Power of Hints for Online Learning with Movement Costs

Aditya Bhaskara
University of Utah
a.bhaskara@utah.edu

Ashok Cutkosky
Boston University
ashok@cutkosky.com

Ravi Kumar
Google
ravi.k53@gmail.com

Manish Purohit
Google
mpurohit@google.com

Abstract

We consider the online linear optimization problem with *movement costs*, a variant of online learning in which the learner must not only respond to cost vectors c_t with points x_t in order to maintain low regret, but is also penalized for movement by an additional cost $\|x_t - x_{t+1}\|^{1+\epsilon}$ for some $\epsilon \geq 0$. Classically, simple algorithms that obtain the optimal $\sqrt{T}$ regret already are very stable and do not incur a significant movement cost. However, recent work has shown that when the learning algorithm is provided with weak "hint" vectors that have a positive correlation with the costs, the regret can be significantly improved to $\log T$. In this work, we study the stability of such algorithms, and provide matching upper and lower bounds showing that incorporating movement costs results in intricate tradeoffs between $\log T$ when $\epsilon \geq 1$ and $\sqrt{T}$ regret when $\epsilon = 0$.

1 Introduction

Online linear optimization (OLO) is a classical learning formulation with a rich history and a variety of applications [28]. In this setting, at each time step t, an online algorithm has to respond with a vector x_t before a linear cost vector c_t is revealed to it; the algorithm incurs an additive cost $\langle c_t, x_t \rangle$. The performance of the algorithm after T steps is measured in terms of regret, i.e., the cost of the best single vector it could have committed to in hindsight. The OLO formulation captures many natural problems such as recommendation systems, portfolio selection, and matrix completion among others. See the excellent textbooks [4, 11, 25] for more details.

Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

The regret bounds for standard OLO versions are well understood and a tight bound of $\Theta(\sqrt{T})$ is known [4, 16, 28]. There have been several works studying conditions under which it is possible to improve this bound, such placing restrictions on the family of cost functions [12], on the domain [15], or by assuming that the losses do not change significantly with time [23, 7]. The work of Rakhlin and Sridharan [22] began the study of OLO under additional information in the form of hints about the cost function before it is revealed to the algorithm. More recently [8, 13], it was shown that if a hint vector that is only mildly correlated with the cost vector is available at each time step, then a regret of $O(\log T)$ can be achieved. This result was generalized to $O(\sqrt{B} \log T)$ regret, where B is the number of time steps where the hints are bad, i.e., poorly correlated with the cost vector [2]. Thus, hints can be powerful and can yield poly-logarithmic regret in standard OLO.

Movement costs. In this work we study the role of hints for an important variant of OLO. In this variant, at each time step t, in addition to the standard cost $\langle c_t, x_t \rangle$, the algorithm incurs a *movement cost* that that depends on the distance between its previous response x_{t-1} and current response x_t. This formulation (sometimes called *smoothed OLO*) thus penalizes the algorithm for large changes to its responses. Previous work[1] on smoothed OLO has focused on movement costs of the form $\|x_t - x_{t-1}\|^2$ [6, 9, 18, 19] as well as of the form $\|x_t - x_{t-1}\|$ [1, 5]. OLO with movement costs has many applications in online logistic regression [10], multi-task machine learning [27], data center power management [20], EV charging [17], smart grids [26], and autonomous driving [24].

Movement costs add a new layer of complexity to online optimization. While it is easy to see that the standard algorithms for OLO such as FTRL incur only a total movement of $O(\sqrt{T})$, this additional term is too large if we are aiming for a better regret guarantee (as in the settings above where we have hints). Thus it is

[1] Some of the work focuses on the competitive ratio instead of regret.

natural to ask, how much can hints help in the OLO problem with movement costs? Note that even if the hints are perfect (i.e., the algorithm knows c_t *before* choosing x_t), the optimal algorithm with movement costs is still not obvious. Our goal is to understand these questions better: we consider movement costs of the general form $\|x_t - x_{t-1}\|^{1+\epsilon}$ for $\epsilon \geq 0$ and study the interplay between the quality/nature of the hints, ϵ, and regret. In fact, we consider two hint settings: one in which the hint only provides directional information about the cost vector and another in which the hint has information about both the direction *and* the length of the cost vector. We will show that interestingly, these two settings lead to different optimal regret bounds.

Our results. We start with an informal summary of our contributions.

For the case when the hint only gives information about the direction of the cost vector (i.e., length information is *unavailable* in the hint), we show a regret bound of $\tilde{O}(\sqrt{B} + T^{\frac{1}{2+\epsilon}})$ whenever $\epsilon > 0$ (Theorem 3.8). Here, B is the number of time steps where the hint direction is poorly correlated with that of the cost vector. We also show a matching $\Omega(T^{\frac{1}{2+\epsilon}})$ lower bound on the regret (Theorem 4.3).

For the case when information about the length of the cost vector is *available* as part of the hint, we show a regret bound of $\tilde{O}(\sqrt{B} + T^{\frac{1-\epsilon}{2}})$ for $\epsilon \in [0, 1)$ and a regret bound of $\tilde{O}(\sqrt{B}\log T)$ when $\epsilon \geq 1$ (Theorem 3.9). Here, B is the number of time steps where either the hint is poorly correlated with the cost vector or the length information is bad. We complement this result by showing a matching lower bound of $\Omega(T^{\frac{1-\epsilon}{2}})$ on the regret for $\epsilon \in [0, 1)$ (Theorem 4.2).

Our results thus show two intriguing aspects about OLO with movement costs: (i) poly-logarithmic regret is not possible without length hints, and (ii) even when length hints are available, poly-logarithmic regret is achievable if and only if $\epsilon \geq 1$. This phenomenon is absent in the standard setting of OLO with hints (where a directional hint suffices [2]).

2 Preliminaries

We consider the online linear optimization problem with movement costs. In this setting, at each time $t \in [T]$, an algorithm selects a point $x_t \in \mathbb{R}^d,\ \|x_t\| \leq 1$ and then an adversary reveals a cost vector $c_t \in \mathbb{R}^d$. (Throughout, $\|\cdot\|$ refers to the ℓ_2 norm unless otherwise indicated.) We assume that $\|c_t\| \leq 1,\ \forall t \in [T]$. The algorithm incurs a *cost* of $\langle c_t, x_t\rangle \in [-1, 1]$ as well as a *movement cost* that measures the distance between x_t and x_{t-1}. We consider movement costs of the form $\gamma \cdot \|x_t - x_{t-1}\|^{1+\epsilon}$ for some fixed constants $\gamma, \epsilon \geq 0$. Then, the *total cost* incurred by the algorithm is given by $\sum_t \left(\langle c_t, x_t\rangle + \gamma\|x_t - x_{t-1}\|^{1+\epsilon}\right)$. The regret of the algorithm for some point $u \in \mathbb{R}^d$ is defined as the difference between the total cost incurred by the algorithm and the cost incurred by the fixed point u at all time steps:

$$\mathcal{R}(u, \vec{c}) = \sum_{t=1}^{T} \langle c_t, x_t - u\rangle + \sum_{t=1}^{T-1} \gamma\|x_{t+1} - x_t\|^{(1+\epsilon)}.$$

The *regret* of the algorithm is defined as the worst-case regret over all points in the unit ball.

$$\mathcal{R}(\vec{c}) = \sup_{\substack{u \in \mathbb{R}^d \\ \|u\| \leq 1}} \mathcal{R}(u, \vec{c}).$$

We consider the setting where an algorithm has access to a *hint* regarding the cost vector c_t *before* it needs to respond with x_t. We consider two settings. In the first setting, the hint at time t is a unit vector $d_t \in \mathbb{R}^d,\ \|d_t\| = 1$ that provides information regarding the *direction* of the cost vector c_t. In the second setting, along with the directional hint d_t, the algorithm also has access to $\lambda_t \in \mathbb{R}^+$ that provides information regarding the *length* of the cost vector c_t. In both the settings we allow the hints to be arbitrarily related to the cost vector and can even be generated adversarially. We aim to design algorithms that guarantee good regret whenever the hints provide meaningful information about the cost vectors while maintaining worst-case regret guarantees.

