Online Learning with Non-Convex Losses and Non-Stationary Regret

Xiang Gao
University of Minnesota

Xiaobo Li
University of Minnesota

Shuzhong Zhang
University of Minnesota

Abstract

In this paper, we consider online learning with non-convex loss functions. Similar to Besbes et al. [2015] we apply non-stationary regret as the performance metric. In particular, we study the regret bounds under different assumptions on the information available regarding the loss functions. When the gradient of the loss function at the decision point is available, we propose an online normalized gradient descent algorithm (ONGD) to solve the online learning problem. In another situation, when only the value of the loss function is available, we propose a bandit online normalized gradient descent algorithm (BONGD). Under a condition to be called *weak pseudo-convexity* (WPC), we show that both algorithms achieve a cumulative regret bound of $O(\sqrt{T+V_T T})$, where V_T is the total temporal variations of the loss functions, thus establishing a sublinear regret bound for online learning with non-convex loss functions and non-stationary regret measure.

1 Introduction

Online convex optimization (OCO) has been studied extensively in the literature. In OCO, at each period $t \in \{1, 2, \ldots, T\}$, an online player chooses a feasible strategy x_t from a decision set $\mathcal{X} \subset \mathbf{R}^n$, and suffers a loss given by $f_t(x_t)$, where $f_t(\cdot)$ is a convex loss function. One key feature of the OCO is that the player must make a decision for period t without knowing the loss function $f_t(\cdot)$. The performance of an OCO algorithm is usually measured by the *stationary regret*, which compares the accumulated loss suffered by the player with the loss suffered by the best fixed strategy. Specifically, the stationary regret is defined as

$$\text{Regret}_T^S(\{x_t\}_1^T) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^*), \qquad (1)$$

where x^* is one best fixed decision in hindsight, i.e. $x^* \in \arg\min_{x\in\mathcal{X}} \sum_{t=1}^{T} f_t(x)$. Several sub-linear cumulative regret bounds measured by stationary regret have been established in various papers in the literature. For example, Zinkevich [2003] proposed an online gradient descent algorithm which achieves an regret bound of order $O(\sqrt{T})$ for convex loss functions. The order of the regret can be further improved to $O(\log T)$ if the loss functions are strongly convex (see Hazan et al. [2007]). Moreover, the bounds are shown to be tight for the OCO with convex / strongly convex loss functions respectively in Abernethy et al. [2009]. One important extension of the OCO is the so-called *bandit online convex optimization*, where the online player is only supposed to know the function value $f_t(x_t)$ at x_t, instead of the entire function $f_t(\cdot)$. In particular, when the player can only observe the function value at a single point, Flaxman et al. [2005] established an $O(T^{3/4})$ regret bound for general convex loss functions by constructing a zeroth-order approximation of the gradient. Assuming that the loss functions are smooth, the regret bound can be improved to $O(T^{2/3})$ by incorporating a self-concordant regularizer (see Saha and Tewari [2011]). Alternatively, if multiple points can be inquired at the same time, Agarwal et al. [2010] showed that the regrets can be further improved to $O(T^{1/2})$ and $O(\log T)$ for convex / strongly convex loss functions respectively.

As suggested by its name, the loss functions in the OCO are assumed to be convex. Only a handful of papers studied online learning with non-convex loss functions. In the existing works, most of the time heuristic algorithms were proposed (e.g. Gasso et al. [2011], Ertekin et al. [2011]) without focussing on establishing sublinear regret bounds. There are a few noticeable exceptions though. Hazan and Kale [2012] developed an algorithm that achieves $O(T^{1/2})$ and $O(T^{2/3})$ regret bounds for the full information and bandit settings respectively, by assuming the loss functions to be submodular. Zhang et al. [2015] showed that an $O(T^{2/3})$ regret bound still holds if the loss functions are in

Proceedings of the 21^{st} International Conference on Artificial Intelligence and Statistics (AISTATS) 2018, Lanzarote, Spain. PMLR: Volume 84.

the form of composition between a non-increasing scalar function and a linear function.

The stationary regret requires the benchmark strategy to remain unchanged throughout the periods. This assumption may not be relevant in some of the applications. Recently, a new performance metric known as the *non-stationary regret* was proposed by Besbes et al. [2015]. The non-stationary regret compares the cumulative losses of the online player with the losses of the best possible responses:

$$\text{Regret}_T^{NS}(\{x_t\}_1^T) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x_t^*), \quad (2)$$

where $x_t^* \in \arg\min_{x\in\mathcal{X}} f_t(x)$. Clearly, the non-stationary regret is never less than the stationary regret. Besbes et al. [2015] proves that if there is no restriction on the changes of the loss functions, then the non-stationary regret is linear in T regardless of the strategies. To obtain meaningful bounds, the authors assumed that the temporal change of the sequence of the function $\{f_t\}_1^T$ is bounded. Specifically, the loss functions are assumed to be taken from the set

$$\mathcal{V} := \left\{ \{f_1, f_2, \ldots, f_T\} : \sum_{t=1}^{T-1} \|f_t - f_{t+1}\| \le V_T \right\}, \quad (3)$$

where $\|f_t - f_{t-1}\| = \sup_{x\in\mathcal{X}} |f_t(x) - f_{t-1}(x)|$. For nonzero temporal change V_T, Besbes et al. [2015] then proposes algorithms with sub-linear non-stationary regret bounds: $O(V_T^{1/3}T^{2/3})$, $O(V_T^{1/2}T^{1/2})$, and $O(V_T^{1/3}T^{2/3})$ respectively, for the cases where loss functions are: convex with noisy gradients, strongly convex with noisy gradients, and strongly convex with noisy function values. (Note that $V_T > 0$ is assumed in the bounds.) More recently, Yang et al. [2016] also studied the non-stationary regret bounds for OCO. They proposed an uncertainty set $\mathcal{S}_T^p$ of the sequence of functions in which the worst-case variation of the optimal solution x_t^* of $f_t(\cdot)$ (referred to as the path variation) is bounded:

$$\mathcal{S}_T^p := \{\{f_1, f_2, \ldots, f_T\} : \max_{x_t^* \in \arg\min_{x\in\mathcal{X}} f_t(x)} \sum_{t=1}^{T-1} \|x_t^* - x_{t+1}^*\| \le V_T \Bigg\}.$$

They then proved some upper and lower bounds for the several different feedback structure and functional classes (within the convex function class). In particular, they showed that some existing algorithms (with some modification) can achieve the V_T ,$\sqrt{TV_T}$ and $\sqrt{TV_T}$ for true gradient (with smooth condition), noisy gradient and two-point bandit feedback, respectively. (Note that $V_T > 0$ is assumed in the bounds.)

