

Non-convex Online Optimization With an Offline Oracle

Sohrab Madani, smadani2@illinois.edu

May 10, 2019

Abstract

In this project, we will look at the problem of online optimization in the non-convex setting, assuming that the player has access to an offline oracle. As we will see, it has recently been proven that it is possible to achieve an $O(\frac{1}{\sqrt{T}})$ bound for the expected average regret. The algorithm involves the well-known Follow the Perturbed Leader algorithm, with a slightly strengthened oracle.

1 Introduction

Online optimization refers to the set of problems that can be modeled as follows. The notations follow the main reference [1], which is referred to as the default paper in throughout this report.

- For some dimension $d > 0$, a decision set $\mathcal{W} \subset \mathbb{R}^d$. \mathcal{W} is assumed to be bounded with a diameter $\leq D$ (in the sense of ℓ_1 -norm).
- Set of possible loss functions $\mathcal{L} \subset \mathbb{R}^{\mathcal{W}}$. Here we assume only that \mathcal{L} is G -Lipschitz for some $G > 0$, and that for all $\ell \in \mathcal{L}$, $\|\ell\|_{\infty} \leq B$ for some $B > 0$.
- We also make the assumption that D and G are polynomial in d .
- The game involves a player and an adversary, and it proceeds as follows:
 - At time $t = 1$, the player chooses an arbitrary $w_1 \in \mathcal{W}$.
 - At time $t > 1$, the player chooses w_t according to losses up to $t - 1$.
 - At time t , and after the player chooses w_t the adversary chooses a function $\ell_t \in \mathcal{L}$, and the player suffers the loss $\ell_t(w_t)$.
 - The game continues until we reach $t = T$.
- The goal is for the player to be able to minimize the expected average regret, which is defined as the difference between the expected sum of losses from $t = 1$ to $t = T$, and that of *single best decision in hindsight*. More precisely, regret is defined as

$$\text{regret} = \sum_{t=1}^T \ell_t(w_t) - \inf_{w \in \mathcal{W}} \sum_{t=1}^T \ell_t(w). \quad (1)$$

- Sample Complexity is defined as $\min\{T : \mathbb{E}(\frac{\text{regret}}{T}) < \epsilon\}$, i.e. the minimum T to reach an ϵ bound for the expected average regret.

As we saw in the course notes Theorem 11.1, assuming \mathcal{L} and using gradient descent to find w_t results in a $\frac{1}{\sqrt{T}}$ bound for the expected average regret.

We assume that the player has access to two oracles. First, a value oracle, where given the input $(w, \ell) \in \mathcal{W} \times \mathcal{L}$, it outputs $\ell(w)$, and second, an offline optimization oracle, where given a sequence of loss functions ℓ_1, \dots, ℓ_k and some vector $\sigma \in \mathbb{R}^d$, it outputs

$$\min_{w \in \mathcal{W}} \left[\left(\sum_{i=1}^k \ell_i(w) \right) - \sigma^\top w \right]. \quad (2)$$

Note. the extra assumption of the oracle being able to *perturb* the minimization using a linear noise is the extra assumption that this paper makes which makes it possible to prove the $\frac{1}{\sqrt{T}}$ bound for the expected average regret, by bounding the difference between consecutive decisions of the player.

We measure the complexity in terms of **oracle complexity**, which is defined as the sum of sample complexity, and sum of calls to the value oracle, and offline optimization oracle.

2 Main Algorithm

2.1 Follow the Leader

Follow the leader is a well-known and well-studied algorithm in online optimization. The algorithm works as follows. At time $t = 1$, the player chooses an arbitrary $w \in \mathcal{W}$, and for $t > 1$, he chooses $w_t = \arg \min_{w \in \mathcal{W}} [\ell_1(w) + \dots + \ell_{t-1}(w)]$, which is somewhat similar to the ERM algorithm, in the sense that the player is trying to minimize the regret with respect to the losses he has seen so far.

2.2 Follow the Regularized Leader

The standard Follow the Leader has instability issues which prevents the algorithm to reach a sub-linear expected average regret ($\leq o(T)$). This can be seen with the example where $d = 2$, $\ell_t(w_t) = z_t^\top w_t$, $z_1 = (0.5, 0)$, and for $t > 1$, z_i alternates between $(0, 1)$ and $(1, 0)$. Now with a wrong initiation w_1 , one can check that the algorithm always predicts corresponding to the worst loss. In convex settings, this issue is dealt with by adding a regularizer function to the objective, i.e.,

$$\min_{w \in \mathcal{W}} \left[\sum_{i=1}^{t-1} \ell_i(w) \right] + R(w) \quad (3)$$

where R is the regularizer function. This will not work in our case, since ℓ_t is not assumed to be convex.

