Proceedings of Machine Learning Research vol 98:1–15, 2019 30th International Conference on Algorithmic Learning Theory

Online Linear Optimization with Sparsity Constraints

Jun-Kun Wang JIMWANG@GATECH.EDU
Georgia Institute of Technology

Chi-Jen Lu CJLU@IIS.SINICA.EDU.TW
Academia Sinica

Shou-De Lin SDLIN@CSIE.NTU.EDU.TW
National Taiwan University

Editors: Aurélien Garivier and Satyen Kale

Abstract

We study the problem of online linear optimization with sparsity constraints in the semi-bandit setting. It can be seen as a marriage between two well-known problems: the online linear optimization problem and the combinatorial bandit problem. For this problem, we provide an algorithm which is efficient and achieves a sublinear regret bound. Moreover, we extend our results to two generalized settings, one with delayed feedbacks and one with costs for receiving feedbacks. Finally, we conduct experiments which show the effectiveness of our methods in practice.

Keywords: Online learning, bandit, sparsity

1. Introduction

Consider the online prediction problem in which a learner must make repeated predictions in the following way. In each round, the learner first observes a context or feature vector and must then make a prediction. After that, the learner receives a corresponding loss as well as some feedback before moving on to the next round. This problem models many practical applications such as recommendation, portfolio selection, and time series prediction. However, there are situations in which it is infeasible for the learner to access all the features due to some resource constraints. For example, consider the scenario in which there are a huge number of sensors deployed, while the objective is to perform event detection based on the sensor data. The value read from each sensor serves as a feature, and because of the bandwidth and energy limitation, it would be better to obtain only a subset of them at each time. Then the success of an algorithm depends crucially on how to select the sensors (features) and update the predictor. More examples can be found in Zolghadr et al. (2013).

Such a task has been formulated as the following problem Kale (2014); Foster et al. (2016), for the case of linear predictors. For a total of T rounds, the learner must iteratively predict the label y_t of a sample $\mathbf{x}_t$ arriving in round t. Each sample has d features, but the learner is only allowed to query k of them, with $k < d$, and must make the prediction based only on them. After each prediction, the true label is observed and the learner suffers some corresponding loss. The learner's goal is to minimize the regret, which is defined as the cumulated loss of the learner minus that of the best fixed predictor based on k features. Note that the learner at each step only receives some limited information, and the setting can be seen as a "semi-bandit" one Audibert et al. (2014); Neu and Bartók (2016), which differs from the full-information one in the following way. In the full-information setting, the learner gets to see all the d features of the sample at each step, while in the semi-bandit setting, the learner only sees a small set of $k < d$ features determined by the action chosen at that step. With such limited information in the semi-bandit setting, the task of the learner becomes harder. In particular, the learner has to face the dilemma of exploration versus exploitation.

When the labels are real numbers and the loss is the square loss, this problem turns out to be computationally hard. More precisely, as shown by Foster et al. (2016), no algorithm running in time polynomial in T can achieve a regret bound of about $T^{1-\delta}$, for a constant $\delta > 0$, unless $\mathsf{NP} \subseteq \mathsf{BPP}$. To avoid the issue of computational intractability, we consider using the linear loss to replace the square loss. That is, for a feature vector $\mathbf{x}$ with label y, the loss of a linear predictor $\mathbf{w}$ is defined as $-y \cdot \langle \mathbf{x}, \mathbf{w} \rangle$. Such a loss seems appropriate for binary classification, with $y \in \{-1, 1\}$, because without the minus sign, the product $y \cdot \langle \mathbf{x}, \mathbf{w} \rangle$ is the correlation between the label and the prediction, which corresponds naturally to a reward. We can place this problem in the more general framework of online linear optimization, with $-y\mathbf{x}$ as the loss vector and the feasible set containing the linear predictor $\mathbf{w}$'s. However, as our problem requires an additional sparsity constraint $\|\mathbf{w}\|_0 \leq k$, our feasible set becomes nonconvex, which prevents us from applying standard algorithms based on "mirror descent" or "follow the regularized leader" (see, e.g., Bubeck and Cesa-Bianchi (2012)). In fact, similar issues arise in the problem of combinatorial bandits Cesa-Bianchi and Lugosi (2012) for which a different approach based on "follow the perturbed leader" algorithm Kalai and Vempala (2005) has been proposed Neu and Bartók (2016). Unfortunately, we cannot apply such results either, as their feasible sets consist of only binary strings (indicating which subset of arms are selected), while ours contains real vectors. Furthermore, the algorithms in Cesa-Bianchi and Lugosi (2012); Foster et al. (2016) do not help as their time and space complexity both grow proportionally to d^k, which is prohibitively large even for moderate values of d and k. Therefore, it is not clear if it is possible to have an efficient algorithm achieving a sublinear regret for our problem.

Our results. We answer this question affirmatively. More precisely, we consider the general problem of online sparse linear optimization, in which the feasible set consists of $\mathbf{w} \in \mathbb{R}^d$ satisfying a sparsity constraint $\|\mathbf{w}\|_0 \leq k$ as well as an L_b-norm constraint $\|\mathbf{w}\|_b \leq 1$, for $b \geq 1$. For this problem, we provide an efficient algorithm achieving a regret of about $\sqrt{T}$ and having time and space complexity scaling only linearly in d instead of in d^k. The algorithm works for any $b \geq 1$ and $k \geq 1$, Moreover, we extend our results to two generalized scenarios, one with delayed feedbacks, and one with different costs for accessing different components of the loss vectors. Finally, we perform experiments which show that our algorithm also work well empirically on real-world data sets.

Related works. While we are not aware of previous works on our problem of online sparse linear optimization, there have been several related works for the problem of learning predictors based on subset of features, in addition to Foster et al. (2016). Cesa-Bianchi et al. (2011); Hazan and Koren (2012) considered a related problem in the batch setting, in which the learner in the training phase can only query a small subset of features of each training example, but the final hypothesis for testing can depend on all the features. Their algorithms do not work for us as our problem requires always making predictions from subsets of features. Amin et al. (2015) studied an online problem about prediction with expert advice under budget constraints, which can be seen as a relaxation from the bandit setting towards the full-information one. This setting appears easier than ours as the learner can access additional information not related to the expert it chooses to follow, which allows the decoupling of exploration from exploitation. Zolghadr et al. (2013) considered a setting in which the loss of a learner in each round is defined as its prediction loss plus the cost of querying features, which is different from ours. Kale et al. (2017) considered the square loss function under the contraint that $\|\mathbf{w}\|_0 \leq k$. Yet, they made additional assumptions like restricted isometry property to avoid the computational hardness. Finally, Wang et al. (2013) studied the task of binary classification, but they provided a mistake bound rather than a regret bound, and their method seems specific to the case of 0-1 loss.

