JMLR: Workshop and Conference Proceedings vol 30 (2013) 1–18

Beating Bandits in Gradually Evolving Worlds

Chao-Kai Chiang[1,2] CHAOKAI@IIS.SINICA.EDU.TW
Chia-Jung Lee[1] LEECJ@IIS.SINICA.EDU.TW
Chi-Jen Lu[1] CJLU@IIS.SINICA.EDU.TW
[1] *Institute of Information Science, Academia Sinica, Taipei, Taiwan.*
[2] *Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan.*

Abstract

Consider the online convex optimization problem, in which a player has to choose actions iteratively and suffers corresponding losses according to some convex loss functions, and the goal is to minimize the regret. In the full-information setting, the player after choosing her action can observe the whole loss function in that round, while in the bandit setting, the only information the player can observe is the loss value of that action. Designing such bandit algorithms appears challenging, as the best regret currently achieved for general convex loss functions is much higher than that in the full-information setting, while for strongly convex loss functions, there is even a regret lower bound which is exponentially higher than that achieved in the full-information setting. To aim for smaller regrets, we adopt a relaxed two-point bandit setting in which the player can play two actions in each round and observe the loss values of those two actions. Moreover, we consider loss functions parameterized by their deviation D, which measures how fast they evolve, and we study how regrets depend on D. We show that two-point bandit algorithms can in fact achieve regrets matching those in the full-information setting in terms of D. More precisely, for convex loss functions, we achieve a regret of $O(\sqrt{D})$, while for strongly convex loss functions, we achieve a regret of $O(\ln D)$, which is much smaller than the $\Omega(\sqrt{D})$ lower bound in the traditional bandit setting.

Keywords: Online Convex Optimization, Regret, Deviation, Multi-Point Bandit.

1. Introduction

A fundamental problem in machine learning is the online convex optimization problem, in which a player has to make repeated decisions for a number of T rounds in the following way. In round t, the player chooses an action x_t from a convex feasible set $\mathcal{K} \subseteq \mathbb{R}^n$, and then suffers a loss of $f_t(x_t)$ according to some convex loss function $f_t : \mathcal{K} \to \mathbb{R}$. In the full information setting, the player gets to know the entire loss function f_t after choosing the action, while in the bandit setting, the player knows only the loss value $f_t(x_t)$ of the action. The goal of the player is to minimize her regret, defined as the difference between the total loss she suffers and that of the best fixed action in hindsight.

Many results are known in the full information setting. For general convex loss functions, a regret of $O(\sqrt{nT})$ can be achieved (Zinkevich, 2003), while for strongly convex loss

functions, a smaller regret of $O(n \ln T)$ becomes possible (Hazan et al., 2007). These two results, as well as many others, considered only the worst-case scenario, in which the loss functions have no pattern or are even generated in a malicious way. However, the environments we are in may not always be adversarial, so a research direction is to identify natural patterns or properties of loss functions and to design online algorithms with smaller regrets for them. For loss functions which are linear (and can be seen as vectors), Hazan and Kale (2008) considered a measure called variation, defined as $V = \sum_{t=1}^{T} \|f_t - \mu\|_2^2$, where μ is the average of the loss functions, and they provided an algorithm achieving a regret of $O(\sqrt{V})$. In another work, Hazan and Kale (2009a) considered the online portfolio management problem (Cover, 1991) and achieved a logarithmic regret in terms of a similar measure. Note that loss functions with small variation can be seen as basically centered around their average, which models a stationary environment with loss functions coming from some fixed distribution. Chiang et al. (2012) introduced a more general measure called deviation which models a dynamic environment that usually evolves gradually, including examples such as weather conditions and stock markets. More precisely, Chiang et al. (2012) considered not only linear functions but also convex functions, and defined the deviation as

$$D = \sum_{t=1}^{T} \max_{x \in \mathcal{K}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_2^2 , \tag{1}$$

using the convention that f_0 is the all-0 function, where $\nabla f_\tau(x)$ denotes the gradient of f_τ at x. With this, they provided algorithms achieving a regret of $O(\sqrt{D})$ for convex functions and a smaller regret of $O(n \ln D)$ for strongly convex loss functions. Since one can show that $D \le O(V)$ but not the other way around (Chiang et al., 2012), results with regrets in terms of D are arguably stronger than those in terms of V.

The bandit setting appears much more challenging. For linear functions, Abernethy et al. (2008) achieved a regret of $O(n\sqrt{\vartheta T \ln T})$ using a somewhat involved method of ϑ-self-concordant barriers, while Bubeck et al. (2012) slightly improved the regret to $O(n\sqrt{T \ln T})$ but with an inefficient algorithm. For general convex functions, the best regret currently achieved is $O(T^{2/3}(\ln T)^{1/3})$ by Saha and Tewari (2011), which is far from the $O(\sqrt{nT})$ regret achieved in the full-information setting. Even worse, for strongly convex functions, there is actually an $\Omega(\sqrt{T})$ regret lower bound in the bandit setting (Jamieson et al., 2012), compared to the $O(n \ln T)$ regret upper bound achievable in the full information case. For linear functions with variation V, Hazan and Kale (2009b) achieved a regret of $O(n\sqrt{\vartheta V \ln T})$, but no such result is known for convex functions or strongly convex ones. None is known either for loss functions with small deviation, even for linear functions.

Our goal is to have bandit algorithms for loss functions with small deviation, but it turns out to be difficult as we discuss next. The standard approach for designing a bandit algorithm is to run a full-information algorithm and replace the information it needs by estimated one. For loss functions with small deviation, we would like to apply this to the full-information algorithm of (Chiang et al., 2012), and what it needs in round t is the gradient of the loss function at the action it plays, denoted as ℓ_t. To have a bandit algorithm, a natural attempt is to replace ℓ_t by an estimator g_t using bandit information, which would achieve regrets in terms of $\sum_t \|g_t - g_{t-1}\|_2^2$. However, this deviation of the estimated gradients can be large even when the deviation of the true gradients is small. The

reason is that in most previous works, such as (Flaxman et al., 2005; Abernethy et al., 2008; Abernethy and Rakhlin, 2009; Bubeck et al., 2012), the estimator g_t typically takes the form of $c_t u_t$ for some value $c_t \in \mathbb{R}$ and some vector u_t sampled independently in each round from a set which spans $\mathbb{R}^n$. As a result, u_t and u_{t-1} are very different with high probability, and so are g_t and g_{t-1}, even when ℓ_t and ℓ_{t-1} are close. A possible way around this is to use estimators of a different form. For linear loss functions with small variation, with loss functions centering around their average, Hazan and Kale (2009b) considered estimators of the form $g_t = c_t u_t + \tilde{\mu}_t$ where $\tilde{\mu}_t$ is an estimator of the average, and their success relies on the fact that the average can be estimated accurately with high probability by an online algorithm. This suggests us to use estimators of the form $g_t = c_t u_t + \tilde{g}_{t-1}$ where $\tilde{g}_{t-1}$ is an estimator of ℓ_{t-1}, but it is not clear if it is possible to have an accurate estimator for each ℓ_{t-1} with high probability as each loss function may only appear once. Another issue is the choice of the exploration scheme. Take that of (Flaxman et al., 2005) as an example. In each round, it explores randomly in a neighborhood of diameter δ in order to get a good estimator, but this adds to the regret a term (corresponding to the length of the estimator) which is proportional to $1/\delta^2$ as well as a term which is proportional to δ. Then no good choice of δ can lead to a regret characterized by D instead of by T.

To avoid some of the difficulties, we consider the relaxed two-point bandit setting of (Agarwal et al., 2010), in which one can play two actions, instead of just one, in a given round and get to know their respective loss values, while their average is counted as the loss of that round. In fact, such a relaxation is necessary if we want to achieve a regret comparable to that in the full-information setting for strongly convex functions (Hazan et al., 2007), according to the lower bound of (Jamieson et al., 2012). In such a two-point bandit setting, Agarwal et al. (2010) showed that regrets close to those in the full-information setting can indeed be achieved: $O(n^2\sqrt{T})$ for convex functions and $O(n^2 \ln T)$ for strongly convex functions. One may wonder if their results can be generalized to having regrets characterized by the more refined measure D, instead of simply by T, just as those of (Chiang et al., 2012) in the full-information setting.

