Proceedings of Machine Learning Research vol 336:1–39, 2026 39th Annual Conference on Learning Theory

# Self-Concordant Perturbations for Linear Bandits

**Lucas Lévy** LUCAS.LEVY@INRIA.FR
*École Polytechnique and University of Oxford*

**Jean-Lou Valeau** JEANLOU.VALEAU@ENSAE.FR
*ENSAE Paris and University of Oxford*

**Arya Akhavan** ARYA.AKHAVAN@STATS.OX.AC.UK
*University of Oxford and École Polytechnique*

**Patrick Rebeschini** PATRICK.REBESCHINI@STATS.OX.AC.UK
*University of Oxford*

**Editors:** Steve Hanneke and Tor Lattimore

## Abstract

We consider the adversarial linear bandits setting and present a unified algorithmic framework that bridges Follow-the-Regularized-Leader (FTRL) and Follow-the-Perturbed-Leader (FTPL) methods, extending the known connection between them from the full-information setting. Within this framework, we introduce self-concordant perturbations, a family of probability distributions that mirror the role of self-concordant barriers previously employed in the FTRL-based SCRIBLE algorithm. Using this idea, we design a novel FTPL-based algorithm that combines self-concordant regularization with efficient stochastic exploration. Our approach achieves a regret of $\mathcal{O}(d\sqrt{n \ln n})$ on both the $d$-dimensional hypercube and the $\ell_2$ ball. On the $\ell_2$ ball, this matches the rate attained by SCRIBLE. For the hypercube, this represents a $\sqrt{d}$ improvement over these methods and matches the optimal bound up to logarithmic factors.

**Keywords:** Bandit problems, online optimization, self-concordant barriers

## 1. Introduction

We study *online linear optimization under bandit feedback*, where an agent sequentially selects actions based on the information available from previous steps. The loss of each action varies over time and is unknown to the agent at the moment of decision. At the end of each round, only the loss associated to the chosen action is revealed. The agent's aim is to minimize the *regret*, the difference between the agent's cumulative loss and that of the best fixed action in hindsight. In this setting, two prominent families of algorithms have been widely studied: *Follow-the-Regularized-Leader* (FTRL) (Abernethy et al., 2008) and *Follow-the-Perturbed-Leader* (FTPL) (Kalai and Vempala, 2005). In the full-information setting, i.e., when the whole loss function is observed by the agent at the end of each round, Abernethy et al. (2014, 2016) introduced a general framework, *Gradient-Based Prediction Algorithm* (GBPA), which encompasses both FTRL and FTPL and enables a unified analysis.

Many algorithms designed for the bandit setting adapt techniques from the full-information setting but rely on estimators of the loss function rather than its exact value. Such estimators must be constructed from limited feedback, which makes *exploration*, deliberately selecting potentially suboptimal actions to gather information, essential (Cesa-Bianchi and Lugosi, 2006). Stochastic exploration mechanisms include playing from a fixed exploration distribution with small probability (Auer et al., 2002), or stochastically shifting the solution given by FTRL (Flaxman et al., 2005).

An important instance of FTRL, *Self-Concordant Regularization in Bandit Learning* (SCRIBLE) was introduced by Abernethy et al. (2012). The algorithm relies on self-concordant barriers (Nemirovskii, 1996), a class of convex functions that exhibit a controlled divergence at the boundary of their domain and satisfy a form of local strong convexity with respect to the local norm induced by their Hessian. These properties make self-concordant barriers well adapted to the geometry of the action set and explain their use as regularizers in SCRIBLE. This algorithm attains a regret bound of $\mathcal{O}(d^{3/2}\sqrt{n \ln n})$ for arbitrary convex body action sets $K \subset \mathbb{R}^d$, where $n$ is the number of rounds, or *horizon*. When the action set is the $d$-dimensional $\ell_2$ ball, the bound improves to $\mathcal{O}(d\sqrt{n \ln n})$. Moreover, whenever a closed-form expression of the barrier is available, which holds for many natural convex sets, the algorithm admits an implementation with per-round computational complexity polynomial in $d$.

For linear bandits, Dani et al. (2007) established a minimax lower bound of $\Omega(d\sqrt{n})$ on the regret for arbitrary convex bodies. Subsequent work closed this gap in terms of rates: Bubeck et al. (2012) proposed an FTRL-based algorithm achieving $\mathcal{O}(d\sqrt{n \ln n})$ regret in general, and introduced OSMD, which attains sharper bounds of $\mathcal{O}(d\sqrt{n})$ for the $d$-dimensional hypercube and $\mathcal{O}(\sqrt{dn \ln n})$ for the $\ell_2$ ball. More recently, Hazan and Karnin (2016) gave an efficient algorithm achieving $\mathcal{O}(d\sqrt{n \ln n})$ regret for arbitrary convex bodies using spanning ellipsoids for exploration, and van der Hoeven et al. (2018) achieved a similar rate using an exploration scheme based on John's ellipsoid. Despite these advances, existing optimal-rate algorithms rely on carefully tailored preprocessing schemes, and do not arise naturally from the self-concordant barrier framework. This motivates the question of whether self-concordant–based methods can be extended beyond SCRIBLE to achieve near-optimal regret while retaining a simple, perturbation-based exploration mechanism.

We study this question through FTPL, which selects its action by solving a randomly perturbed linear program. While FTPL can be reduced to FTRL in the full-information linear setting (Abernethy et al., 2014), the distinction is crucial in bandits: FTRL requires an explicit exploration mechanism, while FTPL is inherently stochastic, and samples its action from the extreme points of the action set. This observation motivates FTPL-based methods as a natural vehicle for efficient exploration, provided the perturbations replicate the effects of self-concordant regularization.

We make the following contributions.

1. We introduce BANDITS-GBPA, a unifying framework for linear bandit algorithms that includes both FTRL- and FTPL-based approaches. BANDITS-GBPA extends the GBPA framework of Abernethy et al. (2014, 2016), originally developed for the full-information setting, to the bandit setting. While this was also the goal of Abernethy et al. (2015), their work is limited to the multi-armed bandit case, i.e., when the action set is finite. Unlike in the full-information setting, our framework does not imply the inclusion of FTPL within FTRL, but it provides a conceptual structure allowing a common analysis.

2. Building on this framework, we introduce the *Self-Concordant FTPL* (SC-FTPL) algorithm, which incorporates both self-concordant regularization, and an FTPL sampling scheme. A central element of our approach is the identification and analysis of a family of perturbation distributions that we call *self-concordant perturbations*. These distributions replicate, within an FTPL framework, the properties that self-concordant barriers provide in FTRL, while defining a sampling scheme enabling more exploration.

3. We construct self-concordant perturbations for two canonical action sets, and we analyze their impact on SC-FTPL. These constructions serve not merely as domain-specific algorithmic

designs, but as concrete examples showing how perturbations can recover self-concordant barrier properties. When the action set is the $d$-dimensional hypercube, we show that SC-FTPL produces lower-variance loss vector estimates, and achieves a regret of $\mathcal{O}(d\sqrt{n \ln n})$, matching the minimax lower bound established by Dani et al. (2007) up to a logarithmic factor. This represents an improvement of a factor $\sqrt{d}$ over SCRIBLE and matches the performance of OSMD (Bubeck et al., 2012), without requiring an additional domain-specific sampling scheme. For the $\ell_2$ ball however, SC-FTPL achieves suboptimal regret in $\mathcal{O}(d\sqrt{n \ln n})$, matching the performance of SCRIBLE.

Our proofs combine three types of techniques. First, we leverage standard results and decomposition methods from the bandit literature (Abernethy et al., 2008; Bubeck and Cesa-Bianchi, 2012) to handle the core regret arguments. Second, we perform careful computations to identify suitable self-concordant barriers and to bound the norms of the estimators. While sometimes intricate, these are mostly technical. Finally, the analysis of SC-FTPL on the $\ell_2$ ball requires a novel argument to handle the absence of a uniform almost-sure bound on the local norms. We control the growth of the cumulative estimator via a carefully constructed supermartingale and stopping time arguments. This technique may be of independent interest for other sequential decision making problems with similar challenges.

The rest of the paper is organized as follows. Section 2 formalizes the linear bandits setting. Then, in Section 3, we present our main results: the SC-FTPL algorithm, self-concordant perturbations, and regret bounds on the hypercube and $\ell_2$ ball. In Section 4, we present the unifying BANDITS-GBPA framework. Section 5 reviews key properties of self-concordant barriers and motivates SC-FTPL's sampling and estimation schemes. In Section 6, we detail the analysis of SC-FTPL for the hypercube and the $\ell_2$ ball. Finally, in Section 7, we discuss some open research directions.

## 2. Problem Setting

We now formalize the *Adversarial Linear Bandits* framework, mentioned in the introduction. An instance of this problem is defined by an *horizon* $n \in \mathbb{N}$, an *action set* $K \subset \mathbb{R}^d$, and an unknown sequence of loss vectors $y_1, \ldots, y_n \in \mathbb{R}^d$. In each round $t \in [n] := \{1, ..., n\}$, the learner chooses, possibly at random, an action $a_t \in K$. Then, they observe the loss associated to their chosen action $\langle y_t, a_t \rangle$. The learner's decision can depend on an exogenous source of randomness and the history of previous feedback: $a_1, \langle y_1, a_1 \rangle, \ldots, a_{t-1}, \langle y_{t-1}, a_{t-1} \rangle$, but not on the current loss vector $y_t$.

On top of that, we make the following assumptions. First, the action set $K$ is a convex body, i.e., a compact convex set with non-empty interior. Moreover, the losses are bounded: for all loss vectors $y$ and action $a \in K$, $|\langle y, a \rangle| \leq 1$.

The quantity of interest is the difference between the learner's cumulated loss and the cumulated loss of the best action in-hindsight, called the regret. We define the learner's *regret with respect to some competitor* $u \in K$ as

$$R_n(u) := \mathbb{E}\Big[\sum_{t=1}^{n} \langle y_t, a_t \rangle\Big] - \sum_{t=1}^{n} \langle y_t, u \rangle \,, \tag{1}$$

where the expectation is taken with respect to the randomness in the actions of the learner. The learner's *regret* is then defined as $R_n := \sup_{u \in K} R_n(u)$.

## 3. Main Results

In this section, we introduce our main contribution, the *Self-Concordant Follow-the-Perturbed-Leader* (SC-FTPL) algorithm, along with its regret guarantees for the hypercube and the $\ell_2$ ball. We begin by defining *self-concordant perturbations*. Their definition assumes familiarity with $\vartheta$-self-concordant barriers. Additional details and properties of such barriers are deferred to Section 5.1.