In this paper, we consider online non-convex optimization with non-stationary regret as the performance metric. To the best of our knowledge, such a combination had not been studied before. For each period t, even after the decision x_t is made, the online player is not assumed to know the function $f_t(\cdot)$; instead, only some partial information regarding the loss at x_t is revealed. Specifically, only $\nabla f(x_t)$ (in the first-order setting) or $f(x_t)$ (the zeroth-order setting) is available to the player. Similar to Yang et al. [2016], we define the uncertainty set $\mathcal{S}_T$ of the sequence of functions as follows:

$$\mathcal{S}_T := \{\{f_1, f_2, \ldots, f_T\} : \exists x_t^* \in \arg\min_{x\in\mathcal{X}} f_t(x), \; t = 1, ..., T, \text{ s.t. } \sum_{t=1}^{T-1} \|x_t^* - x_{t+1}^*\| \le V_T \Bigg\}.$$

Note that $\mathcal{S}_T^p \subseteq \mathcal{S}_T$ for the same V_T. In particular, consider a static sequence $f_t(\cdot) = f(\cdot), t = 1, .., T$ where $f(\cdot)$ has multiple optimal solutions. This sequence would clearly belong to $\mathcal{S}_T$ even when $V_T = 0$. However, V_T would have to be linear in T in order for this sequence to be in $\mathcal{S}_T^p$. We propose the Online Normalized Gradient Descent (ONGD) and the novel Bandit Online Normalized Gradient Descent (BONGD) algorithms for the first-order setting and the zeroth-order setting respectively. For the loss functions satisfying (4) and a condition to be introduced later, we show that these two algorithms both achieve $O(\sqrt{T + V_T T})$ regret bound. Compared to the regret bounds in Yang et al. [2016], our regret bound for the first-order setting is worse but is the same for the zeroth-order setting. Note however, that our loss functions are non-convex and we use a weaker version of variational constraints.

Regarding non-convex objective function, a related work in the literature is Hazan et al. [2015], where the authors propose a Normalized Gradient Descent (NGD) method for solving an optimization model with the so-called *strictly locally quasi-convex* (SLQC) function as the objective. They further show that the NGD converges to an ϵ-optimal minimum within $O(1/\epsilon^2)$ iterations. This paper generalizes the results in Hazan et al. [2015] in the following aspects:

- Hazan et al. [2015] considers an optimization model, while this paper considers an online learning model.
- Hazan et al. [2015] assumes the objective function to be strictly locally quasi-convex (SLQC). In this paper we introduce the notion of weak pseudo-convexity (WPC), which will be shown to be a weaker condition than the SLQC. We show that the regret bounds hold if the objective function is weak pseudo-convex.
- Hazan et al. [2015] considers only the first-order setting, while our proposed BONGD algorithm works for the bandit (zeroth-order) setting as well.

The rest of the paper is organized as follows. Section 2 presents some preparations including the assumptions and notations. In Section 3, we present the ONGD algorithm and prove its regret bound. In Section 4, we present the

BONGD algorithm for the zeroth-order setting and show its regret bound under some assumptions. Finally, we conclude the paper in Section 5.

2 Problem Setup

In this section, we present the assumptions underlying our online learning model and introduce some notations that will be used in the paper. Let $\mathcal{X} \subset \mathbf{R}^n$ be a convex decision set that is known to the player. For every period $t \in \{1, 2, \ldots, T\}$, the loss function is $f_t(\cdot)$. Throughout the paper, we assume that $\mathcal{X} \subset \mathbf{R}^n$ is bounded, i.e., there exists $R > 0$ such that $\|x\| \leq R$ for all $x \in \mathcal{X}$. We present the following definitions regarding the loss functions.

Definition 1 (Bounded Gradient) *A function $f(\cdot)$ is said to have bounded gradient if there exists a finite positive value M such that for all $x \in \mathcal{X}$, it holds that $\|\nabla f(x)\| \leq M$.*

Note that if $f(\cdot)$ has bounded gradient, then it is also Lipschitz continuous with Lipschitz constant M on the set $\mathcal{X}$.

Definition 2 (Weak Pseudo-Convexity) *A function $f(\cdot)$ is said to be **weakly pseudo-convex** (WPC) if there exists $K > 0$ such that*

$$f(x) - f(x^*) \leq K \frac{\nabla f(x)^\top (x - x^*)}{\|\nabla f(x)\|},$$

holds for all $x \in \mathcal{X}$, with the convention that $\frac{\nabla f(x)}{\|\nabla f(x)\|} = 0$ if $\nabla f(x) = 0$, where x^* is one optimal solution, i.e., $x^* \in \arg\min_{x \in \mathcal{X}} f(x)$.

Here we discuss some implications of the weak pseudo-convexity. If a differentiable function $f(\cdot)$ is Lipschitz continuous and pseudo-convex, then we have (see similar derivation in Nesterov [2004])

$$f(x) - f(y) \leq M \frac{\nabla f(x)^\top (x - y)}{\|\nabla f(x)\|},$$

for all y, x with $f(x) \geq f(y)$, where M is Lipschitz constant. Therefore, we can simply let $K = M$, and the function is also weakly pseudo-convex. Moreover, as another example, the star-convex function proposed by Nesterov and Polyak [2006] is weakly pseudo-convexity.

