and achieve an equilibrium with high probability. The proof is provided in the appendix.

4.1. Implication to GANs

In particular, we consider the case where the x-th player is a *generator*, who produces synthetic samples (e.g. images), whereas the y-th player acts as a *discriminator* by assigning

scores to samples reflecting the probability of being generated from the true distribution. Formally, by choosing a parameter $x \in \mathcal{X}$ and drawing a random noise z, the x-th player produces a sample denote $G_x(z)$. Conversely, the y-th player chooses a parameter $y \in \mathcal{Y}$ and assign the score $D_y(G_x(z)) \in [0, 1]$ to the sample $G_x(z)$. The function F usually corresponds to the log-likelihood of mistakenly assigning an high score to a synthetic example and vice versa. It is reasonable to assume that F is Lipschitz and bounded w.r.t. the network parameters. As a result, efficient convergence to GANs is established by assuming an access to an offline oracle.

5. Discussion

Our work establishes a computational equivalence between online and statistical learning in the non-convex setting. We shed light on the hardness result of (Hazan and Koren, 2016) by demonstrating that online learning is significantly more difficult than statistical learning only when no structure is assumed.

One interesting direction for further investigation is to refine the comparison model and study the polynomial dependencies more carefully. One obvious question is to understand the gap in terms of the horizon parameter T between the regret bounds for the one-dimensional and the multidimensional settings.

Acknowledgements

We thank Karan Singh for recognizing a bug in our original proof and several discussions. We also thank Alon Cohen and Roi Livni for fruitful discussions. Elad Hazan acknowledges funding from NSF award Number 1704860.

References

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1): 1–122, 2012.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge university press, 2006. ISBN 9780511546921. doi: 10.1017/CBO9780511546921.

Nicolò Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the Generalization Ability of Online Learning Algorithms for Pairwise Loss Functions. *IEEE Transactions on Information Theory*, 50:2050—-2057, 2004. URL http://arxiv.org/abs/1305.2505.

Alon Cohen and Tamir Hazan. Following the perturbed leader for online structured learning. In *International Conference on Machine Learning*, pages 1034–1042, 2015.

Luc Devroye, Gábor Lugosi, and Gergely Neu. Prediction by random-walk perturbation. In *Conference on Learning Theory*, pages 460–473, 2013.

Miroslav Dudik, Nika Haghtalab, Haipeng Luo, Robert E. Schapire, Vasilis Syrgkanis, and Jennifer Wortman Vaughan. Oracle-efficient online learning and auction design.

In *Annual Symposium on Foundations of Computer Science - Proceedings*, 2017. ISBN 9781538634646. doi: 10.1109/FOCS.2017.55.

Alon Gonen and Shai Shalev-Shwartz. Fast Rates for Empirical Risk Minimization of Strict Saddle Problems. In Ohad Shamir and Satyen Kale, editors, *Proceedings of the 2017 Conference on Learning Theory*, pages 1043—-1063. PMLR, 2017. URL http://arxiv.org/abs/1701.04271.

James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

Elad Hazan. Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016. ISSN 2167-3888. doi: 10.1561/2400000013. URL http://ocobook.cs.princeton.edu/OCObook.pdfhttp://www.nowpublishers.com/article/Details/OPT-013.

Elad Hazan and Satyen Kale. Online submodular minimization. *Journal of Machine Learning Research*, 13(Oct):2903–2922, 2012.

Elad Hazan and Tomer Koren. The Computational Power of Optimization in Online Learning. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 128–141. ACM, 2016. URL https://arxiv.org/pdf/1504.02089.pdf.

Elad Hazan, Karan Singh, and Cyril Zhang. Efficient regret minimization in non-convex games. In *International Conference on Machine Learning*, pages 1433–1441, 2017.

Adam Kalai and Santosh Vempala. Efficient Algorithms for Online Decision Problems. *Journal of Computer and System Sciences*, 2004.

Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On Convergence and Stability of GANs. *arXiv preprint arXiv:1705.07215*, 2017. URL https://github.com/kodalinaveen3/DRAGANhttp://arxiv.org/abs/1705.07215.

Dale Schuurmans and Martin A Zinkevich. Deep learning games. In *Advances in Neural Information Processing Systems*, pages 1678–1686, 2016.

Shai Shalev-Shwartz. Online Learning and Online Convex Optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2011. ISSN 1935-8237. doi: 10.1561/2200000018. URL http://www.nowpublishers.com/article/Details/MAL-018.

Tim Van Erven, Wojciech Kotłowski, and Manfred K Warmuth. Follow the leader with dropout perturbations. In *Conference on Learning Theory*, pages 949–974, 2014.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.