We answer this affirmatively. That is, we provide two-point bandit algorithms which achieve regrets close to the full-information ones in (Chiang et al., 2012). For linear functions, our regret is $O(n^{3/2}\sqrt{D})$. For convex functions, our regret is $O(n^2\sqrt{D + \ln T})$, which becomes $O(n^2\sqrt{D})$ when $D \geq \Omega(\ln T)$. For strongly convex functions, our regret is $O(n^2(\ln(D + \ln T))$, which becomes $O(n^2 \ln D)$ when $D \geq \Omega(\ln T)$. Note that the dependencies on D in our regret bounds match those of (Chiang et al., 2012) in the full-information setting. Compared to the regrets of (Agarwal et al., 2010) in the two-point bandit setting, we recover their results in the extreme case with $D = \Omega(T)$, but our regrets become much smaller when D is much smaller than T. The contrast to regrets in the (one-point) bandit setting is even sharper, as the regret of (Saha and Tewari, 2011) for convex functions is substantially higher than ours, while for strongly convex functions, there is actually an $\Omega(\sqrt{D})$ lower bound[1] which is exponentially higher than our upper bound. Moreover, all of our algorithms are simple and efficient, which again demonstrates the power of two-point bandit algorithms. Finally, as our algorithms are based on the full-information ones of

1. Such a lower bound can be easily modified from that of (Jamieson et al., 2012).

(Chiang et al., 2012), we inherit their nice property that all our algorithms can be derived from one single meta algorithm and their regrets can all be analyzed in one single framework.

2. Preliminaries

Let $\mathbb{N}$ denote the set of positive integers and $\mathbb{R}$ the set of real numbers. For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \cdots, n\}$ and $\mathbb{R}^n$ the set of n-dimensional vectors over $\mathbb{R}$. For vectors x, $y \in \mathbb{R}^n$, denote the inner product of x and y by $\langle x, y \rangle$ and the Euclidean norm of x by $\|x\|_2$. For a convex set $\mathcal{X} \subseteq \mathbb{R}^n$ and some $y \in \mathbb{R}^n$, let $\mathcal{P}_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \|x - y\|_2$, which we call the projection of y onto $\mathcal{X}$. Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be the set of standard basis for $\mathbb{R}^n$.

We consider the *online convex optimization problem* with *two-point bandit feedback*, in which an online algorithm must play T rounds in the following way. In each round t, it plays two actions w_t and w'_t from a convex feasible set $\mathcal{K} \subseteq \mathbb{R}^n$, and after that, it receives the loss information $f_t(w_t)$ and $f_t(w'_t)$ and suffers a loss of $\frac{1}{2}\left(f_t(w_t) + f_t(w'_t)\right)$ according to some convex loss function $f_t : \mathcal{K} \to \mathbb{R}$. The goal is to minimize the *expected regret*:

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{2}\left(f_t(w_t) + f_t\left(w'_t\right)\right)\right] - \min_{\pi \in \mathcal{K}} \sum_{t=1}^{T} f_t(\pi), \tag{2}$$

which is the expected total loss of the online algorithm minus that of the best offline algorithm playing a fixed action $\pi \in \mathcal{K}$ for all T rounds, where the expectation is over the randomness of the algorithm.

As in (Flaxman et al., 2005), we assume that the feasible set satisfies the condition that $r\mathbb{B} \subseteq \mathcal{K} \subseteq R\mathbb{B}$, for some positive constants $r \leq R$, where $\mathbb{B} = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ is the unit ball centered at $\mathbf{0}$. We assume that each loss function f_t has bounded gradient $\|\nabla f_t(x)\|_2 \leq G$ for any $x \in \mathcal{K}$, where $\nabla f_t(x)$ denotes the gradient of f_t at x, and note that this implies the G-Lipschitz condition: $|f_t(x) - f_t(y)| \leq G \|x - y\|_2$ for any $x, y \in \mathcal{K}$. As in previous works, we also assume that each loss function is λ-smooth:

$$\|\nabla f_t(x) - \nabla f_t(y)\|_2 \leq \lambda \|x - y\|_2 . \tag{3}$$

In addition, we will also consider loss functions which are *H-strongly convex*. Formally, a function $f : \mathcal{K} \to \mathbb{R}$ is called H-strongly convex, for some $H > 0$, if

$$\forall x, y \in \mathcal{K} : f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{H}{2} \|x - y\|_2^2 . \tag{4}$$

Finally, we will need the following two simple facts, which we prove in Appendix A.

Proposition 1 *(a) For $m \in \mathbb{N}$ and $a_1, \ldots, a_m \in \mathbb{R}$, $\left(\sum_{t=1}^{m} a_t\right)^2 \leq m \sum_{t=1}^{m} a_t^2$. (b) For $n \in \mathbb{N}$ and $x, y \in \mathbb{R}^n$, $\|x + y\|_2^2 \leq 2\|x\|_2^2 + 2\|y\|_2^2$.*

3. Meta Algorithm

All the algorithms in the coming sections are based on the following Meta algorithm, given in Algorithm 1. It is in turn based on the full-information algorithm of (Chiang et al., 2012), which follows the gradient descent algorithm to update $x_{t+1} = \mathcal{P}_{\mathcal{X}}(x_t - \eta \ell_t)$ after seeing ℓ_t,

but plays $\hat{x}_{t+1} = \mathcal{P}_{\mathcal{X}}(x_{t+1} - \eta\ell_t)$ instead in round $t+1$, where $\ell_t = \nabla f_t(\hat{x}_t)$. The idea is that in the case of small deviation, one could use ℓ_t as an approximation of the next ℓ_{t+1}, and play $\hat{x}_{t+1}$ which moves further in the direction of $-\ell_t$, and this can indeed be shown to achieve regrets in terms of the deviation $\sum_t \|\ell_t - \ell_{t-1}\|_2^2$. In the bandit setting, we do not have ℓ_t available, and the standard approach is to feed the full-information algorithm with an estimator for ℓ_t using the bandit information. An easy way to estimate ℓ_t based on that of (Agarwal et al., 2010) is to choose a standard basis vector $\mathbf{e}_{i_t}$ randomly, play two actions $w_t = \hat{x}_t + \delta\mathbf{e}_{i_t}$ and $w'_t = \hat{x}_t - \delta\mathbf{e}_{i_t}$, compute $v_{t,i_t} = \frac{1}{2\delta}(f_t(w_t) - f_t(w'_t))$, and use $\tilde{g}_t = (nv_{t,i_t})\mathbf{e}_{i_t}$ as the estimator. It can be shown that $\mathbb{E}[\tilde{g}_t]$ is close to ℓ_t. If we feed this estimator $\tilde{g}_t$ to the algorithm of (Chiang et al., 2012), we obtain regret bounds in terms of $\sum_t \|\tilde{g}_t - \tilde{g}_{t-1}\|_2^2$, which unfortunately may be much larger than deviation. The reason is that even when $\|\ell_t - \ell_{t-1}\|_2^2$ is small, $\|\tilde{g}_t - \tilde{g}_{t-1}\|_2^2 = \left\|(nv_{t,i_t})\mathbf{e}_{i_t} - (nv_{t-1,i_{t-1}})\mathbf{e}_{i_{t-1}}\right\|_2^2$ may be large if $i_t \neq i_{t-1}$. Thus, in our algorithm, we only follow the idea of (Agarwal et al., 2010) up to computing v_{t,i_t}, in our first three steps, and then we use different estimators. Our key observation is that in the regret term $\|\ell_t - \ell_{t-1}\|_2^2$ of (Chiang et al., 2012), ℓ_t comes from using gradient descent to update x_{t+1}, while ℓ_{t-1} comes from using it as an approximation of ℓ_t to move from x_t to $\hat{x}_t$. Therefore, we distinguish the two different uses and compute two different estimators for them, as shown in step 4 of our algorithm, with g_t as an estimator of ℓ_t which needs to have $\mathbb{E}[g_t]$ close to ℓ_t, and with $\hat{g}_t$ as an approximation of ℓ_{t+1}. Note that g_t and $\hat{g}_t$ computed there are obtained from $\hat{g}_{t-1}$ by modifying only its i_t'th entry, where $\hat{g}_{\tau,i}$ denotes the i'th entry of the vector $\hat{g}_\tau$. Then we do the the update in step 5, which can be seen as that of (Chiang et al., 2012) using the estimators g_t and $\hat{g}_t$. The parameter η_t is the learning rate, which will be chosen differently for different classes of loss functions in the following sections.

