Journal of Machine Learning Research 1 (2000) 1-48 Submitted 4/00; Published 10/00

Expected Regret and Pseudo-Regret are Equivalent When the Optimal Arm is Unique

Daron Anderson ANDERSD3@TCD.IE
Department of Computer Science and Statistics
Trinity College Dublin
Ireland

Douglas Leith DOUG.LEITH@TCD.IE
Department of Computer Science and Statistics
Trinity College Dublin
Ireland

April 2022

Abstract

In online linear optimisation with stochastic losses it is common to bound the pseudo-regret of an algorithm rather than the expected regret. This is attributed to the expected fluctuations for i.i.d sums making expected regret bounds better than $\Omega(\sqrt{T})$ impossible. In this paper we show that when there is a unique optimal action and the action set X is a polytope the difference between pseudo-regret and expected regret is $o(1)$. This means that the existing upper bounds on pseudo-regret in the literature can immediately be extended to also upper bound the expected regret. Our results are independent of the algorithm used to select the actions and apply equally to the bandit and full-information settings.

1. Introduction

In online linear optimisation we choose an action x_t from the domain $X \subset \mathbb{R}^d$ and then experience loss $\ell_t \cdot x_t$. In the full information setting we observe the full loss vector $\ell_t \in \mathbb{R}^d$ after selecting action x_t, while in the bandit setting we only observe the loss $\ell_t \cdot x_t \in \mathbb{R}$ (which gives a natural generalization of multi-armed bandits).When the loss vectors are stochastic then we would like to minimise the *expected regret* $\mathbb{E}[R_T] = \mathbb{E}[\sum_{t=1}^T \ell_t \cdot x_t - \sum_{t=1}^T \ell_t \cdot x_T^*]$, where the expectation is over the loss vectors. Here, $x_T^* \in \arg\min\{\sum_{t=1}^T \ell_t \cdot x : x \in X\}$ depends on the realisation $\ell_1, \ell_2, \ldots, \ell_T$ of the loss vectors.

Unfortunately, the expected regret can be $\Omega\big(\sqrt{T}\big)$ and so many classical performance bounds are instead expressed in terms of the *pseudo-regret* $\mathbb{P}[R_T] = \mathbb{E}[\sum_{t=1}^T \ell_t \cdot x_t] - T\mathbb{E}[\ell_t]y^*$ with $y^* \in \arg\min\{\mathbb{E}[\ell_t]\cdot x : x \in X\}$. For example, early work by Dani et al. (2008) established an $O(\log(T)/\Delta)$ upper bound on the pseudo-regret with bandit feedback, where Δ is the difference between the smallest and second smallest elements of $\mathbb{E}[\ell_t]$. This bound was subsequently improved by Abbasi-Yadkori et al. (2011) while Auer and Chiang (2016) shows that a player can simultaneously achieve $O(\log(T)/\Delta)$ pseudo-regret for stochastic losses and $O(\sqrt{T})$ pseudo-regret for adversarial losses. In

the full information setting, Anderson and Leith (2020a); Mourtada and Gaïffas (2019) establish an $O(1/\Delta)$ upper bound on the pseudo-regret when the loss vectors are stochastic and X is the simplex.

Since the benchmark y^* used in the pseudo-regret minimises the average loss $\mathbb{E}[\ell_t] \cdot x$, $x \in X$ it does not depend on the specific realisation $\ell_1, \ell_2, \ldots, \ell_T$. This simplifies analysis but comes at the price that $\mathbb{P}[R_T] \leq \mathbb{E}[R_T]$. Hence, an upper bound on pseudo-regret $\mathbb{P}[R_T]$ need not imply an upper bound on expected regret $\mathbb{E}[R_T]$. Indeed, it was often believed in the bandit literature that the expected regret and pseudo-regret differ by quantities of order $O(\sqrt{T})$. In this paper we show that in the favourable case of a unique optimal action and action set X a polytope this difference vanishes i.e. $0 \leq \mathbb{E}[R_T] - \mathbb{P}[R_T] \leq o(1)$. This means that the existing upper bounds on pseudo-regret in the literature can immediately be extended to also upper bound the expected regret. Our results are independent of the algorithm used to select the actions and apply equally to the bandit and full-information settings.

1.1 Previous Bounds on Expected Regret

While most of the stochastic online linear optimisation literature suggests that useful bounds on the expected regret are impossible and so focuses on the pseudo-regret, there are some notable exceptions where bounds on expected regret are indeed obtained. Gaillard et al. (2014) consider a new variant of the Prod algorithm on the simplex, with different learning rates for each arm and full information feedback. Their Theorem 11 shows that for i.i.d cost vectors the algorithm gives $\mathbb{E}[R_T] \leq O(1)$ provided the expected cost vector has a unique minimiser. Corollary 11 of Huang et al. (2017) says that the Follow-The-Leader algorithm with full information feedback gives $\mathbb{E}[R_T] \leq O(1)$ when the cost vectors are i.i.d, the action domain is a polytope and the expected cost vector has a unique minimiser. For strongly convex domains their Theorem 5 implies that for i.i.d cost vectors running Follow-The-Leader gives $\mathbb{E}[R_T] \leq O(\log T)$. Wei and Luo (2018) consider problems on the simplex under bandit feedback. In Section 4.2 they give an algorithm with the standard $O(\sqrt{T \log T})$ regret bound against adaptive adversaries, which specialises to give $\mathbb{E}[R_T] \leq O(\log(T)/\Delta)$ in the case where the cost vectors satisfy the martingale [1] type inequality $\mathbb{E}[a_t(j) - a_t(j^*) \,|\, a_1, \ldots, a_{t-1}] > \Delta$ for some $\Delta > 0$ and fixed j^* and all $j \neq j^*$. For strongly-convex cost functions it is well-known we can get $O(\log T)$ regret by running online gradient descent (Hazan et al., 2007) or follow-the-leader (Cesa-Bianchi and Lugosi (2006) Section 3.2). If the functions are generated by an i.i.d sequence we immediately have $\mathbb{E}[R_N] \leq O(\log T)$.