2. Preliminaries

Let us first introduce some notations which we will use later. For a positive integer d, let $[d]$ denote the set $\{1, \ldots, d\}$. For a vector $\mathbf{x} \in \mathbb{R}^d$ and for $i \in [d]$, we let $\mathbf{x}_i$ denote the i'th component of $\mathbf{x}$, and when we need to use the notation $\mathbf{x}_t$ for a time-indexed vector in $\mathbb{R}^d$, we will write $\mathbf{x}_{t,i}$ for the the i'th component of $\mathbf{x}_t$. We let $\|\mathbf{x}\|_b$ denote the L_b-norm of a vector $\mathbf{x}$. For a condition $\mathcal{C}$, we use the indicator function $\mathbb{1}_{\mathcal{C}}$ to give the value 1 if $\mathcal{C}$ holds and 0 otherwise.

As discussed in the introduction, we are interested in the online sparse linear optimization problem in which the learner can only choose actions from a feasible set consisting of "sparse" vectors. In particular, we will consider feasible sets of the form $\{\mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\|_0 \leq k \text{ and } \|\mathbf{w}\|_b \leq C\}$, for $b \in [1, \infty]$ and $C > 0$. For simplicity of presentation, we will discuss only the case with $C = 1$, as it is straightforward to extend our results to the general case, and we let $\mathcal{K}$ denote such a feasible set. Formally, we consider the online linear optimization problem, in which the learner must play the following game for a total of T rounds. In round t, the learner must first choose an action $\mathbf{w}_t$ from some feasible set $\mathcal{K}$. For this choice, the learner suffers some loss $\langle \boldsymbol{\theta}_t, \mathbf{w}_t \rangle$, according to some loss vector $\boldsymbol{\theta}_t$. After that, the learner receives the feedbacks $\boldsymbol{\theta}_{t,i}$, for every i such that $\mathbf{w}_{t,i} \neq 0$. Without loss of generality, assume that each $\|\boldsymbol{\theta}_t\|_\infty \leq 1$ (again, our results can be easily extended to the more general case with $\|\boldsymbol{\theta}_t\|_{b'} \leq C'$, for any positive b' and C'). Note that the problem of online prediction with limited feature access, discussed in the introduction, can be cast as this problem, with the loss vector defined as $\boldsymbol{\theta}_t = -y_t \mathbf{x}_t$, for the feature vector $\mathbf{x}_t$ and its label y_t. To measure the performance of the learner, a common way is to compare its expected total loss to that of the best fixed action $\mathbf{w}_* \in \mathcal{K}$ in hindsight. The difference is called the *regret* of the learner, which is

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle \boldsymbol{\theta}_t, \mathbf{w}_t \rangle\right] - \min_{\mathbf{w}_* \in \mathcal{K}} \sum_{t=1}^{T} \langle \boldsymbol{\theta}_t, \mathbf{w}_* \rangle,$$

where the expectation above is over the randomness used by the learner.

3. Follow the Perturbed Sparse Leaders

As mentioned, we consider the feasible set

$$\mathcal{K} = \{\mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\|_0 \leq k \text{ and } \|\mathbf{w}\|_b \leq 1\},$$

for $b \in [1, \infty]$. Our algorithm was inspired by that of Neu and Bartók (2016), which is based on the "follow the perturbed leader" algorithm. However, their results do not apply here, since they have a different feasible set, with binary vectors only, and their loss vectors have only nonnegative

Algorithm 1 Follow the perturbed sparse leader

for $t = 1$ **to** T **do**
1: Play $\mathbf{w}_t$ computed according to (2).
2: Receive feedback $\boldsymbol{\theta}_{t,i}$ for $i \in Q_t$.
3: Call Algorithm 2 with input (i, t, M) to get $h_{t,i}$, $\forall i$.
4: Construct $\hat{\boldsymbol{\theta}}_t$, with $\hat{\boldsymbol{\theta}}_{t,i} = \boldsymbol{\theta}_{t,i} \cdot h_{t,i} \cdot \mathbb{1}_{i \in Q_t}$.

components. In fact, some of their key lemmas rely heavily on these two assumptions, and we have to develop new analyses for our different setting.

Our algorithm is summarized in Algorithm 1, with the parameters

$$\begin{cases} \eta = \sqrt{(k^{(b-1)/b} \log d)/(d^2 T \log T)}, \\ \gamma = c_0 d \eta \log T, \\ M = \lceil (d \ln T)/(k\gamma) \rceil, \end{cases} \tag{1}$$

where c_0 is a small enough constant to make $\gamma < 1$ for any T, and we use the convention that $(b-1)/b = 1$ for $b = \infty$. Formally, our algorithm does the following in round t. First, it samples a random perturbation $\mathbf{Z}_t$ from the two-sided exponential distribution, with the density function

$$f(\mathbf{z}) = e^{-\|\mathbf{z}\|_1}/2^d$$

for $\mathbf{z} \in \mathbb{R}^d$. Then it computes the perturbed leader:

$$\hat{\mathbf{w}}_t \in \min_{\mathbf{w} \in \mathcal{K}} \left\langle \mathbf{w}, \eta \sum_{\tau=1}^{t-1} \hat{\boldsymbol{\theta}}_\tau - \mathbf{Z}_t \right\rangle,$$

where each $\hat{\boldsymbol{\theta}}_\tau$ is our estimator for $\boldsymbol{\theta}_\tau$ to be described later in (3). Such a $\hat{\mathbf{w}}_t$ can be found efficiently, as guaranteed by the following, which we prove in Subsection 3.1. Notice that in each round t, a new sample of $\mathbf{Z}_t$ is drawn from $f(\mathbf{z})$ for computing the update.

Lemma 1 *For any* $\mathbf{v} \in \mathbb{R}^d$*, the optimization problem* $\min_{\mathbf{w} \in \mathcal{K}} \langle \mathbf{w}, \mathbf{v} \rangle$ *can be solved in polynomial time* $O(d \log k)$.

Our algorithm does not simply play this $\hat{\mathbf{w}}_t$ in round t; instead, it plays

$$\mathbf{w}_t = \begin{cases} \hat{\mathbf{w}}_t, & \text{with probability } 1 - \gamma; \\ \text{uniformly random } \mathbf{w} \in \bar{\mathcal{K}}, & \text{with probability } \gamma; \end{cases} \tag{2}$$

where $\bar{\mathcal{K}}$ contains exactly those $\mathbf{w} \in \mathcal{K}$ such that for some $Q \subseteq [d]$ of size k, $\mathbf{w}$ takes the value $1/k^{1/b}$ at those dimensions in Q and the value 0 elsewhere. By choosing $\mathbf{w}_t$ this way, our algorithm is guaranteed to receive each $\boldsymbol{\theta}_{t,i}$ with a good probability. More precisely, by letting

$$Q_t \equiv \{i : \mathbf{w}_{t,i} \neq 0\} \text{ and } q_{t,i} \equiv \Pr[i \in Q_t \mid \mathcal{F}_{t-1}],$$

where $\mathcal{F}_{t-1}$ denotes the σ-algebra generated by the random events up to round $t-1$, we have

$$q_{t,i} \geq \gamma k/d \text{ for every } i.$$

Algorithm 2 (i, t, M) Geometric resampling

for $n = 1, 2, \ldots, M$ **do**

1: Sample $\mathbf{w}_t^{(n)}$ independently according to the distribution of $\mathbf{w}_t$ in (2).