To see this, consider the 1-D case where the loss function is of the form $(\max(wx, 0) - y)^2$, where $x \in [-1, 1]$ and $y \in [0, 1]$. Now let us assume that we have added a regularization term $\eta \|w\|^2$ to the objective function. Because of the $\max(wx, 0)$ term, it does not matter to

the loss function *how much* negative wx is (for instance $wx = -\epsilon$ is the same as $wx = -1$). this means that if x 's are mostly distributed away from zero, the ℓ_2 regularization does not motivate the algorithm to find solutions near to zero.

2.3 Follow the Perturbed Leader

In this variation, a linear random noise function is added to the objective function so as to ensure stability. More specifically, the objective function in this case becomes

$$\min_{w \in \mathcal{W}} \left[\sum_{i=1}^{t-1} \ell_i(w) \right] - \sigma^\top w \quad (4)$$

where $\sigma \in \mathbb{R}^d$ is the noise vector. We assume that coordinates of σ all have the i.i.d exponential distribution with parameter η . The proposed algorithm of the paper is as follows:

Algorithm 1 Non-Convex Follow the Perturbed Leader

Parameter $\eta > 0$

for $t = 0$ to T **do**

 Draw i.i.d random vector σ_t with distribution $\exp(\eta)^d$

 Predict at time t :

$$w_t \in \arg \min \left[\left(\sum_{i=1}^{t-1} \ell_i(w) \right) - \sigma_t^\top w \right]$$

end for

which \in is used (instead of $=$) as there may be more than one minimizer for the objective function of the oracle. The following is the main theorem of the paper.

Theorem 1. The Oracle Complexity of Algorithm 1 is $\text{poly}(d, \frac{1}{\epsilon})$.

Note that the

3 Proof Sketch

We break down the proof in to several steps, and go through each separately.

3.1 Reduction to Oblivious Setting

The assumption here is that the adversary observes the player's choice and then outputs a loss function. This is referred to as the non-oblivious setting, as opposed to the oblivious setting, where all the choices of the adversary are made before the game starts (i.e. ℓ_1, \dots, ℓ_T are predetermined). According the Lemma 4.1 in [3], any bound on the expected average regret of a game in the oblivious setting hold asymptotically true for the non-oblivious game. Therefore, for our purpose, we can assume that the adversary is oblivious. this also allows us to draw one single noise vector in the beginning and use it for every round (as opposed to drawing a fresh sample each time).

3.2 Reduce to bounding consecutive decisions

In [4], the authors prove the following lemma, which links the problem of bounding the expected average regret to stability and initial error:

$$\mathbb{E}(\text{regret}_T) \leq \mathbb{E}[R(w^*) - R(w_1)] + \sum_{t=1}^T \mathbb{E}[\ell_t(w_t) - \ell_t(w_{t+1})] \quad (5)$$

where $R(w)$ is the regularizer in the Follow the Regularized Leader setting, and w^* is the best decision in hindsight, i.e. $w^* = \arg \min\{\sum_{t=1}^T \ell_t(w) : w \in \mathcal{W}\}$.

In our case, $R(w) = \sigma^\top w$. Therefore we have

$$\mathbb{E}|R(w) - R(w_1)| = \mathbb{E}|\sigma^\top(w^* - w_1)| \leq \mathbb{E}[\|\sigma\|_\infty \|w^* - w_1\|_1] \leq \eta^{-1}(\log d + 1)D,$$

where we have used the fact that the diameter of \mathcal{W} is $\leq D$, and that the expected value of the maximum of d i.i.d exponential random variables with parameter η is bounded by $\eta^{-1}(\log d + 1)$.

From the G -Lipschitz property of the functions, for the terms inside the sum, we have $|\ell_t(w_t) - \ell_t(w_{t+1})| \leq G\|w_t - w_{t+1}\|_1$. Therefore the only thing that remains is to bound two consecutive decisions.