**Definition 1 (Self-Concordant Perturbation)** *Let $K \subset \mathbb{R}^d$ be a convex body, $\vartheta > 0$, and $\mathcal{D}$ be a probability distribution on $\mathbb{R}^d$, absolutely continuous with respect to the Lebesgue measure. Let $\phi_K : \mathbb{R}^d \to \mathbb{R}$ be the support function of $K$. We say that $\mathcal{D}$ is a $\vartheta$-*self-concordant perturbation *for $K$ if there exists $\mathcal{R}$ a $\vartheta$-self-concordant barrier on $K$ such that*

$$\nabla \mathcal{R}^*(\theta) = \mathbb{E}_{\xi \sim \mathcal{D}}[\nabla \phi_K(\theta + \xi)] \quad \text{for all } \theta \in \mathbb{R}^d ,$$

*where $\mathcal{R}^*$ is the Fenchel conjugate of $\mathcal{R}$. In this case, we say that $\mathcal{D}$* replicates *$\mathcal{R}$.*

Note that this definition involves the derivative of $\phi_K$, the support function of $K$, which is not necessarily defined on $\mathbb{R}^d$. However, because $\phi_K$ is convex, it is differentiable almost everywhere. Moreover, $\mathcal{D}$ is absolutely continuous, so we have that $\mathbb{P}_{\xi \sim \mathcal{D}}(\phi_K \text{ differentiable in } \theta + \xi) = 1$ for all $\theta \in \mathbb{R}^d$, and thus $\mathbb{E}_{\xi \sim \mathcal{D}}[\nabla \phi_K(\theta + \xi)]$ is well-defined.

For additional intuition on the behavior of self-concordant perturbations, we note in Appendix B that they are heavy-tailed.

**The Self-Concordant FTPL Algorithm** We now define the *Self-Concordant FTPL* (SC-FTPL) algorithm. The learner maintains a cumulative estimate of the loss vectors $\hat{Y}_{t-1} = \sum_{s=1}^{t-1} \hat{y}_s$, initialized at 0. At each round $t \in [n]$, the algorithm follows a FTPL scheme: it samples $\xi_t \sim \mathcal{D}$, where $\mathcal{D}$ is a self-concordant perturbation, and selects

$$a_t = \arg\min_{a \in K} \langle a, \eta \hat{Y}_{t-1} - \xi_t \rangle .$$

After observing the scalar loss $\langle y_t, a_t \rangle$, the learner constructs an estimator $\hat{y}_t$ of $y_t$, using the second-moment matrix $Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]$, where $\mathbb{E}_{t-1}[\cdot]$ is the conditional expectation given past actions $a_1, \ldots, a_{t-1}$. The full procedure is specified in Algorithm 1.

**Algorithm 1** Self-Concordant FTPL (SC-FTPL)

**Require:** A $\vartheta$-self-concordant perturbation $\mathcal{D}$ for $K$ and a learning rate $\eta > 0$

1: Set $\hat{Y}_0 = 0$.
2: **for** $t = 1, \ldots, n$ **do**
3: Sample $\xi_t \sim \mathcal{D}$ independently from the past.
4: Play action $a_t = \arg\min_{a \in K} \langle a, \eta \hat{Y}_{t-1} - \xi_t \rangle$.
5: Receive punctual loss $\langle y_t, a_t \rangle$.
6: Compute estimator $\hat{y}_t = Q_t^{-1} a_t \langle y_t, a_t \rangle$, where $Q_t := \mathbb{E}_{t-1}[a_t a_t^\top]$.
7: Update $\hat{Y}_t = \hat{Y}_{t-1} + \hat{y}_t$.
8: **end for**

As we will see in Section 5, this construction inherits properties of self-concordant barriers used by SCRIBLE, while introducing a new sampling mechanism: SC-FTPL draws actions that concentrate on extreme points of $K$, thus exploring more effectively the available action space.

**Performance on the Hypercube and $\ell_2$ Ball** We analyze SC-FTPL in details on two canonical action sets: the hypercube $[-1,1]^d$ and the unit $\ell_2$ ball $\mathbb{B}^d := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. In both cases, we construct explicit self-concordant perturbations and derive the following regret guarantees.

**Theorem 2 (Regret of SC-FTPL on the Hypercube)** *Let $\mathcal{D}$ be the $d$-self-concordant perturbation for $[-1,1]^d$ defined in Proposition 7. Assume that $\frac{n}{\ln n} \geq 36\, d$. Then,* SC-FTPL *on the hypercube with perturbation distribution $\mathcal{D}$ and learning rate $\eta = \sqrt{\frac{3 \ln n}{n}}$ has regret bounded by*

$$R_n \leq \frac{2\sqrt{3}}{3} d\sqrt{n \ln n} + 2\,.$$

This result highlights the main advantage of our approach. While the previous SCRIBLE approach suffers a regret of $\mathcal{O}(d^{3/2}\sqrt{n})$ on the hypercube due to the suboptimal exploration, our FTPL-based approach improves this dependence by a factor of $\sqrt{d}$. This matches the minimax optimal rate for linear bandits on the hypercube up to logarithmic factors.

**Theorem 3 (Regret of SC-FTPL on the $\ell_2$ Ball)** *Let $\mathcal{D}$ be the 1-self-concordant perturbation for $\mathbb{B}^d$ defined in Proposition 9. Assume that $\frac{n}{\ln n} \geq \max(2d^2, 64)$. Then, there exists a universal constant $C > 0$ such that* SC-FTPL *on the $\ell_2$ ball with perturbation distribution $\mathcal{D}$ and learning rate $\eta = \frac{1}{d}\sqrt{\frac{4 \ln n}{5\, n}}$ has regret bounded by*

$$R_n \leq \sqrt{5}\, d\sqrt{n \ln n} + 2 + C\,\frac{\ln^3 n}{d^2}\,.$$

A more detailed analysis of SC-FTPL in this particular case, as well as the construction of the self-concordant perturbations appear in Section 6.

## 4. Gradient-Based Prediction

In this section, we present the unified GBPA framework introduced by Abernethy et al. (2014) in the full-information setting, and then define the BANDITS-GBPA algorithm.

### 4.1. Follow-the-Leader Style Algorithms

First, we place ourselves in the *full-information setting*. This setting is different from the bandit setting introduced in Section 2 in that the whole loss vector $y_t$ is observed at the end of round $t$, instead of the punctual loss $\langle y_t, a_t\rangle$. For all $t \geq 1$, let $Y_{t-1} := \sum_{s=1}^{t-1} y_s$ be the cumulative loss before time $t$.

An instance of FTRL is defined by a convex *regularizer* $\psi : K \to \mathbb{R}$[1]. At each time $t \in [n]$, the action chosen by FTRL solves the penalized optimization problem

$$a_t \in \operatorname*{arg\,min}_{a \in K} \left\{\langle a, Y_{t-1}\rangle + \psi(a)\right\}\,, \tag{2}$$

1. In the literature, the regularizer is usually scaled by a *learning rate* $\eta > 0$ (Orabona, 2026). The tuning of the learning rate is crucial in the analysis of FTRL in order to enjoy sublinear regret. However, for the sake of brevity, we do not include it in this first part and simply consider that it is included in the regularizer $\psi$.

where ties are resolved arbitrarily. *Follow-the-Perturbed-Leader* (FTPL) adds regularization via a stochastic perturbation (Kalai and Vempala, 2005). Given a probability distribution $\mathcal{D}$ on $\mathbb{R}^d$ called the *perturbation distribution*, at each time $t \in [n]$, FTPL selects the action

$$a_t \in \arg\min_{a \in K} \langle a, Y_{t-1} - \xi_t \rangle \,, \tag{3}$$

where $\xi_t \sim \mathcal{D}$ is independent from the past, and ties are resolved arbitrarily.

In the full-information setting, these two procedures have been shown to be part of a more general framework called *Gradient-Based Prediction Algorithm* (GBPA) by Abernethy et al. (2014, 2016). In GBPA, we give ourselves a differentiable function $\Phi : \mathbb{R}^d \to \mathbb{R}$, called a *potential*, such that $\operatorname{Im} \nabla\Phi \subseteq K$. At each step $t \in [n]$, the action played by GBPA is $a_t = \nabla\Phi(-Y_{t-1})$.

Indeed, an instance of FTRL with strictly convex regularizer $\psi$ is recovered by choosing $\Phi = \psi^*$. For all $\theta \in \mathbb{R}^d$, we then have $\nabla\Phi(\theta) = \arg\min_{a \in K}\{\langle a, -\theta\rangle + \psi(a)\}$, so the actions selected by GBPA and FTRL coincide.

Likewise, consider FTPL with an integrable perturbation distribution $\mathcal{D}$, and define $\Phi(\theta) = \mathbb{E}_{\xi\sim\mathcal{D}}[\phi_K(\theta + \xi)]$, where $\phi_K$ is the support function of $K$. The potential $\Phi$ is well defined, and if $\phi_K(\theta + \xi)$ is differentiable with probability one, we can swap the expectation and gradient (Bertsekas, 1973), which yields

$$\nabla\Phi(\theta) = \mathbb{E}_{\xi\sim\mathcal{D}}[\nabla\phi_K(\theta + \xi)] = \mathbb{E}_{\xi\sim\mathcal{D}}\big[\arg\min_{a \in K} \langle a, -\theta - \xi\rangle\big] \,. \tag{4}$$

Therefore, the action chosen by GBPA with potential $\Phi$ is the expectation of the action chosen by FTPL with perturbation distribution $\mathcal{D}$.

Note that for FTRL , the reciprocal holds due to the properties of the Fenchel-Legendre conjugate, and every instance of GBPA can be described as an instance of FTRL. This is not the case for FTPL, and some regularizers cannot be replicated by perturbations. Therefore, the relation between FTPL and FTRL is a strict inclusion in the full-information setting. Hofbauer and Sandholm (2002, Proposition 2.2) show that the function $\psi : x \mapsto -\sum_{i=1}^d \ln(x_i)$ defined on the probability simplex $\Delta^d := \{x \in \mathbb{R}_+^d : \|x\|_1 = 1\}$ does not admit a representation of the form of Equation (4). Generally, characterizing which regularizers admit an equivalent perturbation is challenging, with prior work largely focused on the simplex (Abernethy et al., 2015; Kim and Tewari, 2019).

### 4.2. Gradient-Based Prediction for Linear Bandits

We now present BANDITS-GBPA, extending GBPA from the full-information setting to the linear bandit setting. In the full information setting, GBPA relies on $Y_{t-1}$ to pick an action. In the bandit case, however, the learner only knows the past scalar loss $\langle y_s, a_s\rangle, s < t$. It must estimate $Y_{t-1}$ through randomization, and use this estimate in the GBPA step. Thus, BANDITS-GBPA needs two additional ingredients:

1. a *sampling scheme*, which randomizes the chosen action to enable exploration. Formally, a sampling scheme is a mapping $S$ from $K$ to $\mathcal{P}(K)$ the set of probability distributions on $K$, such that the chosen action $a_t$ at time $t$ will be sampled from the distribution $S(\nabla\Phi(-\hat{Y}_{t-1}))$.
2. an *estimation scheme*, which constructs an estimate of the unobserved loss vector from the observed scalar feedback. Formally, it is a function $E : \mathbb{R} \times K \times K \to \mathbb{R}^d$ mapping the observed scalar loss, the sampled action, and the expected action to the estimation of the loss vector, i.e., $\hat{y}_t = E(\langle y_t, a_t\rangle, a_t, \nabla\Phi(-\hat{Y}_{t-1}))$.