The existing literature therefore creates the confusing situation where on the one hand most papers say that obtaining bounds on the expected regret is impossible while on the other hand a few papers do manage to bound the expected regret but they do this without commenting on whether this is surprising or not and without dwelling on the assumptions. The present paper is a first step towards clarifying the situation.

1. Martingales are slightly more general than i.i.d sequences. Many of the common concentration inequalities used for i.i.d sequences generalise immediately to martingales.

2. Notation

For vectors $x \in \mathbb{R}^d$ we write $\|x\|$ for the Euclidean norm. By a *polytope* we mean the convex hull of a finite set. For every polytope $\mathcal{P}$ there exists a unique finite set $\mathcal{V}$ of *vertices* such that $\mathcal{P}$ is the convex hull of $\mathcal{V}$ and not the convex hull of any proper subset of $\mathcal{V}$. For any convex set $X \subset \mathbb{R}^d$ and $v \in X$ the *normal cone* is the set $N_X(v) = \{u \in \mathbb{R}^d : u \cdot (x - v) \leq 0 \text{ for all } x \in X\}$. For any $w \in \mathbb{R}^d$ we have $-w \in N_X(v)$ if and only if v is a minimiser of $w \cdot x$ over X.

Throughout $\ell_1, \ell_2, \ldots \in \mathbb{R}^d$ is an i.i.d sequence of cost vectors with $\mathbb{E}[\ell_t] = \bar{\ell}$. The point $y^* \in \arg\min\{\bar{\ell} \cdot x : x \in X\}$ minimises the expected cost vector and $x_T^* \in \arg\min\left\{\sum_{t=1}^T \ell_t \cdot x : x \in X\right\}$ minimises the realised cost vectors. Note y^* is fixed but x_T^* varies between instances of the problem. Since a linear function on a polytope is minimised at a vertex we can always assume x_T^* and y^* are vertices when X is a polytope. The expected regret is:

$$\mathbb{E}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot x_t - \min_{x \in X} \sum_{t=1}^T \ell_t \cdot x\right] = \mathbb{E}\left[\max_{x \in X}\left\{\sum_{t=1}^T \ell_t \cdot x_t - \sum_{t=1}^T \ell_t \cdot x\right\}\right]$$

and the pseudo-regret is:

$$\mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot x_t\right] - \min_{x \in X} \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot x\right] = \max_{x \in X} \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot x_t - \sum_{t=1}^T \ell_t \cdot x\right]$$

By Jensen's inquality, $\mathbb{E}[R_T] \geq \mathbb{P}[R_T]$.

3. Preliminaries

Lemma 1 shows that, even though $\mathbb{E}[R_T]$ and $\mathbb{E}[P_T]$ are defined in terms of on the action sequence, the gap $\mathbb{E}[R_T] - \mathbb{E}[P_T]$ depends only on the cost vectors:

Lemma 1 We have $\mathbb{E}[R_T] - \mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*)\right]$.

Proof By definition $\mathbb{E}[R_T] - \mathbb{P}[R_T]$ equals

$$\mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot (x_t - x_T^*)\right] - \mathbb{E}\left[\sum_{t=1}^T \bar{\ell} \cdot (x_t - y^*)\right] = \mathbb{E}\left[\sum_{t=1}^T (\ell_t - \bar{\ell}) \cdot x_t\right] + \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*)\right]$$

To see the first part is zero, recall each action x_{n+1} is chosen based only on $\ell_1, \ldots, \ell_n$. Hence x_{n+1} is a function of $\ell_1, \ldots, \ell_n$ and since all $\ell_1, \ell_2, \ldots$ are independent we see each x_t is independent of ℓ_t and independent of $\ell_t - \bar{\ell}$. Thus the expectation distributes over the product and we have

$$\mathbb{E}\left[\sum_{t=1}^T (\ell_t - \bar{\ell}) \cdot x_t\right] = \sum_{t=1}^T \mathbb{E}\big[(\ell_t - \bar{\ell}) \cdot x_t\big] = \sum_{t=1}^T \mathbb{E}\big[\ell_t - \bar{\ell}\big] \cdot \mathbb{E}[x_t] = \sum_{t=1}^T \big(\mathbb{E}[\ell_t] - \bar{\ell}\big) \cdot \mathbb{E}[x_t].$$

Since each $\mathbb{E}[\ell_t] = \bar{\ell}$ the above is zero. ■

Next we use Lemma 1 to bound the gap $\mathbb{E}[R_T] - \mathbb{P}[R_T]$ in terms of the probability that the realised minimiser coincides with the expected minimiser:

Lemma 2 For $D_1 = \max\{\|x-y\|_1 : x,y \in X\}$ and $D = \max\{\|x-y\| : x,y \in X\}$ we have the bounds

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq \sqrt{\mathbb{E}\|x_T^* - y^*\|^2}\sqrt{\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|^2} \leq D\sqrt{P(x_T^* \neq y^*)}\sqrt{\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|^2}$$

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq \sqrt{\mathbb{E}\|x_T^* - y^*\|_1^2}\sqrt{\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|_\infty^2} \leq D_1\sqrt{P(x_T^* \neq y^*)}\sqrt{\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|_\infty^2}.$$