Proposition 1 *If $f(\cdot)$ is star-convex and smooth with bounded gradient in $\mathcal{X}$, then $f(\cdot)$ is weakly pseudo-convex.*

The proof of Proposition 1 can be found in the supplementary file. We next introduce a property that is essentially the same as the SLQC property introduced in Hazan et al. [2015].

Definition 3 (Acute Angle) *Gradient of $f(\cdot)$ is said to satisfy the **acute angle** condition if there exists a positive value Z such that*

$$\begin{aligned} \cos(\nabla f(x), x - x^*) &= \frac{\nabla f(x)^\top (x - x^*)}{\|\nabla f(x)\| \cdot \|x - x^*\|} \\ &\geq Z > 0, \end{aligned}$$

holds for all $x \in \mathcal{X}$, with the convention that $\frac{\nabla f(x)}{\|\nabla f(x)\|} = 0$ if $\nabla f(x) = 0$, where x^* is one optimal solution, i.e., $x^* \in \arg\min_{x \in \mathcal{X}} f(x)$.

The following proposition shows that the acute angle condition together with the Lipschitz continuity implies the weak pseudo-convexity.

Proposition 2 *If $f(\cdot)$ has bounded gradient and satisfies the acute angle condition, then $f(\cdot)$ is weakly pseudo-convex.*

The proof of Proposition 2 can be found in the supplementary file. The class of weakly pseudo-convex functions certainly go beyond the acute angle condition. For example, below is another class of functions satisfying the WPC.

Proposition 3 *If $f(\cdot)$ has bounded gradient and satisfy the α-homogeneity with respect to its minimum, i.e., there exists $\alpha > 0$ satisfying*

$$f(t(x - x^*) + x^*) - f(x^*) = t^\alpha (f(x) - f(x^*)),$$

for all $x \in \mathcal{X}$ and $t \geq 0$ where $x^* = \arg\min_{x \in \mathcal{X}} f(x)$, then $f(\cdot)$ is weak pseudo-convex.

The proof of Proposition 3 can be found in the supplementary file. Proposition 3 suggests that all non-negative homogeneous polynomial satisfies WPC with respect to 0. Take $f(x) = (x_1^2 + x_2^2)^2 + 10(x_1^2 - x_2^2)^2$ as an example. It is easy to verify that $f(\cdot)$ satisfies the condition in Proposition 3, and thus is weakly pseudo-convex. In Figure 1, the curvature of $f(x)$ and a sub-level set of this function are plotted. The function is not quasi-convex since the sub-level set is non-convex. However, this function satisfies the acute-angle condition in 3.

Note that if $f_i(x)$ is α_i-homogeneous with respect to the shared minimum x^* for all $1 \leq i \leq I$ with $\alpha_i \geq \alpha > 0$, and the gradient of f_i is uniformly bounded over a set $\mathcal{X}$, then $\sum_{i=1}^I f_i(x)$ is WPC. As a result, we can construct functions that are WPC but do not satisfy the acute-angle condition. Consider a two-dimensional function $f(x) = x_1^2 + |x_2|^{3/2}$, and suppose that $\mathcal{X}$ is the unit disc centered at the origin. Clearly, $f(x)$ is differentiable and Lipchitz continuous in $\mathcal{X}$. Also, it is the sum of a 2-homogeneous function and a $3/2$-homogeneous function with a shared minimum $(0, 0)$. Thus $f(x)$ is WPC. We compute that

$$\begin{aligned} \cos(\nabla f(x), x - x^*) &= \frac{\nabla f(x)^\top (x - x^*)}{\|\nabla f(x)\| \cdot \|x - x^*\|} \\ &= \frac{2x_1^2 + \frac{3}{2}|x_2|^{3/2}}{\sqrt{(4x_1^2 + \frac{9}{4}|x_2|)(x_1^2 + x_2^2)}}. \end{aligned}$$

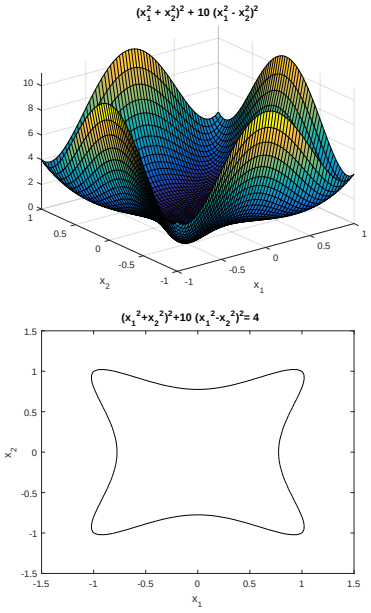

Figure 1: Plot of a WPC function that is not quasi-convex

Consider a parameterized path $(x_1, x_2) = (t^{1/2}, t^{2/3})$ with $t > 0$. On this path, we have

$$\begin{aligned}\cos(\nabla f(x), x - x^*) &= \frac{2x_1^2 + \frac{3}{2}|x_2|^{3/2}}{\sqrt{(4x_1^2 + \frac{9}{4}|x_2|)(x_1^2 + x_2^2)}} \\ &= \frac{7t}{2\sqrt{(4t + \frac{9}{4}t^{2/3})(t + t^{4/3})}} \\ &= \frac{7t^{1/6}}{2\sqrt{(4t^{1/3} + \frac{9}{4})(1 + t^{1/3})}}.\end{aligned}$$

Therefore, along the path, as t approaches to 0, we have $\cos(\nabla f(x), x - x^*) \to 0$. This example shows that a WPC function may fail to satisfy the acute angle condition.

As we mentioned before, in order to establish some sublinear non-stationary regret bound, we need to confine the loss functions $\{f_t\}_1^T$ in a unified manner. Therefore, we introduce the uncertainty set of the loss functions $\mathcal{S}_T$, as a set of admissible loss functions where their total variation of the minimizers are bounded by V_T.