The resulting procedure is summarized in Algorithm 2. This template encompasses several existing linear bandits algorithm, including SCRIBLE.

**Algorithm 2** Gradient-Based Prediction for Linear Bandits (BANDITS-GBPA)

**Require:** A differentiable potential function $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that $\operatorname{Im} \nabla\Phi \subseteq K$, a sampling scheme $S : K \to \mathcal{P}(K)$ and an estimation scheme $E : \mathbb{R} \times K \times K \to \mathbb{R}^d$.

1: Set $\hat{Y}_0 = 0$.
2: **for** $t = 1, \ldots, n$ **do**
3: Let $x_t = \nabla\Phi(-\hat{Y}_{t-1})$.
4: Sample action $a_t \sim S(x_t)$ independently from the past.
5: Observe punctual loss $\langle y_t, a_t \rangle$.
6: Compute estimator $\hat{y}_t = E(\langle y_t, a_t \rangle, a_t, x_t)$.
7: Update $\hat{Y}_t = \hat{Y}_{t-1} + \hat{y}_t$.
8: **end for**

We also define a notion of unbiasedness for the sampling and estimation schemes. For $S : K \to \mathcal{P}(K)$ a sampling scheme and $E : \mathbb{R} \times K \times K \to \mathbb{R}^d$ an estimation scheme, we say that $(S, E)$ is *unbiased* if for all $x \in K, y \in \mathbb{R}^d$,

$$\mathbb{E}_{a\sim S(x)}\left[a\right] = x \qquad \text{and} \qquad \mathbb{E}_{a\sim S(x)}\left[E(\langle y, a\rangle, a, x)\right] = y\,. \tag{5}$$

We will subsequently show that both SCRIBLE and SC-FTPL are unbiased instances of BANDITS-GBPA.

## 5. From Self-Concordant Regularization to Self-Concordant Perturbations

We contextualize SC-FTPL by recalling key properties of self-concordant barriers, including the local strong convexity of their dual and the controlled divergence near the boundary of their domain, reviewing the SCRIBLE algorithm and its suboptimal sampling scheme, and motivating our improved sampling and estimation strategies.

### 5.1. Self-Concordant Barriers

Self-concordant barriers, originally introduced for interior-point methods, are a class of functions extensively studied in Nemirovskii (1996). In this section, we define self-concordant barriers and present some of their key properties.

**Definition 4 (Self-Concordant Barrier)** *Let $K$ be a convex set with non-empty interior and $\mathcal{R} : \operatorname{int} K \to \mathbb{R}$. The function $\mathcal{R}$ is a* self-concordant function *on $K$ if*

- *it is three times continuously differentiable and convex,*
- *for all $(x_n)_n \in \operatorname{int} K^{\mathbb{N}}$ such that $x_n$ converges to the boundary of $K$, $\mathcal{R}(x_n) \to +\infty$,*
- *for all $x \in \operatorname{int} K$ and $h \in \mathbb{R}^d$, $|D^3\mathcal{R}(x)[h,h,h]| \leq 2(D^2\mathcal{R}(x)[h,h])^{3/2}$.*

*Let $\vartheta > 0$. $\mathcal{R}$ is a* $\vartheta$-self-concordant barrier *if it is a self-concordant function and if for all $x \in \operatorname{int} K$ and $h \in \mathbb{R}^d$, $|D\mathcal{R}(x)[h]| \leq \sqrt{\vartheta\, D^2\mathcal{R}(x)[h,h]}$.*

*We have denoted $D^k\mathcal{R}(x)[h_1, \ldots, h_k]$ the $k$-th derivative of $\mathcal{R}$ at $x$ applied to the directions $h_1, \ldots, h_k$.*

For $\mathcal{R}$ a $\vartheta$-self-concordant barrier on $K$ and $x \in \operatorname{int} K$, the Hessian $\mathcal{R}$ in $x$, $\nabla^2\mathcal{R}(x)$ is positive definite. Thus, we can define the local norm with respect to the Hessian of $\mathcal{R}$ in $x$

$$\|y\|_{\nabla^2\mathcal{R}(x)} \coloneqq \sqrt{y^\top \nabla^2\mathcal{R}(x)\, y} \tag{6}$$

for all $y \in \mathbb{R}^d$. We also define the the open unit Dikin ellipsoid, which is the unit ball for this local norm:

$$W(x) \coloneqq \left\{ y \in \mathbb{R}^d : \|y - x\|_{\nabla^2\mathcal{R}(x)} < 1 \right\} .$$

For all $x \in \operatorname{int} K$, it holds that $W(x) \subset K$. Thus, the local geometry defined by the Hessian of $\mathcal{R}$ stretches $K$ such that the Dikin ellipsoid centered in $x$, is always inside $K$.

We now introduce two important properties of self-concordant barriers that justify their use as regularizers. First, for $\mathcal{R}$ a self-concordant function, its Fenchel conjugate $\mathcal{R}^*$ is also a self-concordant function and satisfies a local quadratic control of its Bregman divergence: for any $x, y \in \operatorname{dom}(\mathcal{R}^*)$ such that $\|y - x\|_{\nabla^2\mathcal{R}^*(x)} \leq \frac{1}{2}$,

$$B_{\mathcal{R}^*}(y, x) \leq \|y - x\|^2_{\nabla^2\mathcal{R}^*(x)}. \tag{7}$$

This inequality can be viewed as a local strong convexity property of $\mathcal{R}$,[2] and follows from standard results on self-concordant functions (Nemirovskii, 1996). A detailed proof is given in Appendix C.1. Second, if $\mathcal{R}$ is a $\vartheta$-self-concordant barrier on a convex set $K$ with non-empty interior, we can control the rate at which the barrier diverges near the boundary of $K$. For all $x, y \in \operatorname{int} K$,

$$\mathcal{R}(y) \leq \mathcal{R}(x) - \vartheta \ln(1 - \pi_x(y)), \tag{8}$$

where $\pi_x(y) \coloneqq \inf\{t > 0 : x + t^{-1}(y - x) \in K\}$ is the *Minkowski function in* $x$.

### 5.2. Self-Concordant Regularization

The *Self-Concordant Regularization in Bandits Learning* (SCRIBLE) algorithm, introduced by Abernethy et al. (2012), can be described as a particular algorithm in the BANDITS-GBPA framework. It is defined by a $\vartheta$-self-concordant barrier $\mathcal{R}$ on $K$ and a learning rate $\eta > 0$. Then, its potential is given by $\Phi = (\frac{1}{\eta}\mathcal{R})^*$ and given $x_t \in K$, the sampling distribution $S(x_t)$ is defined as the uniform distribution over the $2d$ poles of the Dikin ellipsoid $W(x_t)$.

The self-concordant barrier $\mathcal{R}$ thus plays a dual role: it defines both the potential and the sampling scheme through its Hessian. The properties of the Dikin ellipsoid ensure that $a_t \in K$, while simultaneously providing sufficient, though suboptimal, exploration to estimate $y_t$.

For an optimal learning rate $\eta$, SCRIBLE achieves regret $R_n = \mathcal{O}(d\sqrt{\vartheta n \ln n})$. Following recent works on universal self-concordant barriers (Lee and Yue, 2021), we know that every convex body $K \subset \mathbb{R}^d$ admits a $\vartheta \leq d$ barrier, yielding a $\mathcal{O}(d^{3/2}\sqrt{n \ln n})$ bound. For the $\ell_2$ ball, which admits a 1-self-concordant barrier, the regret improves to $\mathcal{O}(d\sqrt{n \ln n})$, in contrast to the hypercube, where no improvement beyond $\vartheta = n$ is possible (Nesterov and Nemirovskii, 1994). Hence, the regret of SCRIBLE is suboptimal by a $\sqrt{d}$ factor. As we will discuss next, this suboptimality stems from the variance of the estimator induced by Dikin-ellipsoid sampling.

2. For comparison, a convex, differentiable function $f$ is $\mu$-strongly-convex with respect to some norm $\|\cdot\|$ if and only if $B_{f^*}(x, y) \leq \frac{1}{2\mu}\|x - y\|_*^2$ for all $x, y$.

### 5.3. Self-Concordant FTPL

SC-FTPL uses the same potential as SCRIBLE but employs a different sampling scheme. Whereas SCRIBLE sampled from the poles of the Dikin ellipsoid, SC-FTPL uses the sampling naturally induced by FTPL, which is inherently randomized.

We can check that this is indeed an instance of BANDITS-GBPA, for the potential $\Phi = (\frac{1}{\eta}\mathcal{R})^*$ where $\mathcal{R}$ is the self-concordant barrier replicated by $\mathcal{D}$. The unbiasedness of the sampling scheme then follows from the definition of self-concordant perturbations.

Since SC-FTPL shares the same potential as SCRIBLE, its regret analysis relies on the same local strong convexity and barrier-growth properties, which we exploit in the theorem below.

**Theorem 5** *Let $K \subset \mathbb{R}^d$ be a convex body, $\eta > 0$, and $\mathcal{D}$ be a $\vartheta$-self-concordant perturbation for $K$, such that $\mathbb{E}_{\xi\sim\mathcal{D}}[\nabla\phi_K(\theta+\xi)\nabla\phi_K(\theta+\xi)^\top]$ is non-singular for all $\theta \in \mathbb{R}^d$. Then,* SC-FTPL *with learning rate $\eta$ and perturbation $\mathcal{D}$ has its regret bounded by*

$$R_n \leq \frac{\vartheta \ln n}{\eta} + \eta \sum_{t=1}^{n} \mathbb{E}[\|\hat{y}_t\|_t^2] + 2\,,$$

*provided that $2\eta\|\hat{y}_t\|_t \leq 1$ almost surely for all $t \in [n]$, with $\|\cdot\|_t = \|\cdot\|_{\nabla^2\mathcal{R}^*(-\eta\hat{Y}_{t-1})}$ where $\mathcal{R}$ is the self-concordant barrier replicated by $\mathcal{D}$.*

This theorem is adapted from Abernethy et al. (2012). We present here a proof sketch and defer the full proof to Appendix C.2.

**Proof sketch** We first establish that SC-FTPL is an instance of BANDITS-GBPA with potential $\Phi = (\frac{1}{\eta}\mathcal{R})^*$, where $\mathcal{R}$ is the self-concordant barrier replicated by $\mathcal{D}$. This allows us to use a standard regret decomposition in bandits literature, which upper-bounds the regret of SC-FTPL by the sum of the initial potential gap and cumulative stability terms.