Algorithm 1 META algorithm

Let $\mathcal{X} = (1-\mu)\mathcal{K}$. Let $x_1 = \hat{x}_1 = \mathbf{0}$ and $\hat{g}_0 = \mathbf{0}$.

In round $t \in [T]$:

1: Choose i_t uniformly from $[n]$.

2: Play two actions $w_t = \hat{x}_t + \delta\mathbf{e}_{i_t}$ and $w'_t = \hat{x}_t - \delta\mathbf{e}_{i_t}$.

3: Observe partial information $f_t(w_t)$ and $f_t(w'_t)$. Let $v_{t,i_t} = \frac{1}{2\delta}(f_t(w_t) - f_t(w'_t))$.

4: Compute

$$g_t = n\left(v_{t,i_t} - \hat{g}_{t-1,i_t}\right)\mathbf{e}_{i_t} + \hat{g}_{t-1} \quad \text{and} \quad \hat{g}_t = \left(v_{t,i_t} - \hat{g}_{t-1,i_t}\right)\mathbf{e}_{i_t} + \hat{g}_{t-1}.$$

5: Update

$$x_{t+1} = \mathcal{P}_{\mathcal{X}}(x_t - \eta_t g_t) \quad \text{and} \quad \hat{x}_{t+1} = \mathcal{P}_{\mathcal{X}}(x_{t+1} - \eta_{t+1}\hat{g}_t).$$

Next, we derive a general regret bound for our algorithm. Following (Agarwal et al., 2010), we consider a smaller feasible set $\mathcal{X} = (1-\mu)\,\mathcal{K}$ for $\hat{x}_t$, with $\mu = \delta/r$, so that w_t and w'_t played in step 2 are feasible points in $\mathcal{K}$, according to Observation 3.2 of (Flaxman et al., 2005). As in (Agarwal et al., 2010), we can choose an arbitrarily small $\delta > 0$, which is the

advantage one can have in the two-point bandit setting[2]; for our purpose, any δ such that $\delta(\lambda GRn^2/r) \leq o(1/T)$ suffices. Similarly to (Agarwal et al., 2010), to bound the regret of our algorithm against $\bar{\pi} = \arg\min_{x\in\mathcal{K}} \sum_{t=1}^{T} f_t(x)$, which is the best fixed action in $\mathcal{K}$, it suffices to bound the regret according to the actions $\hat{x}_t$'s against $\pi = \arg\min_{x\in\mathcal{X}} \sum_{t=1}^{T} f_t(x)$, which is the best fixed action in $\mathcal{X}$. This is established by the following lemma, which we prove in Appendix B.

Lemma 2 $\sum_{t=1}^{T} \left(\frac{1}{2}\left(f_t(w_t) + f_t(w'_t)\right) - f_t(\bar{\pi})\right) \leq \sum_{t=1}^{T} \left(f_t(\hat{x}_t) - f_t(\pi)\right) + o(1)$.

This allows us to turn our attention to bound the sum $\sum_{t=1}^{T} \left(f_t(\hat{x}_t) - f_t(\pi)\right)$. Recall that $\ell_t = \nabla f_t(\hat{x}_t)$ and let $\ell_{t,i}$ denote the i'th entry of the vector ℓ_t which equals $\nabla_i f_t(\hat{x}_t)$, where $\nabla_i f_t(\hat{x}_t)$ denotes the i'th entry of $\nabla f_t(\hat{x}_t)$. Note that $f_t(\hat{x}_t) - f_t(\pi)$ is at most $\langle \ell_t, \hat{x}_t - \pi\rangle$ for convex f_t and at most $\langle \ell_t, \hat{x}_t - \pi\rangle - \frac{H}{2}\left\|\hat{x}_t - \pi\right\|_2^2$ for H-strongly convex f_t. Thus, we have the following.

Lemma 3 *Let $C_t = 0$ for convex f_t and $C_t = \frac{H}{2}\left\|\hat{x}_t - \pi\right\|_2^2$ for H-strongly convex f_t. Then we have $f_t(\hat{x}_t) - f_t(\pi) \leq \langle \ell_t, \hat{x}_t - \pi\rangle - C_t$.*

Next, recall that the update rule of our algorithm can be seen as that of Chiang et al. (2012) using g_t as an estimator of the gradient ℓ_t and using $\hat{g}_t$ as an approximation of ℓ_{t+1}. Then we have the following from (Chiang et al., 2012); for completeness, we provide the proof in Appendix C.

Lemma 4 *Let $S_t = \eta_t \left\|g_t - \hat{g}_{t-1}\right\|_2^2$, $A_t = \frac{1}{2\eta_t}\left\|\pi - x_t\right\|_2^2 - \frac{1}{2\eta_t}\left\|\pi - x_{t+1}\right\|_2^2$, and $B_t = \frac{1}{2\eta_t}\left\|x_{t+1} - \hat{x}_t\right\|_2^2 + \frac{1}{2\eta_t}\left\|\hat{x}_t - x_t\right\|_2^2$. Then we have $\langle g_t, \hat{x}_t - \pi\rangle \leq S_t + A_t - B_t$.*

To connect Lemma 3 with Lemma 4, we rely on the following, which we prove in Appendix D. Note that this justifies our use of g_t as an estimator of ℓ_t.

Lemma 5 $\mathbb{E}\left[\langle \ell_t, \hat{x}_t - \pi\rangle\right] \leq \mathbb{E}\left[\langle g_t, \hat{x}_t - \pi\rangle\right] + o(1/T)$.

Finally, by taking expectation on the bound in Lemma 2 and combining the bounds in the previous three lemmas, we have the following theorem.

Theorem 6 *The expected regret of the* Meta *algorithm is*

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(\frac{1}{2}\left(f_t(w_t) + f_t(w'_t)\right) - f_t(\bar{\pi})\right)\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\left(S_t + A_t - B_t - C_t\right)\right] + o(1).$$

In the following sections, we will consider different classes of loss functions and instantiate the Meta algorithm accordingly, and then concrete regret bounds will be derived. Note that the key term in the regret bound of Theorem 6 is the sum of $S_t = \eta_t \left\|g_t - \hat{g}_{t-1}\right\|_2^2$, as it is related to the deviation according to the following lemma, which we prove in Appendix E.

2. This is because the estimator g_t now can have a bounded length, unlike the (one-point) bandit setting in which the estimator's length and consequently the regret grows proportionally to $1/\delta^2$.

Lemma 7 *For any $t \in [T]$, let α_t be the smallest integer such that $0 \le \alpha_t < t$ and $i_\tau \neq i_t$ for any $\alpha_t < \tau < t$, and let $\hat{D}_t = (\ell_{t,i_t} - \ell_{\alpha_t,i_t})^2$. Then, $\|g_t - \hat{g}_{t-1}\|_2^2 \le n^2 \hat{D}_t + o(1/T)$.*

Note that $\hat{D}_t$ is related to the difference between two gradients $t - \alpha_t$ rounds away, which is related to the deviation accumulated through those rounds. Thus, to have a small $\hat{D}_t$, we would like to have α_t close to t. This leads us to adopt the exploration scheme of (Agarwal et al., 2010) but modify it to sample each $\mathbf{e}_i$ with equal probability, so that $t - \alpha_t$ can be shown to have a small expected value.

4. Linear Loss Functions

In this section, we consider linear loss functions. The deviation of such a sequence of loss functions according to (1) now turns into $D = \sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_2^2 = \sum_{t=1}^T \|f_t - f_{t-1}\|_2^2$, where we let $f_0 = \ell_0$ be the all-0 vector $\mathbf{0}$. To instantiate the META algorithm, we set $\eta_t = \eta$ for all t, for some η be chosen later.

To bound the expected regret of our algorithm, we know from Theorem 6 that it suffices to bound

$$\mathbb{E}\left[\sum_{t=1}^T (S_t + A_t - B_t - C_t)\right] \le \mathbb{E}\left[\sum_{t=1}^T S_t\right] + \mathbb{E}\left[\sum_{t=1}^T A_t\right],$$

as $B_t \ge 0$ and $C_t = 0$ for linear functions. Note that with $A_t = \frac{1}{2\eta}\|\pi - x_t\|_2^2 - \frac{1}{2\eta}\|\pi - x_{t+1}\|_2^2$, we have by telescoping that

$$\sum_{t=1}^T A_t = \frac{1}{2\eta}\|\pi - x_1\|_2^2 - \frac{1}{2\eta}\|\pi - x_{T+1}\|_2^2 \le \frac{R^2}{2\eta},$$

as $\|\pi - x_1\|_2^2 \le R^2$ and $\|\pi - x_{T+1}\|_2^2 \ge 0$. It remains to bound $\mathbb{E}\left[\sum_{t=1}^T S_t\right]$.