3 Algorithms

Both the algorithms we present have a common structure: an *outer learner* that uses the hint provided at each step along with the output (denoted $\overline{x}_t$) of an *inner learner* to produce the prediction x_t for that step. After the cost vector is revealed, an appropriately defined "surrogate" loss function $\ell_t(\cdot)$ is provided to the inner learner, which it uses to come up with the next prediction $\overline{x}_{t+1}$. The inner learner turns out to be a simple FTRL (Follow The Regularized Leader) procedure, and our assumptions on the quality of the hints will ensure that the loss functions provided to the inner learner are strongly convex in most iterations. The inner learner is presented in Algorithm 2. While this structure is similar to some of the prior work on OLO with hints [2, 8], the main novelty is in choosing how far to move along a hint, captured by a hyperparameter r. This is because we now incur a cost for movement, thus moving may not always result in an improved objective value, e.g., when the cost vector has a small length.

In our description of the outer learner, unlike in [2, 14], we do not maintain and update the parameter r adaptively. Instead, we will use a meta-algorithm that runs multiple copies of Algorithm 1, one for each choice of r. Using a combination procedure developed in [3], it follows that the meta-algorithm has regret that can be upper bounded by the regret of the copy with the best possible value of r (with a logarithmic overhead; see Section 3.4 for a formal statement). Thus in what follows, we assume a fixed value for the parameter r, and show that there exists a setting of this parameter that yields a low regret.

Recall that each hint at time t consists of a directional vector $d_t, \|d_t\| = 1$ and, if available, a length hint $\lambda_t \in [0, 1]$. In the following we analyze Algorithm 1 in both the settings. Before we proceed with the analysis, we first establish some basic properties of the loss function and the inner learner.

Algorithm 1 Outer learner.

Input: Distance parameter $r \geq 1$
Initialize inner learner with parameter r
for $t = 1 \dots T$ **do**
  Receive hint: direction d_t; if available, length λ_t
  if λ_t is available **then**
    $h_t \leftarrow \lambda_t d_t$
  else
    $h_t \leftarrow d_t$
  end if
  Receive $\overline{x}_t$ from inner learner
  $x_t \leftarrow \overline{x}_t + \frac{(\|\overline{x}_t\|^2 - 1)}{2r} h_t$
  Play x_t and receive cost c_t
  if λ_t is available **then**
    $\mu_t \leftarrow \max\left(0, \frac{\langle c_t, h_t \rangle}{r} - 2\gamma 3^{1+\epsilon} \left(\frac{\lambda_t}{r}\right)^{1+\epsilon}\right)$
  else
    $\mu_t \leftarrow \max(0, \frac{\langle c_t, h_t \rangle}{r})$
  end if
  Define $\ell_t(x) = \langle c_t, x \rangle + \frac{\mu_t}{2}(\|x\|^2 - 1)$
  Send function $\ell_t(x)$ to inner learner
end for

Algorithm 2 Inner learner.

Input: Initial regularization parameter r
$\overline{x}_1 \leftarrow 0$
$\psi(x) \leftarrow \frac{r}{2}\|x\|^2$
for $t = 1 \dots T$ **do**
  Send $\overline{x}_t$
  Receive loss function $\ell_t(x)$
  $\overline{x}_{t+1} \leftarrow \operatorname{argmin}_{\|x\| \leq 1} \left\{\psi(x) + \sum_{\tau=1}^{t} \ell_t(x)\right\}$
end for

3.1 Basic properties

We first show properties that link the quality of the hints to the true regret and the surrogate loss ℓ_t.

Lemma 3.1. *The surrogate loss $\ell_t(\cdot)$ and the values $\overline{x}_t$ and x_t defined in Algorithm 1 satisfy the following:*

1. *if $\mu_t = 0$, then for any $\epsilon \geq 0$: $\langle c_t, x_t \rangle + 2 \cdot 3^{1+\epsilon}\gamma\|x_t - \overline{x}_t\|^{1+\epsilon} \leq \ell_t(\overline{x}_t) + \frac{|\langle c_t, h_t \rangle|}{2r} + 3^{1+\epsilon}\gamma\frac{\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}}$.*
2. *For all u in the unit ball, $\ell_t(u) \leq \langle c_t, u \rangle$.*
3. *$\|x_t\| \leq 1$, for all t.*

Proof. 1. By definition of x_t, we have

$$\begin{aligned}\langle c_t, x_t \rangle &= \langle c_t, \overline{x}_t \rangle + \langle c_t, h_t \rangle \cdot \frac{(\|\overline{x}_t\|^2 - 1)}{2r} \\ &= \ell_t(\overline{x}_t) + \langle c_t, h_t \rangle \cdot \frac{(\|\overline{x}_t\|^2 - 1)}{2r} \\ &\leq \ell_t(\overline{x}_t) + \frac{|\langle c_t, h_t \rangle|}{2r},\end{aligned}$$

where the first equality follows since $\ell_t(\overline{x}_t) = \langle c_t, \overline{x}_t \rangle$ if $\mu_t = 0$. Also,

$$\|x_t - \overline{x}_t\| = \frac{\|h_t\|(1 - \|\overline{x}_t\|^2)}{2r} \leq \frac{\|h_t\|}{2r}.$$

2. This is immediate from the definition.
3. Since $r \geq 1$, we have

$$\begin{aligned}\|x_t\| &\leq \|\overline{x}_t\| + \frac{1}{2}(1 - \|\overline{x}_t\|^2) \\ &\leq \sup_{z \in [0,1]} \left\{z + \frac{1 - z^2}{2}\right\} \leq 1.\end{aligned}$$

□

Next we show a simple fact that will be useful later. We defer the proof to the Appendix.

Lemma 3.2. *Let $b, c > 0$, and $a_1, \dots, a_T$ be arbitrary non-negative numbers. Then,*

$$\sum_{t=1}^{T} \frac{a_t}{b + c\sum_{\tau=1}^{t} a_\tau} \leq \frac{1}{c} \log\left(1 + \frac{c\sum_{t=1}^{T} a_t}{b}\right).$$

Finally, we argue that the inner learner, Algorithm 2, inherently incurs both a small value of $\sum_{t=1}^{T} \ell_t(\overline{x}_t) - \ell_t(u)$ as well as a small movement cost, so long as μ_t is not small.

Lemma 3.3. *Let $\overline{x}_t$ and ℓ_t be as defined in Algorithm 1, and let $\epsilon \geq 0$ and $r \geq 1$ be parameters. For any $\alpha \in (0, 1]$, define*

$$S := S_\alpha = \left\{t \mid \mu_t \leq \frac{\alpha\|c_t\|^2}{2r}\right\}.$$

Then for any $\|u\| \le 1$, Algorithm 2 satisfies:

$$\sum_{t=1}^{T} \ell_t(\overline{x}_t) - \ell_t(u) \le \frac{r}{2} + \frac{2\sum_{t\in S}\|c_t\|^2}{r} + \frac{4r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right).$$

Further, the movement cost can be bounded as

$$\sum_{t=1}^{T-1}\|\overline{x}_{t+1}-\overline{x}_t\|^{1+\epsilon} \le 2^{1+\epsilon}\frac{\sum_{t\in S}\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}} + 2^{1+\epsilon}\left(\frac{2}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}} T^{\frac{1-[\epsilon]_1}{2}},$$

where $[\epsilon]_1 = \min(1,\epsilon)$.