Definition 4 *The uncertainty set of functions $\mathcal{S}_T$ is defined as*

$$\mathcal{S}_T := \left\{\{f_1, f_2, \ldots, f_T\} : \exists x_t^* \in \arg\min_{x\in\mathcal{X}} f_t(x), \; t = 1, ..., T, \;\text{ s.t. } \sum_{t=1}^{T-1} \|x_t^* - x_{t+1}^*\| \le V_T\right\}. \tag{4}$$

where $V_T \ge 0$.

In the zeroth-order setting to be discussed in Section 4, only the function value is available. Therefore, some randomized approaches are needed in the algorithm. To account for this situation, we introduce the expected non-stationary regret for an algorithm that outputs a random sequence $\{x_t\}_1^T$ in the performance metric.

Definition 5 *The expected non-stationary regret for a randomized algorithm $\mathcal{A}$ is defined as*

$$ERegret_T^{NS}(\{x_t\}_1^T) = \mathsf{E}\left[\sum_{t=1}^{T}(f_t(x_t) - f_t(x_t^*))\right], \tag{5}$$

where the expectation is taken over the filtration generated by the random sequence $\{x_t\}_1^T$ produced by $\mathcal{A}$.

3 The First-Order Setting

In this section, we assume that for each period, the gradient information at the current point is available to the online player after the decision is made. Specifically, at each period t, the sequence of the events is as follows:

1. The online player chooses a strategy x_t;
2. The online player receives the feedback $\nabla f_t(x_t)$;
3. Regret $f_t(x_t) - f_t(x_t^*)$ incurs (but is not necessarily known to the online player).

We propose the Online Normalized Gradient Descend algorithm (ONGD) in this setting. The normalized gradient descent method was first proposed in Nesterov [2004] which can be applied to solve the pseudo-convex minimization problem. The ONGD algorithm uses the first-order information $\nabla f_t(x_t)$ to compute the normalized vector $\nabla f_t(x_t)/\|\nabla f_t(x_t)\|$ as the search direction. Similar to the standard gradient method, it moves along that search direction with a specific stepsize $\eta > 0$ and then projects the point back to the decision set $\mathcal{X}$; see Algorithm 1 for the details. Note that in Algorithm 1, $\prod_{\mathcal{X}}(y) := \arg\min_{x\in\mathcal{X}} \|y - x\|$ is the projection operator. The main result is shown in the following theorem which claims an $O(\sqrt{T + V_T T})$ non-stationary regret bound for ONGD.

Algorithm 1 Online Normalized Gradient Descent

Input: feasible set $\mathcal{X}$, # time period T
Initialization: $x_1 \in \mathcal{X}$
for $t = 1$ **to** T **do**
  chooses x_t and receives the feedback $g_t = \nabla f_t(x_t)$
  if $\|g_t\| > 0$ **then**
    $x_{t+1} = \prod_{\mathcal{X}}\left(x_t - \eta \frac{g_t}{\|g_t\|}\right)$
  else
    $x_{t+1} = x_t$
  end if
end for

Theorem 1 *Let V_T be as defined in (4). For any sequence of loss functions $\{f_t\}_1^T \in \mathcal{S}_T$ where f_t is weakly pseudo-convex with common constant K, let the stepsize $\eta = \sqrt{\frac{4R^2+6RV_T}{T}}$. Then, the following regret bound holds for ONGD:*

$$\begin{aligned} \text{Regret}_T^{NS}(\{x_t\}_1^T) &\leq K\sqrt{T(R^2+1.5RV_T)} \\ &= O(\sqrt{T+V_T T}). \end{aligned}$$

Proof: Let $x_t^*, t = 1, .., T$ be the sequence of optimal solutions satisfying the condition in Definition 4, and $z_t := \|x_t - x_t^*\|$. Then we have:

$$\begin{aligned} z_{t+1}^2 &= \|x_{t+1} - x_{t+1}^*\|^2 \\ &= \|x_{t+1} - x_t^*\|^2 + \|x_t^* - x_{t+1}^*\|^2 \\ &\quad +2(x_{t+1} - x_t^*)^\top (x_t^* - x_{t+1}^*) \\ &\leq \left\| \prod_{\mathcal{X}} \left(x_t - \eta \frac{g_t}{\|g_t\|} \right) - x_t^* \right\|^2 + 6R\|x_t^* - x_{t+1}^*\| \\ &\leq \left\| x_t - \eta \frac{g_t}{\|g_t\|} - x_t^* \right\|^2 + 6R\|x_t^* - x_{t+1}^*\| \\ &= z_t^2 + \eta^2 - 2\eta \frac{g_t^\top (x_t - x_t^*)}{\|g_t\|} + 6R\|x_t^* - x_{t+1}^*\|. \end{aligned}$$

By rearranging terms and multiplying K on both sides we have

$$K\frac{g_t^\top (x_t - x_t^*)}{\|g_t\|} \leq \frac{K}{2\eta}\left(z_t^2 - z_{t+1}^2 + \eta^2 + 6R\|x_t^* - x_{t+1}^*\| \right). \tag{6}$$

By Definition 2, noting that $g_t = \nabla f_t(x_t)$, we have

$$\begin{aligned} & f_t(x_t) - f_t(x_t^*) \\ \leq\ & K\frac{\nabla f_t(x_t)^\top (x_t - x_t^*)}{\|\nabla f_t(x_t)\|} \\ \leq\ & \frac{K}{2\eta}\left(z_t^2 - z_{t+1}^2 + \eta^2 + 6R\|x_t^* - x_{t+1}^*\| \right). \end{aligned}$$

Summing these inequalities from $t = 1, ..., T$, we have

$$\begin{aligned} & \text{Regret}_T^{NS}(\{x_t\}_1^T) \\ \leq\ & \frac{K}{2\eta}\left(z_1^2 - z_{T+1}^2 + T\eta^2 + 6R\sum_{t=1}^T \|x_t^* - x_{t+1}^*\| \right) \\ \leq\ & \frac{K}{2\eta}\left(4R^2 + T\eta^2 + 6RV_T \right). \end{aligned}$$

As a result, by noting $\eta = \sqrt{\frac{4R^2+6RV_T}{T}}$, we have

$$\begin{aligned} \text{Regret}_T^{NS}(\{x_t\}_1^T) &\leq K\sqrt{T(R^2+1.5RV_T)} \\ &= O(\sqrt{T+V_T T}). \end{aligned}$$

□

4 The Zeroth-Order Setting

In the previous section, it is assumed that the gradient information is available, which may not be the case in some applications. Such exceptions include the multi-armed bandit problem, dynamic pricing and Bayesian optimization. Therefore, in this section, we consider the setting where the online player only receives the function value $f_t(x_t)$, instead of the gradient $\nabla f_t(x_t)$, as the feedback.