Proof Lemma 1 gives $\mathbb{E}[R_T] - \mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*)\right]$. The right-hand-side can be written as $\mathbb{E}\left[\sum_{t=1}^T \bar{\ell}\cdot(y^* - x_T^*)\right] + \mathbb{E}\left[\sum_{t=1}^T(\ell_t - \bar{\ell})\cdot(y^* - x_T^*)\right]$. Since y^* minimises $\bar{\ell}\cdot x$ the first term is nonpositive and we get

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq \mathbb{E}\left[(y^* - x_T^*)\cdot\sum_{t=1}^T(\ell_t - \bar{\ell})\right] \leq \mathbb{E}\left[\|y^* - x_T^*\|\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|\right] \tag{1}$$

By Cauchy-Schwarz for the L^2 inner product $(f,g) = \mathbb{E}_\omega[f(\omega)g(\omega)]$ we have

$$\mathbb{E}\left[\|y^* - x_T^*\|\sum_{t=1}^T\|\ell_t - \bar{\ell}\|\right] \leq \sqrt{\mathbb{E}\|y^* - x_T^*\|^2}\sqrt{\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|^2}.$$

The random variable $\|x_T^* - y^*\|^2$ is at most D^2 and equals zero for $x_T^* = y^*$. Hence the expectation is at most $D^2P(x_T^* \neq y^*)$ and we get $\sqrt{\mathbb{E}\|y^* - x_T^*\|^2} \leq D\sqrt{P(x_T^* \neq y^*)}$. This proves the first bound. The proof for the second bound is similar except on line (1) we use $(y^* - x_T^*)\cdot\sum_{t=1}^T(\ell_t - \bar{\ell}) \leq \|y^* - x_T^*\|_1\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|_\infty$ and proceed as before. ■

The following vector concentration inequality is a special case of Pinelis (1994) Theorem 3.5:

Theorem 1 Suppose $X_1, X_2, \ldots \in \mathbb{R}^d$ are independent random variables with expectation zero and each $\|X_n\| \leq B$. For each $\epsilon \geq 0$ we have

$$P\left(\left\|\sum_{t=1}^T X_T\right\| > \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2TB^2}\right)$$

4. Polytopes

Henceforth $\mathcal{P} \subset \mathbb{R}^d$ is a polytope. The vertices are $\{v_1, v_2, \ldots, v_V\}$. The cost vectors are $\ell_1, \ell_2, \ldots$ with expectation $\bar{\ell}$. We assume $y^* = v_1$ is the unique minimiser of $\bar{\ell}$. This is a notational convenience and can always be achieved by permuting the component labels. For $j \leq V$ write $\Delta_j = \bar{\ell}\cdot(v_j - v_1)$ and $\Delta = \min\{\Delta_j : j > 1\}$.

Define $L_2 = \max\{\|\ell_t\| : t \in \mathbb{N}\}$ and $B_2 = \max\{\|\ell_t - \bar{\ell}\| : t \in \mathbb{N}\}$ almost surely. Write $D = \max\{\|x-y\| : x,y \in \mathcal{P}\}$ for the diameter of the domain. Our main theorem is the following:

Theorem 2 Suppose the cost vectors $\ell_1, \ell_2, \ldots$ satisfy the bounds above and the expected cost vector $\bar{\ell}$ has unique minimiser v_1. Then we have the bounds

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq o(1) \qquad \mathbb{E}[R_T] - \mathbb{P}[R_T] \leq 25D^2B_2^2/\Delta.$$

We prove Theorem 2 in several stages. Recall that since v_1 is the unique minimiser $-\bar{\ell}$ is normal to $\mathcal{P}$ at v_1 but not normal at $v_2, \dots, v_V$. More specifically, we have that:

Lemma 3 *(Anderson and Leith, 2020b, Lemma 12)* For each $j = 2, 3, \dots, V$ we have

$$\theta_j = \min\left\{ \frac{\bar{\ell} \cdot u}{\|\bar{\ell}\|\|u\|} : u \in N_{\mathcal{P}}(v_j) \right\} \geq \frac{1/2}{1 + D^2\|\bar{\ell}\|^2/\Delta_j^2} - 1$$

where $N_{\mathcal{P}}(v_j)$ is the normal cone to $\mathcal{P}$ at v_j.

Geometrically speaking, since $\frac{\bar{\ell}\cdot u}{\|\bar{\ell}\|\|u\|}$ is the cosine of the angle between $\bar{\ell}$ and u, Lemma 3 bounds how far $-\bar{\ell}$ is from being normal at v_j in terms of the gap Δ_j.

Next we derive a condition on the noise that makes $\sum_{t=1}^T \ell_t$ have v_1 as unique minimiser.:

Lemma 4 Suppose $\Big\| \sum_{t=1}^T (\ell_t - \bar{\ell}) \Big\| < \frac{T\|\bar{\ell}\|}{5}\sqrt{2(\min\theta_j + 1)}$. Then we have $x_T^* = v_1$.