2: **if** $\mathbf{w}_{t,i}^{(n)} \neq 0$ **then** break the loop.

3: Return $h_{t,i} = n$.

It remains to specify our estimator $\hat{\boldsymbol{\theta}}_t$ for $\boldsymbol{\theta}_t$. One possibility would be to set $\hat{\boldsymbol{\theta}}_{t,i} = \boldsymbol{\theta}_{t,i} \cdot \frac{1}{q_{t,i}} \cdot \mathbb{1}_{i \in Q_t}$, which has the desirable property that $\mathbb{E}[\hat{\boldsymbol{\theta}}_{t,i} \mid \mathcal{F}_{t-1}] = \boldsymbol{\theta}_{t,i}$, because $\mathbb{E}[\mathbb{1}_{i \in Q_t} \mid \mathcal{F}_{t-1}] = q_{t,i}$. However, to use this estimator, the learner needs the value $q_{t,i}$, which seems hard to determine as it does not appear to have a closed form. Fortunately, we can adopt the geometric resampling approach of Neu and Bartók (2016) to approximate $\frac{1}{q_{t,i}}$ by another number $h_{t,i}$, which is described in Algorithm 2. Then we choose our estimator $\hat{\boldsymbol{\theta}}_t$ by setting

$$\hat{\boldsymbol{\theta}}_{t,i} = \boldsymbol{\theta}_{t,i} \cdot h_{t,i} \cdot \mathbb{1}_{i \in Q_t}, \tag{3}$$

for each i (note that $\hat{\boldsymbol{\theta}}_{t,i} = 0$ for $i \notin Q_t$). This estimator is almost unbiased, as shown in the following, which we prove in Subsection 3.2.

Remark: We would like to emphasize that the step on line 1 of Algorithm 2 is about sampling a new $\hat{\mathbf{w}}_t$ first, by computing the perturbed leader again, and then sampling a $\mathbf{w}_t$ by (2). The computational time complexity of Algorithm 2 is $O(Md)$, assuming that the oracle of uniformly sampling a $\mathbf{w} \in \bar{\mathcal{K}}$ is available.

Lemma 2 *For any i, $\left|\mathbb{E}[\hat{\boldsymbol{\theta}}_{t,i} \mid \mathcal{F}_{t-1}] - \boldsymbol{\theta}_{t,i}\right| \leq 1/T$.*

Finally, our algorithm can achieve a regret bound of about $\sqrt{T}$, as guaranteed by the following theorem, which we prove in Subsection 3.3.

Theorem 3 *The regret of Algorithm 1 is at most $\mathcal{O}(\sqrt{\alpha T \log T})$ for $\alpha = k^{(b-1)/b} d^2 \log d$.*

3.1. Proof of Lemma 1

For simplicity, let us assume that the dimensions are arranged to have $|\mathbf{v}_1| \geq |\mathbf{v}_2| \geq \cdots \geq |\mathbf{v}_d|$. For the case of $\mathcal{K} = \mathcal{K}_1$, it is well known that one can have the minimizer $\mathbf{w}$ with $\mathbf{w}_i = -\mathrm{sign}(\mathbf{v}_i)\mathbb{1}_{i=1}$ for every i. For the case of $\mathcal{K} = \mathcal{K}_\infty$, it is easy to check that one can have the minimizer $\mathbf{w}$ with $\mathbf{w}_i = -\mathrm{sign}(\mathbf{v}_i)\mathbb{1}_{i \leq k}$ for every i.

Now, let us consider the case of $\mathcal{K}$ with $b \in (1, \infty)$. For $\mathbf{v} \in \mathbb{R}^d$ and $Q \subseteq [d]$, let $\mathbf{v}_Q$ denote the projection of $\mathbf{v}$ to those dimensions in Q. Then for any $\mathbf{v} \in \mathbb{R}^d$, and any $\mathbf{w} \in \mathcal{K}$ with $Q = \{i : \mathbf{w}_i \neq 0\}$, we know by Hölder's inequality that $\langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{w}_Q, \mathbf{v}_Q \rangle \geq -\|\mathbf{w}\|_b \cdot \|\mathbf{v}_Q\|_a$, for $a = b/(b-1)$. Moreover, one can have $\langle \mathbf{w}_Q, \mathbf{v}_Q \rangle = -\|\mathbf{w}\|_b \cdot \|\mathbf{v}_Q\|_a$, when $|\mathbf{w}_i|^b / \|\mathbf{w}\|_b^b = |\mathbf{v}_i|^a / \|\mathbf{v}\|_a^a$ and $\mathbf{w}_i \mathbf{v}_i \leq 0$ for every $i \in Q$. Thus, to find $\mathbf{w} \in \mathcal{K}$ which minimizes $\langle \mathbf{w}, \mathbf{v} \rangle$, we first let $Q = [k]$ so that $\|\mathbf{v}_Q\|_a \geq \|\mathbf{v}_{Q'}\|_a$ for any $Q' \subseteq [d]$ with $|Q'| \leq k$. Then we let $\hat{\mathbf{w}}_i = -\mathrm{sign}(\mathbf{v}_i)|\mathbf{v}_i|^{a/b}\mathbb{1}_{i \in Q}$, and choose $\mathbf{w}_i = \hat{\mathbf{w}}_i / \|\hat{\mathbf{w}}\|_b$. Clearly, we have $\mathbf{w} \in \mathcal{K}$ and $\langle \mathbf{w}, \mathbf{v} \rangle = -\|\mathbf{v}_Q\|_a \leq -\|\mathbf{v}_{Q'}\|_a \leq \langle \mathbf{w}', \mathbf{v} \rangle$ for any $\mathbf{w}' \in \mathcal{K}$ with $Q' = \{i : \mathbf{w}'_i \neq 0\}$ as $|Q'| \leq k$. The procedure clearly can be done in polynomial time (e.g. $O(d \log k)$) by using some efficient algorithms in sorting.