As mentioned above, the zeroth-order (or bandit) setting has been studied in the OCO literature. The main technique in the OCO literature (see Flaxman et al. [2005] for example) is to construct a zeroth-order approximation of the gradient of a smoothed function. That smoothed function is often created by integrating the original loss function with a chosen probability distribution. By querying some random samples of the function value according to a probability distribution, the player is able to create an unbiased zeroth-order approximation of the gradient of the smoothed function. This is, however, not applicable in our online normalized gradient descent algorithm since what we need is the direction of the gradient. Therefore, we shall first develop a new type of zeroth-order oracle that can approximate the gradient direction without averaging multiple samples of gradients when the norm of the gradient is not too small.

To proceed, we require some additional conditions on the loss function.

Definition 6 (Error Bound) *There exists $D > 0$ and $\gamma > 0$ such that*

$$\|x - x_t^*\| \leq D\|\nabla f_t(x)\|^\gamma,$$

for all $x \in \mathcal{X}$, $1 \leq t \leq T$, where x_t^* is the optimal solution to $f_t(\cdot)$, i.e., $x_t^* = \arg\min_{x\in\mathcal{X}} f_t(x)$.

Since $\mathcal{X}$ is a compact set, the error bound condition is essentially the requirement for a unique optimal solution and no local minimum.

Definition 7 (Lipschitz Gradient) *There exists a positive number L, such that*

$$\|\nabla f_t(x) - \nabla f_t(y)\| \leq L\|x - y\|,$$

for all $x, y \in \mathcal{X}$, where $1 \leq t \leq T$.

Note that ... We introduce some notations that will be used in subsequent analysis.

$S(n)$:	the unit sphere in $\mathbf{R}^n$;
$m(A)$:	the measure of set $A \subset \mathbf{R}^n$;
β_n:	the area of the unit sphere $S(n)$;
dS_n:	the differential unit on the unit sphere $S(n)$;
$\mathbf{1}_A(x)$:	the indicator function of set A;
$\text{sign}(\cdot)$:	the sign function.

Before we present the main results, several lemmas are in order. The first lemma considers some geometric properties of the unit sphere.

Lemma 1 *For any non-zero vector $d \in \mathbf{R}^n$ and $\delta < 1$, let S_δ^x be defined as*

$$S_\delta^x := \left\{ v \in S(n) | \text{ s.t. } |d^\top v| < \delta^2 \right\}.$$

If $\|d\| \geq \delta$, then there exists a constant $C_n > 0$, such that

$$m(S_\delta^x) < C_n \delta.$$

Proof: We have

$$m(S_\delta^x) = \int\limits_{v \in S(n) \cap S_\delta^x} dS_n.$$

By the symmetry of $S(n)$, we may assume w.l.o.g. that $d = (0, \ldots, 0, \|d\|)^\top$. Let $a = \frac{\delta^2}{\|d\|}$. Since $a < 1$, we have

$$\begin{aligned}
& m(S_\delta^x) \\
= & \int\limits_{v \in S(n)} \mathbf{1}_{\left\{-\frac{\delta^2}{\|d\|} \leq v_n \leq \frac{\delta^2}{\|d\|}\right\}}(v) dS_n \\
= & 2 \int\limits_{1-a^2 \leq v_1^2 + \cdots + v_{n-1}^2 \leq 1} \frac{1}{\sqrt{1 - v_1^2 - \cdots - v_{n-1}^2}} dv_1 \cdots dv_{n-1} \\
= & 2 \int\limits_{\sqrt{1-a^2} \leq r \leq 1} \frac{r^{n-2}}{\sqrt{1-r^2}} dr \cdot dS_{n-1} \\
= & 2\beta_{n-1} \int\limits_{\sqrt{1-a^2} \leq r \leq 1} \frac{r^{n-2}}{\sqrt{1-r^2}} dr \\
\leq & 2\beta_{n-1} \int\limits_{\sqrt{1-a^2} \leq r \leq 1} \frac{1}{\sqrt{1-r^2}} dr \\
= & 2\beta_{n-1} \left(\frac{\pi}{2} - \arcsin(\sqrt{1-a^2}) \right) \\
= & 2\beta_{n-1} (\arcsin a) < 2\beta_{n-1} \frac{\pi}{2} a = \pi \beta_{n-1} \frac{\delta^2}{\|d\|} \leq \pi \beta_{n-1} \delta.
\end{aligned}$$

By setting $C_n = \pi \beta_{n-1}$, the desired result follows. □

The next lemma leads to an unbiased first-order estimator of the direction of a vector.

Lemma 2 *Suppose $d \in \mathbf{R}^n$, and $d \neq 0$. Then,*

$$\int\limits_{v \in S(n)} \text{sign}(d^\top v) v dS_n = P_n \frac{d}{\|d\|},$$

where P_n is a constant.