It is worthy to mention that it is possible to get a bound using the diameter of the decision set \mathcal{W} : $\|w_t - w_{t+1}\|_1 \leq D$. This would yield

$$\mathbb{E}((\text{regret})_T) \leq \eta^{-1}(\log d + 1)D + TGD$$

which, when divided by D , is clearly not sub-linear. Therefore, we have to find a smarter bound. Hopefully, we will be able to find a bound depending on η in both terms to be able to choose an optimum value.

3.3 The Final Step

The final step is to bound $\|w_t - w_{t+1}\|_1$. The paper does this in two steps, once in 1-D and once in the general d dimension case. The techniques used here are very similar, and we will go over the proof of the 1-D case here.

The paper uses the following lemma to find a bound: **Lemma 1.** For any two functions f_1 and $f_2 : \mathcal{W} \rightarrow \mathbb{R}$, and vectors σ_1 and $\sigma_2 \in \mathbb{R}^d$, let $i(\sigma_i) \in \arg \min\{f_i(w) - \sigma_i\}$. Then we have

$$f(w_1(\sigma_1)) - f(w_2(\sigma_2)) \leq \sigma^\top(w_1(\sigma_1) - w_2(\sigma_2))$$

where $f := f_1 - f_2$ and $\sigma = \sigma_1 - \sigma_2$.

The proof is straightforward using the definitions.

This lemma is then used to prove that for every t and a fixed σ , $w_t(\sigma + 2G)$ (where G is the Lipschitz constant of loss functions) cannot be smaller than $w_{t+1}(\sigma)$. The proof actually works for any $\sigma + \alpha G$ where $\alpha > 1$.

However, I think the authors over-complicated the proof a little. Here I give a simpler proof. Note that

$$w_t(\sigma + 2G) = \arg \min_w \sum_{i=1}^{t-1} \ell_i(w) - \sigma^\top w - 2Gw$$

and

$$w_{t+1}(\sigma) = \arg \min_w \sum_{i=1}^t \ell_i(w) - \sigma^\top w$$

so that the difference between the two objective functions is

$$\text{function for } t+1 - \text{function for } t = l_t(w) + 2G. \quad (6)$$

Now starting at $w_t(\sigma + 2G)$ for the second minimization, adding any value will increase the second term in 6 increases more than the first term could ever decrease. This is because the loss function ℓ_t is G -Lipschitz. Therefore, the second minimization has to have a minimizer $\leq w_t(\sigma + 2G)$, which proves the claim. We could make the exact same argument to deduce that $w_{t+1}(\sigma + 2G) \geq w_t(\sigma)$. Now defining $w_{\max}(\sigma) = \max(w_t(\sigma), w_{t+1}(\sigma))$ (and similarly $w_{\min}(\sigma)$), we have

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

We will now use our recent observation together with the properties of the exponential distribution to find an appropriate bound. Consider $w_{\min}(\sigma)$. Let A be the event that $\sigma \leq 2G$. Then

$$\mathbb{E}w_{\min}(\sigma) = \mathbb{E}(w_{\min}(\sigma)|A)\mathbb{P}(A) + \mathbb{E}(w_{\min}(\sigma)|\bar{A})(1 - \mathbb{P}(A))$$

A direct calculation gives $\mathbb{P}(A) = 1 - \exp(-2\eta G)$, so we have

$$\mathbb{E}w_{\min}(\sigma) \geq (1 - \exp(-2\eta G))\mathbb{E}(w_{\max}(\sigma)) + \mathbb{E}(w_{\min}(\sigma)|\bar{A})\exp(-2\eta G)$$

But from the observation that we made, we have $\mathbb{E}(w_{\min}(\sigma)|\bar{A}) = \mathbb{E}(w_{\min}(\sigma)|\sigma > 2G) = \mathbb{E}(w_{\min}(\sigma' + 2G)) \geq \mathbb{E}(w_{\max}(\sigma))$. Rearranging and using $\exp(x) \geq 1 + x$ gives

$$\mathbb{E}\|w_t(\sigma) - w_{t+1}(\sigma)\| \leq 2\eta DG. \quad (7)$$

Combining this with what we had before gives

$$\mathbb{E}((\text{regret})_T) \leq \eta^{-1}(\log d + 1)D + 2\eta TGD.$$

To minimize this, we choose η such that the two terms are equal, and hence

$$\mathbb{E}((\text{regret})_T)/T < 4D\sqrt{\log(d) + 1}\sqrt{\frac{G}{T}} \quad (8)$$

proving the claim made in Theorem 1 in the 1-D case.