3.2. Proof of Lemma 2

According to the definition,

$$\mathbb{E}[\hat{\boldsymbol{\theta}}_{t,i} \mid \mathcal{F}_{t-1}] = \boldsymbol{\theta}_{t,i} \cdot \mathbb{E}[h_{t,i} \mid \mathcal{F}_{t-1}] \cdot q_{t,i},$$

as $h_{t,i}$ is independent of Q_t. Moreover,

$$\begin{aligned} & \mathbb{E}\left[h_{t,i} \mid \mathcal{F}_{t-1}\right] \\ = \; & \sum_{n=1}^{M} n(1-q_{t,i})^{n-1} q_{t,i} + M(1-q_{t,i})^M \\ = \; & \frac{1}{q_{t,i}} \left(1-(1-q_{t,i})^M\right) \end{aligned}$$

by a routine calculation. As a result, we have

$$\left| \mathbb{E}[\hat{\boldsymbol{\theta}}_{t,i} \mid \mathcal{F}_{t-1}] - \boldsymbol{\theta}_{t,i} \right| \leq (1-q_{t,i})^M \leq e^{-Mq_{t,i}} \leq 1/T,$$

as $M = \lceil (d \ln T)/(k\gamma) \rceil \geq (1/q_{t,i}) \ln T$.

3.3. Proof of Theorem 3

Let $\mathbf{w}_*$ be the best offline predictor in $\mathcal{K}$. By definition, the regret of our algorithm is

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle \mathbf{w}_t, \boldsymbol{\theta}_t \rangle\right] - \sum_{t=1}^{T} \langle \mathbf{w}_*, \boldsymbol{\theta}_t \rangle = \sum_{t=1}^{T} \mathbb{E}\left[\langle \mathbf{w}_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle\right],$$

which can be decomposed as

$$\sum_{t=1}^{T} \mathbb{E}\left[\langle \mathbf{w}_t - \hat{\mathbf{w}}_t, \boldsymbol{\theta}_t \rangle\right] + \sum_{t=1}^{T} \mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle\right]. \tag{4}$$

The first sum in (4) is at most $2k^{(b-1)/b}\gamma T$ because for each t, $\hat{\mathbf{w}}_t \neq \mathbf{w}_t$ with probability γ and in that case $\langle \mathbf{w}_t - \hat{\mathbf{w}}_t, \boldsymbol{\theta}_t \rangle \leq \|\mathbf{w}_t - \hat{\mathbf{w}}_t\|_1 \cdot \|\boldsymbol{\theta}_t\|_\infty \leq 2k^{(b-1)/b}$ since $\|\boldsymbol{\theta}_t\|_\infty \leq 1$ and $\|\mathbf{w}_t - \hat{\mathbf{w}}_t\|_1 \leq \|\mathbf{w}_t\|_1 + \|\hat{\mathbf{w}}_t\|_1 \leq 2k^{(b-1)/b}$, where we use Holder's inequality to get the upper bound of $\|\mathbf{w}\|_1$ as $\|\mathbf{w}\|_b \leq 1$ for any $\mathbf{w} \in \mathcal{K}$. To bound the second sum in (4), we follow Neu and Bartók (2016) and use the help of a virtual algorithm that (i) uses a time-independent perturbation vector and (ii) is allowed to peek one round ahead into the future. More precisely, the virtual algorithm first draws the perturbation vector $\tilde{\mathbf{Z}}$ according to the two-sided exponential distribution (independently from those used by our algorithm), and then in each round t it plays

$$\tilde{\mathbf{w}}_t \in \arg\min_{\mathbf{w} \in \mathcal{K}} \left\langle \mathbf{w}, \eta \sum_{\tau=1}^{t} \hat{\boldsymbol{\theta}}_\tau - \tilde{\mathbf{Z}} \right\rangle.$$

Our key lemma is the following.

Lemma 4 *Given t and $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{t-1}$, let $U_t(\mathbf{z})$ denote an arbitrary element of*

$$\arg\min_{\mathbf{w}\in\mathcal{K}} \left\langle \mathbf{w}, \eta \sum_{\tau=1}^{t-1} \hat{\boldsymbol{\theta}}_\tau - \mathbf{z} \right\rangle.$$

Then,

$$\sum_{t=1}^{T} \mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle\right] \leq 2k^{(b-1)/b} + \frac{2}{\eta} k^{(b-1)/b} \ln d +$$
$$\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - U_t(\mathbf{z} - \eta\hat{\boldsymbol{\theta}}_t), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z}) \mid \mathcal{F}_{t-1}\right]\right]$$

where in the last line, the outer expectation is over the randomness up to round $t-1$, which determines $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{t-1}$, and the inner expectation is over that of round t, which determines $\hat{\boldsymbol{\theta}}_t$.

Proof Our proof follows the approach taken by Kalai and Vempala (2005); Neu and Bartók (2016). Let us first consider any fixed t, and note that we can express $\mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle\right]$ as

$$\mathbb{E}\left[\mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle \mid \mathcal{F}_{t-1}\right]\right], \tag{5}$$

where the outer expectation is over the randomness up to round $t-1$ and the inner expectation is over that of round t, which determines $\hat{\mathbf{w}}_t$ and $\hat{\boldsymbol{\theta}}_t$. Now consider any fixed realization of randomness up to round $t-1$, denoted as $\mathbf{r}_{t-1}$, which in turn fixes $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{t-1}$. Observe that the inner expectation in (5) equals to the value

$$\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle df(\mathbf{z}), \tag{6}$$

and we can see it as a fixed value which is determined by $\mathbf{r}_{t-1}$ but independent of the random variables $\hat{\mathbf{w}}_t$ and $\hat{\boldsymbol{\theta}}_t$. Then, let $\mathbb{E}_t[\cdot]$ denote the expectation over the distribution of $\hat{\boldsymbol{\theta}}_t$ conditioned on $\mathbf{r}_{t-1}$, and let us decompose (6) as

$$\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - \mathbf{w}_*, \boldsymbol{\theta}_t - \mathbb{E}_t[\hat{\boldsymbol{\theta}}_t] \rangle df(\mathbf{z}) + \tag{7}$$
$$\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - \mathbf{w}_*, \mathbb{E}_t[\hat{\boldsymbol{\theta}}_t] \rangle df(\mathbf{z}). \tag{8}$$

We claim that (7) is at most $2k^{(b-1)/b}/T$, because for any $\mathbf{w} \in \mathcal{K}$,

$$\begin{aligned} &\left\langle \mathbf{w} - \mathbf{w}_*, \boldsymbol{\theta}_t - \mathbb{E}_t[\hat{\boldsymbol{\theta}}_t] \right\rangle \\ &\leq \|\mathbf{w} - \mathbf{w}_*\|_1 \cdot \left\|\boldsymbol{\theta}_t - \mathbb{E}_t[\hat{\boldsymbol{\theta}}_t]\right\|_\infty \\ &\leq 2k^{(b-1)/b}/T, \end{aligned}$$

using the fact that $\|\mathbf{w} - \mathbf{w}_*\|_1 \leq \|\mathbf{w}\|_1 + \|\mathbf{w}_*\|_1 \leq 2k^{(b-1)/b}$ and the bound $|\boldsymbol{\theta}_{t,i} - \mathbb{E}_t[\hat{\boldsymbol{\theta}}_{t,i}]| \leq 1/T$ from Lemma 2. On the other hand, (8) equals

$$\mathbb{E}_t\left[\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - \mathbf{w}_*, \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z})\right],$$

which can be decomposed as

$$\mathbb{E}_t\left[\int_{\mathbb{R}^d}\langle U_t(\mathbf{z})-U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t),\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\right]+$$
$$\mathbb{E}_t\left[\int_{\mathbb{R}^d}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\right].$$