Proof Since x_T^* minimises $w = \sum_{t=1}^T \ell_t$ we can assume $x_T^* = v_j$ for some $j \in \{1, \dots, V\}$. Hence $u = -w$ is in the normal cone $N_{\mathcal{P}}(v_j)$. We claim $\frac{\bar{\ell}\cdot u}{\|\bar{\ell}\|\|u\|} < \theta_j$ for $j \neq 1$ and so, by Lemma 3, $j = 1$. To that end write $\frac{\bar{\ell}\cdot u}{\|\bar{\ell}\|\|u\|} = \frac{1}{2}\big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{w}{\|w\|}\big\|^2 - 1$ to see

$$\frac{\bar{\ell}\cdot u}{\|\bar{\ell}\|\|u\|} < \theta_j \iff \frac{1}{2}\Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{w}{\|w\|}\Big\|^2 - 1 < \theta_j \iff \Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{w}{\|w\|}\Big\| < \sqrt{2(\theta_j + 1)}. \tag{2}$$

To prove the right-hand-side of (2) write $w = T\bar{\ell} - E$ for $E = \sum_{t=1}^T (\bar{\ell} - \ell_t)$ to get

$$\begin{aligned}
\Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{w}{\|w\|}\Big\| &= \Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{T\bar{\ell} - E}{\|T\bar{\ell} - E\|}\Big\| = \Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{T\bar{\ell}}{\|T\bar{\ell} - E\|} + \frac{E}{\|T\bar{\ell} - E\|}\Big\| \\
&\leq \Big\|\frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{T\bar{\ell}}{\|T\bar{\ell} - E\|}\Big\| + \Big\|\frac{E}{\|t\bar{\ell} - E\|}\Big\| = \|\bar{\ell}\|\Big|\frac{1}{\|\bar{\ell}\|} - \frac{1}{\|\bar{\ell} - E/T\|}\Big| + \frac{\|E\|}{\|T\bar{\ell} - E\|}
\end{aligned} \tag{3}$$

By assumption $\|E\| < (T\|\bar{\ell}\|/5)\sqrt{2(\min\theta_j + 1)}$. Since $\frac{\bar{\ell}\cdot u}{\|\bar{\ell}\|\|u\|}$ is the cosine of the angle between $\bar{\ell}$ and u we have $\min\theta_j \in [0, 1]$ and $\sqrt{2(\min\theta_j + 1)} \leq 2$ and $\|E\| < 2T\|\bar{\ell}\|/5$. Hence $\|T\bar{\ell} - E\| \geq T\|\bar{\ell}\| - \|E\| \geq 3T\|\bar{\ell}\|/5$ and the second term of (3) has

$$\frac{\|E\|}{\|T\bar{\ell} - E\|} \leq \frac{\|E\|}{3T\|\bar{\ell}\|/5} \leq \frac{(T\|\bar{\ell}\|/5)\sqrt{2(\min\theta_j + 1)}}{3T\|\bar{\ell}\|/5} \leq \frac{\sqrt{2(\min\theta_j + 1)}}{3}$$

By convexity the first term of (3) has

$$\|\bar{\ell}\|\Big|\frac{1}{\|\bar{\ell}\|} - \frac{1}{\|\bar{\ell} - E/T\|}\Big| \leq \|\bar{\ell}\|\frac{|\|\bar{\ell}\| - \|\bar{\ell} - E/T\||}{\min\{\|\bar{\ell}\|^2, \|\bar{\ell} - E/T\|^2\}} \leq \|\bar{\ell}\|\frac{\|E/T\|}{\min\{\|\bar{\ell}\|^2, \|\bar{\ell} - E/T\|^2\}}$$

Since $\|E/T\| \leq 3\|\bar{\ell}\|/5$ the denominator is at least $(3\|\bar{\ell}|/5)^2$ and we get

$$\|\bar{\ell}\|\Big|\frac{1}{\|\bar{\ell}\|} - \frac{1}{\|\bar{\ell} - E/T\|}\Big| \leq \|\bar{\ell}\|\frac{\|E/T\|}{(3\|\bar{\ell}\|/5)^2} = \|\bar{\ell}\|\frac{T\|E\|}{(3T\|\bar{\ell}\|/5)^2} = \frac{25}{9}\frac{\|E\|}{T\|\bar{\ell}\|} \leq \frac{5}{9}\sqrt{2(\min\theta_j + 1)}$$

Combine the bounds for the terms of (3) to get

$$\left\| \frac{\bar{\ell}}{\|\bar{\ell}\|} - \frac{w}{\|w\|} \right\| \leq \left(\frac{1}{3} + \frac{5}{9} \right) \sqrt{2(\min \theta_j + 1)} \leq \sqrt{2(\min \theta_j + 1)} \leq \sqrt{2(\theta_j + 1)}.$$

This proves the right-hand-side of (2) as required. ■

Proof of Theorem 2 Lemma 2 says

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq D \sqrt{P(x_T^* \neq v_1)} \sqrt{\mathbb{E} \left\| \sum_{t=1}^T (\ell_t - \bar{\ell}) \right\|^2} \tag{4}$$

To bound the second factor $\sqrt{\mathbb{E} \left\| \sum_{t=1}^T (\ell_t - \bar{\ell}) \right\|^2}$ use Theorem 1 to get $P\left(\left\| \sum_{t=1}^T (\bar{\ell} - \ell_t) \right\| > \epsilon \right) \leq 2 \exp\left(-\epsilon^2 / 2TB_2^2 \right)$ for each $\epsilon \geq 0$. Thus $P\left(\left\| \sum_{t=1}^T (\bar{\ell} - \ell_t) \right\|^2 > \epsilon \right) \leq 2 \exp\left(-\epsilon / 2TB_2^2 \right)$ and Lemma 8 (see Appendix) gives

$$\mathbb{E}\left[\left\| \sum_{t=1}^T (\bar{\ell} - \ell_t) \right\|^2 \right] \leq \int_0^\infty P\left(\left\| \sum_{t=1}^T (\bar{\ell} - \ell_t) \right\|^2 > \epsilon \right) dt \leq 2 \int_0^\infty \exp\left(-\frac{\epsilon}{2TB_2^2} \right) dt = 4TB_2^2$$

and so $\sqrt{\mathbb{E} \left\| \sum_{t=1}^T (\ell_t - \bar{\ell}) \right\|^2} \leq 2\sqrt{T} B_2$.