Proof. First, observe that Algorithm 2 is performing the classic FTRL step with the fixed regularizer $\frac{r}{2}\|x\|^2$. Further, each loss ℓ_t is clearly μ_t-strongly convex. Therefore, by [21, Theorem 1], (using time-varying norms $\|x\|_t^2 = \frac{\mu_t}{2}\|x\|^2$) we have:

$$\sum_{t=1}^{T}\ell_t(\overline{x}_t) - \ell_t(u) \le \frac{r}{2}\|u\|^2 + \sum_{t=1}^{T}\frac{\|\nabla\ell_t(\overline{x}_t)\|^2}{2r + 2\sum_{\tau=1}^{t}\mu_\tau}.$$

Now, observe that $\nabla\ell_t(\overline{x}_t) = c_t + \mu_t\overline{x}_t$. Further, since $r \ge 1$ and $\|h_t\| \le 1$, by definition of μ_t we must have $\mu_t \le \|c_t\|$. Thus since $\|\overline{x}_t\| \le 1$, we have $\|\nabla\ell_t(\overline{x}_t)\| \le 2\|c_t\|$. Plugging in this bound yields

$$\begin{aligned}\sum_{t=1}^{T}\ell_t(\overline{x}_t) - \ell_t(u) &\le \frac{r}{2}\|u\|^2 + \sum_{t=1}^{T}\frac{2\|c_t\|^2}{r+\sum_{\tau=1}^{t}\mu_\tau}\\ &\le \frac{r}{2} + \sum_{t\in S}\frac{2\|c_t\|^2}{r} + \sum_{t\notin S}\frac{2\|c_t\|^2}{r+\sum_{\tau\notin S,\tau\le t}\mu_\tau}\\ &\le \frac{r}{2} + \frac{2\sum_{t\in S}\|c_t\|^2}{r} + \sum_{t\notin S}\frac{2\|c_t\|^2}{r+\sum_{\tau\notin S,\tau\le t}\frac{\alpha\|c_\tau\|^2}{2r}}\end{aligned}$$

using Lemma 3.2,

$$\le \frac{r}{2} + \frac{2\sum_{t\in S}\|c_t\|^2}{r} + \frac{4r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right).$$

To bound the movement, we appeal to [21, Lemma 7 and Theorem 1], which imply that if $\overline{x}_t$ is defined as $\overline{x}_t = \operatorname{argmin}\frac{r}{2}\|x\|^2 + \sum_{\tau=1}^{t-1}\ell_\tau(x)$ for all t, then

$$\|\overline{x}_{t+1}-\overline{x}_t\| \le \frac{\|\nabla\ell_t(\overline{x}_t)\|}{r+\sum_{\tau=1}^{t}\mu_\tau} \le \frac{2\|c_t\|}{r+\sum_{\tau=1}^{t}\mu_\tau}.$$

Thus,

$$\begin{aligned}\sum_{t=1}^{T-1}\|\overline{x}_{t+1}-\overline{x}_t\|^{1+\epsilon} &\le 2^{1+\epsilon}\sum_{t=1}^{T-1}\frac{\|c_t\|^{1+\epsilon}}{\left(r+\sum_{\tau=1}^{t}\mu_\tau\right)^{1+\epsilon}}\\ &\le 2^{1+\epsilon}\sum_{t\in S}\frac{\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}} + 2^{1+\epsilon}\sum_{t\notin S}\frac{\|c_t\|^{1+\epsilon}}{\left(r+\sum_{\tau\notin S,\tau\le t}\mu_\tau\right)^{1+\epsilon}}\\ &\le 2^{1+\epsilon}\frac{\sum_{t\in S}\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}} + 2^{1+\epsilon}\sum_{t\notin S}\frac{\|c_t\|^{1+\epsilon}}{\left(r+\frac{\alpha}{2r}\sum_{\tau\notin S,\tau\le t}\|c_t\|^2\right)^{1+\epsilon}}.\end{aligned}$$

Let us bound this second sum. First, observe that each term in the sum is at most 1. Therefore if we replace the exponents $1+\epsilon$ with $1+[\epsilon]_1$ we can only increase the sum (recall $[\epsilon]_1 = \min(1,\epsilon)$). Next, apply Hölder's inequality with $p = \frac{2}{1+[\epsilon]_1}$ and $q = \frac{2}{1-[\epsilon]_1}$. This yields

$$\begin{aligned}&\sum_{t\notin S}\frac{\|c_t\|^{1+[\epsilon]_1}}{\left(r+\frac{\alpha}{2r}\sum_{\tau\notin S,\tau\le t}\|c_\tau\|^2\right)^{1+[\epsilon]_1}}\\ &\le \left(\sum_{t\notin S}\frac{\|c_t\|^2}{r+\frac{\alpha}{2r}\sum_{\tau\notin S,\tau\le t}\|c_\tau\|^2}\right)^{\frac{1+[\epsilon]_1}{2}} \times \left(\sum_{t\notin S}\frac{1}{\left(r+\frac{\alpha}{2r}\sum_{\tau\notin S,\tau\le t}\|c_\tau\|^{1+1}\right)^{\frac{1+[\epsilon]_1}{1-[\epsilon]_1}}}\right)^{\frac{1-[\epsilon]_1}{2}}\\ &\le \left(\sum_{t\notin S}\frac{\|c_t\|^2}{r+\frac{\alpha}{2r}\sum_{\tau\notin S,\tau\le t}\|c_\tau\|^2}\right)^{\frac{1+[\epsilon]_1}{2}} \frac{T^{\frac{1-[\epsilon]_1}{2}}}{r^{\frac{1+[\epsilon]_1}{2}}}\end{aligned}$$

using Lemma 3.2,

$$\le \left(\frac{2r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}} \frac{T^{\frac{1-[\epsilon]_1}{2}}}{r^{\frac{1+[\epsilon]_1}{2}}}.$$

□

3.2 Length hints are unavailable

Using these basic properties, we are in a position to begin analyzing the regret of Algorithm 1. We first consider the case when the length hints are *not* available to the outer learner. To start, we describe some simple consequences from the definition of μ_t.

Lemma 3.4. *When length hints are* unavailable *in Algorithm 1, we have*

1. *If $\mu_t > 0$, then $\langle c_t, x_t\rangle + 2\gamma 3^{1+\epsilon}\|x_t - \overline{x}_t\|^{1+\epsilon} \le \ell_t(\overline{x}_t) + 3^{1+\epsilon}\frac{\gamma}{r^{1+\epsilon}}$.*
2. *If $\langle c_t, d_t\rangle \ge \alpha\|c_t\|^2$, then we have $\mu_t \ge \frac{\alpha}{2r}\|c_t\|^2$.*

Proof. 1. Using the definition of ℓ_t and x_t and using the fact $\mu_t > 0$, we have:

$$\langle c_t, x_t\rangle = \ell_t(\overline{x}_t)$$

$$\langle c_t, x_t\rangle + 2\gamma(3\|x_t - \overline{x}_t\|)^{1+\epsilon} \le \ell_t(\overline{x}_t) + 2\gamma\frac{(3\|h_t\|)^{1+\epsilon}}{(2r)^{1+\epsilon}}$$

$$\leq \ell_t(\overline{x}_t) + 3^{1+\epsilon}\frac{\gamma}{r^{1+\epsilon}}.$$

2. By definition of μ_t and using $\langle c_t, d_t\rangle \geq \alpha\|c_t\|^2$, we have $\mu_t \geq \frac{\alpha\|c_t\|^2}{r} \geq \frac{\alpha\|c_t\|^2}{2r}$. □

We now proceed to prove our first bound on the final regret of Algorithm 1. Specifically, we show that when length hints are not available, the regret scales as roughly $\tilde{O}\left(\sqrt{B} + T^{\frac{1}{2+\epsilon}}\right)$. In particular, when $B = 0$, the regret grows asymptotically slower than the $\sqrt{T}$ bound we would expect without *any* hints.