We know from Lemma 7 that $S_t = \eta\|g_t - \hat{g}_{t-1}\|_2^2 \le \eta n^2 \hat{D}_t + o(1/T)$, where $\hat{D}_t = (\ell_{t,i_t} - \ell_{\alpha_t,i_t})^2$, which is related to the difference between two loss functions several rounds away, instead of just between two consecutive ones as used by the definition of deviation. To bridge the gap, we need the following.

Lemma 8 *For $t \in [T]$ and $i \in [n]$, let $\rho_{t,i} = \max\{\tau_2 - \tau_1 : 0 \le \tau_1 < t \le \tau_2 \le T$ and $i_\tau \neq i$ for any $\tau_1 < \tau < \tau_2\}$. Then,*

$$\sum_{t=1}^T \hat{D}_t \le \sum_{t=1}^T \sum_{i=1}^n \rho_{t,i} (\ell_{t,i} - \ell_{t-1,i})^2.$$

Proof From the definition, α_t is the most recent round before round t such that $i_{\alpha_t} = i_t$, or $\alpha_t = 0$ if there is no such round. Then for any $t \in [T]$ and $i \in [n]$ such that $i_t = i$, we can rewrite $\hat{D}_t = (\ell_{t,i} - \ell_{\alpha_t,i})^2$ as

$$\left(\sum_{\tau=\alpha_t+1}^t (\ell_{\tau,i} - \ell_{\tau-1,i})\right)^2 \le (t - \alpha_t) \sum_{\tau=\alpha_t+1}^t (\ell_{\tau,i} - \ell_{\tau-1,i})^2 = \sum_{\tau=\alpha_t+1}^t \rho_{\tau,i} (\ell_{\tau,i} - \ell_{\tau-1,i})^2,$$

where the inequality follows from Proposition 1(a), and the equality follows from the fact that $\rho_{\tau,i} = (t - \alpha_t)$ for any $\alpha_t + 1 \le \tau \le t$. Therefore,

$$\sum_{t=1}^{T} \hat{D}_t = \sum_{i=1}^{n} \sum_{t:i_t=i} \hat{D}_t \le \sum_{i=1}^{n} \sum_{t:i_t=i} \sum_{\tau=\alpha_t+1}^{t} \rho_{\tau,i} \left(\ell_{\tau,i} - \ell_{\tau-1,i}\right)^2 \le \sum_{i=1}^{n} \sum_{\tau=1}^{T} \rho_{\tau,i} \left(\ell_{\tau,i} - \ell_{\tau-1,i}\right)^2,$$

where the last inequality holds since for any i, the intervals $[\alpha_t + 1, t]$, for t with $i_t = i$, have no intersection, according to the definition of α_t. ■

With this lemma, we can have the following, which links regret to deviation.

Lemma 9 $\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] \le 2nD$.

Proof From Lemma 8, we know that

$$\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] \le \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E}\left[\rho_{t,i} \left(\ell_{t,i} - \ell_{t-1,i}\right)^2\right] = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E}\left[\rho_{t,i}\right] \left(\ell_{t,i} - \ell_{t-1,i}\right)^2,$$

where the last equality follows from the fact that the gradient of a linear function does not depend on where it is taken, so each $(\ell_{t,i} - \ell_{t-1,i})^2$ is a fixed value independent of the randomness of the expectation. It remains to bound $\mathbb{E}[\rho_{t,i}]$. Recall the definition of $\rho_{t,i}$, and suppose $\rho_{t,i} = \tau_2 - \tau_1$ where $0 \le \tau_1 < t \le \tau_2 \le T$ and $i_\tau \ne i$ for $\tau_1 < \tau < \tau_2$. Then we can write the random variable $\rho_{t,i}$ as the sum of two random variables $t - \tau_1$ and $\tau_2 - t$, and observe that both can be bounded by a geometric random variable, denoted by Z, with $\Pr[Z = k] = (1/n)(1 - 1/n)^{k-1}$ for $k \ge 1$ and $\mathbb{E}[Z] = n$; in fact, $t - \tau_1 = \min\{Z, t\} \le Z$ and $\tau_2 - t = \min\{Z - 1, T - t\} \le Z$. Thus,

$$\mathbb{E}\left[\rho_{t,i}\right] = \mathbb{E}\left[t - \tau_1\right] + \mathbb{E}\left[\tau_2 - t\right] \le 2n. \tag{5}$$

Consequently, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] \le \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E}\left[\rho_{t,i}\right] \left(\ell_{t,i} - \ell_{t-1,i}\right)^2 \le 2n \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\ell_{t,i} - \ell_{t-1,i}\right)^2 = 2nD.$$

■

Since $\mathbb{E}\left[\sum_{t=1}^{T} S_t\right] \le \eta n^2 \mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] + o(1) \le 2\eta n^3 D + o(1)$, we can conclude that the expected regret of our algorithm is at most

$$2\eta n^3 D + \frac{R^2}{2\eta} + o(1) \le O\left(Rn^{3/2}\sqrt{D}\right),$$

by choosing $\eta = R/\sqrt{n^3 D}$, which gives us the following.

Theorem 10 *Suppose the loss functions are linear and have deviation D. Then the expected regret of our algorithm is at most $O(Rn^{3/2}\sqrt{D})$.*

5. Convex Loss Functions

In this section we consider convex loss functions. The deviation D of loss functions is measured by (1), which is $\sum_{t=1}^{T} \max_{x\in\mathcal{K}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_2^2$. To instantiate the Meta algorithm for such loss functions, we again set $\eta_t = \eta$ for all t, for some η to be chosen later.

To bound the expected regret of our algorithm, we know from Theorem 6 that it suffices to bound

$$\mathbb{E}\left[\sum_{t=1}^{T}(S_t + A_t - B_t - C_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} S_t\right] + \mathbb{E}\left[\sum_{t=1}^{T} A_t\right] - \mathbb{E}\left[\sum_{t=1}^{T} B_t\right], \tag{6}$$

as $C_t = 0$ for convex functions. As in Section 4, we have $\sum_{t=1}^{T} A_t \le \frac{R^2}{2\eta}$. On the other hand, we will need the help of $\mathbb{E}\left[\sum_{t=1}^{T} B_t\right]$ here. The following lemma from (Chiang et al., 2012) provides a lower bound for it; for completeness, we give the proof in Appendix F.

Lemma 11 $\mathbb{E}\left[\sum_{t=1}^{T} B_t\right] \ge \frac{1}{4\eta}\mathbb{E}\left[\sum_{t=1}^{T} \|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right] - O(1)$.

Next, to bound $\mathbb{E}\left[\sum_{t=1}^{T} S_t\right]$, we again turn to bound $\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right]$, as $S_t \le \eta n^2 \hat{D}_t + o(1/T)$. Note that unlike a linear function, the gradient of a convex function now depends on where the gradient is taken, and Lemma 9, which works for linear functions, does not work here for convex functions. As in previous works, we assume that each f_t satisfies the λ-smoothness condition given in (3), and note that according to the discussion in Section 5 of (Chiang et al., 2012), the smoothness condition is in fact necessary in order to achieve a regret bound in terms of deviation. To obtain a cleaner bound, let us assume that the parameters λ and R are constants, while the parameters T and D are large, with $T, D \ge n, \lambda, R$. Our key lemma is the following.

Lemma 12 $\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] \le O(n^2 D) + O(n \ln T) \cdot \mathbb{E}\left[\sum_{t=1}^{T} \|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right]$.