The first term is $\Phi^*(u) - \min \Phi^* = \eta^{-1}(\mathcal{R}(u) - \min \mathcal{R})$, which diverges as $u$ approaches the boundary of $K$. To establish a bound independent of $u$, we use a standard shrinkage argument and evaluate the regret at point $u' = (1 - n^{-1})u + n^{-1}x^*$ where we bound $\mathcal{R}(u')$ by $\vartheta \ln n$ using (8).

The stability terms are $\mathbb{E}[B_\Phi(-\hat{Y}_t, -\hat{Y}_{t-1})]$. By the local strong convexity of self-concordant barriers (7), the potential $\Phi$ is locally smooth with respect to the local norm. Provided $\eta$ is small enough, we can bound the Bregman divergence by $\eta^2\|\hat{y}_t\|_t^2$, which completes the proof. ■

The same regret bound holds for the SCRIBLE algorithm. Crucially, this theorem shows that the regret is controlled by two quantities: the self-concordance parameter $\vartheta$ and the sum of the variances of loss vector estimators $\hat{y}_t$ measured in the local norms $\|\cdot\|_t$. Thus, improving regret reduces to designing sampling and estimation schemes that produce small $\mathbb{E}[\|\hat{y}_t\|_t^2]$.

**Motivations for FTPL Sampling** In SCRIBLE, the variance of the estimator in the local norm is bounded by $d^2$, which leads to the suboptimal regret $\mathcal{O}(d\sqrt{\vartheta n \ln n})$. The goal of SC-FTPL is to reduce this variance through a different sampling and estimation scheme. More generally, the variance of the loss estimator reflects how effectively the sampling strategy explores the action set by spreading actions across directions. Without stochastic exploration, accurate estimation of the loss vector is impossible and sublinear regret cannot be achieved (Cesa-Bianchi and Lugosi, 2006). Hence, an effective sampling scheme must exploit the available geometry of $K$ around the current point $x_t$.

The Dikin ellipsoid sampling used by SCRIBLE does not do so optimally, an intuition provided by Abernethy and Rakhlin (2009). When $x_t$ lies near the boundary of $K$, the Dikin ellipsoid $W(x_t)$ shrinks in order to remain centered at $x_t$ and contained in $K$. In contrast, FTPL selects actions by solving a linear optimization problem over $K$, resulting in actions at extreme points of the set. This boundary-oriented sampling allows FTPL sampling to exploit the full space in $K$ and, as we show later for the hypercube, can lead to lower variance of the estimator and sharper regret bounds.

**SC-FTPL Estimation Scheme** Let us now turn to the estimation scheme. We construct an unbiased estimator $\hat{y}_t$ for $y_t \in \mathbb{R}^d$ using only the observed bandit feedback $\langle y_t, a_t \rangle$, where $a_t \sim S(x_t)$. Following Bubeck et al. (2012), a natural candidate is

$$\hat{y}_t = Q_t^{-1} a_t \langle a_t, y_t \rangle, \quad \text{with} \quad Q_t = \mathbb{E}_{t-1}[a_t a_t^\top], \tag{9}$$

which is unbiased by construction. Its validity requires $Q_t$ to be invertible, which does not hold for general self-concordant perturbations. A counterexample is given in Appendix D. To address this, the next proposition gives a sufficient condition ensuring $Q_t$ is positive definite.

**Proposition 6** *Let $K \subset \mathbb{R}^d$ be a convex body and $D$ be an absolutely continuous probability distribution on $\mathbb{R}^d$. Suppose that $\operatorname{supp}(D) = \mathbb{R}^d$. Then, for all $\theta \in \mathbb{R}^d$, the matrix*

$$Q = \mathbb{E}_{\xi \sim D}[\nabla \phi_K(\theta + \xi) \nabla \phi_K(\theta + \xi)^\top]$$

*is positive definite.*

**Proof sketch** The matrix $Q$ is positive semi-definite by construction. To establish non-singularity, we show $u^\top Q u > 0$ for any $u \neq 0$. Because $D$ has full support and is absolutely continuous, this is equivalent to showing that the set $Z^c := \{y \in \mathbb{R}^d : \nabla \phi_K(y) \notin u^\perp\}$ has non-zero Lebesgue measure. Geometrically, $\nabla \phi_K(y)$ is the element of $K$ that maximizes $\langle \cdot, y \rangle$. We construct an open convex cone $C$ such that for any $y \in C$, $\nabla \phi_K(y)$ cannot lie in the subspace $u^\perp$, because its projection onto $u$ dominates its projection onto $u^\perp$. As $C \subset Z^c$ is open and non-empty, $Z^c$ has non-zero measure, concluding the proof. For all missing details, see Appendix D. ■

In the next section, we will show that for the hypercube and the $\ell_2$ ball, one can explicitly choose self-concordant perturbations with full support, which ensures $Q_t$ remains non-singular.

## 6. Analysis of SC-FTPL for Two Specific Action Sets

In this last section, we study SC-FTPL on two particular action sets: the hypercube and the unit $\ell_2$ ball. For each of them, we construct a self-concordant perturbation and detail the steps that lead to the regret bound established in Theorem 2 and Theorem 3. Finally, we provide a time complexity analysis of SC-FTPL in both of these cases.

### 6.1. Hypercube Case

Let $K = [-1, 1]^d$. We consider the *entropic barrier* of Bubeck and Eldan (2015), which is defined for all convex bodies as the Fenchel conjugate of the function

$$\mathcal{R}^*(\theta) := \ln \int_K \exp\langle \theta, x \rangle \, dx \,. \tag{10}$$

We know that for all convex bodies, the entropic barrier is a $d$-self-concordant barrier (Chewi, 2023). For the hypercube, the self-concordance parameter $\vartheta = d$ is optimal, as there does not exist any $\vartheta$-self-concordant barrier with $\vartheta < d$ (Nesterov and Nemirovskii, 1994, Proposition 2.3.6).

For the hypercube, while the entropic barrier $\mathcal{R}$ has no closed-form expression, its conjugate $\mathcal{R}^*$ and its conjugate gradient are

$$\mathcal{R}^*(\theta) = \sum_{i=1}^{d} \ln \frac{2\sinh\theta_i}{\theta_i}, \qquad \nabla\mathcal{R}^*(\theta) = [L(\theta_i)]_{i=1}^{d}$$

for all $\theta \in \mathbb{R}^d$, where $L : t \mapsto \coth(t) - 1/t$.

**Proof** The exponential of the entropic barrier's conjugate is

$$\exp \mathcal{R}^*(\theta) = \int_{[-1,1]^d} e^{\langle\theta,x\rangle} dx = \prod_{i=1}^{d} \int_{-1}^{1} e^{\theta_i x_i} dx_i = \prod_{i=1}^{d} \frac{e^{\theta_i} - e^{-\theta_i}}{\theta_i} .$$

Taking the logarithm yields the stated expression. For the gradient, notice that $\mathcal{R}^*$ is additively separable and that the derivative of $t \mapsto \ln(2\sinh(t)/t) = \ln 2 + \ln\sinh(t) - \ln t$ is $L$. ■

Hence, $\nabla\mathcal{R}^*$ is component-wise separable. Similarly, the gradient of the support function is $\nabla\phi_K : \theta \mapsto [\mathrm{sgn}(\theta_i)]_{i=1}^{d}$. This separability allows us to construct easily a perturbation that replicates $\mathcal{R}$ on $[-1,1]^d$.

**Proposition 7** *There exists a probability distribution over $\mathbb{R}^d$ whose marginal distributions are independent and have probability density function*

$$f : t \in \mathbb{R} \mapsto \frac{1}{2t^2} - \frac{1}{2\sinh^2(t)} .$$

*This is a $d$-self-concordant perturbation for $[-1,1]^d$.*

The proof of this proposition is quite straightforward and is deferred to Appendix E.1. Now that we have exhibited a self-concordant perturbation for $[-1,1]^d$, we upper-bound the variance of the local norm of the estimators of SC-FTPL, in order to obtain an upper bound on the regret.

**Proposition 8** *Let $\mathcal{D}$ be the self-concordant perturbation defined in Proposition 7 and $\mathcal{R}$ be the entropic barrier on the hypercube. Consider the* SC-FTPL *algorithm run with perturbation $\mathcal{D}$ and any learning rate $\eta > 0$. Then, for all $t \in [n]$, the estimator $\hat{y}_t$ satisfies*

$$\mathbb{E}_{t-1}[\|\hat{y}_t\|_t^2] \leq \frac{1}{3}d \qquad \textit{and} \qquad \|\hat{y}_t\|_t^2 \leq 3\,d \;\; \textit{a.s.,}$$

*where $\|\cdot\|_t = \|\cdot\|_{\nabla^2\mathcal{R}^*(-\eta\hat{Y}_{t-1})}$.*

The proof is mainly computational and is provided in Appendix E.2. Crucially, the bound of the variance of the local norm of the estimator improves by a $3d$ factor over the SCRIBLE's $d^2$ bound. This confirms that, on the hypercube, SC-FTPL achieves more accurate loss vector estimation. Now that we bounded the variance of the local norm of the estimator, we can establish Theorem 2.

**Proof sketch of Theorem 2.** The second inequality of Proposition 8 and the assumption on $n$ and $d$ ensure $\eta\|\hat{y}_t\|^2 \leq 1/2$ almost surely. Theorem 5 then applies, yielding a regret bound in terms of $\vartheta = d$ and $\mathbb{E}[\|\hat{y}_t\|_t^2] \leq d/3$. Combining these and optimizing $\eta$ gives the stated regret bound. Full details are in Appendix E.3. ■

This theorem bounds the regret of SC-FTPL on the hypercube in $\mathcal{O}(d\sqrt{n \ln n})$ for an optimal choice of $\eta$. This bound improves over SCRIBLE by a factor of $\sqrt{d}$ and matches the lower bound established by Dani et al. (2007) up to logarithmic factors. Finally, we provide a complexity analysis of SC-FTPL on the hypercube in Appendix E.4, and prove that it has a per-round complexity in $\mathcal{O}(d)$. This matches the complexity of SCRIBLE and OSMD in the same setting.

### 6.2. $\ell_2$ Ball Case

Let us now consider the case where the action set is $\mathbb{B}^d$ the unit $\ell_2$ ball. We consider the *log-barrier* on the ball given by

$$\mathcal{R}(x) \coloneqq -\ln(1 - \|x\|_2^2)\,, \quad x \in \operatorname{int}\mathbb{B}^d\,.$$

This is a 1-self-concordant barrier on $\mathbb{B}^d$. Here, both $\mathcal{R}$ and $\phi_{\mathbb{B}^d}$ are spherically invariant, i.e., depend only on the norm of their argument. Exploiting this symmetry, we construct an explicit perturbation replicating $\mathcal{R}$ on $\mathbb{B}^d$:

**Proposition 9** *Let $\mathbf{T}$ be a random vector in $\mathbb{R}^d$ following a multivariate $t$-distribution with location 0, scale matrix $I_d$ and $d+1$ degrees of freedom. Let $U$ be a random variable following a uniform distribution on $[0,1]$, independent from $\mathbf{T}$. Then, the distribution of*

$$\xi = \frac{\mathbf{T}}{\sqrt{d+1}\,U}$$

*is a 1-self-concordant perturbation for $\mathbb{B}^d$.*

While the computations leading to the proof of Proposition 9 reduce to a one-dimensional radial calculation, identifying the perturbation that exactly reproduces the barrier geometry is non trivial and, to the best of our knowledge, novel. These computations are detailed in Appendix F. Now that we have exhibited a self-concordant perturbation for $\mathbb{B}^d$, we derive bounds on the estimators of SC-FTPL for this choice of perturbation.