For the general case in d dimensions, the authors tried proving similar results for each coordinate, and combine them in $\|\cdot\|_1$. However, the above method is not applicable to coordinates directly. I believe this is for $d > 1$, \mathbb{R}^d is not a totally ordered set, and therefore it is not possible to repeat the argument we made for 6 here. Instead, the authors make use of the assumption that $\|\ell_t\|_\infty \leq B$, and essentially repeat the previous argument. This results in a $O(T^{\frac{-1}{3}})$ bound for the expected average regret, a result which is slightly less powerful than the one for the 1-D case.

4 Improving the General Case

In a more recent paper [2], the authors have improved upon the result of [1] for the general case. It is important to note that they use the exact same algorithm as Algorithm 1, and they improve the proof for the d dimensions. The general idea of separating the coordinates, and defining w_{\max} and w_{\min} remains the same. However, unlike [1], they are able to leverage the Lipschitz continuity of ℓ_t .

The key idea of the paper is condition the expected value of $\|w_t - w_{t+1}\|_1$ on the event that the i th coordinate of $w_t - w_{t+1}$, i.e. $|w_t(i) - w_{t+1}(i)|$, is $\leq 10d\|w_t - w_{t+1}\|_1$, which means separating coordinates where that one contributes to at least $\frac{1}{10}$ of the norm.

More specifically, they prove that

$$\mathbb{E}(\text{regret}_T) \leq O\left(\eta^{-1}Dd\log(d) + \eta d^2 DG^2 T\right). \quad (9)$$

For appropriate choice of η , we get expected average regret $\leq O(DG\sqrt{d\log d/T})$, yielding the $1/\sqrt{T}$ bound. They also provide an extra result assuming that the oracle has an approximation error α , meaning that $f(w)$ for the w that the oracle returns is within the α neighborhood of the minimum value of f , making the result more appealing for practical purposes. They show that under this extra assumption, the bound in 9 changes to

$$\mathbb{E}(\text{regret}_T) \leq O\left(\eta^{-1}Dd\log(d) + \eta d^2 DG^2 T + \alpha T\right), \quad (10)$$

introducing a linear term α , meaning that if the oracle is inaccurate, the inaccuracy directly translates to a constant error in the bound for expected average regret (which does not vanish as $T \rightarrow \infty$).

5 Simulation

We consider the one dimensional example where $\ell_t(w) = \sin(\pi(w - t))$, and $w \in \mathcal{W} = [0, 1]$, and $t \in [0, 1]$. This means that the possible domain for the sin function for different values of t , is the set of intervals of length 1 that are a subset of $[-1, 1]$, which clearly shows that ℓ_t is generally not convex. However, we do have that $\sin(\pi(w - t))$ is G -Lipschitz with $G = \pi$. Therefore the setup of this example fits the online optimization problem we have considered in previous sections. The proposed algorithm (Algorithm 1) is run for this problem, and the results for average regret are compared to the theory bounds derived earlier.

As it can be seen, the overall behaviour in the simulation is similar to the behaviour of the theoretical bound. This also demonstrates the claim in [2] that the $O(\sqrt{1/T})$ bound is optimal.

Figure 2 also shows the example where a random linear noise is added to the oracle, causing fluctuations, but keeping the overall trend, as expected.