By combining all the bounds derived so far, we can obtain an upper bound on the inner expectation in (5), for fixed t and $\mathbf{r}_{t-1}$. With such a bound, by taking an expectation over $\mathbf{r}_{t-1}$ and a summation over t, we obtain

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\langle\hat{\mathbf{w}}_t-\mathbf{w}_*,\boldsymbol{\theta}_t\rangle\right] \;\leq\; & 2k^{(b-1)/b}+\\ & \sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}_t\left[\int_{\mathbb{R}^d}\langle U_t(\mathbf{z})-U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t),\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\right]\right]+\\ & \sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}_t\left[\int_{\mathbb{R}^d}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\right]\right].\end{aligned} \tag{9}$$

Note that the last sum in (9) equals

$$\mathbb{E}\left[\sum_{t=1}^{T}\int_{\mathbb{R}^d}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\right],$$

with the expectation taken over the distribution of $\hat{\boldsymbol{\theta}}_1,\ldots,\hat{\boldsymbol{\theta}}_T$, by noting that each $U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)$ is independent of $\hat{\boldsymbol{\theta}}_{t+1},\ldots,\hat{\boldsymbol{\theta}}_T$. Then Lemma 4 follows from the following.

Proposition 5 *For any $\hat{\boldsymbol{\theta}}_1,\ldots,\hat{\boldsymbol{\theta}}_T$,*

$$\sum_{t=1}^{T}\int_{\mathbb{R}^d}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\leq\frac{2}{\eta}k^{(b-1)/b}\ln d.$$

Proof Taking the standard approach for analyzing the "follow the perturbed leader" algorithm, one can show that

$$\begin{aligned}&\sum_{t=1}^{T}\int_{\mathbb{R}^d}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\\ &\quad=\;\int_{\mathbb{R}^d}\sum_{t=1}^{T}\langle U_t(\mathbf{z}-\eta\hat{\boldsymbol{\theta}}_t)-\mathbf{w}_*,\hat{\boldsymbol{\theta}}_t\rangle df(\mathbf{z})\\ &\quad\leq\;\frac{1}{\eta}\,\mathbb{E}[\langle\tilde{\mathbf{Z}},\tilde{\mathbf{w}}_1\rangle],\end{aligned}$$

where $\tilde{\mathbf{Z}}$ is the perturbation vector sampled by the virtual algorithm and $\tilde{\mathbf{w}}_1$ is what it plays in the first round (see for example the proof of Lemma 7 in Neu and Bartók (2016)). However, to bound the expectation $\mathbb{E}[\langle \tilde{\mathbf{Z}}, \tilde{\mathbf{w}}_1 \rangle]$, we do not find any existing bound, such as that in Neu and Bartók (2016), which we can use directly, as our $\tilde{\mathbf{Z}}$ is sampled from the two-sided exponential distribution and our $\tilde{\mathbf{w}}_1 \in \mathcal{K}$ is under a different constraint and has real-valued entries. Still, we can modify the proof of Lemma 14 in Neu and Bartók (2016) to bound the expectation in the following way.

Let $\tilde{\mathbf{Z}}_1^*, \dots, \tilde{\mathbf{Z}}_d^*$ denote the permutation of $\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_d$ such that $\tilde{\mathbf{Z}}_i^* \geq \tilde{\mathbf{Z}}_{i+1}^*$ for every i. Then for $a = b/(b-1)$,

$$\langle \tilde{\mathbf{Z}}, \tilde{\mathbf{w}}_1 \rangle \leq \left(\sum_{i=1}^{k} \left| \tilde{\mathbf{Z}}_i^* \right|^a \right)^{1/a} \|\tilde{\mathbf{w}}_1\|_b \leq \left(\sum_{i=1}^{k} \left| \tilde{\mathbf{Z}}_i^* \right|^a \right)^{1/a}$$

by Hölder's inequality. Let

$$Y = \left(\sum_{i=1}^{k} |\tilde{\mathbf{Z}}_i^*|^a \right)^{1/a},$$

which is a nonnegative random variable. Then for any $A \geq 0$, we have

$$\mathbb{E}[Y] = \int_0^\infty \Pr[Y > y] dy \leq A + \int_A^\infty \Pr[Y > y] dy,$$

where

$$\Pr[Y > y] = \Pr\left[\sum_{i=1}^{k} \left| \tilde{\mathbf{Z}}_i^* \right|^a > y^a \right] \leq \Pr\left[\left| \tilde{\mathbf{Z}}_1^* \right| > \frac{y}{k^{1/a}} \right]$$

which is at most

$$\sum_i \Pr\left[\left| \tilde{\mathbf{Z}}_i \right| > \frac{y}{k^{1/a}} \right] = d e^{-y/k^{1/a}}.$$

Therefore, we have

$$\mathbb{E}[Y] \leq A + d \int_A^\infty e^{-y/k^{1/a}} dy = A + (dk^{1/a}) e^{-A/k^{1/a}}$$

which is at most

$$k^{1/a} \ln d + k^{1/a} \leq 2k^{1/a} \ln d$$

by choosing $A = k^{1/a} \ln d$. As $\mathbb{E}[\langle \tilde{\mathbf{Z}}, \tilde{\mathbf{w}}_1 \rangle] \leq \mathbb{E}[Y]$ and $a = b/(b-1)$, Proposition 5 follows. ■

Our remaining task is to bound the last sum in Lemma 4. For this, we rely on the following lemma.

Lemma 6 *For any t,*

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - U_t(\mathbf{z} - \eta \hat{\boldsymbol{\theta}}_t), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z}) \mid \mathcal{F}_{t-1} \right] \leq 2\eta d^2.$$

Proof Recall that $f(\mathbf{z}) = e^{-\|\mathbf{z}\|_1}/2^d$. Fix any t and any realization of randomness up to round $t-1$, denoted as $\mathbf{r}_{t-1}$, which in turn fixes $\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{t-1}$. Let $\mathbb{E}_t[\cdot]$ denote the expectation over the randomness in round t conditioned on $\mathbf{r}_{t-1}$, and let $\Pr_t[\cdot]$ denote the corresponding conditional probability (conditioned on $\mathbf{r}_{t-1}$). Note that for any $\hat{\boldsymbol{\theta}}_t$,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}) - U_t(\mathbf{z} - \eta\hat{\boldsymbol{\theta}}_t), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z}) \\
&= \int_{\mathbb{R}^d} \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z}) - \int_{\mathbb{R}^d} \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z} + \eta\hat{\boldsymbol{\theta}}_t) \\
&= \int_{\mathbb{R}^d} \left(1 - e^{\|\mathbf{z}\|_1 - \|\mathbf{z}+\eta\hat{\boldsymbol{\theta}}_t\|_1}\right) \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle df(\mathbf{z}) \\
&\leq \int_{\mathbb{R}^d} \Big| \|\mathbf{z}\|_1 - \|\mathbf{z} + \eta\hat{\boldsymbol{\theta}}_t\|_1 \Big| \Big| \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle \Big| \, df(\mathbf{z}) \\
&\leq \eta \|\hat{\boldsymbol{\theta}}_t\|_1 \int_{\mathbb{R}^d} \Big| \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle \Big| \, df(\mathbf{z}).
\end{aligned}$$