**Proposition 10** *Let $\mathcal{D}$ be the 1-self-concordant perturbation defined in Proposition 9 and $\mathcal{R}$ be the log-barrier on $\mathbb{B}^d$. Consider Algorithm 1 run with perturbation distribution $\mathcal{D}$ and any learning rate $\eta > 0$. Then, for all $t \in [n]$, the estimator $\hat{y}_t$ satisfies the following inequalities*

1. $\mathbb{E}_{t-1}[||\hat{y}_t||_t^2] \leq \frac{5}{4}d^2$,
2. $||\hat{y}_t||_t^2 \leq d^2\eta\,\|\hat{Y}_{t-1}\|_2 + 4\,d^2$ *almost surely, and*
3. $\mathbb{E}_{t-1}[||\hat{y}_t||_2^2] \leq d^2\eta\,\|\hat{Y}_{t-1}\|_2 + 2\,d^2$.

*where* $\|\cdot\|_t = \|\cdot\|_{\nabla^2\mathcal{R}^*(-\eta\hat{Y}_{t-1})}$.

**Proof sketch** Fix $t \in [n]$ and let $\theta = -\eta \hat{Y}_{t-1}$. By the spherical symmetry of $\mathcal{D}$, $Q_t$ is invariant under rotations that fix $\theta$, which implies that it admits the decomposition $Q_t = q\, P_\theta + q_\perp\, P_{\theta^\perp}$, where $P_\theta$ and $P_{\theta^\perp}$ are the orthogonal projection matrices onto $\theta$ and $\theta^\perp$, respectively. We then show, using properties of the multivariate $t$-distribution, that both coefficients are positive and satisfy $q \geq 1/d$ and $q_\perp \geq 1/(d(\|\theta\|+1))$. Thus $Q_t^{-1} = q^{-1} P_\theta + q_\perp^{-1} P_{\theta^\perp}$.

Substituting $Q_t^{-1}$ and $\nabla^2 \mathcal{R}^*(\theta)$ into the expressions of $\mathbb{E}_{t-1}[\|\hat{y}_t\|_t^2]$, $\|\hat{y}_t\|_t^2$, and $\mathbb{E}_{t-1}[\|\hat{y}_t\|_2^2]$, each quantity reduces to a one–dimensional inequality in $\|\theta\|$. Real analysis computations then yields the three stated bounds. The full proof, including the derivation of the bounds on $q$ and $q_\perp$, is deferred to Appendix F.2. ■

From these bounds, we derive Theorem 3, which bounds the regret of SC-FTPL on the $\ell_2$ ball.

**Proof sketch of Theorem 3.** The main difficulty is that, unlike the hypercube case, there is no uniform almost-sure bound on $\|\hat{y}_t\|_t$ independent of $\hat{Y}_{t-1}$. As a result, the regret decomposition of Theorem 5, which requires $\eta\|\hat{y}_t\|_t \leq 1/2$ for all $t$, cannot be applied directly.

To address this issue, we work on the high-probability event

$$E \coloneqq \{\forall t \in [n],\, \eta\|\hat{y}_t\|_t \leq 1/2\}\,.$$

Using the almost-sure bound of Proposition 10, we show that $E$ holds whenever $\sup_{t<n} \|\hat{Y}_t\|^2$ remains below a threshold $1/C_n$, where $C_n = o(n^{-5/2})$ is an explicit quantity arising from the analysis and for which the resulting failure probability is negligible compared to the main regret term.

To control the growth of $\|\hat{Y}_t\|^2$, we use the $\ell_2$-norm variance bound of Proposition 10 to construct an appropriate supermartingale. Applying Ville's inequality yields a tail bound of order $C_n n^2$ on the probability that $\sup_{t<n} \|\hat{Y}_t\|^2 > 1/C_n$.

We then define the stopping time

$$\tau \coloneqq \inf\{t < n : \|\hat{Y}_t\|^2 > 1/C_n\} \wedge n\,,$$

and decompose the regret at time $\tau$. On the event $\{\tau = n\}$, the condition $\eta\|\hat{y}_t\|_t \leq 1/2$ holds throughout, and an optional stopping argument allows us to apply the same regret analysis as in Theorem 5. Using $\vartheta = 1$ and the local norm variance bound from Proposition 10, this yields the leading term $d\sqrt{5n\ln n} + 2$ for an optimal value of the learning rate $\eta$.

On the complementary event $\{\tau < n\}$, the regret is trivially bounded by $2n$, which multiplied by the failure probability contributes a lower-order term $\mathcal{O}(C_n n^3) = o(\sqrt{n})$. Combining both parts concludes the proof. Full details are given in Appendix F.3. ■

Compared to the hypercube setting, our results on the $\ell_2$ ball are weaker in two aspects. First, we were unable to establish an $\mathcal{O}(d)$ bound on the variance of the local norm of $\hat{y}_t$. Our analysis yields a quadratic dependence on $d$, matching the SCRIBLE estimator, which leads to the $\mathcal{O}(d\sqrt{n\ln n})$ regret in Theorem 3. Second, there is no uniform almost-sure bound on $\|\hat{y}_t\|_t$ independent of $\hat{Y}_{t-1}$. This justify the need for the Ville's inequality and optional stopping time arguments we developed in the proof of Theorem 3. These arguments introduce an additional term in the regret bound. While this additional term is asymptotically negligible when compared to the main $\mathcal{O}(d\sqrt{n\ln n})$ term, it may still be important for small values of $d$.

We include a discussion of the computational cost of SC-FTPL on the $\ell_2$-ball in Appendix F.4. We show that it has a per-round complexity in $\mathcal{O}(d)$: both sampling the perturbation $\xi$ as defined in Proposition 9 and computing $a_t$ are $\mathcal{O}(d)$. Then, computing the loss estimator $\hat{y}_t$ can be done in $\mathcal{O}(d)$ using numerical integration and leveraging the structure of $Q_t^{-1}$. This per-round computational cost matches the complexity of SCRIBLE and of van der Hoeven et al. (2018)'s algorithm on the $\ell_2$-ball.

## 7. Future Directions

Our work open up to several unexplored research directions. First, we have shown that the regret of SC-FTPL on the $\ell_2$ ball is bounded in $\mathcal{O}(d\sqrt{n \ln n})$ for our choice of self-concordant perturbation. Whether this rate is tight for FTPL introduced by any perturbation scheme remains an open question.

Second, we have not established the existence of self-concordant perturbations for general convex bodies. It remains unclear whether such perturbations exist beyond specific domains. A promising direction is to investigate the entropic barrier, a universal self-concordant barrier defined for any convex body (Bubeck and Eldan, 2015), and determine whether it systematically admits a corresponding perturbation distribution.

Finally, Hazan and Levy (2014) showed how self-concordant regularization can be used beyond linear bandits, when the loss functions are convex. It would be natural to explore whether the perturbation-based perspective developed here can yield similar benefits in that setting.

## Acknowledgments

L. Lévy, J.-L. Valeau, A. Akhavan, and P. Rebeschini were funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number EP/Y028333/1).

## References

Jacob Abernethy and Alexander Rakhlin. Beating the adaptive bandit with high probability. In *Proceedings of The 22nd Annual Conference on Learning Theory*, 2009.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory*, 2008.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Transactions on Information Theory*, 58(7):4164–4175, 2012.

Jacob Abernethy, Chansoo Lee, Abhinav Sinha, and Ambuj Tewari. Online linear optimization via smoothing. In *Proceedings of The 27th Conference on Learning Theory*, 2014.

Jacob Abernethy, Chansoo Lee, and Ambuj Tewari. Fighting bandits with a new kind of smoothness. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Jacob Abernethy, Chansoo Lee, and Ambuj Tewari. Perturbation techniques in online learning and optimization. In Tamir Hazan, George Papandreou, and Daniel Tarlow, editors, *Perturbations, Optimization, and Statistics*, pages 233–264. The MIT Press, 2016.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Dimitri P. Bertsekas. Stochastic optimization problems with nondifferentiable cost functionals. *Journal of Optimization Theory and Applications*, 12:218–231, 1973.

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

Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of The 28th Conference on Learning Theory*, 2015.

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Proceedings of the 25th Annual Conference on Learning Theory*, 2012.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Sinho Chewi. The entropic barrier is n-self-concordant. In Ronen Eldan, Bo'az Klartag, Alexander Litvak, and Emanuel Milman, editors, *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2020-2022*, pages 209–222. Springer, 2023.

Varsha Dani, Sham M Kakade, and Thomas Hayes. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2005.

Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 7th edition, 2007.

Elad Hazan and Zohar Karnin. Volumetric spanners: An efficient exploration basis for learning. *Journal of Machine Learning Research*, 17(119):1–34, 2016.

Elad Hazan and Kfir Y. Levy. Bandit convex optimization: Towards tight bounds. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Daniel Herrera-Esposito and Johannes Burge. Projected normal distribution: Moment approximations and generalizations. *arXiv preprint arXiv:2506.17461*, 2025.

Josef Hofbauer and William H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Baekjin Kim and Ambuj Tewari. On the optimality of perturbations in stochastic and adversarial multi-armed bandit problems. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Samuel Kotz and Saralees Nadarajah. *Multivariate t-distributions and their applications*. Cambridge university press, 2004.

Yin Tat Lee and Man-Chung Yue. Universal barrier is n-self-concordant. *Mathematics of Operations Research*, 46(3):1129–1148, 2021.

Arkadii Nemirovskii. *Interior-point polynomial time methods in convex programming*. Lecture Notes, 1996.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2026.

Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

Dirk van der Hoeven, Tim van Erven, and Wojciech Kotłowski. The many faces of exponential weights in online learning. In *Proceedings of the 31st Conference On Learning Theory*, 2018.

## Appendix A. Notation

In this section, we collect the main pieces of notation used in this paper.

**Convex Analysis Notation** Let $K \subset \mathbb{R}^d$ be a convex body, i.e., a non-empty compact convex set. We denote $\phi_K : \theta \mapsto \max_{x \in K} \langle x, \theta \rangle$ its *support function* and for all $x \in \mathbb{R}^d$, $\pi_x : y \in \mathbb{R}^d \mapsto \inf\{t > 0 : x + t^{-1}(y - x) \in K\}$ its *Minkowski function in* $x$.