Theorem 3.5. *Suppose length hints are* unavailable *in Algorithm 1, $\epsilon \geq 0$, and that all but B indices t satisfy $\langle c_t, d_t\rangle \geq \alpha\|c_t\|^2$. Then with $r = 1 + \max\left(\left(\frac{\gamma 6^{1+\epsilon}\alpha T}{\log(1+T)}\right)^{\frac{1}{2+\epsilon}}, \sqrt{\alpha B/\log(1+T)}\right)$ for all $\|u\| \leq 1$,*

$$\mathcal{R}(u,\vec{c}) \leq O\left((6^{1+\epsilon}\gamma T)^{\frac{1}{2+\epsilon}}\left(\frac{\log T}{\alpha}\right)^{\frac{1+\epsilon}{2+\epsilon}} + (1+6^{1+\epsilon}\gamma)\sqrt{\frac{B\log T}{\alpha}} + 6^{1+\epsilon}\gamma\left(\frac{\log T}{\alpha}\right)^{\frac{1+[\epsilon]_1}{2}} T^{\frac{1-[\epsilon]_1}{2}}\right).$$

Proof. Let S be the set of indices for which $\mu_t < \frac{\alpha\|c_t\|^2}{2r}$ as in Lemma 3.3. Notice that by Lemma 3.4(2) and the definition of B, we have $|S| \leq B$.

$$\begin{aligned}
\mathcal{R}(u,\vec{c}) &= \sum_{t=1}^{T}\langle c_t, x_t - u\rangle + \sum_{t=1}^{T-1}\gamma\|x_{t+1}-x_t\|^{1+\epsilon} \\
&\leq \sum_{t=1}^{T}\langle c_t, x_t - u\rangle + \gamma 3^{1+\epsilon}\left(\sum_{t=1}^{T-1}\|\overline{x}_{t+1}-\overline{x}_t\|^{1+\epsilon} + \sum_{t=1}^{T-1}\|x_t-\overline{x}_t\|^{1+\epsilon} + \sum_{t=1}^{T-1}\|x_{t+1}-\overline{x}_{t+1}\|^{1+\epsilon}\right) \\
&\leq \sum_{t=1}^{T}\langle c_t, x_t - u\rangle + \gamma 3^{1+\epsilon}\sum_{t=1}^{T-1}\|\overline{x}_{t+1}-\overline{x}_t\|^{1+\epsilon} + 2\gamma 3^{1+\epsilon}\sum_{t=1}^{T-1}\|x_t-\overline{x}_t\|^{1+\epsilon}
\end{aligned}$$

applying Lemma 3.1(1, 2) and Lemma 3.4(1),

$$\leq \sum_{t=1}^{T}(\ell_t(\overline{x}_t) - \ell_t(u) + \frac{\gamma 3^{1+\epsilon}}{r^{1+\epsilon}}) + \sum_{t\in S}\frac{|\langle c_t, h_t\rangle|}{2r} + \gamma 3^{1+\epsilon}\sum_{t=1}^{T-1}\|\overline{x}_{t+1}-\overline{x}_t\|^{1+\epsilon}$$

applying Lemma 3.3,

$$\leq \frac{r}{2} + \frac{2\sum_{t\in S}\|c_t\|^2}{r} + \frac{4r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right) + \frac{\gamma 3^{1+\epsilon}T}{r^{1+\epsilon}} + \sum_{t\in S}\left(\frac{|\langle c_t, h_t\rangle|}{2r} + 6^{1+\epsilon}\gamma\frac{\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}}\right) + 6^{1+\epsilon}\gamma\left(\frac{2}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}} T^{\frac{1-[\epsilon]_1}{2}}$$

since $|S| \leq B$,

$$\leq \frac{r}{2} + \frac{2B}{r} + \frac{4r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right) + \frac{\gamma 3^{1+\epsilon}T}{r^{1+\epsilon}} + \frac{B}{2r} + 6^{1+\epsilon}\gamma\frac{B}{r^{1+\epsilon}} + 6^{1+\epsilon}\gamma\left(\frac{2}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}} T^{\frac{1-[\epsilon]_1}{2}}.$$

Now in the above bound we set $r = 1 + \max\left(\left(\frac{\gamma 6^{1+\epsilon}\alpha T}{\log(1+T)}\right)^{\frac{1}{2+\epsilon}}, \sqrt{\alpha B/\log(1+T)}\right)$ to obtain the final regret bound. □

3.3 Length hints are available

Next, we turn to the setting in which length hints λ_t are available to the outer learner. Our first task is to prove an analogue of Lemma 3.4 that characterizes the relationship between the modified definition of μ_t and the quality of the hints.

Lemma 3.6. *When length hints are* available *in Algorithm 1, we have*

1. *If $\mu_t > 0$, then $\langle c_t, x_t\rangle + 2\gamma 3^{1+\epsilon}\|x_t - \overline{x}_t\|^{1+\epsilon} \leq \ell_t(\overline{x}_t)$.*
2. *If $\langle c_t, \lambda_t d_t\rangle \geq \alpha\|c_t\|^2$, $\lambda_t^2 \leq \beta\|c_t\|^2$, and $r \geq 1$, then so long as $r^{[\epsilon]_1} \geq \frac{4\gamma 3^{1+\epsilon}\beta^{\frac{1+[\epsilon]_1}{2}}}{\alpha\|c_t\|^{1-[\epsilon]_1}}$, we have $\mu_t \geq \frac{\alpha}{2r}\|c_t\|^2$.*

Proof. (1) Using the definition of x_t and ℓ_t, we have:

$$\begin{aligned}
\langle c_t, x_t\rangle &= \langle c_t, \overline{x}_t\rangle + \langle c_t, h_t\rangle\cdot\frac{(\|\overline{x}_t\|^2 - 1)}{2r} \\
&\leq \ell_t(\overline{x}_t) + (\frac{\langle c_t, h_t\rangle}{r} - \mu_t)\frac{(\|\overline{x}_t\|^2-1)}{2} \\
&= \ell_t(\overline{x}_t) + 2\gamma 3^{1+\epsilon}\frac{\|h_t\|^{1+\epsilon}}{2r^{1+\epsilon}}(\|\overline{x}_t\|^2 - 1) \\
&\leq \ell_t(\overline{x}_t) - 3^{1+\epsilon}\gamma\frac{(\|h_t\|(1-\|\overline{x}_t\|^2))^{1+\epsilon}}{r^{1+\epsilon}}.
\end{aligned}$$

Since $\|\overline{x}_t\| \leq 1$ and $x_t - \overline{x}_t = \frac{(\|\overline{x}_t\|^2-1)h_t}{2r}$, we have $\|x_t - \overline{x}_t\|^{1+\epsilon} \leq \frac{(\|h_t\|(1-\|\overline{x}_t\|^2))^{1+\epsilon}}{2r^{1+\epsilon}}$ and so the conclusion follows.