To bound the first factor Lemma 4 says

$$P(x_T^* \neq v_1) \leq P\left(\left\| \sum_{t=1}^T (\ell_t - \bar{\ell}) \right\| \geq \frac{T\|\bar{\ell}\|}{5} \sqrt{2(\min \theta_j + 1)} \right).$$

Theorem 1 says the right-hand-side is at most

$$2 \exp\left(- \left(\frac{T\|\bar{\ell}\|}{5} \sqrt{2(\min \theta_j + 1)} \right)^2 \frac{1}{2TB_2^2} \right) = 2 \exp\left(- \frac{\|\bar{\ell}\|^2 (\min \theta_j + 1)}{25 B_2^2} T \right).$$

To bound the exponent use Lemma 3 to get

$$\theta_j + 1 \geq \frac{1/2}{1 + D^2 \|\bar{\ell}\|^2 / \Delta_j^2} \geq \frac{1/2}{1 + D^2 \|\bar{\ell}\|^2 / \Delta^2}$$

and so

$$P(x_T^* \neq v_1) \leq 2 \exp\left(\frac{1}{1 + D^2 \|\bar{\ell}\|^2 / \Delta^2} \frac{\|\bar{\ell}\|^2}{50 B_2^2} T \right).$$

Now combine the bounds for the factors of (4) to get

$$\begin{aligned} \mathbb{E}[R_T] - \mathbb{P}[R_T] &\leq 2\sqrt{2} D B_2 \sqrt{T} \exp\left(\frac{1}{1 + D^2 \|\bar{\ell}\|^2 / \Delta^2} \frac{\|\bar{\ell}\|^2}{100 B_2^2} T \right) \\ &= 2\sqrt{2} D B_2 \sqrt{T} e^{-AT} \text{ for } A = \frac{1}{1 + D^2 \|\bar{\ell}\|^2 / \Delta^2} \frac{\|\bar{\ell}\|^2}{100 B_2^2}. \end{aligned}$$

Lemma 7 (see Appendix) says $\sqrt{T}e^{-AT} \leq 1/\sqrt{2eA}$ and the above gives

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq \frac{2\sqrt{2}DB_2}{\sqrt{2eA}} = \frac{2DB_2}{\sqrt{e}}\sqrt{\left(1 + \frac{D^2\|\bar{\ell}\|^2}{\Delta^2}\right)\frac{100B_2^2}{\|\bar{\ell}\|^2}}$$
$$= \frac{20DB_2^2}{\sqrt{e}}\sqrt{\frac{1}{\|\bar{\ell}\|^2} + \frac{D^2}{\Delta^2}} \leq \frac{20DB_2^2}{\sqrt{e}}\left(\frac{1}{\|\bar{\ell}\|} + \frac{D}{\Delta}\right) = \frac{20}{\sqrt{e}}\frac{D^2B_2^2}{\Delta} + \frac{20DB_2^2}{\sqrt{e}\|\bar{\ell}\|}.$$

To put the second term in the correct form recall $\Delta = \bar{\ell} \cdot (v_j - v_1)$ for some $j > 1$. Hence $\Delta \leq \|\bar{\ell}\|\|v_j - v_1\| \leq D\|\bar{\ell}\|$ and so $1/\|\bar{\ell}\| \leq D/\Delta$ and the second term is at most $2D^2B_2^2/\sqrt{e}\Delta$. Hence the above gives

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \leq \frac{40D^2B_2^2}{\sqrt{e}\Delta} \leq 25\frac{D^2B_2^2}{\Delta}.$$

as claimed. ■

5. Examples

5.1 A Unique Minimizer is Necessary

The following example demonstrates that the assumption of a unique minimiser is necessary in the sense that if multiple minimisers are allowed $\mathbb{E}[R_T] - \mathbb{P}[R_T]$ might be $\Omega(\sqrt{T})$ i.e. same size as the expected fluctuation $\mathbb{E}\left\|\sum_{t=1}^T(\ell_t - \bar{\ell})\right\|$ between the actual and expected cost vectors.

The *d-simplex* is the set $\mathcal{S} = \{x \in \mathbb{R}^d : \mathbf{1} \cdot x = 1 \text{ and all } x(j) \geq 0\}$ for $\mathbf{1} = (1, 1, \ldots, 1)$. We write $e_1, e_2, \ldots, e_d \in \mathbb{R}^d$ for the standard basis vectors.

Lemma 5 There exists an i.i.d sequence $\ell_1, \ell_2, \ldots \in \mathbb{R}^3$ of cost vectors with $\bar{\ell} = (2, 0, 0)$ such that when the domain is the 3-simplex we have $\mathbb{E}[R_T] - \mathbb{P}[R_T] \geq \sqrt{T}/10$ for T sufficiently large.