Proof: By the symmetry of $S(n)$, again we may assume $d = (0, \ldots, 0, \|d\|)^\top$, and

$$\int\limits_{v \in S(n)} \text{sign}(d^\top v) v dS_n = 2 \int\limits_{v \in S(n)} \mathbf{1}_{v_n \geq 0}(v) v dS_n.$$

Notice that if $v \in S(n)$, then $\boldsymbol{u} = (-v_1, -v_2, \ldots, -v_{n-1}, v_n)^\top$ is also in $S(n)$. As a result, the above integral will be on the direction of $\frac{d}{\|d\|} = (0, 0, \ldots, 0, 1)^\top$, and its length is given by

$$\begin{aligned}
& 2 \int\limits_{v \in S(n)} \mathbf{1}_{v_n \geq 0}(v) v_n dS_n \\
= & 2 \int\limits_{0 \leq v_1^2 + \cdots + v_{n-1}^2 \leq 1} \sqrt{1 - v_1^2 - \cdots - v_{n-1}^2} dS_n \\
= & 2 \int\limits_{0 \leq v_1^2 + \cdots + v_{n-1}^2 \leq 1} \frac{\sqrt{1 - v_1^2 - \cdots - v_{n-1}^2}}{\sqrt{1 - v_1^2 - \cdots - v_{n-1}^2}} dv_1 \ldots dv_{n-1} \\
= & 2 \int\limits_{0 \leq r \leq 1} r^{n-2} dr dS_{n-1} \\
= & \frac{2\beta_{n-1}}{n-1} := P_n.
\end{aligned}$$

□

Using the previous lemmas, we have the following result which constructs a zeroth-order estimator for the normalized gradient.

Theorem 2 *Suppose $f(x)$ has Lipschitz gradient and $\|\nabla f(x)\| \geq \delta$ at x. Let $\epsilon = \frac{\delta^2}{L}$. Then we have*

$$\left\| \mathsf{E}_{S(n)} \left[\text{sign}(f(x + \epsilon v) - f(x)) v \right] - Q_n \frac{\nabla f(x)}{\|\nabla f(x)\|} \right\| \leq 2 D_n \delta$$

where v is a random vector uniformly distributed over $S(n)$, and $Q_n = \frac{P_n}{\beta_n}$ and $D_n = \frac{C_n}{\beta_n}$.

Proof: By Definition 7, we have

$$\begin{aligned}
& |f(x + \epsilon v) - f(x) - \epsilon \nabla f(x)^\top v| \leq \frac{\epsilon L}{2} \|v\|^2 \iff \\
& \nabla f(x)^\top v - \frac{\epsilon}{2} L \leq \frac{f(x + \epsilon v) - f(x)}{\epsilon} \leq \nabla f(x)^\top v + \frac{\epsilon}{2} L.
\end{aligned}$$

Since $|\nabla f(x)^\top v| \geq \delta^2$ for $v \in S(n) \setminus S_\delta^x$, if we let $\epsilon = \frac{\delta^2}{L}$, we have

$$\nabla f(x)^\top v - \frac{\delta^2}{2} \leq \frac{f(x + \epsilon v) - f(x)}{\epsilon} \leq \nabla f(x)^\top v + \frac{\delta^2}{2}.$$

Thus,

$$\begin{aligned}
& \text{sign}\left(\nabla f(x)^\top v \right) = \text{sign}\left(\nabla f(x)^\top v - \frac{\delta^2}{2} \right) \\
\leq & \text{sign}\left(\frac{f(x + \epsilon v) - f(x)}{\epsilon} \right) \leq \text{sign}\left(\nabla f(x)^\top v + \frac{\delta^2}{2} \right) \\
= & \text{sign}\left(\nabla f(x)^\top v \right),
\end{aligned}$$

implying $\text{sign}(\nabla f(x)^\top v) = \text{sign}\left(\frac{f(x + \epsilon v) - f(x)}{\epsilon} \right)$. There-

fore,

$$\begin{aligned}
&\beta_n \mathsf{E}_{S(n)}\left[\text{sign}(f(x+\epsilon v)-f(x))v\right] \\
=& \int\limits_{v\in S(n)\setminus S_\delta^x} [\text{sign}(f(x+\epsilon v)-f(x))v]\, dS(n) \\
&+ \int\limits_{v\in S_\delta^x} [\text{sign}(f(x+\epsilon v)-f(x))v]\, dS(n) \\
=& \int\limits_{v\in S(n)\setminus S_\delta^x} \left[\text{sign}(\nabla f(x)^\top v)v\right] dS(n) \\
&+ \int\limits_{v\in S_\delta^x} [\text{sign}(f(x+\epsilon v)-f(x))v]\, dS(n) \\
=& \int\limits_{v\in S(n)} \left[\text{sign}(\nabla f(x)^\top v)v\right] dS(n) \\
&- \int\limits_{v\in S_\delta^x} \left[\text{sign}(\nabla f(x)^\top v)v\right] dS(n) \\
&+ \int\limits_{v\in S_\delta^x} [\text{sign}(f(x+\epsilon v)-f(x))v]\, dS(n) \\
=& \ P_n \frac{\nabla f(x)}{\|\nabla f(x)\|} - \int\limits_{v\in S_\delta^x} \left[\text{sign}(\nabla f(x)^\top v)v\right] dS(n) \\
&+ \int\limits_{v\in S_\delta^x} [\text{sign}(f(x+\epsilon v)-f(x))v]\, dS(n),
\end{aligned}$$

where the last equality is due to Lemma 2.

Putting the estimations together, we have

$$\begin{aligned}
&\left\| \mathsf{E}_{S(n)}\left[\text{sign}(f(x+\epsilon v)-f(x))v\right] - \frac{P_n}{\beta_n}\frac{\nabla f(x)}{\|\nabla f(x)\|} \right\| \\
\le& \ \frac{1}{\beta_n} \int\limits_{v\in S_\delta^x} \left\| \text{sign}(\nabla f(x)^\top v)v \right\| dS(n) \\
&+ \frac{1}{\beta_n} \int\limits_{v\in S_\delta^x} \|\text{sign}(f(x+\epsilon v)-f(x))v\|\, dS(n) \\
\le& \ \frac{2m(S_\delta^x)}{\beta_n} \le \frac{2C_n\delta}{\beta_n}.
\end{aligned}$$

Note that $Q_n = \frac{P_n}{\beta_n}$ and $D_n = \frac{C_n}{\beta_n}$, the theorem is proved. □

Based on Theorem 2, for a given $\delta > 0$ we have a zeroth-order estimator for the normalized gradient given as:

$$G_t(x_t, v_t) = \frac{\text{sign}(f_t(x_t+\epsilon v_t) - f_t(x_t))}{Q_n} v_t, \qquad (7)$$

where $\epsilon = \delta^2/L$ and v_t is an uniformly distributed random vector over $S(n)$. Theorem 2 implies that the distance between the estimator and the normalized gradient can be controlled up to a factor of δ. Essentially, the Bandit Online Normalized Gradient Descent (BONGD) algorithm replaces the normalized gradient by $G_t(x_t, v_t)$ in the ONGD algorithm.