Proof We know from Lemma 8 that

$$\sum_{t=1}^{T} \hat{D}_t \le \sum_{t=1}^{T}\sum_{i=1}^{n} \rho_{t,i} \left(\ell_{t,i} - \ell_{t-1,i}\right)^2.$$

By definition, $(\ell_{t,i} - \ell_{t-1,i})^2 = (\nabla_i f_t(\hat{x}_t) - \nabla_i f_{t-1}(\hat{x}_{t-1}))^2$, which does not correspond to a term in deviation because the gradients are taken at different points. To relate it to deviation, let $\hat{\ell}_{t-1} = \nabla f_{t-1}(\hat{x}_t)$ with $\hat{\ell}_{t-1,i} = \nabla_i f_{t-1}(\hat{x}_t)$, and rewrite $(\ell_{t,i} - \ell_{t-1,i})^2$ as $(\ell_{t,i} - \hat{\ell}_{t-1,i} + \hat{\ell}_{t-1,i} - \ell_{t-1,i})^2$, which is at most $2(\ell_{t,i} - \hat{\ell}_{t-1,i})^2 + 2(\hat{\ell}_{t-1,i} - \ell_{t-1,i})^2$ by Proposition 1(a). Then

$$\mathbb{E}\left[\sum_{t=1}^{T} \hat{D}_t\right] \le 2\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \rho_{t,i}\left(\ell_{t,i} - \hat{\ell}_{t-1,i}\right)^2\right] + 2\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \rho_{t,i}\left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2\right]. \tag{7}$$

The first expectation in (7) is now related to deviation since $(\ell_{t,i} - \hat{\ell}_{t-1,i})^2 = (\nabla_i f_t(\hat{x}_t) - \nabla_i f_{t-1}(\hat{x}_t))^2$, with the two gradients taken at the same point. However, unlike in the case

of linear functions, $(\ell_{t,i} - \hat{\ell}_{t-1,i})^2$ is now itself a random variable, which depends on the randomness of the expectation and has correlation with $\rho_{t,i}$. To overcome this problem, we use the upper bound

$$\left(\ell_{t,i} - \hat{\ell}_{t-1,i}\right)^2 \leq \|\nabla f_t(\hat{x}_t) - \nabla f_{t-1}(\hat{x}_t)\|_2^2 \leq D_t,$$

where $D_t = \max_{x \in \mathcal{K}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_2^2$ is a fixed value. Then the first expectation in (7) can be bounded from above by

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \rho_{t,i} D_t\right] = \sum_{t=1}^{T}\sum_{i=1}^{n} \mathbb{E}\left[\rho_{t,i}\right] D_t \leq \sum_{t=1}^{T}\sum_{i=1}^{n} (2n) D_t = 2n^2 D,$$

where the first inequality uses the inequality (5) from Section 4, and the second equality uses the fact that $D = \sum_{t=1}^{T} D_t$.

The second expectation in (7) is slightly harder to bound. As before, the complication comes from the correlation between the two random variables $\rho_{t,i}$ and $(\hat{\ell}_{t,i} - \ell_{t,i})^2$, but here we do not have a fixed upper bound for $(\hat{\ell}_{t,i} - \ell_{t,i})^2$ which is good enough. Instead, we turn to bound $\rho_{t,i}$. Let $\bar{\rho} = 4n \ln T$ and let Q denote the bad event that $\rho_{t,i} > \bar{\rho}$ for some $t \in [T]$ and $i \in [n]$, which only happens with probability

$$\Pr[Q] \leq Tn \left(1 - \frac{1}{n}\right)^{4n \ln T} \leq Tn \cdot e^{-4 \ln T} = \frac{n}{T^3}.$$

Then the second expectation in (7) can be expressed as

$$\Pr[\neg Q] \cdot \mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \rho_{t,i} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2 \,\middle|\, \neg Q\right] + \Pr[Q] \cdot \mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \rho_{t,i} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2 \,\middle|\, Q\right],$$

where the first term is at most

$$\Pr[\neg Q] \cdot \bar{\rho} \cdot \mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2 \,\middle|\, \neg Q\right] \leq \bar{\rho} \cdot \mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2\right],$$

and the second term is at most

$$\Pr[Q] \cdot T \cdot \mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{n} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2 \,\middle|\, Q\right].$$

Note that by the definition of $\hat{\ell}_t$ and by the λ-smoothness condition,

$$\sum_{i=1}^{n} \left(\hat{\ell}_{t,i} - \ell_{t,i}\right)^2 = \|\nabla f_t(\hat{x}_{t+1}) - \nabla f_t(\hat{x}_t)\|_2^2 \leq \lambda^2 \cdot \|\hat{x}_{t+1} - \hat{x}_t\|_2^2,$$

with $\|\hat{x}_{t+1} - \hat{x}_t\|_2 \leq 2R$. Thus, the second expectation in (7) is at most

$$\bar{\rho} \cdot \lambda^2 \cdot \mathbb{E}\left[\sum_{t=0}^{T-1} \|\hat{x}_{t+1} - \hat{x}_t\|_2^2\right] + \frac{n}{T^3} \cdot T^2 \lambda^2 4R^2 \leq O(n \ln T) \cdot \mathbb{E}\left[\sum_{t=1}^{T} \|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right] + o(1).$$

Finally, by combining the bounds for the two expectations in (7), we have the lemma. ■

With this lemma, we obtain

$$\mathbb{E}\left[\sum_{t=1}^{T} S_t\right] \leq O(\eta n^4 D) + O(\eta n^3 \ln T) \cdot \mathbb{E}\left[\sum_{t=1}^{T} \|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right] + o(1).$$

For some $\eta \leq O(1/(n^2\sqrt{D + \ln T}))$, we can have $O(\eta n^3 \ln T) \leq \frac{1}{4\eta}$ so that the second term above is at most $\mathbb{E}\left[\sum_{t=1}^{T} B_t\right] + O(1)$ by Lemma 11, and the expected regret of our algorithm, according to (6), can be bounded from above by

$$O\left(\eta n^4 D + \frac{1}{\eta}\right) \leq O\left(n^2\sqrt{D + \ln T}\right).$$

As a result, we have the following theorem.

Theorem 13 *When the loss functions are convex and have deviation D, the expected regret of our algorithm is at most $O\left(n^2\sqrt{D + \ln T}\right)$, where the hiding constant factor is a small polynomial of the constants λ and R.*

6. Strongly Convex Loss Functions

In this section we consider H-strongly convex functions. That is, we suppose that for some constant $H > 0$, each loss function f_t is H-strongly convex, so that

$$f_t(\hat{x}_t) - f_t(\pi) \leq \langle \ell_t, \hat{x}_t - \pi \rangle - \frac{H}{2} \|\pi - \hat{x}_t\|_2^2. \tag{8}$$

The deviation D of the loss functions is again measured by (1). To instantiate the Meta algorithm for such loss functions, now we choose the learning rate

$$\eta_t = 1 \Bigg/ \left(1 + \frac{H}{2} + \frac{H}{2\gamma} \sum_{\tau=1}^{t-1} \|g_\tau - \hat{g}_{\tau-1}\|_2^2\right),$$

with $\gamma = 5n^2G^2$ so that $\gamma \geq \|g_t - \hat{g}_{t-1}\|_2^2$ for any $t \in [T]$.[3] It is easy to verify that η_{t+1} can be computed at the end of round t for updating $\hat{x}_{t+1}$, as g_t and $\hat{g}_{t-1}$ are available then.

To bound the expected regret of our algorithm, we know from Theorem 6 that it suffices to bound

$$\mathbb{E}\left[\sum_{t=1}^{T} (S_t + A_t - B_t - C_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} (S_t + A_t - C_t)\right] - \mathbb{E}\left[\sum_{t=1}^{T} B_t\right],$$

where $C_t = \frac{H}{2} \|\pi - \hat{x}_t\|_2^2$ for H-strongly convex functions. With the help of such C_t, we can reduce the regret down to only logarithmic in D. Our key lemma is the following, and we give the proof in Appendix G, which follows closely a similar one in (Chiang et al., 2012).

3. From Lemma 7, we know that $\|g_t - \hat{g}_{t-1}\|_2^2 \leq n^2 \|\ell_t - \ell_{\alpha_t}\|_2^2 + o(1/T)$, which by Proposition 1(b) is at most $n^2 \left(2\|\ell_t\|_2^2 + 2\|\ell_{\alpha_t}\|_2^2\right) + o(1/T) \leq 5n^2G^2$, as each gradient is assumed to have L_2-norm at most G.

Lemma 14 $\sum_{t=1}^{T}(S_t + A_t - C_t) \le \frac{4\gamma}{H}\ln\left(1 + \frac{H}{2\gamma}\sum_{t=1}^{T}\|g_t - \hat{g}_{t-1}\|_2^2\right) + O(1)$.