Moreover, since $\hat{\boldsymbol{\theta}}_{t,j} = 0$ if $j \notin Q_t$, we have

$$\begin{aligned}
\int_{\mathbb{R}^d} \Big| \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle \Big| \, df(\mathbf{z}) &\leq \sum_{j \in Q_t} \int_{\mathbb{R}^d} |(U_t(\mathbf{z}))_j| |\hat{\boldsymbol{\theta}}_{t,j}| dz \\
&\leq \sum_{j \in Q_t} q_{t,j} h_{t,j},
\end{aligned}$$

where $(U_t(\mathbf{z}))_j$ denotes the j'th component of the vector $U_t(\mathbf{z})$. This implies that

$$\begin{aligned}
&\mathbb{E}_t \left[\|\hat{\boldsymbol{\theta}}_t\|_1 \int_{\mathbb{R}^d} \Big| \langle U_t(\mathbf{z}), \hat{\boldsymbol{\theta}}_t \rangle \Big| \, df(\mathbf{z}) \right] \\
&\leq \mathbb{E}_t \left[\sum_{i \in Q_t} q_{t,i} h_{t,i}^2 \right] + \mathbb{E}_t \left[\sum_{i \in Q_t} \sum_{j \neq i} h_{t,i} q_{t,j} h_{t,j} \right]. \qquad (10)
\end{aligned}$$

The first term in (10) can be written as

$$\sum_i \Pr_t[i \in Q_t] \cdot \mathbb{E}_t[q_{t,i} h_{t,i}^2] = \sum_i q_{t,i}^2 \mathbb{E}_t[h_{t,i}^2] \leq 2d,$$

as Q_t is independent of $h_{t,i}$ given $\mathbf{r}_{t-1}$ and $\mathbb{E}_t[h_{t,i}^2]$ equals

$$\sum_{n=1}^{M} n^2 (1 - q_{t,i})^{n-1} q_{t,i} + M^2 (1 - q_{t,i})^M \leq \frac{2}{q_{t,i}^2}$$

by a routine calculation. The second term in (10) equals

$$\sum_i \Pr_t[i \in Q_t] \cdot \sum_{j \neq i} \mathbb{E}_t[h_{t,i}] q_{t,j} \mathbb{E}_t[h_{t,j}]$$

as Q_t, $h_{t,i}$, and $h_{t,j}$, for $j \neq i$, are independent of each other given $\mathbf{r}_{t-1}$ (we compute $h_{t,i}$ and $h_{t,j}$ using independent samples), which is at most

$$\sum_i q_{t,i} \sum_{j \neq i} \frac{1}{q_{t,i}} q_{t,j} \frac{1}{q_{t,j}} \leq d^2,$$

The lemma follows as the sum in (10) is at most $2d + d^2 \leq 2d^2$. ■

By combining all the bounds above, we can conclude that the regret of our algorithm is at most

$$2\gamma k^{(b-1)/b}T + 2k^{(b-1)/b} + \frac{2}{\eta}k^{(b-1)/b}\ln d + 2\eta d^2 T,$$

which gives the stated bound in Theorem 3 for our choice of parameters.

■

4. Extensions

Our algorithms can also be extended to work in more general settings. Here we consider two such examples, one with delayed feedbacks and one with costs for receiving feedbacks.

4.1. Delayed feedbacks

Consider the scenario in which the feedbacks may be delayed, instead of being received right away. This has been considered previously by Quanrud and Khashabi (2015) in the full-information setting and Cesa-Bianchi et al. (2018) in a adversarial bandit setting, and here we study it in our semi-bandit setting. Formally, the feedback $\boldsymbol{\theta}_{t,i}$ for $i \in Q_t$ for round t is delivered at the end of round $t + D_t - 1$, for some $D_t \geq 1$, and let $D = \sum_{t=1}^{T} D_t$. Hence, in the standard setting with no delays, $D_t = 1$ and $D = T$.

In this delay setting, before round t, only the feedbacks from some subset S_t of previous rounds are available, so we can only compute $\hat{\boldsymbol{\theta}}_\tau$ for $\tau \in S_t$. Thus, we modify Algorithm 1 by choosing the perturbed leader according to $\hat{\mathbf{w}}_t \in \arg\min_{\mathbf{w}\in\mathcal{K}}\langle \mathbf{w}, \eta\sum_{\tau\in S_t}\hat{\boldsymbol{\theta}}_\tau - \mathbf{Z}_t\rangle$, while the rest is the same. The resulting algorithm achieves the following regret bound, which we prove in the supplementary.

Theorem 7 *By choosing the parameters $\eta = \sqrt{(k^{(b-1)/b}\log d)/(d^2 D\log D)}$ and $\gamma = 2d\eta\log D$, the regret of the new algorithm is at most $\mathcal{O}(\alpha\sqrt{D\log D})$, for $\alpha = d\sqrt{k^{(b-1)/b}\log d}$.*

4.2. Knapsack constraints

Consider the scenario that receiving feedback $\boldsymbol{\theta}_{t,i}$ incurs some cost c_i, which is independent of the round t but may depend on the dimension i. The learner knows the costs and has a budget B in each round, which limits the total feedback costs affordable in each round. More precisely, the feasible set now becomes

$$\mathcal{V} \equiv \left\{ \mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\|_b \leq 1, \sum_{i:\mathbf{w}_i \neq 0} c_i \leq B \right\}.$$

We would like to modify Algorithm 1 to work for this new setting. Given the different feasible set $\mathcal{V}$, the learner now faces a different optimization problem: find $\hat{\mathbf{w}} \in \mathcal{V}$ to minimize $\Psi(\mathbf{w}) = \langle \mathbf{w}, \boldsymbol{\Theta}\rangle$, for $\boldsymbol{\Theta} = \eta\sum_{\tau=1}^{t-1}\hat{\boldsymbol{\theta}}_\tau - \mathbf{Z}_t$ in round t. Following the proof of Lemma 1, it suffices to solve the problem: find $\hat{J} \subseteq [d]$ to maximize $\Phi(J) = \|\boldsymbol{\Theta}_J\|_a$ subject to $\sum_{i\in J} c_i \leq B$, for the number a such that $1/a + 1/b = 1$.