Let $f$ be a closed, proper convex function on $\mathbb{R}^d$. Its *Fenchel conjugate* is $f^* : \theta \in \mathbb{R}^d \mapsto \sup_{x \in \mathbb{R}^d} \langle x, \theta \rangle - f(x)$. We refer to Rockafellar (1970) for an extensive treatment of support functions and Fenchel duality. If $f$ is differentiable, we define the *Bregman divergence with respect to* $f$ for $x, y \in \operatorname{dom} f$ as $B_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$.

**Bandit Notation** We denote $n \in \mathbb{N}$ the horizon and $K$ the action set. For all round $t \in [n] := \{1, \ldots, n\}$, the loss vector is denoted $y_t$ and the learner's action $a_t$.

Let $(\mathcal{F}_t)_{0 \le t \le n}$ be the natural filtration defined by $\mathcal{F}_t = \sigma(a_s, s \le t)$, and let $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t]$ be the conditional expectation given the first $t$ rounds.

For SC-FTPL, $\mathcal{D}$ denotes the $\vartheta$-self-concordant perturbation and $\eta$ the learning rate. At round $t$, $\xi_t$ is the perturbation sampled from $\mathcal{D}$, $Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]$, $\hat{y}_t = Q_t^{-1} a_t \langle y_t, a_t \rangle$ is the loss estimator, and $\hat{Y}_t = \sum_{s=1}^t \hat{y}_s$ is the cumulated estimate.

Finally, $\mathcal{R}$ denotes the self-concordant barrier replicated by $\mathcal{D}$, and

$$\| \cdot \|_t = \| \cdot \|_{\nabla^2 \mathcal{R}^*(-\eta \hat{Y}_{t-1})}$$

is the local norm induced by the Hessian of $\mathcal{R}^*$ in $-\eta \hat{Y}_{t-1}$.

## Appendix B. Heavy-Tailed Nature of Self-Concordant Perturbations

In this section, we highlight an interesting property of self-concordant perturbations that is not needed for the main proofs but provides valuable intuition. Specifically, these perturbations are heavy-tailed, which helps explain why they naturally induce strong exploration in bandit algorithms. Understanding this behavior sheds light on the role of the perturbation in SC-FTPL and related algorithms.

**Proposition 11** *Let $K \subset \mathbb{R}^d$ be a convex body and $\mathcal{D}$ be a self-concordant perturbation for $K$. Then $\mathcal{D}$ does not have a finite first moment.*

The proof of Proposition 11 makes use of the following lemma from Bertsekas (1973) that justifies exchanging the gradient and expectation operators.

**Lemma 12 (Bertsekas, 1973)** *Let $h : \mathbb{R}^d \to \mathbb{R}$ be a convex function and $\xi$ be a random vector on $\mathbb{R}^d$, whose distribution is absolutely continuous w.r.t the Lebesgue measure. Assume that for all $x \in \mathbb{R}^d$, $\mathbb{E}|h(x - \xi)| < \infty$ and define $H(x) = \mathbb{E}[h(x - \xi)]$. Then $h$ is differentiable almost everywhere, $H$ is differentiable everywhere and for all $x \in \mathbb{R}^d$,*

$$\nabla H(x) = \mathbb{E}[\nabla h(x - \xi)] \,.$$

**Proof of Proposition 11.** We proceed by contradiction. Assume that $\mathbb{E}\|\xi\| < +\infty$ with $\xi \sim \mathcal{D}$. For all $\theta \in \mathbb{R}^d$,

$$\mathbb{E}|\phi_K(\theta + \xi)| \leq \big(\|\theta\| + \mathbb{E}\|\xi\|\big) \sup_{x \in K} \|x\| < +\infty\,.$$

From Lemma 12, it follows that for all $\theta \in \mathbb{R}^d$,

$$\nabla \mathbb{E}[\phi_K(\theta + \xi)] = \mathbb{E}[\nabla \phi_K(\theta + \xi)] = \nabla \mathcal{R}^*(\theta)\,.$$

So there exists $C \in \mathbb{R}$ such that

$$\mathcal{R}^*(\theta) = \mathbb{E}[\phi_K(\theta + \xi)] + C \quad \text{for all } \theta \in \mathbb{R}^d.$$

The support function $\phi_K$ is $L$-Lipschitz with $L = \sup_{x \in K} \|x\|$, so for all $\theta, \xi \in \mathbb{R}^d$,

$$\phi_K(\theta + \xi) - \phi_K(\theta) \geq -L\|\xi\|\,.$$

Taking the expectation over $\xi \sim \mathcal{D}$ yields

$$\mathcal{R}^*(\theta) - \phi_K(\theta) \geq -L\,\mathbb{E}\|\xi\| + C\,, \quad \forall \theta \in \mathbb{R}^d\,. \tag{11}$$

For all $x \in K$, we have

$$\begin{aligned}\mathcal{R}^*(\nabla \mathcal{R}(x)) &= \langle \nabla \mathcal{R}(x), x\rangle - \mathcal{R}(x) \\ &\leq \phi_K(\nabla \mathcal{R}(x)) - \mathcal{R}(x)\,.\end{aligned}$$

Hence $(\mathcal{R}^* - \phi_K)(\nabla \mathcal{R}(x))$ goes to $-\infty$ when $\mathcal{R}(x)$ goes to $+\infty$. Because $\mathcal{R}$ is unbounded from above on $K$, this directly contradicts (11). So $\mathcal{D}$ does not have a finite first moment. ■

This property highlights an important distinction between defining self-concordant perturbations at the level of gradients versus potentials. In Abernethy et al. (2014), the *replication* property between a perturbation distribution $\mathcal{D}$ and a regularizer $\psi$ is defined at the potential level, i.e.,

$$\psi^*(\theta) = \mathbb{E}_{\xi \sim \mathcal{D}}[\phi_K(\theta + \xi)], \quad \theta \in \mathbb{R}^d. \tag{12}$$

By exchanging the gradient and expectation, they then derive the equality of the action in expectation,

$$\nabla \psi^*(\theta) = \mathbb{E}_{\xi \sim \mathcal{D}}[\nabla \phi_K(\theta + \xi)], \quad \theta \in \mathbb{R}^d. \tag{13}$$

This second equality can be called *replication property at the gradient level*. In Definition 1 however, we define self-concordant perturbations using the replication at the gradient level. Proposition 11 explains this: since $\mathcal{D}$ lacks a finite first moment, the expectation $\mathbb{E}[\phi_K(\theta + \xi)]$ may fail to exist. However, because $K$ is bounded, the expectation $\mathbb{E}[\nabla \phi_K(\theta + \xi)]$ is always well-defined.

## Appendix C. Proof of Self-Concordance Results

### C.1. Proof of Equation (7)

**Proof** From (Nemirovskii, 1996, 2.2), we know that if $\mathcal{R}$ is a self-concordant function, then so is $\mathcal{R}^*$ and that for all $x \in \operatorname{int} K$, $y \in W(x)$, it holds that

$$\mathcal{R}(x + y) \leq \mathcal{R}(x) + \nabla \mathcal{R}(x)^\top y + \rho(\|x - y\|_x) \tag{14}$$

where $\rho(t) = -\ln(1-t) - t$. Because it is also a self-concordant function, (14) holds for $\mathcal{R}^*$, and we have that if $\|y - x\|_{\nabla^2 \mathcal{R}^*(x)} < 1$, then

$$\mathcal{R}^*(x+y) \leq \mathcal{R}^*(x) + \nabla \mathcal{R}^*(x)^\top y + \rho(\|x-y\|_{\nabla^2 \mathcal{R}^*(x)}).$$

Thus, all that remains is to prove that $\rho(t) \leq t^2$ for $t \leq 1/2$. Let $f(t) = t^2 - \rho(t)$. Then, $f'(t) = 2t - \frac{1}{1-t} + 1 = \frac{t(2t-1)}{t-1}$. For all $t \leq 1/2$, we have that $f'(t) \geq 0$, hence $f(t) \geq f(0) = 0$. This proves that that $\rho(t) \leq t^2$ for all $t \leq 1/2$, which concludes the proof. ■

### C.2. Proof of Theorem 5

The proof of Theorem 5 makes use of the following lemma, which bounds the regret of BANDITS-GBPA when the sampling and estimation schemes are unbiased, following a standard regret decomposition in the bandit literature. For instance, a similar, though not identical, statement can be found in Bubeck and Cesa-Bianchi (2012, Section 5).

**Lemma 13** *Let $K$ be a convex set. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be convex, differentiable and such that $\operatorname{Im} \nabla\Phi \subseteq K$. Then, Algorithm 2 with potential $\Phi$ and unbiased sampling and estimation schemes satisfies, for all $u \in K$,*

$$R_n(u) \leq \Phi^*(u) - \min_K \Phi^* + \sum_{t=1}^n \mathbb{E}\big[B_\Phi(-\hat{Y}_t, -\hat{Y}_{t-1})\big]\,.$$

**Proof of Lemma 13.** The first step of the proof is to notice that, because of the unbiasedness of sampling and estimation schemes, it holds that for all $t \in [n]$,

$$\mathbb{E}_{t-1}[\langle y_t, a_t - u\rangle] = \langle y_t, x_t - u\rangle = \mathbb{E}_{t-1}[\langle \hat{y}_t, x_t - u\rangle]\,.$$

It follows that the regret can be written as

$$\begin{aligned}
R_n(u) &= \mathbb{E}\Big[\sum_{t=1}^n \langle \hat{y}_t, x_t - u\rangle\Big] \\
&= \sum_{t=1}^n \mathbb{E}\big[\langle \hat{Y}_t - \hat{Y}_{t-1}, \nabla\Phi(-\hat{Y}_{t-1})\rangle\big] - \sum_{t=1}^n \mathbb{E}\big[\langle \hat{y}_t, u\rangle\big] \\
&= \sum_{t=1}^n \mathbb{E}\big[-\Phi(-\hat{Y}_t) + \Phi(-\hat{Y}_{t-1}) + B_\Phi(-\hat{Y}_t, -\hat{Y}_{t-1})\big] - \mathbb{E}\big[\langle \hat{Y}_n, u\rangle\big]\,,
\end{aligned}$$

where the second equality follows from the definition of $x_t$ and the third equality follows from the definition of the Bregman divergence. By telescoping the $\Phi(-\hat{Y}_t)$ terms, we obtain

$$R_n(u) = \mathbb{E}\big[-\Phi(-\hat{Y}_n) + \Phi(-\hat{Y}_0)\big] + \sum_{t=1}^n \mathbb{E}\big[B_\Phi(-\hat{Y}_t, -\hat{Y}_{t-1})\big] - \mathbb{E}\big[\langle \hat{Y}_n, u\rangle\big]\,. \tag{15}$$