From Lemma 7, we know that $\|g_t - \hat{g}_{t-1}\|_2^2 \le n^2\hat{D}_t + o(1/T)$. Furthermore, as in Section 5, we assume that each f_t satisfies the λ-smoothness condition given in (3), so we can use the upper bound for $\mathbb{E}\left[\sum_{t=1}^{T}\hat{D}_t\right]$ in Lemma 12. As before, to obtain a cleaner bound, we assume that the parameters λ, R, G, H are all constants, and the parameters T, D are large, with $T, D \ge n, \lambda, R, G, H$. Then since the logarithm function is concave,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} S_t + A_t - C_t\right] &\le \frac{4\gamma}{H}\ln\left(1 + \frac{H}{2\gamma}\mathbb{E}\left[\sum_{t=1}^{T}\|g_t - \hat{g}_t\|_2^2\right]\right) + O(1)\\ &\le \frac{4\gamma}{H}\ln\left(O(n^2D) + O(n\ln T)\cdot\mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right]\right),\end{aligned}$$

by Lemma 12. If $\mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right] \le O(1)$, we immediately have

$$\mathbb{E}\left[\sum_{t=1}^{T} S_t + A_t - C_t\right] \le \frac{4\gamma}{H}\ln\left(O(n^2D + n\ln T)\right) \le O(\gamma\ln(D + \ln T)).$$

Thus, let us assume otherwise. Then, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} S_t + A_t - C_t\right] &\le \frac{4\gamma}{H}\ln\left(O(n^2D + n\ln T)\cdot\mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right]\right)\\ &\le O(\gamma\ln(D + \ln T)) + \frac{4\gamma}{H}\ln\mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right].\end{aligned}$$

The second term above may be large, and to cancel it, we rely on the following lemma, which we prove in Appendix H.

Lemma 15 $\mathbb{E}\left[\sum_{t=1}^{T} B_t\right] \ge \frac{1}{4}\mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right] - O(1)$.

Let $W = \mathbb{E}\left[\sum_{t=1}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\right]$, and note that $\frac{1}{4}W \ge \frac{4\gamma}{H}\ln W$ when $W \ge (\frac{\gamma}{H})^c$ for some constant c, which implies that $\frac{4\gamma}{H}\ln W - \frac{1}{4}W \le O(\gamma\ln\gamma)$. Thus, we can conclude that

$$\mathbb{E}\left[\sum_{t=1}^{T} S_t + A_t - C_t - B_t\right] \le O\left(\gamma\ln\left(D + \ln T\right)\right) + O\left(\gamma\ln\gamma\right) \le O(n^2\ln(D + \ln T)).$$

As a result, we have the following theorem.

Theorem 16 *When the loss functions are H-strongly convex and have deviation D, the expected regret of our algorithm is at most $O(n^2\ln(D + \ln T))$, where the hiding constant factor is a small polynomial of the constants λ, R, G, and $1/H$.*

References

Jacob Abernethy and Alexander Rakhlin. Beating the adaptive bandit with high probability. In *COLT*, 2009.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *COLT*, pages 263–274, 2008.

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

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. ISBN 0521833787.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. *Journal of Machine Learning Research - Proceedings Track*, 23:41.1–41.14, 2012.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. *Journal of Machine Learning Research - Proceedings Track*, 23:6.1–6.20, 2012.

Thomas Cover. Universal portfolios. *Mathematical Finance*, 1:1–19, 1991.

Abraham Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA*, pages 385–394, 2005.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. In *COLT*, pages 57–68, 2008.

Elad Hazan and Satyen Kale. On stochastic and worst-case models for investing. In *NIPS*, pages 709–717, 2009a.

Elad Hazan and Satyen Kale. Better algorithms for benign bandits. In *SODA*, pages 38–47, 2009b.

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

Kevin G. Jamieson, Robert D. Nowak, and Benjamin Recht. Query complexity of derivative-free optimization. In *NIPS*, 2012.

Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. *Journal of Machine Learning Research - Proceedings Track*, 15:636–642, 2011.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pages 928–936, 2003.

Appendix A. Proof of Proposition 1

To prove (a), we let u be the all-1 vector and v the vector $(a_1, \ldots, a_m)$, and by the Cauchy-Schwarz inequality, we have $\left(\sum_{t=1}^m a_t\right)^2 = \langle u, v\rangle^2 \leq \|u\|_2^2 \|v\|_2^2 = m \sum_{t=1}^m a_t^2$. To prove (b), simply note that $2\|x\|_2^2 + 2\|y\|_2^2 - \|x+y\|_2^2 = \|x-y\|_2^2 \geq 0$.

Appendix B. Proof of Lemma 2

Observe that it suffices to prove that both $\sum_{t=1}^T \left(\frac{1}{2}\left(f_t(w_t) + f_t(w_t')\right) - f_t(\hat{x}_t)\right) \leq o(1)$ and $\sum_{t=1}^T \left(f_t(\pi) - f_t(\bar{\pi})\right) \leq o(1)$ hold.

First, from the G-Lipschitz condition, $f_t(w_t) - f_t(\hat{x}_t) \leq G\|w_t - \hat{x}_t\|_2 \leq G\delta$, and similarly, $f_t(w_t') - f_t(\hat{x}_t) \leq G\delta$. Thus

$$\sum_{t=1}^T \left(\frac{1}{2}\left(f_t(w_t) + f_t(w_t')\right) - f_t(\hat{x}_t)\right) \leq TG\delta \leq o(1).$$

Next, following the idea in (Flaxman et al., 2005), we know that as a convex function, $f_t((1-\mu)\bar{\pi}) = f_t((1-\mu)\bar{\pi} + \mu\mathbf{0}) \leq (1-\mu) f_t(\bar{\pi}) + \mu f_t(\mathbf{0}) = f_t(\bar{\pi}) + \mu\left(f_t(\mathbf{0}) - f_t(\bar{\pi})\right)$, where the second term is at most $\mu GR \leq o(1/T)$. By summing over t, we have

$$\sum_{t=1}^T f_t(\pi) \leq \sum_{t=1}^T f_t((1-\mu)\bar{\pi}) \leq \sum_{t=1}^T f_t(\bar{\pi}) + o(1),$$

where the first inequality holds since $(1-\mu)\bar{\pi} \in \mathcal{X}$ and $\pi = \arg\min_{x\in\mathcal{X}} \sum_{t=1}^T f_t(x)$. This implies that $\sum_{t=1}^T \left(f_t(\pi) - f_t(\bar{\pi})\right) \leq o(1)$, and we have the lemma.

Appendix C. Proof of Lemma 4

We remark that our proof is a simplification of the more general one given in (Chiang et al., 2012) with $\mathcal{R}_t(x) = \frac{1}{2\eta_t}\|x\|_2^2$.

Let us rewrite $\langle g_t, \hat{x}_t - \pi\rangle$ as $\langle g_t, \hat{x}_t - x_{t+1}\rangle + \langle g_t, x_{t+1} - \pi\rangle$ which equals

$$\langle g_t - \hat{g}_{t-1}, \hat{x}_t - x_{t+1}\rangle + \langle \hat{g}_{t-1}, \hat{x}_t - x_{t+1}\rangle + \langle g_t, x_{t+1} - \pi\rangle. \tag{9}$$

The first term above is at most $\|g_t - \hat{g}_{t-1}\|_2 \|\hat{x}_t - x_{t+1}\|_2$, by the Cauchy-Schwarz inequality, which is at most $\eta_t \|g_t - \hat{g}_{t-1}\|_2^2 = S_t$ by the next proposition.

Proposition 17 $\|\hat{x}_t - x_{t+1}\|_2 \leq \eta_t \|g_t - \hat{g}_{t-1}\|_2$.