(2) Using the conditions on $h_t = \lambda_t d_t$ and r, we have:

$$\frac{\langle c_t, h_t\rangle}{r} - 2\gamma 3^{1+\epsilon}\frac{\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}} \geq \frac{\alpha\|c_t\|^2}{r} - 2\gamma 3^{1+\epsilon}\frac{\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}}$$

using $r \geq 1$ and $\|h_t\| \leq 1$,

$$\geq \frac{\alpha\|c_t\|^2}{r} - 2\gamma 3^{1+\epsilon}\frac{\|h_t\|^{1+[\epsilon]_1}}{r^{1+[\epsilon]_1}}$$
$$\geq \frac{\alpha\|c_t\|^2}{r} - 2\gamma\frac{3^{1+\epsilon}\beta^{\frac{1+[\epsilon]_1}{2}}\|c_t\|^{1+[\epsilon]_1}}{r^{1+[\epsilon]_1}} \geq \frac{\alpha}{2r}\|c_t\|^2. \qquad \square$$

Now, we have all the tools needed to bound our final regret bound when using length hints. This result will show that when both directional and length hints are available, we can obtain regret $\tilde{O}\left(\sqrt{B} + T^{\frac{1-[\epsilon]_1}{2}}\right)$, where $[\epsilon]_1 = \min(1, \epsilon)$. Notice that there is an "elbow" effect at $\epsilon = 1$ at which the regret becomes logarithmic and stops improving.

Theorem 3.7. *Suppose length hints are* available *in Algorithm 1 and all but B indices t satisfy both $\langle c_t, \lambda_t d_t\rangle \geq \alpha\|c_t\|^2$ and $\lambda_t^2 \leq \beta\|c_t\|^2$. Then there exists a setting of the parameter $r > 1 + \frac{4\gamma\beta}{\alpha}$ such that for all $\|u\| \leq 1$,*

$$\mathcal{R}(u, \vec{c}) \leq O\left(\frac{\gamma(\beta^{\frac{(1+[\epsilon]_1)^2}{4}} + \beta^{\frac{3+[\epsilon]_1}{4}})(\log(T))^{\frac{1+[\epsilon]_1}{2}}T^{\frac{1-[\epsilon]_1}{2}}}{\alpha^{\frac{3+[\epsilon]_1}{2}}} + (1+6^{1+\epsilon}\gamma)\frac{\sqrt{B\log(T)}}{\sqrt{\alpha}} + \left(\frac{3^{1+\epsilon}\gamma\beta}{\alpha} + \frac{1}{\alpha}\right)\log(T)\right).$$

Proof. Let S be the set of indices for which $\mu_t < \frac{\alpha\|c_t\|^2}{2r}$ as in Lemma 3.3. By Lemma 3.6, for any r we can decompose S into disjoint sets S_B and S_r, where every index in S_B satisfies either $\langle c_t, \lambda_t d_t\rangle < \alpha\|c_t\|^2$ or $\lambda_t^2 > \beta\|c_t\|^2$, and every index in S_r satisfies both $\langle c_t, \lambda_t d_t\rangle \geq \alpha\|c_t\|^2$ and $\lambda_t^2 \leq \beta\|c_t\|^2$, and also satisfies $r^{[\epsilon]_1} < \frac{4\cdot 3^{1+\epsilon}\gamma\beta^{\frac{1+[\epsilon]_1}{2}}}{\alpha\|c_t\|^{1-[\epsilon]_1}}$. By assumption, we have $|S_B| \leq B$. Note that since $r > \frac{4\gamma\beta}{\alpha}$, S_r is empty whenever $\epsilon \geq 1$.

Following exactly the same argument as the proof of Theorem 3.5, we have,

$$\mathcal{R}(u, \vec{c}) \leq \sum_{t=1}^{T}\langle c_t, x_t - u\rangle + 3^{1+\epsilon}\gamma\sum_{t=1}^{T-1}\|\overline{x}_{t+1} - \overline{x}_t\|^{1+\epsilon} + 2\cdot 3^{1+\epsilon}\gamma\sum_{t=1}^{T-1}\|x_t - \overline{x}_t\|^{1+\epsilon}$$

using Lemma 3.1 and Lemma 3.6,

$$\leq \sum_{t=1}^{T}(\ell_t(\overline{x}_t) - \ell_t(u)) + \sum_{t=1}^{T-1}3^{1+\epsilon}\gamma\|\overline{x}_{t+1} - \overline{x}_t\|^{1+\epsilon} + \sum_{t\in S}\left(\frac{|\langle c_t, h_t\rangle|}{2r} + \frac{3^{1+\epsilon}\gamma\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}}\right)$$

using Lemma 3.3,

$$\leq \frac{r}{2} + \frac{2\sum_{t\in S}\|c_t\|^2}{r} + \frac{4r}{\alpha}\log\left(1 + \frac{\alpha T}{2r^2}\right) + \sum_{t\in S}\left(\frac{|\langle c_t, h_t\rangle|}{2r} + \frac{3^{1+\epsilon}\gamma\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}} + 6^{1+\epsilon}\gamma\frac{\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}}\right) + 6^{1+\epsilon}\gamma\left(\frac{2}{\alpha}\log\left(1 + \frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}}T^{\frac{1-[\epsilon]_1}{2}}.$$

Now, we break the sums over S up into sums over S_B and S_r. First, let us consider the sum over indices in S_B. Using $|S_B| \leq B$, we have:

$$\sum_{t\in S_B}\frac{2\|c_t\|^2}{r} + \frac{|\langle c_t, h_t\rangle|}{2r} + \frac{\gamma(3\|h_t\|)^{1+\epsilon}}{r^{1+\epsilon}} + \frac{\gamma(6\|c_t\|)^{1+\epsilon}}{r^{1+\epsilon}} \leq \frac{5B}{2r} + \frac{2\cdot 6^{1+\epsilon}\gamma B}{r^{1+\epsilon}}$$

Now, let us consider the case that $\epsilon \geq 1$. In this scenario, we must have S_r is empty, so that $S = S_B$. Therefore the overall regret is bounded as:

$$\mathcal{R}(u, \vec{c}) \leq \frac{r}{2} + \frac{4r}{\alpha}\log\left(1 + \frac{\alpha T}{2r^2}\right) + \frac{5B}{2r} + \frac{2\cdot 6^{1+\epsilon}\gamma B}{r^{1+\epsilon}} + 6^{1+\epsilon}\gamma\left(\frac{2}{\alpha}\log\left(1 + \frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}}T^{\frac{1-[\epsilon]_1}{2}}.$$

And substituting the setting $r = 1 + \frac{4\cdot 3^{1+\epsilon}\gamma\beta}{\alpha} + \sqrt{\alpha B/\log(1+T)}$ completes the proof.

Thus, for the rest of the proof we consider $\epsilon < 1$, so that $[\epsilon]_1 = \epsilon$. In this case, we must take into account the sum over S_r. Notice that the indices in S_r satisfy $\|c_t\| \leq \frac{3^{\frac{1+\epsilon}{1-\epsilon}}(4\gamma)^{\frac{1}{1-\epsilon}}\beta^{\frac{1+\epsilon}{2-2\epsilon}}}{\alpha^{\frac{1}{1-\epsilon}}r^{\frac{\epsilon}{1-\epsilon}}}$. Also, we have $|\langle c_t, h_t\rangle| \leq \|c_t\|\|h_t\| \leq \sqrt{\beta}\|c_t\|^2$, and $\|h_t\|^{1+\epsilon} \leq \beta^{\frac{1+\epsilon}{2}}\|c_t\|^{1+\epsilon}$. Thus (coarsely bounding $|S_r| \leq T$):

$$\sum_{t\in S_r}\frac{2\|c_t\|^2}{r} \leq \frac{2\cdot 3^{\frac{2+2\epsilon}{1-\epsilon}}T(4\gamma)^{\frac{2}{1-\epsilon}}\beta^{\frac{1+\epsilon}{1-\epsilon}}}{\alpha^{\frac{2}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}.$$