Because $\Phi$ is proper, closed, and convex, we have $\Phi = \Phi^{**}$. Therefore,

$$\Phi(-\hat{Y}_0) = \Phi(0) = \sup_{x\in K}(-\Phi^*(x)) = \min_{x\in K} \Phi^*(x)\,,$$

and

$$\Phi(-\hat{Y}_n) = \sup_{a \in K} \left\{ \langle a, -\hat{Y}_n \rangle - \Phi^*(a) \right\} \geq \langle u, -\hat{Y}_n \rangle - \Phi^*(u) \,.$$

It follows that

$$-\Phi(-\hat{Y}_n) - \langle u, \hat{Y}_n \rangle \leq \Phi^*(u) \quad \text{almost surely,}$$

Combining this last inequality into (15) concludes the proof. ■

**Proof of Theorem 5.** We first prove that Self-Concordant FTPL is an unbiased instance of Bandits-GBPA. Its potential is given by $\Phi = (\frac{1}{\eta}\mathcal{R})^*$, where $\mathcal{R}$ denotes the self-concordant barrier replicated by $\mathcal{D}$. By properties of the Fenchel conjugate, we have

$$\Phi(\theta) = \frac{1}{\eta}\mathcal{R}^*(\eta\,\theta) \quad \text{and} \quad \Phi^*(x) = \frac{1}{\eta}\mathcal{R}(x)$$

Hence, for all $t \in [d]$,

$$\begin{aligned}
\mathbb{E}_{t-1}[a_t] &= \mathbb{E}_{\xi_t}\left[\nabla\phi_K(\xi_t - \eta\,\hat{Y}_{t-1})\right] \\
&= \nabla\mathcal{R}^*(-\eta\,\hat{Y}_{t-1}) \\
&= \nabla\Phi(-\hat{Y}_{t-1})\,,
\end{aligned}$$

where the second equality follows from the definition of self-concordant perturbations. This shows that the sampling scheme is unbiased. Moreover,

$$\mathbb{E}_{t-1}[\hat{y}_t] = Q_t^{-1} \cdot \mathbb{E}_{t-1}\left[a_t a_t^\top\right] y_t = y_t \,.$$

Thus, Self-Concordant FTPL satisfies the unbiasedness condition of Bandits-GBPA, and Lemma 13 yields the regret decomposition

$$R_n(u) \leq \Phi^*(u) - \min_{x \in K} \Phi^*(x) + \sum_{t=1}^{n} \mathbb{E}\left[B_\Phi(-\hat{Y}_t, -\hat{Y}_{t-1})\right], \tag{16}$$

for all $u \in K$. Expressing (16) in terms of $\mathcal{R}$, we obtain

$$R_n(u) \leq \frac{\mathcal{R}(u) - \min_{x \in K} \mathcal{R}(x)}{\eta} + \frac{1}{\eta}\sum_{t=1}^{n} \mathbb{E}\left[B_{\mathcal{R}^*}(-\eta\hat{Y}_t, -\eta\hat{Y}_{t-1})\right]. \tag{17}$$

Let $x^* = \arg\min_{x \in K} \mathcal{R}(x)$ and $\kappa \in (0,1)$, and define $u' = \kappa x^* + (1-\kappa)u$. Then $\pi_{x^*}(u') = (1-\kappa)\pi_{x^*}(u) \leq 1 - \kappa$. By Equation (8), it follows that

$$\mathcal{R}(u') - \mathcal{R}(x^*) \leq -\vartheta \ln \kappa \,.$$

We now decompose the regret as

$$
\begin{aligned}
R_n(u) &= \mathbb{E}\Big[\sum_{t=1}^{n}\langle y_t, a_t - u' + u' - u\rangle\Big] \\
&= R_n(u') + \sum_{t=1}^{n}\langle y_t, u' - u\rangle \\
&\le \frac{\mathcal{R}(u') - \mathcal{R}(x^*)}{\eta} + \frac{1}{\eta}\sum_{t=1}^{n}\mathbb{E}[B_{\mathcal{R}^*}(-\eta\hat{Y}_t, -\eta\hat{Y}_{t-1})] + \kappa\sum_{t=1}^{n}\langle y_t, x^* - u\rangle \\
&\le -\frac{1}{\eta}\vartheta\ln\kappa + \frac{1}{\eta}\sum_{t=1}^{n}\mathbb{E}[B_{\mathcal{R}^*}(-\eta\hat{Y}_t, -\eta\hat{Y}_{t-1})] + 2n\kappa\,, \qquad (18)
\end{aligned}
$$

where the first inequality uses (17) in $u'$. Since (18) holds for any $\kappa \in (0, 1)$, taking $\kappa = \frac{1}{n}$ gives

$$
R_n(u) \le \frac{1}{\eta}\vartheta\ln n + \frac{1}{\eta}\sum_{t=1}^{n}\mathbb{E}[B_{\mathcal{R}^*}(-\eta\hat{Y}_t, -\eta\hat{Y}_{t-1})] + 2\,. \qquad (19)
$$

It remains to bound the Bregman divergence term. Since $\|\eta\hat{y}_t\|_t \le 1/2$ almost surely for all $t \in [n]$, we can apply the local smoothness inequality of (7), which yields

$$
\begin{aligned}
R_n(u) &\le \frac{1}{\eta}\vartheta\ln n + \frac{1}{\eta}\sum_{t=1}^{n}\mathbb{E}[\| - \eta\hat{Y}_t + \eta\hat{Y}_{t-1}\|^2_{\nabla^2\mathcal{R}^*(-\eta\hat{Y}_{t-1})}] + 2 \\
&= \frac{1}{\eta}\vartheta\ln n + \eta\sum_{t=1}^{n}\mathbb{E}[\|\hat{y}_t\|_t^2] + 2\,.
\end{aligned}
$$

■

## Appendix D. Estimation Scheme of SC-FTPL

### D.1. Proof of Proposition 6

**Proof** Let $\theta \in \mathbb{R}^d$. The matrix $Q$ is symmetric and positive semi-definite by construction. To prove non-singularity, it suffices to show that $Q$ is positive definite, i.e., $u^\top Q u > 0$ for all $u \in \mathbb{R}^d \backslash \{0\}$. Consider an arbitrary non-zero vector $u$. We have that

$$
u^\top Q u = \mathbb{E}_{\xi\sim D}[(u^\top \nabla\phi_K(\theta + \xi))^2]\,.
$$

Since the integrand is non-negative, the expectation is zero if and only if $u^\top\nabla\phi_K(\theta+\xi) = 0$ almost surely with respect to $D$. Since $D$ is absolutely continuous with full support on $\mathbb{R}^d$, this condition is equivalent to the set $Z = \{y \in \mathbb{R}^d : \nabla\phi_K(y) \in u^\perp\}$ having full Lebesgue measure in $\mathbb{R}^d$. Thus, in order to prove $u^\top Q u > 0$, it suffices to show that $\mathbb{R}^d \setminus Z$ has non-zero Lebesgue measure.

Since $K$ is bounded and has non-empty interior, there exists $x \in K$ and $\varepsilon, M > 0$ such that $B(x, \varepsilon) \subset K \subset B(0, M)$. Without loss of generality, we may assume $\langle u, x\rangle \ge 0$. Define the cone

$$
C = \left\{ y \in \mathbb{R}^d : \langle y, \frac{u}{\|u\|}\rangle > \frac{2M}{\varepsilon}\|p_{u^\perp}(y)\| \right\}\,,
$$

where $p_{u^\perp}$ denotes the orthogonal projection onto $u^\perp$. The set $C$ is open and non-empty, and thus has non-zero Lebesgue measure. We now show that $C \cap Z = \emptyset$.

Let $y \in C$. Consider the point $z = x + \varepsilon \frac{u}{\|u\|} \in K$. It holds that

$$\begin{aligned}\langle y, z\rangle &= \langle p_u(y), p_u(x)\rangle + \langle p_{u^\perp}(y), p_{u^\perp}(x)\rangle + \varepsilon\langle y, \frac{u}{\|u\|}\rangle \\ &> -\|p_{u^\perp}(y)\|\,\|p_{u^\perp}(x)\| + 2M\|p_{u^\perp}(y)\| \\ &\geq M\,\|p_{u^\perp}(y)\|\end{aligned}$$

Moreover, for any $h \in K \cap u^\perp$, observe that

$$\langle y, h\rangle = \langle p_{u^\perp}(y), h\rangle \leq M\,\|p_{u^\perp}(y)\|\,.$$

Combining these results, we strictly have $\langle z, y\rangle > \sup_{h\in K\cap u^\perp}\langle h, y\rangle$. Since

$$\nabla\phi_K(y) \in \arg\max_{w\in K}\langle w, y\rangle,$$

it follows that $\nabla\phi_K(y) \notin u^\perp$. Therefore, $\langle\nabla\phi_K(y), u\rangle \neq 0$ for all $y \in C$, implying $C \subseteq \mathbb{R}^d \setminus Z$. Since $C$ has non-zero measure, $u^\top Q u > 0$. ■

### D.2. Ill-Definition of the Estimator

The SC-FTPL estimator estimates the loss vector $y_t$ using

$$\hat{y}_t = Q_t^{-1} a_t \langle a_t, y_t\rangle \quad \text{with } Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]\,.$$

For this estimator to be well-defined, the covariance matrix $Q_t$ must be non-singular. We established in Proposition 6 that if the perturbation distribution $\mathcal{D}$ has full support, then $Q_t$ is invertible. Furthermore, in Section 6, when $K$ is the hypercube or the $\ell_2$ ball, we explicitly constructed self-concordant perturbations satisfying this condition.

A natural theoretical question is whether the self-concordance property alone is sufficient to guarantee the non-singularity of $Q_t$. In this section, we show that the answer is negative. We provide a counterexample of a valid self-concordant perturbation that results in a singular covariance matrix.

Consider the hypercube $K = [-1, 1]^d$ with $d \geq 2$ and the univariate density function

$$f(t) = \frac{1}{2t^2} - \frac{1}{2\sinh^2 t}\,, \quad t \in \mathbb{R}\,.$$

Proposition 7 established that if a random vector $\xi$ has density $p(x) = \prod_{i=1}^d f(x_i)$, i.e., if the coordinates of $\xi$ are independent and have density $f$, then the distribution $\mathcal{D}$ of $\xi$ is a $d$-self-concordant perturbation for $K$. Since $\mathcal{D}$ has full support on $\mathbb{R}^d$, the resulting covariance matrix $Q_t$ is always non-singular.

However, the self-concordance property for the hypercube relies solely on the marginal distributions of $\xi$, not on the independence of its coordinates. We exploit this to construct a degenerate perturbation. Let $\mathcal{D}'$ be an absolutely continuous distribution on $\mathbb{R}^d$ defined by the density:

$$q(x) = 2^{d-1} \cdot \mathbf{1}\left\{x \in (\mathbb{R}_+)^d \cup (\mathbb{R}_-)^d\right\} \cdot \prod_{i=1}^d f(x_i)\,.$$

Because the function $f$ is even, one can verify that the marginal distributions of $\mathcal{D}'$ are identical to those of $\mathcal{D}$, each coordinate has density $f$. Consequently, $\mathcal{D}'$ remains a valid $d$-self-concordant perturbation for $K$.