Proof Let $\phi(w) = \|w - (x_t - \eta_t \hat{g}_{t-1})\|_2^2$ so that $\hat{x}_t = \arg\min_{w\in\mathcal{X}} \phi(w)$. Then by the optimality criterion in convex optimization (see pages 139–140 of (Boyd and Vandenberghe, 2004)), we have $\langle \nabla\phi(\hat{x}_t), x_{t+1} - \hat{x}_t\rangle \geq 0$, with $\nabla\phi(\hat{x}_t) = 2(\hat{x}_t - (x_t - \eta_t\hat{g}_{t-1}))$, which implies that

$$\langle \hat{x}_t - (x_t - \eta_t\hat{g}_{t-1}), x_{t+1} - \hat{x}_t\rangle \geq 0.$$

Similarly, by letting $\phi(w) = \|w - (x_t - \eta_t g_t)\|_2^2$ so that $x_{t+1} = \arg\min_{w\in\mathcal{X}} \phi(w)$, we have

$$\langle x_{t+1} - (x_t - \eta_t g_t), \hat{x}_t - x_{t+1}\rangle \geq 0.$$

Adding these two inequalities together, we have

$$\langle (\hat{x}_t - x_{t+1}) + \eta_t (\hat{g}_{t-1} - g_t), x_{t+1} - \hat{x}_t \rangle \geq 0,$$

which implies that

$$\langle \hat{x}_t - x_{t+1}, \hat{x}_t - x_{t+1} \rangle \leq \langle \eta_t (\hat{g}_{t-1} - g_t), x_{t+1} - \hat{x}_t \rangle \leq \|\eta_t (\hat{g}_{t-1} - g_t)\|_2 \|\hat{x}_t - x_{t+1}\|_2,$$

by the Cauchy-Schwarz inequality. As $\langle \hat{x}_t - x_{t+1}, \hat{x}_t - x_{t+1} \rangle = \|\hat{x}_t - x_{t+1}\|_2^2$, we can divide both sides of the inequality above by $\|\hat{x}_t - x_{t+1}\|_2$, and the proposition follows. ■

To bound the other two terms in (9), we need the following.

Proposition 18 *Suppose $\eta > 0$, $g \in \mathbb{R}^n$, $u \in \mathcal{X}$, and $v = \arg\min_{x \in \mathcal{X}} \|x - (u - \eta g)\|_2$. Then for any $w \in \mathcal{X}$,*

$$\langle g, v - w \rangle \leq \frac{1}{2\eta} \left(\|w - u\|_2^2 - \|w - v\|_2^2 - \|v - u\|_2^2 \right).$$

Proof Let $\phi(x) = \|x - (u - \eta g)\|_2^2$ so that $v = \arg\min_{x \in \mathcal{X}} \phi(x)$. Then from the optimality criterion, $\langle \nabla \phi(v), w - v \rangle \geq 0$, with $\nabla \phi(v) = 2(v - (u - \eta g)) = 2((v - u) + \eta g)$, which implies that $\langle g, v - w \rangle \leq \frac{1}{\eta} \langle v - u, w - v \rangle$. By a straightforward calculation, $\langle v - u, w - v \rangle = \frac{1}{2}(\|w - u\|_2^2 - \|w - v\|_2^2 - \|v - u\|_2^2)$, and the proposition follows. ■

From Proposition 18, we have

$$\begin{aligned} \langle \hat{g}_{t-1}, \hat{x}_t - x_{t+1} \rangle &\leq \frac{1}{2\eta_t} \left(\|x_{t+1} - x_t\|_2^2 - \|x_{t+1} - \hat{x}_t\|_2^2 - \|\hat{x}_t - x_t\|_2^2 \right) \text{ and} \\ \langle g_t, x_{t+1} - \pi \rangle &\leq \frac{1}{2\eta_t} \left(\|\pi - x_t\|_2^2 - \|\pi - x_{t+1}\|_2^2 - \|x_{t+1} - x_t\|_2^2 \right). \end{aligned}$$

Adding the two inequalities above, we get that $\langle \hat{g}_{t-1}, \hat{x}_t - x_{t+1} \rangle + \langle g_t, x_{t+1} - \pi \rangle$ is at most

$$\frac{1}{2\eta_t} \left(\|\pi - x_t\|_2^2 - \|\pi - x_{t+1}\|_2^2 \right) - \frac{1}{2\eta_t} \left(\|x_{t+1} - \hat{x}_t\|_2^2 + \|\hat{x}_t - x_t\|_2^2 \right) = A_t - B_t.$$

Combining this with $\langle g_t - \hat{g}_{t-1}, \hat{x}_t - x_{t+1} \rangle \leq S_t$ derived before, we have the lemma.

Appendix D. Proof of Lemma 5

Let $v_{t,i} = \frac{1}{2\delta} (f_t(\hat{x}_t + \delta \mathbf{e}_i) - f_t(\hat{x}_t - \delta \mathbf{e}_i))$ so that $v_{t,i_t} = \frac{1}{2\delta} (f_t(w_t) - f_t(w'_t))$ and $g_t = n (v_{t,i_t} - \hat{g}_{t-1,i_t}) \mathbf{e}_{i_t} + \hat{g}_{t-1}$.

Let us first consider any fixed choice of $i_{[t-1]} = (i_1, \ldots, i_{t-1})$, which has $\hat{x}_t$, $\ell_t = \nabla f_t(\hat{x}_t)$, and $\hat{g}_{t-1}$ fixed, with i_t still left random. Let $\mathbb{E}_t[\cdot]$ denote the expectation over the random i_t, conditioned on the fixed $i_{[t-1]}$. Note that

$$\mathbb{E}_t[g_t] = \mathbb{E}_t[n v_{t,i_t} \mathbf{e}_{i_t}] - \mathbb{E}_t[(n \hat{g}_{t-1,i_t}) \mathbf{e}_{i_t} - \hat{g}_{t-1}],$$

where the second term above is zero since i_t is chosen uniformly over $[n]$, and the first term above is

$$\mathbb{E}_t\left[nv_{t,i_t}\mathbf{e}_{i_t}\right] = \sum_{i=1}^{n} v_{t,i}\mathbf{e}_i = \sum_{i=1}^{n} \frac{1}{2\delta}\left(f_t(\hat{x}_t + \delta\mathbf{e}_i) - f_t(\hat{x}_t - \delta\mathbf{e}_i)\right)\mathbf{e}_i.$$

Then our goal becomes to show that the above is close to $\ell_t = \nabla f_t(\hat{x}_t)$, and for that it suffices to show that each $v_{t,i}$ is close to $\ell_{t,i} = \nabla_i f_t(\hat{x}_t)$. Note that by Taylor's expansion, $f_t(\hat{x}_t + \delta\mathbf{e}_i) - f_t(\hat{x}_t - \delta\mathbf{e}_i) = \langle \nabla f_t(\xi_{t,i}), 2\delta\mathbf{e}_i\rangle$ for some $\xi_{t,i}$ on the line between $\hat{x}_t + \delta\mathbf{e}_i$ and $\hat{x}_t - \delta\mathbf{e}_i$, which implies that

$$v_{t,i} = \frac{1}{2\delta}\langle \nabla f_t(\xi_{t,i}), 2\delta\mathbf{e}_i\rangle = \nabla_i f_t(\xi_{t,i}).$$

Then by the λ-smoothness assumption, we have

$$|\ell_{t,i} - v_{t,i}| = |\nabla_i f_t(\hat{x}_t) - \nabla_i f_t(\xi_{t,i})| \le \|\nabla f_t(\hat{x}_t) - \nabla f_t(\xi_{t,i})\|_2 \le \lambda\|\hat{x}_t - \xi_{t,i}\|_2 \le \lambda\delta, \quad (10)$$

which implies that

$$\|\ell_t - \mathbb{E}_t[g_t]\|_2^2 = \|\ell_t - \mathbb{E}_t\left[nv_{t,i_t}\mathbf{e}_{i_t}\right]\|_2^2 = \sum_{i=1}^{n}\left(\ell_{t,i} - v_{t,i}\right)^2 \le n\left(\lambda\delta\right)^2,$$

and thus by the Cauchy-Schwarz inequality,

$$\langle \ell_t - \mathbb{E}_t[g_t], \hat{x}_t - \pi\rangle \le \|\ell_t - \mathbb{E}_t[g_t]\|_2 \cdot \|\hat{x}_t - \pi\|_2 \le \sqrt{n}\lambda\delta \cdot 2R \le o(1/T).$$

Finally, let us go back to have $i_{[t-1]} = (i_1, \ldots, i_{t-1})$ randomly chosen, and let $\mathbb{E}_{[t-1]}[\cdot]$ denote the expectation over the randomly chosen $i_{[t-1]}$. Then, we have the lemma as

$$\mathbb{E}\left[\langle \ell_t, \hat{x}_t - \pi\rangle\right] - \mathbb{E}\left[\langle g_t, \hat{x}_t - \pi\rangle\right] = \mathbb{E}_{[t-1]}\left[\langle \ell_t - \mathbb{E}_t[g_t], \hat{x}_t - \pi\rangle\right] \le o(1/T).$$

Appendix E. Proof of Lemma 7

Recall that

$$\|g_t - \hat{g}_{t-1}\|_2^2 = \|n\left(v_{t,i_t} - \hat{g}_{t-1,i_t}\right)\mathbf{e}_{i_t}\|_2^2 = n^2\left(v_{t,i_t} - \hat{g}_{t-1,i_t}\right)^2,$$

and from the definition of α_t, we know that $\hat{g}_{t-1,i_t} = \hat{g}_{\alpha_t,i_t} = v_{\alpha_t,i_t}$. Thus, we have $\left(v_{t,i_t} - \hat{g}_{t-1,i_t}\right)^2 = \left(v_{t,i_t} - v_{\alpha_t,i_t}\right)^2$, and we show next that it is close to $\left(\ell_{t,i_t} - \ell_{\alpha_t,i_t}\right)^2$.