$$\sum_{t\in S_r}\frac{|\langle c_t, h_t\rangle|}{r} \leq \frac{3^{\frac{2+2\epsilon}{1-\epsilon}}\sqrt{\beta}T(4\gamma)^{\frac{2}{1-\epsilon}}\beta^{\frac{1+\epsilon}{1-\epsilon}}}{\alpha^{\frac{2}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}.$$

$$\sum_{t\in S_r}\frac{\gamma\|h_t\|^{1+\epsilon}}{r^{1+\epsilon}} \le \sum_{t\in S_r}\frac{\gamma\beta^{\frac{1+\epsilon}{2}}\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}}$$
$$\le \frac{3^{\frac{(1+\epsilon)^2}{1-\epsilon}}\gamma\beta^{\frac{1+\epsilon}{1-\epsilon}}T(4\gamma)^{\frac{1+\epsilon}{1-\epsilon}}}{\alpha^{\frac{1+\epsilon}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}$$
$$\le \frac{3^{\frac{(1+\epsilon)^2}{1-\epsilon}}\beta^{\frac{1+\epsilon}{1-\epsilon}}T(4\gamma)^{\frac{2}{1-\epsilon}}}{\alpha^{\frac{1+\epsilon}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}.$$
$$\sum_{t\in S_r}\frac{\gamma\|c_t\|^{1+\epsilon}}{r^{1+\epsilon}} \le \frac{3^{\frac{(1+\epsilon)^2}{1-\epsilon}}\beta^{\frac{(1+\epsilon)^2}{2-2\epsilon}}T(4\gamma)^{\frac{2}{1-\epsilon}}}{\alpha^{\frac{1+\epsilon}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}.$$

Putting all this together, we have:

$$\mathcal{R}(u,\vec{c}) \le \frac{4\cdot 6^{\frac{3}{1-\epsilon}}(4\gamma)^{\frac{2}{1-\epsilon}}(\beta^{\frac{(1+\epsilon)^2}{2-2\epsilon}}+\beta^{\frac{3+\epsilon}{2-2\epsilon}})T}{\alpha^{\frac{2}{1-\epsilon}}r^{\frac{1+\epsilon}{1-\epsilon}}}$$
$$+\frac{r}{2}+\frac{2B}{r}+\frac{4r}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)$$
$$+\frac{B}{2r}+\frac{3^{1+\epsilon}\gamma B}{r^{1+\epsilon}}+6^{1+\epsilon}\gamma\frac{B}{r^{1+\epsilon}}$$
$$+6^{1+\epsilon}\gamma\left(\frac{2}{\alpha}\log\left(1+\frac{\alpha T}{2r^2}\right)\right)^{\frac{1+[\epsilon]_1}{2}}T^{\frac{1-[\epsilon]_1}{2}}.$$

Now, the proof is complete by setting $r = 1+\frac{4\cdot 3^{1+\epsilon}\gamma\beta}{\alpha}+\sqrt{\frac{\alpha B}{\log(1+T)}}+6^{\frac{3}{4}}\left(\frac{T\gamma^{\frac{2}{1-\epsilon}}(\beta^{\frac{(1+\epsilon)^2}{2}}+\beta^{\frac{3+\epsilon}{2-2\epsilon}})}{\alpha^{\frac{1+\epsilon}{1-\epsilon}}\log(1+T)}\right)^{\frac{1-\epsilon}{2}}$. □

3.4 A meta-algorithm for selecting r

So far we have provided an algorithm that can takes advantage of hints through a user-supplied parameter r. Intuitively, r measures how much we "trust" each hint. With the correct value for r, we showed that these algorithms can achieve regret bounds matching the lower bounds we will show in Section 4. However, this correct tuning is not available a priori. To address this, we invoke a combination procedure developed in a recent work of [3, Algorithm 4]. This *combiner* procedure is a meta-algorithm that takes as input K online learning algorithms and produces a single (randomized) online algorithm with an expected regret bound only a factor of $\log K$ worse than the best of the K individual regret bounds in *hindsight*.

From Theorems 3.5 and 3.7, the optimal choice of r is always $\le T$. The idea is to instantiate $O(\log T)$ copies of Algorithm 1 one for each value of $r \in \{1, 2, 2^2, \ldots, T\}$.[2] Now using the combiner algorithm, we can obtain the results of Theorems 3.5 and 3.7 with an additional $\log\log T$ multiplicative factor in the regret, *without* needing to tune r based on α, β, or B.

An issue with the argument above is that the combiner of [3] does not account for movement costs. However, we show that their algorithm applies with a mild modification, and an extra additive $O(\gamma \log T \log K)$ in the regret. We outline the main differences below.

Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_K$ are the algorithms being combined. The combiner of [3] runs all the algorithms in parallel (internally); additionally, it maintains a parameter ζ and an index i_t. At time t, the combiner plays the output of algorithm $\mathcal{A}_{i_t}$. Initially, i_0 is chosen uniformly at random from $[K]$ and $\zeta = 1$; here, ζ is a guess for the minimum regret (without movement cost) of the algorithms (in hindsight). Subsequently, the combiner runs $\mathcal{A}_i$'s in parallel, tracking their regret. It maintains an *active set* of algorithms whose regret has not exceeded ζ. As long as the "current choice" i_t is in the active set, the combiner's response is identical to that of $\mathcal{A}_{i_t}$. But if the regret of $\mathcal{A}_{i_t}$ exceeds ζ, the combiner sets i_t to a uniformly random algorithm in the active set. If the active set becomes empty, then ζ is doubled, and all the algorithms are restarted. The analysis proceeds by showing that if the minimum regret in hindsight (over the $\mathcal{A}_i$'s, without the movement cost term) is R, then ζ can only double $O(\log R) \le O(\log T)$ times. Further, for any $\zeta \le R$, only $O(\log K)$ "switches" are needed in expectation before the active set becomes empty and ζ doubles.

Now, to incorporate movement costs, we pursue an identical algorithm, except that we track the regret (with movement cost) for each algorithm, again switching algorithms when this quantity exceeds the guess ζ. The only change to the analysis is that now additional movement cost is incurred when switching algorithms, which is not captured earlier. Fortunately, since ζ doubles at most $O(\log T)$ times and on average only $O(\log K)$ switches occur each time ζ doubles, this additional movement cost is bounded by $O(\gamma \log K \log T)$. Applying this to our setting, we get:

Theorem 3.8. *There exists a (randomized) algorithm for OLO with hints that, having access to only direction hints, achieves the following expected regret bound for every $\alpha \in (0, 1)$:*

$$\mathbb{E}\left[\mathcal{R}(u,\vec{c})\right] \le O(\log\log T)\cdot\left((\gamma T)^{\frac{1}{2+\epsilon}}\left(\frac{\log T}{\alpha}\right)^{\frac{1+\epsilon}{2+\epsilon}}\right.$$
$$+(1+\gamma)\sqrt{\frac{B\log T}{\alpha}}$$
$$\left.+\gamma\left(\frac{\log T}{\alpha}\right)^{\frac{1+[\epsilon]_1}{2}}T^{\frac{1-[\epsilon]_1}{2}}+\gamma\log T\right),$$

[2] As stated, this assumes that the algorithm knows T beforehand. However, this can be overcome as follows: the combiner does not require all the algorithms to begin at $t = 0$. By having the algorithm with $r = 2^i$ start at $t = 2^i$, we can obtain the same regret bound without knowing T.

where $B = |\{t \mid \langle d_t, c_t\rangle < \alpha\|c_t\|^2\}|$.