For the case with $b = 1$, it is known that we only need to consider J of size one. That is, the optimization problem now becomes finding $i \in [d]$ which maximizes $|\boldsymbol{\Theta}_i|$ subject to $c_i \leq B$. This can be easily solved by enumerating through d possible values of i. With this modification, while the rest is the same as Algorithm 1, we have an efficient algorithm which achieves the same regret bound as that in Theorem 3.

For the case with $b \in (1, \infty]$, we now face the knapsack problem: find $\hat{J} \subseteq [d]$ to maximize $\Phi(J) = \sum_{i \in J} |\mathbf{\Theta}_i|^a$ subject to $\sum_{i \in J} c_i \leq B$. Unfortunately, the knapsack problem is known to be NP-hard. Thus, we use an approximation algorithm, which for any given $\epsilon > 0$ finds J with $\Phi(J) \geq (1-\epsilon)\Phi(\hat{J})$ in time $\mathcal{O}(d^3\epsilon^{-1})$ (see e.g. Section 11.8 in Kleinberg and Tardos (2005)). It can be used to find $\mathbf{w}$ with $\Psi(\mathbf{w}) \leq (1-\epsilon)\Psi(\hat{\mathbf{w}})$ (note that $\Psi(\hat{\mathbf{w}}) \leq 0$). Then similarly to Kalai and Vempala (2005), one can show that this results in an extra term of $(1-(1-\epsilon)^T)T$ in the regret. By choosing $\epsilon = 1/T^2$, the regret can be kept in the same order as before, although the time complexity in each round now increases to $\mathcal{O}(d^3T^2)$.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AOARD) under award number No.FA2386-17-1-4038 and from Taiwan Ministry of Science and Technology (MOST) under AI-center of NTU.

References

K. Amin, S. Kale, G. Tesauro, and D. Turaga. Budgeted prediction with expert advice. In *AAAI*, 2015.

J. Audibert, S. Bubeck, and G. Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014.

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

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and Systems Sciences*, 78:1404–1422, 2012.

N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir. Efficient learning with partially observed attributes. *Journal of Machine Learning Research*, 12:2857–2878, 2011.

Nicolo Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Nonstochastic bandits with composite anonymous feedback. *COLT*, 2018.

D. Foster, S. Kale, and H. Karloff. Online sparse linear regression. In *COLT*, 2016.

E. Hazan and T. Koren. Linear regression with limited observation. In *ICML*, 2012.

A. Kalai and S. Vempala. Efficient algorithms for on-line decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

S. Kale. Open problem: Efficient online sparse linear regression. In *COLT*, 2014.

Satyen Kale, Zohar Karnin, Tengyuan Liang, and David Pal. Adaptive feature selection: Computationally efficient online sparse linear regression under rip. *ICML*, 2017.

J. Kleinberg and E. Tardos. Algorithm design. *Addison Wesley*, 2005.

G. Neu and G. Bartók. Importance weighting without importance weights: An efficient algorithm for combinatorial semi-bandits. *Journal of Machine Learning Research*, 17(154):1–21, 2016.

K. Quanrud and D. Khashabi. Online learning with adversarial delays. In *NIPS*, 2015.

J. Wang, P. Zhao, S. Hoi, and R. Jin. Online feature selection and its applications. *IEEE Transactions on Knowledge & Data Engineering*, 26(3):698–710, 2013.

N. Zolghadr, G. Bartók, R. Greiner, A. György, and C. Szepesvári. Online learning with costly features and labels. In *NIPS*, 2013.

Appendix A. Proof of Theorem 7

To bound the regret of the new algorithm, let us compare it with that of Algorithm 1 (in the no-delay setting) using the new choice of η and γ. Let $\mathbf{w}_t$ denote what the new algorithm plays in round t, which is $\hat{\mathbf{w}}_t$ with probability $1-\gamma$ and a random $\mathbf{w} \in \bar{\mathcal{K}}$ having $\|\mathbf{w}\|_0 = k$ with probability γ. Let $\mathbf{w}'_t$ and $\hat{\mathbf{w}}'_t$ denote those of Algorithm 1 corresponding to $\mathbf{w}_t$ and $\hat{\mathbf{w}}_t$. Then the regret difference between these two algorithm is

$$\sum_{t=1}^{T} \mathbb{E}[\langle \mathbf{w}_t - \mathbf{w}'_t, \boldsymbol{\theta}_t \rangle] \leq (1-\gamma) \sum_{t=1}^{T} \mathbb{E}[\langle \hat{\mathbf{w}}_t - \hat{\mathbf{w}}'_t, \boldsymbol{\theta}_t \rangle].$$

Note that the optimization problem for $\hat{\mathbf{w}}_t$, compared to that for $\hat{\mathbf{w}}'_t$, has some $\hat{\boldsymbol{\theta}}_\tau$'s missing due to delays: those with τ in the set $\bar{S}_t \equiv [t-1] \setminus S_t$. Let $\hat{\boldsymbol{\theta}}_{\bar{S}_t} = \sum_{\tau \in \bar{S}_t} \hat{\boldsymbol{\theta}}_\tau$ and let us again use the notation $U_t(\mathbf{z})$ for $\arg\min_{\mathbf{w} \in \mathcal{K}} \left\langle \mathbf{w}, \eta \sum_{\tau=1}^{t-1} \hat{\boldsymbol{\theta}}_\tau - z \right\rangle$. Then for any t, we have the conditional expectation $\mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \hat{\mathbf{w}}'_t, \boldsymbol{\theta}_t \rangle \mid \mathcal{F}_{t-1}\right]$ equal to

$$\int_{\mathbb{R}^d} \langle U_t(\mathbf{z}), \boldsymbol{\theta}_t \rangle df(\mathbf{z}) - \int_{\mathbb{R}^d} \langle U_t(\mathbf{z} + \eta \hat{\boldsymbol{\theta}}_{\bar{S}_t}), \boldsymbol{\theta}_t \rangle df(\mathbf{z}),$$

which, following the proof of Lemma 6, can be upper-bounded by

$$\begin{aligned} &\eta \left\| \hat{\boldsymbol{\theta}}_{\bar{S}_t} \right\|_1 \int_{\mathbb{R}^d} |\langle U_t(\mathbf{z}), \boldsymbol{\theta}_t \rangle| \, df(\mathbf{z}) \\ \leq \quad & \eta \sum_{\tau \in \bar{S}_t} \left\| \hat{\boldsymbol{\theta}}_\tau \right\|_1 \int_{\mathbb{R}^d} |\langle U_t(\mathbf{z}), \boldsymbol{\theta}_t \rangle| \, df(\mathbf{z}) \\ \leq \quad & \eta k^{(b-1)/b} \sum_{\tau \in \bar{S}_t} \|\hat{\boldsymbol{\theta}}_\tau\|_1, \end{aligned}$$

as $\|\boldsymbol{\theta}_t\|_\infty \leq 1$ and $\|U_t(\mathbf{z})\|_1 \leq k^{(b-1)/b}$ with $U_t(\mathbf{z}) \in \mathcal{K}$. Taking the sum over t and the expectation over these $\hat{\boldsymbol{\theta}}_\tau$'s, we obtain

$$\sum_{t=1}^{T} \mathbb{E}\left[\langle \hat{\mathbf{w}}_t - \hat{\mathbf{w}}'_t, \boldsymbol{\theta}_t \rangle\right] \leq \eta k^{(b-1)/b} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{\tau \in \bar{S}_t} \|\hat{\boldsymbol{\theta}}_\tau\|_1\right].$$