Algorithm 2 Bandit Online Normalized Gradient Descent

Input: feasible set $\mathcal{X}$, # time period T, δ
Initialization: $x_1 \in \mathcal{X}$, $\epsilon = \delta^2/L$
for $t = 1$ **to** T **do**
  Sample v_t uniformly over $S(n) \subset \mathbf{R}^n$;
  play x_t and $x_t + \epsilon v_t$;
  receive feedbacks $f_t(x_t)$ and $f_t(x_t + \epsilon v_t)$;
  set $G_t(x_t, v_t) = \frac{\text{sign}(f_t(\boldsymbol{x}_t+\epsilon v_t) - f_t(\boldsymbol{x}_t))}{Q_n} v_t$;
  update $x_{t+1} = \prod_{\mathcal{X}}(x_t - \eta G_t(x_t, v_t))$.
end for

Note that Algorithm 2 actually outputs a random sequence of vectors $\{x_t\}_1^T$; hence the notion of expected non-stationary regret is applicable here. Let us denote $\{\mathcal{F}_t\}_1^T$ be the filtration generated by $\{x_t\}_1^T$. Then v_t is independent of $\mathcal{F}_t$. Note that in Algorithm 2, at each step, it queries the function at another point $x_t + \epsilon v_t$. Therefore, besides its output sequence $\{x_t\}_1^T$, we need to include $\{x_t + \epsilon v_t\}_1^T$ in our regret. We thus define $\text{ERegret}_T^{NS}(\{x_t\}_1^T, \{x_t + \epsilon v_t\}_1^T) = \mathsf{E}[\sum_{t=1}^T (f_t(x_t) - f_t(x_t^*))] + \mathsf{E}[\sum_{t=1}^T (f_t(x_t + \epsilon v_t) - f_t(x_t^*))]$. The following theorem shows that by choosing η and δ appropriately, we can still achieve an $O(\sqrt{T + V_T T})$ expected non-stationary regret bound.

Theorem 3 *Let V_T be defined in* (4)*. Assume that the loss functions have Lipschitz gradients (Definition 7), satisfying the error bound condition (Definition 6) and are weakly pseudo-convex with bounded gradient. For any sequence of loss functions $\{f_t\}_1^T \in \mathcal{S}_T$, applying BONGD with $\eta = Q_n\sqrt{\frac{4R^2+6RV_T}{T}}$ and $\delta = \min\{T^{-\frac{1}{2\gamma}}, T^{-\frac{1}{4}}\}$ where $Q_n = \frac{P_n}{\beta_n}$ and P_n is a constant, the following regret bound holds* $\text{ERegret}_T^{NS}(\{x_t\}_1^T, \{x_t + \epsilon v_t\}_1^T) \le O(\sqrt{T + V_T T})$.

Proof: Let $z_t := \|x_t - x_t^*\|$. Then,

$$\begin{aligned}
& z_{t+1}^2 \\
=& \ \|x_{t+1} - x_{t+1}^*\|^2 \\
=& \ \|x_{t+1} - x_t^*\|^2 + \|x_t^* - x_{t+1}^*\|^2 \\
& +2(x_{t+1} - x_t^*)^\top (x_t^* - x_{t+1}^*) \\
\le& \ \|x_{t+1} - x_t^*\|^2 + 2R\|x_t^* - x_{t+1}^*\| + 4R\|x_t^* - x_{t+1}^*\| \\
=& \ \|\textstyle\prod_{\mathcal{X}}(x_t - \eta G_t(x_t, v_t)) - x_t^*\|^2 + 6R\|x_t^* - x_{t+1}^*\| \\
\le& \ \|x_t - \eta G_t(x_t, v_t) - x_t^*\|^2 + 6R\|x_t^* - x_{t+1}^*\| \\
=& \ z_t^2 + \eta^2\|G_t(x_t, v_t)\|^2 - 2\eta G_t(x_t, v_t)^\top (x_t - x_t^*) \\
& +6R\|x_t^* - x_{t+1}^*\| \\
\le& \ z_t^2 + \frac{\eta^2}{Q_n^2} - 2\eta G_t(x_t, v_t)^\top (x_t - x_t^*) + 6R\|x_t^* - x_{t+1}^*\|.
\end{aligned}$$

By rearranging the terms, we have:

$$\begin{aligned}
& KG_t(x_t, v_t)^\top (x_t - x_t^*) \qquad (8) \\
\le& \ \frac{K}{2\eta}\left(z_t^2 - z_{t+1}^2 + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\| \right).
\end{aligned}$$

Now based on $\|\nabla f_t(x_t)\|$, we have two different cases:

• $\|\nabla f_t(x_t)\| \geq \delta$. In this case, by Theorem 2, we have

$$\|\mathsf{E}[G_t(x_t, v_t)|x_t] - \frac{\nabla f_t(x_t)}{\|\nabla f_t(x_t)\|}\| \leq \frac{2D_n}{Q_n}\delta.$$

Therefore,

$$\begin{aligned}
& f_t(x_t) - f_t(x_t^*) \\
\leq\ & K\frac{\nabla f_t(x_t)^\top (x_t - x_t^*)}{\|\nabla f_t(x_t)\|} \\
\leq\ & K\mathsf{E}[G_t(x_t, v_t)|x_t]^\top (x_t - x_t^*) + \frac{2D_nK}{Q_n}\delta\|x_t - x_t^*\| \\
=\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& + \frac{2D_nK}{Q_n}\delta\|x_t - x_t^*\| \\
\leq\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& + \frac{4D_nK}{Q_n}R\delta. \qquad (9)
\end{aligned}$$