Theorem 3.9. *There exists a randomized algorithm for OLO with hints that, with access to both direction and length hints, achieves the following expected regret bound for any* $\alpha, \beta \in (0,1)$*:*

$$\mathbb{E}\left[\mathcal{R}(u,\vec{c})\right] \le O(\log\log T)\cdot\left((1+\gamma)\frac{\sqrt{B\log T}}{\sqrt{\alpha}} + \left(\frac{\gamma\beta}{\alpha}+\frac{1}{\alpha}\right)\log T + \gamma\left(\frac{\log T}{\alpha}\right)^{\frac{1+[\epsilon]_1}{2}} T^{\frac{1-[\epsilon]_1}{2}} + \gamma\log T\right),$$

where $B = \{t \mid \langle d_t, c_t\rangle < \alpha\|c_t\|^2 \text{ or } \lambda_t^2 > \beta\|c_t\|^2\}|$.

Remark. Using the weaker combination algorithm from [3] along with the same reasoning as above, we note that we can obtain deterministic algorithm, but the regret bounds in Theorems 3.8 and 3.9 incur a multiplicative factor of $O(\log T)$ instead of $O(\log\log T)$.

4 Lower bounds

In this section we show that the algorithms we obtained are essentially optimal. Our lower bound constructions use the following technical lemma that is a simple consequence of the concavity of the logarithm.

Lemma 4.1. *Let* $\delta, \epsilon, \alpha > 0$*, and let* $T \ge 1$ *be some parameter. We have*

$$\delta^{1+\epsilon} - \delta T^{-\alpha} \ge -\frac{\epsilon\cdot T^{-\frac{\alpha(1+\epsilon)}{\epsilon}}}{(1+\epsilon)(1+\epsilon)^{1/\epsilon}}.$$

Proof. Using the concavity of the logarithm, we have that for any parameter $Z > 0$,

$$\log\left(\frac{1}{1+\epsilon}\cdot\delta^{1+\epsilon} + \frac{\epsilon}{1+\epsilon}\cdot Z^{\frac{1+\epsilon}{\epsilon}}\right) \ge \frac{1}{1+\epsilon}\log(\delta^{1+\epsilon}) + \frac{\epsilon}{1+\epsilon}\log(Z^{\frac{1+\epsilon}{\epsilon}}).$$

Or equivalently,

$$\frac{\delta^{1+\epsilon} + \epsilon Z^{\frac{1+\epsilon}{\epsilon}}}{1+\epsilon} \ge \delta Z.$$

Setting Z such that $(1+\epsilon)Z = T^{-\alpha}$, we get

$$\delta^{1+\epsilon} - \delta T^{-\alpha} \ge -\epsilon\left(\frac{T^{-\alpha}}{1+\epsilon}\right)^{(1+\epsilon)/\epsilon}. \qquad \square$$

Our first lower bound shows that even if we have a perfect hint at every time t, one cannot achieve polylogarithmic regret if the movement costs are large. Specifically, as long as the exponent of the movement cost function $\epsilon \in [0,1)$, $\Omega(T^{\frac{1-\epsilon}{2}})$ regret is inevitable.

Theorem 4.2. *Let* $\epsilon \in [0,1), \gamma > 0$ *be the movement cost parameters. Then for any (possibly randomized) online algorithm* $\mathcal{A}$*, there exists a sequence* $\{c_t\}$ *of cost vectors such that even with a* perfect *hint at every step, the algorithm incurs a regret* $\Omega(\gamma T^{\frac{1-\epsilon}{2}})$.

Proof. By Yao's minimax principle, it suffices to exhibit a distribution over cost vectors $\{c_t\}_{t=1}^T$ for which any deterministic algorithm $\mathcal{A}$ incurs the desired regret in expectation.

We consider the simple one-dimensional case with

$$c_t = \pm\gamma\cdot T^{-\epsilon/2}\cdot e_1,$$

where the signs are chosen independently and uniformly at random at every time step, and e_1 is a fixed unit vector. Suppose that the hints are perfect, i.e., $d_t = c_t/\|c_t\|$ and $\lambda_t = \|c_t\|$. Let $h_t = \lambda_t d_t$. Now, consider any deterministic algorithm $\mathcal{A}$. At each step, the algorithm makes the choice of x_t knowing $c_1, c_2, \ldots, c_t$. Denote $\delta_t = x_t - x_{t-1}$.

The expected cost incurred by $\mathcal{A}$ at step t is

$$\begin{aligned}&\mathbb{E}[\langle x_t, c_t\rangle + \gamma\left\|x_t - x_{t-1}\right\|^{1+\epsilon}]\\ &\quad= \mathbb{E}[\langle x_{t-1}, c_t\rangle + \langle\delta_t, c_t\rangle + \gamma\left\|\delta_t\right\|^{1+\epsilon}].\end{aligned}$$

But since c_t has a random sign that is unknown to the algorithm when x_{t-1} is played, we have $\mathbb{E}[\langle x_{t-1}, c_t\rangle] = 0$. By Cauchy–Schwarz, we have $\langle\delta_t, c_t\rangle \ge -\|c_t\|\|\delta_t\| \ge -\gamma T^{-\epsilon/2}\|\delta_t\|$. Substituting these terms, the expected cost incurred by $\mathcal{A}$ at step t is at least

$$\mathbb{E}[\langle x_t, c_t\rangle + \gamma\left\|\delta_t\right\|^{1+\epsilon}] \ge \mathbb{E}[-\gamma T^{-\epsilon/2}\left\|\delta_t\right\| + \gamma\left\|\delta_t\right\|^{1+\epsilon}].$$

If $\epsilon = 0$, the RHS is zero. If $\epsilon > 0$, we apply Lemma 4.1 with $\alpha = \epsilon/2$ and $\delta = \|\delta_t\|$ to obtain

$$\begin{aligned}\mathbb{E}[\langle x_t, c_t\rangle + \gamma\left\|\delta_t\right\|^{1+\epsilon}] &\ge -\frac{\gamma\epsilon\cdot T^{-\frac{(1+\epsilon)}{2}}}{(1+\epsilon)(1+\epsilon)^{1/\epsilon}}\\ &\ge -\frac{\gamma\epsilon\cdot T^{-\frac{(1+\epsilon)}{2}}}{2(1+\epsilon)},\end{aligned}$$

where the second inequality holds since $(1+\epsilon)^{1/\epsilon} \ge 2$ for $\epsilon \in (0,1)$. Since the above inequality holds at each time step $t \le T$, the total expected cost incurred by the algorithm is at least

$$-\frac{\gamma\epsilon}{2(1+\epsilon)}\cdot T^{\frac{1-\epsilon}{2}}.$$

On the other hand, the optimum point in hindsight (which is a unit vector along the direction $-\sum_t c_t$), incurs an expected loss of

$$-\mathbb{E}\left[\left\|\sum_t c_t\right\|\right] = -\sqrt{\frac{2}{\pi}}\gamma T^{\frac{1-\epsilon}{2}}.$$

(The constant comes from the expected value of the magnitude of a Gaussian.[3]) This implies the desired expected regret bound for any $\epsilon \in [0, 1)$. □

Our next lower bound shows that in the case of $\epsilon > 0$, having a *length* hint is important for obtaining poly-logarithmic regret, providing a significant qualitative separation between what is possible with and without informative length hints.

Theorem 4.3. *Let $\epsilon > 0$, and let $\gamma = 1$. Then for any online algorithm $\mathcal{A}$, there is a sequence $\{c_t\}$ of cost vectors such that having a perfect* directional *hint at every step still incurs a regret $\Omega\left(T^{\frac{1}{2+\epsilon}}\right)$.*

Proof. The proof once again proceeds using Yao's principle. We consider the following distribution over cost vectors: let $c_t = z_t u_t$, where $u_t = \pm e_1$ (signs chosen independently and uniformly at random at each time step), and z_t is a Bernoulli random variable that is 1 with probability $(1/2) \cdot T^{-\frac{\epsilon}{2+\epsilon}}$, and 0 otherwise.