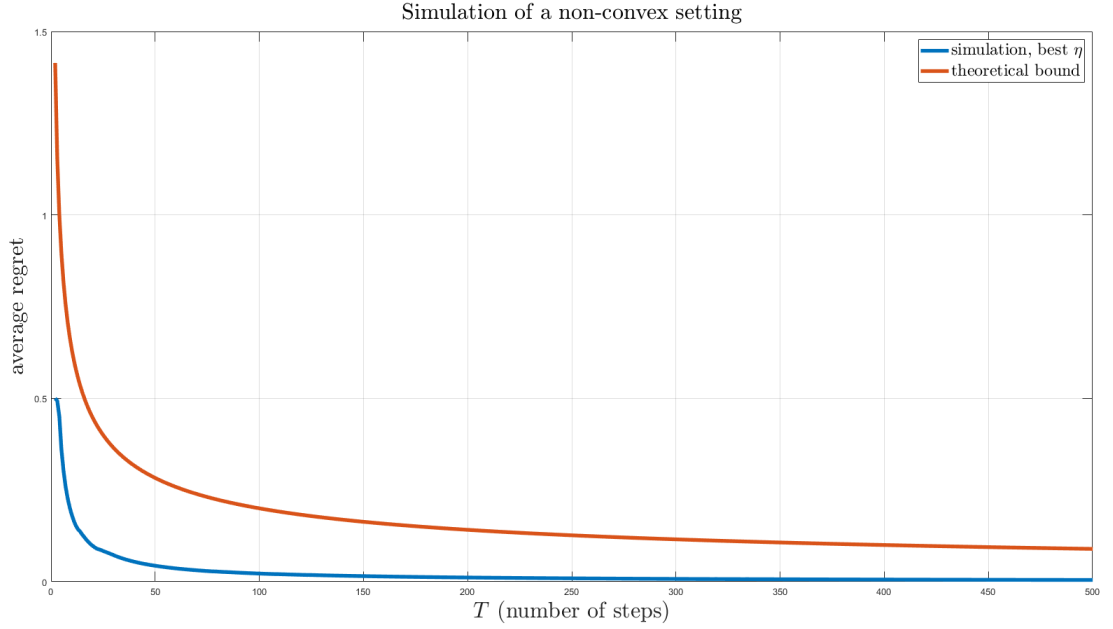


Figure 1: Computer Simulation for a non-convex example, with perfect oracle

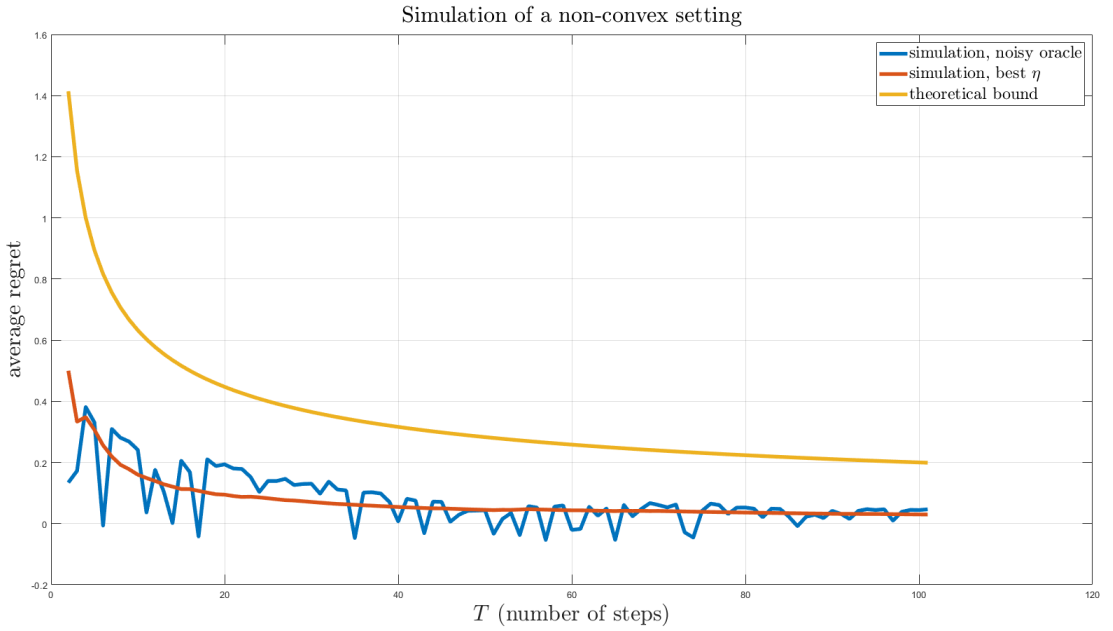


Figure 2: Computer Simulation for a non-convex example, with noisy oracle

References

- [1] N. Agarwal, A. Gonen, E. Hazan, *Learning in Non-convex Games with an Optimization Oracle*, arXiv preprint arXiv:1810.07362 (2018).
- [2] Suggala, A. Sai, and P. Netrapalli. *Online Non-Convex Learning: Following the Perturbed Leader is Optimal*. arXiv preprint arXiv:1903.08110 (2019).
- [3] N. Cesa-Bianchi, G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2006
- [4] Kalai, Adam, and S. Vempala. *Efficient algorithms for online decision problems*, Journal of Computer and System Sciences 71, no. 3 (2005): 291-307.