• $\|\nabla f_t(x_t)\| < \delta$. In this case, by the error bound property (Definition 6) we have

$$\|x_t - x_t^*\| \leq D\|\nabla f_t(x_t)\|^\gamma < D\delta^\gamma.$$

Therefore, due to the boundedness of gradient

$$f_t(x_t) - f_t(x_t^*) \leq M\|x_t - x_t^*\| \leq MD\delta^\gamma, \qquad (10)$$

and

$$\begin{aligned}
0 \leq\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& - K\mathsf{E}[G_t(x_t, v_t)|x_t]^\top (x_t - x_t^*) \\
\leq\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& + K\frac{\beta_n}{Q_n}D\delta^\gamma. \qquad (11)
\end{aligned}$$

Adding (10) with (11), it follows that

$$\begin{aligned}
& f_t(x_t) - f_t(x_t^*) \\
\leq\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& + \left(K\frac{\beta_n}{Q_n}D + MD\right)\delta^\gamma. \qquad (12)
\end{aligned}$$

In view of (9) and (12), if we let $U = \max\left\{\frac{4C_nK}{P_n}R, (K\frac{\beta_n}{Q_n}D + MD)\right\}$, then in either case the following inequality holds:

$$\begin{aligned}
& f_t(x_t) - f_t(x_t^*) \\
\leq\ & \frac{K}{2\eta}\left(\mathsf{E}[z_t^2|x_t] - \mathsf{E}[z_{t+1}^2|x_t] + \frac{\eta^2}{Q_n^2} + 6R\|x_t^* - x_{t+1}^*\|\right) \\
& + U\delta^\gamma.
\end{aligned}$$

Summing these inequalities over $t = 1, ..., T$, we have

$$\begin{aligned}
& \text{ERegret}_T^{NS}(\{x_t\}_1^T, \{x_t + \epsilon v_t\}_1^T) \\
=\ & \mathsf{E}\left[\sum_{t=1}^T (f_t(x_t) + f_t(x_t + \epsilon v_t) - 2f_t(x_t^*))\right] \\
\leq\ & \mathsf{E}\left[\sum_{t=1}^T (2f_t(x_t) - 2f_t(x_t^*) + M\epsilon\|v_t\|)\right] \\
\leq\ & \frac{K}{\eta}\left(\mathsf{E}[z_1^2] - \mathsf{E}[z_{T+1}^2] + T\frac{\eta^2}{Q_n^2} + 6R\sum_{t=1}^T\|x_t^* - x_{t+1}^*\|\right) \\
& + 2TU\delta^\gamma + MT\epsilon \\
\leq\ & \frac{K}{2\eta}\left(4R^2 + T\frac{\eta^2}{Q_n^2} + 6RV_T\right) + 2TU\delta^\gamma + TM\frac{\delta^2}{L}.
\end{aligned}$$

By choosing $\eta = Q_n\sqrt{\frac{4R^2+6RV_T}{T}}$, and $\delta = \min\{T^{-\frac{1}{2\gamma}}, T^{-\frac{1}{4}}\}$, we have

$$\begin{aligned}
& \text{ERegret}_T^{NS}(\{x_t\}_1^T, \{x_t + \epsilon v_t\}_1^T) \\
\leq\ & \frac{2K}{Q_n}\sqrt{T(4R^2 + 6RV_T)} + (2U + \frac{M}{L})\sqrt{T} \\
\leq\ & O(\sqrt{T + V_T T}).
\end{aligned}$$

□

Therefore, under the additional error bound condition (Definition 6) and the Lipschitz continuity of the gradient (Definition 7) on the loss functions (e.g. the function depicted in Figure 1 satisfies all the conditions of Theorem 3), the expected regret of BONGD remains $O(\sqrt{T + V_T T})$, which matches both the upper and lower bound in Yang et al. [2016] for the general Lipschitz continuous convex cost functions and two point bandit feedback. Moreover, the zeroth-order estimator for the normalized gradient could be of interest on its own.

5 Concluding Remarks

In this paper, we considered online learning with non-convex loss functions and non-stationary regret measure, and established $O(\sqrt{T + V_T T})$ regret bounds, where V_T is the total variation of the loss functions, for a gradient-type algorithm and a bandit-type algorithm under some conditions on the non-convex loss function. As a direction for future research, it will be interesting to find out if the same regret bound can still be established without knowing V_T in advance. Moreover, it remains open to extend the results to the setting where the loss functions may be noisy and non-smooth.

References

Jacob Abernethy, Alekh Agarwal, Peter L Bartlett, and Alexander Rakhlin. A stochastic view of optimal regret through minimax duality. *arXiv preprint arXiv:0903.5328*, 2009.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT*, pages 28–40. Citeseer, 2010.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

S. Ertekin, L. Bottou, and C. L. Giles. Nonconvex online support vector machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):368–381, Feb 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.109.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

Gilles Gasso, Aristidis Pappaioannou, Marina Spivak, and Léon Bottou. Batch and online learning algorithms for nonconvex neyman-pearson classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2 (3):28, 2011.

Elad Hazan and Satyen Kale. Online submodular minimization. *The Journal of Machine Learning Research*, 13(1):2903–2922, 2012.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

Elad Hazan, Kfir Levy, and Shai Shalev-Shwartz. Beyond convexity: Stochastic quasi-convex optimization. In *Advances in Neural Information Processing Systems*, pages 1594–1602, 2015.

Yurii Nesterov. *Introductory lectures on convex optimization*, volume 87. Springer Science & Business Media, 2004.

Yurii Nesterov and Boris T Polyak. Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *International Conference on Artificial Intelligence and Statistics*, pages 636–642, 2011.

Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning*, pages 449–457, 2016.

Lijun Zhang, Tianbao Yang, Rong Jin, and Zhi-Hua Zhou. Online bandit learning for a special class of non-convex losses. In *AAAI*, pages 3158–3164, 2015.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. 2003.