Proof Let $B_1, B_2, \ldots$ be i.i.d with each $P(B_t = 1) = P(B_t = -1) = 1/2$. Then for $\ell_t = (2, 0, B_t)$ we have $\bar{\ell} = (2, 0, 0)$. By Lemma 1 we have $\mathbb{E}[R_T] - \mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*)\right]$ where y^* minimises $\bar{\ell}$ and x_T^* minimises $\sum_{t=1}^T \ell_t$ over the simplex. In particular take $y^* = e_2$. In the event that $\frac{1}{\sqrt{T}}\sum_{t=1}^T B_t < -1$ we have $x_T^* = e_3$ and so

$$\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*) = \sum_{t=1}^T \ell_t \cdot (e_2 - e_3) = -\sum_{t=1}^T \ell_t(3) = -\sum_{t=1}^T B_t \geq \sqrt{T}.$$

Thus we get $P\left(\sum_{t=1}^T \ell_t \cdot (y^* - x_T^*) \geq \sqrt{T}\right) \geq P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T B_t < -1\right)$. By the central limit theorem $\frac{1}{\sqrt{T}}\sum_{t=1}^T B_t$ converges in law to a $\mathcal{N}(0, 1)$ distribution as $T \to \infty$. Hence for $X \sim \mathcal{N}(0, 1)$ we have $P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T B_t < -1\right) \to P(X < -1)$. The right-hand-side can be computed numerically as

$0.158655... \geq 0.15$. Hence for T sufficiently large we have $P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} B_t < -1\right) > 1/10$ and so $P\left(\sum_{t=1}^{T} \ell_t \cdot (y^* - x_T^*) \geq \sqrt{T}\right) \geq 1/10$.

By definition of x_T^* and y^* the random variable $\sum_{t=1}^{T} \ell_t \cdot (y^* - x_T^*)$ is nonnegative. Hence the above gives $\mathbb{E}[R_T] - \mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t \cdot (y^* - x_T^*)\right] \geq \sqrt{T}/10$.

■

5.2 A Unique Minimizer Is Not Sufficient

While the assumption of a unique minimizer is necessary for $\mathbb{E}[R_N] - \mathbb{P}[R_N]$ to be small, the following example shows that it is not sufficient i.e. additional conditions, such as a restriction on the class of domains X, are needed.

Lemma 6 Let $\varepsilon > 0$ be arbitrary. There exists a domain and i.i.d sequence of cost vectors with $\mathbb{E}[R_T] - \mathbb{P}[R_T] \geq \dfrac{T^{\frac{1}{2}-\varepsilon}}{20}$ for T sufficiently large.

Proof For some $\alpha \geq 3$ with $\frac{1}{2(\alpha-1)} < \varepsilon$ let the domain $X \subset [-1,1] \times [0,1]$ be $X = X_1 \cap X_2$ for $X_1 = \{(x,y) \in \mathbb{R}^2 : y \geq |x|^\alpha\}$ and $X_2 = \{(x,y) \in \mathbb{R}^2 : y \leq 1\}$. Let the cost vectors be $\ell_t = (B_t, 1)$ for $B_1, B_2, \ldots$ i.i.d with each $P(B_t = 1) = P(B_t = -1) = 1/2$. Write $A_T = \sum_{t=1}^{T} \ell_t = (B(T), T)$ for $B(T) = B_1 + \ldots + B_T$.

Lemma 1 says $\mathbb{E}[R_T] - \mathbb{P}[R_T] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t \cdot (y^* - x_T^*)\right]$ where y^* minimises $\bar{\ell}$ and x_T^* minimises A_T. Since $\bar{\ell} = (0,0)$ we have $y^* = (0,0)$ and it is enough to show $-\mathbb{E}\left[\sum_{t=1}^{T} \ell_t \cdot x_T^*\right]$ is large. We claim x_T^* is the point on the graph

$$x^* = (x, |x|^\alpha) = \left(-\sigma\left(\frac{|B(T)|}{\alpha T}\right)^{\frac{1}{\alpha-1}}, \left(\frac{|B(T)|}{\alpha T}\right)^{\frac{\alpha}{\alpha-1}}\right) \qquad \sigma = \operatorname{sign}\big(B(T)\big) \tag{5}$$

Since $|x|^\alpha$ is strictly convex so is the set X_1. It follows for each $|x| < 1$ and normal direction N to X at the point $(x, |x|^\alpha)$ there is no other point where N is the normal direction. Hence it is enough to show A_T is normal inwards at the point above. The slope at the point $(x, |x|^\alpha)$ of the graph is $-\sigma\alpha|x|^{\alpha-1}$ and the inward normal is along $(\sigma\alpha|x|^{\alpha-1}, 1)$. Rescale to get $(\sigma T\alpha|x|^{\alpha-1}, T)$. For the x value in (5) we have $(\sigma T\alpha|x|^{\alpha-1}, T) = (B(T), T) = A_T$ as claimed.

Next use (5) to compute

$$\begin{aligned}
\sum_{t=1}^{T} \ell_t \cdot x_T^* &= \binom{B(T)}{T} \left(-\sigma \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{1}{\alpha-1}}, \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{\alpha}{\alpha-1}} \right) \\
&= -\sigma B(T) \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{1}{\alpha-1}} + T \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{\alpha}{\alpha-1}} \\
&= -|B(T)| \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{1}{\alpha-1}} + T \left(\frac{|B(T)|}{\alpha T} \right)^{\frac{\alpha}{\alpha-1}} \\
&= -\alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha}} |B(T)|^{1+\frac{1}{\alpha-1}} + \alpha^{\frac{\alpha}{1-\alpha}} T^{1+\frac{\alpha}{1-\alpha}} |B(T)|^{\frac{\alpha}{\alpha-1}} \\
&= -\alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha}} |B(T)|^{\frac{\alpha}{\alpha-1}} + \alpha^{\frac{\alpha}{1-\alpha}} T^{\frac{1}{1-\alpha}} |B(T)|^{\frac{\alpha}{\alpha-1}} \\
&= -\left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha}} |B(T)|^{\frac{\alpha}{\alpha-1}} \qquad (6)
\end{aligned}$$