Let $\varepsilon = |v_{t,i_t} - \ell_{t,i_t}| + |\ell_{\alpha_t,i_t} - v_{\alpha_t,i_t}|$, and note that $\varepsilon \le 2\lambda\delta$ by inequality (10) in Appendix D. Then we can express $\left(v_{t,i_t} - v_{\alpha_t,i_t}\right)^2$ as

$$\left(\left(\ell_{t,i_t} - \ell_{\alpha_t,i_t}\right) + \left(v_{t,i_t} - \ell_{t,i_t}\right) + \left(\ell_{\alpha_t,i_t} - v_{\alpha_t,i_t}\right)\right)^2 \le \left(|\ell_{t,i_t} - \ell_{\alpha_t,i_t}| + \varepsilon\right)^2,$$

which is

$$\left(\ell_{t,i_t} - \ell_{\alpha_t,i_t}\right)^2 + 2\varepsilon|\ell_{t,i_t} - \ell_{\alpha_t,i_t}| + \varepsilon^2 \le \left(\ell_{t,i_t} - \ell_{\alpha_t,i_t}\right)^2 + 8\lambda\delta G + \left(2\lambda\delta\right)^2$$

where the last two terms are both $o(1/(Tn^2))$. Then the lemma follows as

$$\|g_t - \hat{g}_{t-1}\|_2^2 = n^2\left(v_{t,i_t} - v_{\alpha_t,i_t}\right)^2 \le n^2\left(\ell_{t,i_t} - \ell_{\alpha_t,i_t}\right)^2 + o(1/T).$$

Appendix F. Proof of Lemma 11

Recall that $B_t = \frac{1}{2\eta}\|x_{t+1} - \hat{x}_t\|_2^2 + \frac{1}{2\eta}\|\hat{x}_t - x_t\|_2^2$, so

$$\begin{aligned}\sum_{t=1}^{T} B_t &= \frac{1}{2\eta}\|\hat{x}_1 - x_1\|_2^2 + \frac{1}{2\eta}\sum_{t=2}^{T}\left(\|x_t - \hat{x}_{t-1}\|_2^2 + \|\hat{x}_t - x_t\|_2^2\right) + \frac{1}{2\eta}\|x_{T+1} - \hat{x}_T\|_2^2 \\ &\geq \frac{1}{2\eta}\sum_{t=2}^{T}\left(\|x_t - \hat{x}_{t-1}\|_2^2 + \|\hat{x}_t - x_t\|_2^2\right) \\ &\geq \frac{1}{4\eta}\sum_{t=2}^{T}\|\hat{x}_t - \hat{x}_{t-1}\|_2^2\end{aligned}$$

by Proposition 1(b). Then the lemma follows as $\|\hat{x}_1 - \hat{x}_0\|_2^2 \leq R^2 \leq O(1)$.

Appendix G. Proof of Lemma 14

The lemma follows immediately from the following two lemmas.

Lemma 19 $\sum_{t=1}^{T}(A_t - C_t) \leq \sum_{t=1}^{T} S_t + O(1)$.

Proof Note that $\sum_{t=1}^{T} A_t$ can be rearranged as

$$\frac{1}{2\eta_1}\|\pi - x_1\|_2^2 + \sum_{t=1}^{T}\left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)\|\pi - x_{t+1}\|_2^2 - \frac{1}{2\eta_{T+1}}\|\pi - x_{T+1}\|_2^2. \tag{11}$$

The first term above is at most $\left(1 + \frac{H}{2}\right)R^2 = O(1)$, and let us drop the last term. For the second term, note that $\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} = \frac{H}{2\gamma}\|g_t - \hat{g}_{t-1}\|_2^2 \leq \frac{H}{2}$ since $\gamma \geq \|g_t - \hat{g}_{t-1}\|_2^2$, and moreover, $\frac{1}{2}\|\pi - x_{t+1}\|_2^2 = \frac{1}{2}\|\pi - \hat{x}_t + \hat{x}_t - x_{t+1}\|_2^2 \leq \|\pi - \hat{x}_t\|_2^2 + \|\hat{x}_t - x_{t+1}\|_2^2$ by Proposition 1(b). Thus, with $C_t = \frac{H}{2}\|\pi - \hat{x}_t\|_2^2$, we obtain

$$\sum_{t=1}^{T}(A_t - C_t) \leq \sum_{t=1}^{T}\frac{H}{2}\|\hat{x}_t - x_{t+1}\|_2^2 + O(1).$$

Since $\frac{H}{2} \leq \frac{1}{\eta_t}$ and $\|\hat{x}_t - x_{t+1}\|_2^2 \leq \eta_t^2\|\hat{g}_{t-1} - g_t\|_2^2$ by the update rule of $\hat{x}_t$ and x_{t+1} and by Proposition 17, we have

$$\sum_{t=1}^{T}(A_t - C_t) \leq \sum_{t=1}^{T}\eta_t\|g_t - \hat{g}_{t-1}\|_2^2 + O(1) = \sum_{t=1}^{T} S_t + O(1),$$

which proves the lemma. ■

Lemma 20 $\sum_{t=1}^{T} S_t \leq \frac{2\gamma}{H}\ln\left(1 + \frac{H}{2\gamma}\sum_{t=1}^{T}\|g_t - \hat{g}_{t-1}\|_2^2\right)$.

Proof Recall that $S_t = \eta_t \|g_t - \hat{g}_{t-1}\|_2^2$ where $\eta_t = 1 \Big/ \left(1 + \frac{H}{2} + \frac{H}{2\gamma} \sum_{\tau=1}^{t-1} \|g_\tau - \hat{g}_{\tau-1}\|_2^2\right)$. Let $V_0 = 1$ and $V_t = 1 + \frac{H}{2\gamma} \sum_{\tau=1}^{t} \|g_\tau - \hat{g}_{\tau-1}\|_2^2$ for $t \geq 1$. Note that $\eta_t \leq \frac{1}{V_t}$ since $\gamma \geq \|g_t - \hat{g}_{t-1}\|_2^2$. This implies that

$$S_t = \eta_t \|g_t - \hat{g}_{t-1}\|_2^2 = \eta_t \frac{2\gamma}{H} (V_t - V_{t-1}) \leq \frac{2\gamma}{H} \left(1 - \frac{V_{t-1}}{V_t}\right) \leq \frac{2\gamma}{H} \ln \frac{V_t}{V_{t-1}},$$

where the last inequality holds since for any two real numbers $a > b > 0$, $1 - \frac{b}{a} \leq \ln \frac{a}{b}$. Therefore, by summing over t, we have

$$\sum_{t=1}^{T} S_t = \sum_{t=1}^{T} \eta_t \|g_t - \hat{g}_{t-1}\|_2^2 \leq \frac{2\gamma}{H} \sum_{t=1}^{T} \ln \frac{V_t}{V_{t-1}} = \frac{2\gamma}{H} \ln \frac{V_T}{V_0} = \frac{2\gamma}{H} \ln V_T.$$

■

Appendix H. Proof of Lemma 15

Recall that $B_t = \frac{1}{2\eta_t} \|x_{t+1} - \hat{x}_t\|_2^2 + \frac{1}{2\eta_t} \|\hat{x}_t - x_t\|_2^2$, and note that $\eta_t \leq 1$ for every $t \in [t]$. Thus, if we let $\eta = 1$, then $B_t \geq \frac{1}{2\eta} \|x_{t+1} - \hat{x}_t\|_2^2 + \frac{1}{2\eta} \|\hat{x}_t - x_t\|_2^2$ for every $t \in [T]$, and the lemma then follows from Lemma 11 (which works for any $\eta > 0$).