At every step, suppose that the algorithm gets the directional hint $d_t = u_t$ before playing x_t, but the length (i.e., z_t) is revealed only after the algorithm plays x_t. Once again, let us denote $\delta_t = x_t - x_{t-1}$. Then the expected cost at step t is

$$\mathbb{E}[\langle x_t, c_t\rangle + \|\delta_t\|^{1+\epsilon}] = \mathbb{E}[\langle x_{t-1}, c_t\rangle + \langle \delta_t, z_t u_t\rangle + \|\delta_t\|^{1+\epsilon}].$$

Since x_{t-1} is played before the random sign in u_t is chosen, we have $\mathbb{E}[\langle x_{t-1}, c_t\rangle] = 0$. Also, we have $\mathbb{E}[\langle \delta_t, z_t u_t\rangle] = \mathbb{E}[z_t] \cdot \mathbb{E}[\langle \delta_t, u_t\rangle] \geq -(1/2)T^{-\frac{\epsilon}{2+\epsilon}} \|\delta_t\|$.

Thus we have the expected cost incurred by the algorithm $\mathcal{A}$ at step t is

$$\mathbb{E}[\langle x_t, c_t\rangle + \|\delta_t\|^{1+\epsilon}] \geq \mathbb{E}\left[-\frac{T^{-\frac{\epsilon}{2+\epsilon}}}{2}\|\delta_t\| + \|\delta_t\|^{1+\epsilon}\right].$$

Applying Lemma 4.1 with $T^{-\alpha}$ replaced by $\frac{T^{\frac{\epsilon}{2+\epsilon}}}{2}$,

$$\mathbb{E}[\langle x_t, c_t\rangle + \|\delta_t\|^{1+\epsilon}] \geq \frac{-\epsilon}{((1+\epsilon)2)^{1+\frac{1}{\epsilon}}} \cdot T^{-\frac{1+\epsilon}{2+\epsilon}}.$$

Using $(1+\epsilon)^{1/\epsilon} > 1$, $\forall \epsilon > 0$, the RHS above is $\geq -\frac{\epsilon}{2(1+\epsilon)} T^{-\frac{1+\epsilon}{2+\epsilon}}$ Thus, the total cost incurred by the algorithm over all time steps is

$$\geq \frac{-\epsilon}{2(1+\epsilon)} \cdot T^{1-\frac{1+\epsilon}{2+\epsilon}} = \frac{-\epsilon}{2(1+\epsilon)} \cdot T^{\frac{1}{2+\epsilon}}.$$

[3] We are using the fact that the sum converges to a Gaussian distribution, which is true in the limit. But due to the slack in the constants, the desired bound holds for T being a large enough constant.

On the other hand, for the optimum solution, the expected cost is

$$-\mathbb{E}\left[\left\|\sum_t c_t\right\|\right] = -\sqrt{\frac{2}{\pi}}\sqrt{(1/2)T^{1-\frac{\epsilon}{2+\epsilon}}} = -\frac{T^{\frac{1}{2+\epsilon}}}{\sqrt{\pi}}.$$

Thus for any $\epsilon > 0$, the desired claim follows. □

5 Conclusions

We have presented algorithms and lower bounds for online learning with movement costs and hints. We consider algorithms with directional and possibly also length hints and prove that while length hints offer more power, in both cases there is a continuum of bounds depending on the movement cost parameter ϵ. Our results highlight the intriguing consequences of different kinds side information in online learning.

References

[1] Nikhil Bansal, Anupam Gupta, Ravishankar Krishnaswamy, Kirk Pruhs, Kevin Schewior, and Clifford Stein. A 2-competitive algorithm for online convex optimization with switching costs. In *APPROX-RANDOM*, pages 96–109, 2015.

[2] Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Online learning with imperfect hints. In *ICML*, pages 822–831, 2020.

[3] Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Online linear optimization with many hints. In *NeurIPS*, 2020.

[4] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[5] Niangjun Chen, Anish Agarwal, Adam Wierman, Siddharth Barman, and Lachlan L. H. Andrew. Online convex optimization using predictions. In *SIGMETRICS*, pages 191–204, 2015.

[6] Niangjun Chen, Joshua Comden, Zhenhua Liu, Anshul Gandhi, and Adam Wierman. Using predictions in online optimization: Looking forward with an eye on the past. In *SIGMETRICS*, pages 193–206, 2016.

[7] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *COLT*, pages 6.1–6.20, 2012.

[8] Ofer Dekel, Arthur Flajolet, Nika Haghtalab, and Patrick Jaillet. Online learning with a hint. In *NIPS*, pages 5299–5308, 2017.

[9] Gautam Goel, Yiheng Lin, Haoyuan Sun, and Adam Wierman. Beyond online balanced descent: An optimal algorithm for smoothed online optimization. In *NeurIPS*, pages 1873–1883, 2019.

[10] Gautam Goel and Adam Wierman. An online algorithm for smoothed regression and LQR control. In *AISTATS*, pages 2504–2513, 2019.

[11] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

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

[13] Elad Hazan and Nimrod Megiddo. Online learning with prior knowledge. In *COLT*, pages 499–513, 2007.

[14] Elad Hazan, Alexander Rakhlin, and Peter L Bartlett. Adaptive online gradient descent. In *NIPS*, pages 65–72, 2008.

[15] Ruitong Huang, Tor Lattimore, András György, and Csaba Szepesvári. Following the leader and fast rates in online linear prediction: Curved constraint sets and other regularities. *JMLR*, 18(145):1–31, 2017.

[16] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *JCSS*, 71(3):291–307, 2005.

[17] T. Kim, Y. Yue, S. Taylor, and I. Matthews. A decision tree framework for spatiotemporal sequence prediction. In *KDD*, pages 577–586, 2015.

[18] Yingying Li and Na Li. Leveraging predictions in smoothed online convex optimization via gradient-based algorithms. In *NeurIPS*, 2020.

[19] Yingying Li, Guannan Qu, and Na Li. Using predictions in online optimization with switching costs: A fast algorithm and a fundamental limit. In *ACC*, pages 3008–3013, 2018.

[20] M. Lin, Z. Liu, A. Wierman, and L. L. Andrew. Online algorithms for geographical load balancing. In *IGCC*, pages 1–10, 2012.

[21] H Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *JMLR*, 18(1):3117–3166, 2017.

[22] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *COLT*, pages 993–1019, 2013.

[23] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic, constrained, and smoothed adversaries. In *NIPS*, 2011.

[24] J. Rios-Torres and A. A. Malikopoulos. A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps. *IEEE Trans. Intelligent Transportation Systems*, 18(5):1066–1077, 2016.

[25] Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

[26] M. Tanaka. Real-time pricing with ramping costs: A new approach to managing a steep change in electricity demand. *Energy Policy*, 34(18):3634–3643, 2006.

[27] F. Zenke, B. Poole, and S. Ganguli. Continual learning through synaptic intelligence. In *ICML*, pages 3987–3995, 2017.

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

A Proof of Lemma 3.2

Proof. By concavity of the logarithm,

$$\log\left(\frac{b}{c}+\sum_{\tau=1}^{t-1}a_\tau+\frac{ca_t}{b+c\sum_{\tau=1}^{t}a_\tau}\right) \le \log\left(\frac{b}{c}+\sum_{\tau=1}^{t}a_\tau\right).$$

Telescoping this sum yields:

$$\sum_{t=1}^{T}\frac{ca_t}{b+c\sum_{\tau=1}^{t}a_\tau} \le \log\left(1+\frac{c\sum_{t=1}^{T}a_t}{b}\right). \qquad \square$$