In Lemma 5 we showed $P\left(B(T) \leq -\sqrt{T}\right) > 1/10$ for T sufficiently high. When this occurs we have

$$\begin{aligned}
-\sum_{t=1}^{T} \ell_t \cdot x_T^* = \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha}} |B(T)|^{\frac{\alpha}{\alpha-1}} &\geq \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha}} T^{\frac{\alpha}{2(\alpha-1)}} \\
&= \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{1-\alpha} + \frac{\alpha}{2(\alpha-1)}}
\end{aligned}$$

To simplify the exponent write

$$\frac{1}{1-\alpha} + \frac{\alpha}{2(\alpha-1)} = \frac{1}{2(1-\alpha)} + \frac{1}{2(1-\alpha)} - \frac{\alpha}{2(1-\alpha)} = \frac{1}{2(1-\alpha)} + \frac{1-\alpha}{2(1-\alpha)} = \frac{1}{2} - \frac{1}{2(\alpha-1)}$$

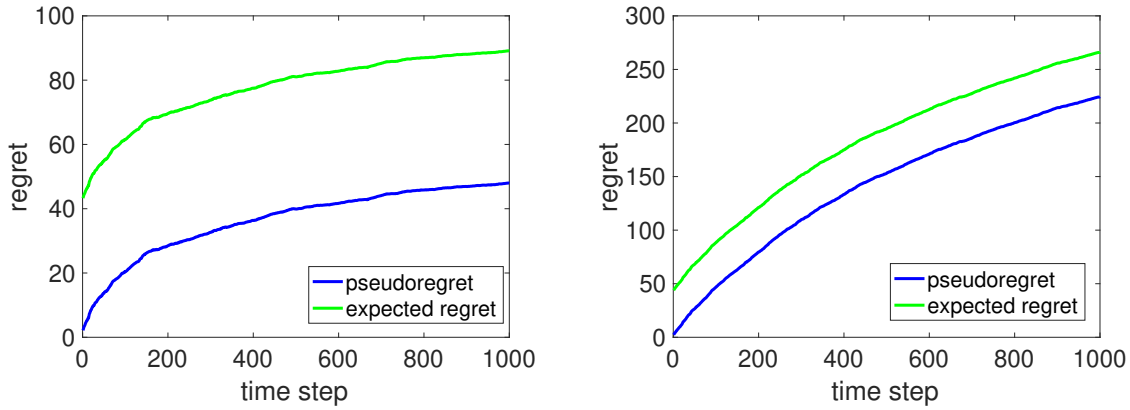
Hence with probability at least $1/10$ we have

$$-\sum_{t=1}^{T} \ell_t \cdot x_T^* \geq \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{2} - \frac{1}{2(\alpha-1)}} \geq \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{2}-\varepsilon}$$

Since $-\sum_{t=1}^{T} \ell_t \cdot x_T^* = -\sum_{t=1}^{T} \ell_t \cdot (y^* - x_T^*)$ is nonnegative we get

$$\mathbb{E}[R_T] - \mathbb{P}[R_T] \geq \frac{1}{10} \left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} T^{\frac{1}{2}-\varepsilon}$$

To see $\alpha^{\frac{1}{1-\alpha}} \to 1$ write $\log \alpha^{\frac{1}{1-\alpha}} = \frac{\log \alpha}{1-\alpha} = -\frac{\log \alpha}{\alpha-1}$. Since $\alpha - 1$ grows faster than $\log \alpha$ we have $\log \alpha^{\frac{1}{1-\alpha}} \to 0$ and $\alpha^{\frac{1}{1-\alpha}} \to 1$. Thus for α sufficiently large we have $\frac{1}{10}\left(1 - \frac{1}{\alpha}\right) \alpha^{\frac{1}{1-\alpha}} > \frac{1}{20}$. For such α we get $\mathbb{E}[R_T] - \mathbb{P}[R_T] \geq \frac{T^{\frac{1}{2}-\varepsilon}}{20}$ as required. ■



(a) Bandit Feedback

(b) Full Information Feedback

Figure 1: Example illustrating constant gap between expected regret and pseudo-regret.

5.3 Numerical Example

We present a brief example illustrating the application of our results. In the usual multi-arm bandit setup the loss incurred by taking action x at time t is $\bar{\ell} \cdot x + \eta_t$ where $\eta_t \in \mathbb{R}$ is i.i.d and action set X is a polytope with unit vectors as vertices. In other words, the loss when arm i is pulled at iteration t is $\bar{\ell}(i) + \eta_t$. Hence, $\arg\min_{x \in X} \sum_{t=1}^{T} \bar{\ell} \cdot x + \eta_t = \arg\min_{x \in X} \sum_{t=1}^{T} \bar{\ell} \cdot x$ since the η_t term in the loss does not depend on x and so the expected regret equals the pseudo-regret.

In the linear bandit setup we instead have loss $(\bar{\ell} + \eta_t) \cdot x = \bar{\ell} \cdot x + \eta_t \cdot x$ where $\eta_t \in \mathbb{R}^d$ is an i.i.d. vector. Due to the product with x in random term $\eta_t \cdot x$, in general $\arg\min_{x \in X} \sum_{t=1}^{T} \bar{\ell} \cdot x + \eta_t \cdot x \neq \arg\min_{x \in X} \sum_{t=1}^{T} \bar{\ell} \cdot x$ and so the expected regret is not equal the pseudo-regret. Nevertheless, our analysis in this paper shows that the difference between the expected regret and pseudo-regret is uniformly bounded. We can therefore apply standard multi-arm bandit analysis techniques for any existing algorithm to derive upper bounds on the pseudo-regret and then use these to also upper bound the expected regret.