Since each $\|\hat{\boldsymbol{\theta}}_\tau\|_1$ is counted at most D_τ times in the last sum and $\mathbb{E}[\|\hat{\boldsymbol{\theta}}_\tau\|_1] \leq \|\boldsymbol{\theta}_\tau\|_1 + d/T \leq 2d$ from Lemma 2, the sum is thus at most $2\eta dk^{(b-1)/b} \sum_{\tau=1}^{T} D_\tau = 2\eta dk^{(b-1)/b} D$.

Thus, we can express the regret of the new algorithm as

$$\sum_{t=1}^{T} \mathbb{E}[\langle \mathbf{w}_t - \mathbf{w}'_t, \boldsymbol{\theta}_t \rangle] + \sum_{t=1}^{T} \mathbb{E}\left[\langle \mathbf{w}'_t - \mathbf{w}_*, \boldsymbol{\theta}_t \rangle\right]$$

where the first sum is at most $2\eta dk^{(b-1)/b} D$ from the discussion above and the second sum is at most

$$\mathcal{O}\left(\gamma k^{(b-1)/b} T + \eta d^2 T + \frac{1}{\eta} k^{(b-1)/b} \ln d\right)$$

according to the proof of Theorem 3. Then the theorem follows with the given choice of η and γ.

Appendix B. Experiments

In addition to our theoretical results, we also perform experiments to verify the effectiveness of our algorithms in practice, which we show next.

First, we compare our Algorithm 1 with two baselines for the constraint set $\mathcal{K}_2$.[1] The first baseline randomly selects a subset of k features before the rounds start, and then it runs the FTPL algorithm using this same subset of features in every round. For the second baseline, we use the best fixed subset of k features, instead of a random one, selected in an offline way. More precisely, we first compute the subset of features which the best offline algorithm would choose, using the algorithm in Lemma 1. Then we use this same subset of features in every round but run FTPL to adapt the predictor dynamically. This is our second baseline, which we call "oracle".

The experiments are conducted on four datasets, all downloaded from the libsvm website. The statistics of the datasets are shown on Table 1. All the datasets except "mnist" have labels in $\{-1, +1\}$. In the "mnist" dataset, the labels are digits from 0 to 9. We choose the difficult 3 vs. 5 classification task, with digit 3 labeled as 1 and digit 5 labeled as -1. Because of the randomness of our algorithms and the baselines, we conduct the experiments five times for each dataset and each budget k, each time with the dataset randomly shuffled. In the experiments, each call to Algorithm 2 is run with $M = 10$ iterations.

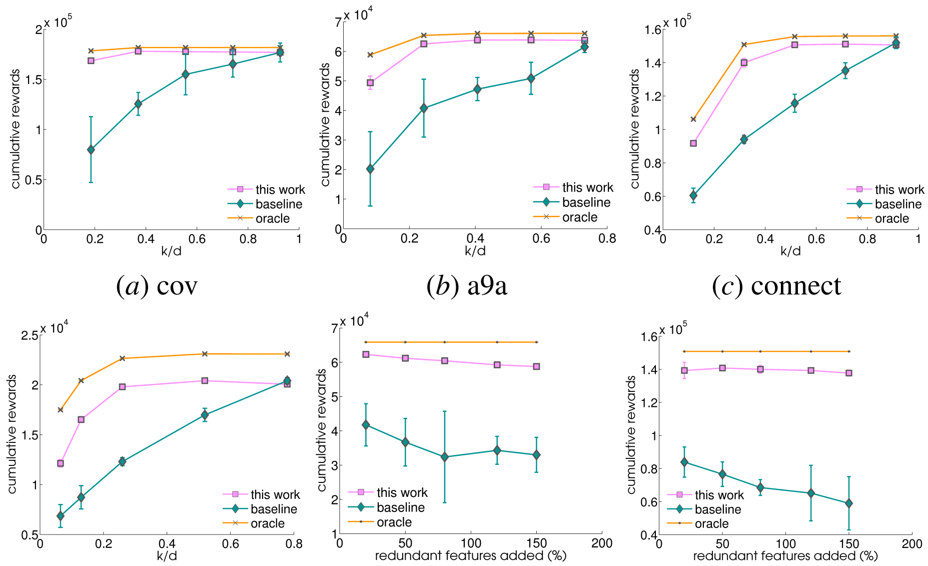
Data	# samples	dimensions
a9a	48,842	123
connect	61,108	126
cov subset	58,100	54
mnist 3 vs. 5	13,454	770

Table 1: Dataset statistics

Figure 1 show the results. For a better illustration, we use the cumulative rewards $\sum_t r_t = \sum_t y_t \langle \mathbf{w}_t, \mathbf{x}_t \rangle$ as the performance measure instead of the cumulative losses. Figure 1 (a)$\sim$(d) shows

1. We are not aware of any relevant work which we can compare fairly to. In fact, we believe that we are the first to provide efficient algorithms for this problem, as existing algorithms all have runtime of the order of d^k, which would take too long to run even for moderate values of d and k.

the graphs of the cumulative rewards v.s. the k/d ratios. Smaller ratios of k/d mean more stringent budget constraints. One can see that our algorithm substantially outperforms the baseline in a wide middle range of k/d ratios. The difference becomes smaller when k/d approaches to 1, as the setting goes toward the full information one. Note that "oracle" has the advantage of an offline algorithm, which can select the subset of features based on all the reward functions, while it is allowed to adapt its predictor through time instead of being constrained to a fixed one. Thus, it may seem unfair to compare with such an algorithm, but our experiments show that our algorithm is still competitive to it. Another experiment is conducted to see the effect of redundant features. Here, redundant features are sampled from the standard normal distribution and then added to the data. The numbers of redundant features are set to be $0.2 \sim 1.5$ times of the original ones. Figure 1 (e)$\sim$(f) show that our algorithm is robust to the redundant features while the baseline (using a random subset) degrades as more redundant features are added.



(a) cov (b) a9a (c) connect

(d) mnist (e) redundant a9a (k=40) (f) redundant connect (k=40)

Figure 1: Performance of FOLLOW THE PERTURBED SPARSE LEADERS. (a)$\sim$(d): Cumulative rewards vs. k/d ratios. (e) and (f): Cumulative rewards vs. number of redundant features added.