Figure 1 shows numerical simulations of a 2-arm linear bandit setup with $d = 2$, X the polytope with vertices $(1, 0)$, $(0, 1)$ and loss vector $\ell_t = (0, -24.9)$ with probabilty 0.5 and $(0, 25.1)$ with probabilty 0.5. Hence $\mathbb{E}[\ell_t] = (0, 0.1)$ and action $y^* = (1, 0)$ yields minimal expected loss $\mathbb{E}[\ell_t] \cdot y^* = 0$ i.e. the pseudo-regret is $E[\sum_{t=1}^{T} \ell_t \cdot x_t] - \mathbb{E}[\ell_t] \cdot y^* = \mathbb{E}[\sum_{t=1}^{T} \ell_t \cdot x_t]$. The expected regret is $E[\sum_{t=1}^{T} \ell_t \cdot (x_t - x_T^*)]$ where x_T^* minimises the sample path sum $\sum_{t=1}^{T} \ell_t \cdot x = \sum_{t=1}^{T} \bar{\ell} \cdot x + \eta_t \cdot x$ and so varies from run to run. Figure 1(a) shows the measured pseudo-regret and expected regret for the UCB algorithm with bandit feedback, while Figure 1(b) shows the corresponding values for the lazy subgradient algorithm of Anderson and Leith (2020a) with full information feedback. It can be seen that, as expected, in both cases the difference between the expected regret and pseudo-regret is uniformly bounded and so the existing analytic results in terms of the pseudo-regret can be immediately generalised to encompass the expected regret.

6. Summary and Conclusions

In this paper we show that when there is a unique optimal action and the action set X is a polytope the difference between pseudo-regret and expected regret is $o(1)$. This means that the existing upper bounds on pseudo-regret in the literature can immediately be extended to also upper bound the expected regret. Our results are independent of the algorithm used to select the actions and apply equally to the bandit and full-information settings. This analysis can be extended to include i.i.d convex loss functions rather than just linear cost functions.

Importantly, while uniqueness of the optimal action is necessary for pseudo-regret and expected regret to be within $o(1)$ of one another, it is not sufficient and additional assumptions are needed. Here we use the additional assumption that the action set is a polytope. While other choices are likely possible, e.g. restriction to a strongly convex action set, we leave this as future work.

Acknowledgements

This work was supported by SFI grant 16/IA/4610(T). We would also like to express our thanks for the helpful comments from the anonymous reviewers.

Appendix

Lemma 7 Let $A \geq 0$. The function $F(X) = \sqrt{X}e^{-AX}$ is maximised at $X = 1/2A$ and $F(1/2A) = 1/\sqrt{2eA}$. The function is increasing on $[0, 1/2A]$ and decreasing on $[1/2A, \infty)$.

Proof Consider the function $G(X) = Xe^{-2AX}$. The derivative $G'(X) = (1 - 2AX)e^{-2AX}$ is positive for before $X \leq 1/2A$ and vanishes for $X = 1/2A$ and is negative for $X \geq 1/2A$. Hence the function is increasing on $[0, 1/2A]$ and decreasing on $[1/2A, \infty)$ and maximised at $X = 1/2A$. Since the square root is monotone, the same holds for $\sqrt{G(X)} = F(X)$. The maximum value is $F(1/2A) = \sqrt{1/2A}e^{-A(1/2A)} = e^{-A(1/2A)}/\sqrt{2A} = e^{-1/2}/\sqrt{2A} = 1/\sqrt{2eA}$ ■

The following fact about computing the expectation in terms of the CDF is well-known. But we were unable to find a suitably general proof in the literature.

Lemma 8 Suppose X is a nonnegative random variable. Then $\mathbb{E}[X] = \int_0^\infty P(X > x)dx$.

Proof The integral can be written as

$$\int_0^\infty P(X > x)dx = \int_0^\infty \mathbb{E}_y\big[\mathbf{1}_{X(y)>x}(y)\big]\, dx = \mathbb{E}_y\left[\int_0^\infty \mathbf{1}_{X(y)>x}(y)dx\right].$$

For fixed y define the function $g(x) = \mathbf{1}_{X(y)>x}(y)$. We have $g(x) = 1$ for all $x > X(y)$ and $g(x) = 0$ elsewhere. Since $X(y)$ is nonnegative that means $g(x)$ is the indicator function of $[0, X(y))$. It follows the inner integral equals $X(y)$ and the above becomes $\mathbb{E}_y[X(y)] = \mathbb{E}[X]$. ■

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *NIPS*, pages 2312–2320, 2011.

Daron Anderson and Douglas Leith. Boptimality of the subgradient algorithm in the stochastic setting. *Arxiv E-prints*, 2020a.

Daron Anderson and Douglas Leith. Universal algorithms: Beyond the simplex. *Arxiv E-prints*, 2020b.

Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 116–120. PMLR, 23–26 Jun 2016.

Nicolò Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA, 2006. ISBN 0521841089.

Varsha Dani, , Thomas Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. *21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland*, pages 355–366, 2008.

Pierre Gaillard, Gilles Stoltz, and Tim Van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, pages 176–196, 2014. URL `https://arxiv.org/pdf/1402.2044.pdf`.

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

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. *Journal of Machine Learning Research*, 18(145):1–31, 2017. URL `http://jmlr.org/papers/v18/17-079.html`.

Jaouad Mourtada and Stéphane Gaïffas. On the optimality of the Hedge algorithm in the stochastic regime. *Journal of Machine Learning Research*, 20:1–28, 2019.

Iosif Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *The Annals of Probability*, 22(4):1679–1706, 10 1994. doi: 10.1214/aop/1176988477. URL `https://doi.org/10.1214/aop/1176988477`.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. *Proc 31st Annual Conference on Learning Theory*, 75, 2018.