Consider the first round of SC-FTPL ($t = 1$) with $\hat{Y}_0 = 0$ and perturbation $\xi_1 \sim \mathcal{D}'$. The action is given by:

$$a_1 = \arg\min_{a \in K} \langle a, -\xi_1 \rangle = (\mathrm{sgn}(\xi_{1,i}))_{1 \le i \le d} \,.$$

By construction of $\mathcal{D}'$, the vector $\xi_1$ lies almost surely in $(\mathbb{R}_+)^d \cup (\mathbb{R}_-)$. Therefore, $a_1$ takes values in $\{\mathbf{1}_d, -\mathbf{1}_d\}$ almost surely. In both cases, the outer product is the full matrix $a_1 a_1^\top = \mathbf{1}\mathbf{1}^\top$. Taking the expectation yields $Q_1 = \mathbf{1}\mathbf{1}^\top$. Thus the matrix $Q_1$ has rank 1 and is therefore singular.

## Appendix E. Proofs of Results on the Hypercube

### E.1. Proof of Proposition 7

**Proof** The function $f$ is non-negative on $\mathbb{R}$ and satisfies $\int_{\mathbb{R}} f = 1$, so it is the density function of some probability distribution over $\mathbb{R}$. Thus, there exists some probability distribution $\mathcal{D}$ on $\mathbb{R}^d$ such that for $\xi \sim \mathcal{D}$, the coordinates $\{\xi_i\}_i$ are independent and identically distributed and have probability density function $f$.

We now prove that $\mathcal{D}$ replicates $\mathcal{R}$. For all $\theta \in \mathbb{R}^d$ and $i \in [d]$,

$$\begin{aligned} \mathbb{E}_{\xi\sim\mathcal{D}}[\nabla\phi_K(\theta+\xi)]_i &= \mathbb{E}[\mathrm{sgn}(\theta_i + \xi_i)] \\ &= \mathbb{P}(\theta_i + \xi_i \ge 0) - \mathbb{P}(\theta_i + \xi_i \le 0) \\ &= 1 - 2F_i(-\theta_i)\,, \end{aligned}$$

where $F_i$ is the cumulative distribution function of $\xi_i$. Notice that $f(t) = \frac{1}{2}L'(t)$. Thus,

$$F_i(t) = \frac{1}{2}(L(t) - L(0)) + F_i(0) = \frac{L(t)+1}{2}$$

and

$$\mathbb{E}_{\xi\sim\mathcal{D}}[\nabla\phi_K(\theta+\xi)]_i = -L(-\theta_i) = L(\theta_i) = \frac{\partial\mathcal{R}^*(\theta)}{\partial\theta_i}\,,$$

which concludes the proof. ■

### E.2. Proof of Proposition 8

**Proof** Let $t \in [n]$. For brevity, denote by $a$ the action chosen by SC-FTPL at time $t$, $x = \mathbb{E}_{t-1}[a]$ and $Q = \mathbb{E}_{t-1}[aa^\top]$. We also denote $H = \nabla^2\mathcal{R}^*(-\eta\hat{Y}_{t-1})$ the matrix such that $\|\cdot\|_t = \|\cdot\|_H$.

**Step 1: Computing the expression of $Q^{-1}$.** We first show that $Q$ is non-singular and compute the expression of $Q$. Let $i, j \in [d]$. If $i \ne j$, then $\xi_i$ is independent from $\xi_j$, thus so are $a_i$ and $a_j$ and we have

$$Q_{i,j} = \mathbb{E}_{t-1}[a_i a_j] = \mathbb{E}_{t-1}[a_i]\mathbb{E}_{t-1}[a_j] = x_i x_j\,.$$

If $i = j$, then $Q_{i,i} = \mathbb{E}[a_i^2]$. Because $a \in \mathrm{extr}([-1,1]^d)$, we have that $a_i^2 = 1$ almost surely, and then $Q_{i,i} = 1$. Thus,

$$Q = xx^\top + \mathrm{Cov}(a)\,, \quad \text{where } \mathrm{Cov}(a) = \mathrm{Diag}([1 - x_i^2]_{i=1}^d).$$

A direct computation shows that $Q$ is non-singular, with

$$Q^{-1} = \Big(\frac{-1}{1+\alpha}(\operatorname{Cov} a)^{-1}xx^\top + I_d\Big)(\operatorname{Cov} a)^{-1}\,,$$

with

$$\alpha \coloneqq x^T(\operatorname{Cov} a)^{-1}x = \sum_{i=1}^d \frac{x_i^2}{(1-x_i^2)}\,.$$

**Step 2: Bounding the local norm variance.** Let $y$ denote the loss vector and $\hat{y} = Q^{-1}a\langle y, a\rangle$ the loss vector estimator. It holds that

$$\begin{aligned}\mathbb{E}_{t-1}[\|\hat{y}\|_t^2] &= \mathbb{E}_{t-1}[\langle y, a\rangle a^\top Q^{-1}HQ^{-1}a\langle a, y\rangle] \\ &\le \mathbb{E}_{t-1}[a^\top Q^{-1}HQ^{-1}a] \\ &= \operatorname{tr}\left(HQ^{-1}\right)\,,\end{aligned}$$

where the inequality follows from the boundedness of the losses and the last equality uses the cyclic invariance of the trace. We now develop the expression of $H$, which is

$$H = \nabla^2\mathcal{R}^*(\nabla\mathcal{R}(x)) = \operatorname{Diag}\left[L'(L^{-1}(x_i))\right]_{1\le i\le d}\,,$$

where

$$L(t) = \coth(t) - \frac{1}{t}$$

is the Langevin function, which is a diffeormorphism between $\mathbb{R}$ and $(-1,1)$. Thus, combining the expression of $H$ and $Q^{-1}$, we get

$$\begin{aligned}\operatorname{tr}(HQ^{-1}) &= \sum_{i=1}^d \left(\frac{-1}{1+\alpha}\frac{x_i^2}{1-x_i^2}+1\right)\frac{1}{1-x_i^2}L'(L^{-1}(x_i)) \\ &\le \sum_{i=1}^d \frac{L'(L^{-1}(x_i))}{1-x_i^2} \qquad (20)\end{aligned}$$

Then, we want to show that

$$g(t) \coloneqq \frac{L'(t)}{1-L(t)^2} \le \frac{1}{3}$$

for all $t\in\mathbb{R}$. Because $g$ is even, it suffices to prove the inequality for $t\ge 0$. Developing the expression of $L$, we have that

$$g(t) = \frac{t^2-\sinh^2 t}{t^2+\sinh^2 t - t\sinh(2t)}\,.$$

For all $t\ge 0$, $t\le \sinh t$, so the numerator is nonpositive. Because $g$ is nonnegative, this implies that the denominator is nonpositive, and thus

$$g(t)\le\frac{1}{3} \iff 3(t^2-\sinh^2 t) \ge t^2+\sinh^2 t - t\sinh(2t)\,.$$

Rearranging terms, we define $h(t) := 2t^2 - 4\sinh^2 t + t\sinh(2t)$ and aim to show that $h(t) \geq 0$. The successive derivatives of $h$ are:

$$\begin{aligned} h'(t) &= 4t - 3\sinh(2t) + 2t\cosh(2t) \\ h''(t) &= 4 - 4\cosh(2t) + 4t\sinh(2t) \\ h^{(3)}(t) &= -4\sinh(2t) + 8t\cosh(2t) \\ h^{(4)}(t) &= 16t\sinh(2t) \end{aligned}$$

For $t \geq 0$, $h^{(4)}(t) \geq 0$. Since $h^{(3)}(0) = 0$, $h^{(3)}(t) \geq 0$ for $t \geq 0$. Similarly, since $h''(0) = 0$, $h''(t) \geq 0$ for all $t \geq 0$, and continuing this logic $h'(t) \geq 0$ and finally $h(t) \geq 0$ for all $t \geq 0$. Hence, the inequality $g(t) \leq 1/3$ holds for all $t \in \mathbb{R}$. Plunging this inequality into (20) gives $\operatorname{tr}(HQ^{-1}) \leq \sum_{i=1}^d 1/3$, and finally

$$\mathbb{E}_{t-1}[\|\hat{y}\|_t^2] \leq \frac{d}{3}$$

**Step 3: Bounding the local norm almost surely.** Finally, we derive the almost-sure bound on $\|\hat{y}\|_t^2$. First, observe that

$$\begin{aligned} \|\hat{y}\|_t^2 &\leq a^\top Q^{-1} H Q^{-1} a \\ &= \sum_{i=1}^d \sum_{j=1}^d a_i [Q^{-1} H Q^{-1}]_{i,j} a_j \\ &\leq \sum_{i,j} |[Q^{-1} H Q^{-1}]_{i,j}| \end{aligned}$$

Plugging in the closed-form expressions of $Q^{-1}$ and $H$ yields

$$\begin{aligned} \|\hat{y}\|_t^2 \leq{}& \frac{1}{(1+\alpha)^2} \left( \sum_{i=1}^d \frac{|x_i|}{1-x_i^2} \right)^2 \sum_{k=1}^d \frac{x_k^2 \, L'(L^{-1}(x_k))}{(1-x_k^2)^2} \\ &+ \frac{2}{1+\alpha} \left( \sum_{i=1}^d \frac{|x_i|}{1-x_i^2} \right) \sum_{k=1}^d \frac{|x_k| \, L'(L^{-1}(x_k))}{(1-x_k^2)^2} + \sum_{k=1}^d \frac{L'(L^{-1}(x_k))}{(1-x_k^2)^2} \end{aligned}$$

Let $C = \sup_{x \in (-1,1)} \frac{L'(L^{-1}(x))}{(1-x^2)^2} \in [0, +\infty]$. We can bound the previous expression by

$$\|\hat{y}\|_t^2 \leq \frac{C}{(1+\alpha)^2} \left( \sum_{i=1}^d \frac{|x_i|}{1-x_i^2} \right)^2 \sum_{k=1}^d x_k^2 + \frac{2C}{1+\alpha} \left( \sum_{i=1}^d \frac{|x_i|}{1-x_i^2} \right) \sum_{k=1}^d |x_k| + dC \tag{21}$$

By the Cauchy–Schwarz inequality,

$$\sum_{k=1}^d |x_k| \leq \sqrt{d \sum_{k=1}^d x_k^2}, \quad \text{and} \quad \sum_{i=1}^d \frac{|x_i|}{1-x_i^2} \leq \sqrt{\sum_{i=1}^d \frac{x_i^2}{1-x_i^2}} \sqrt{\sum_{i=1}^d \frac{1}{1-x_i^2}} = \sqrt{\alpha(d+\alpha)}\,.$$

Plugging theses inequalities into (21) gives

$$\|\hat{y}\|_t^2 \leq \frac{C\alpha(d+\alpha)}{(1+\alpha)^2} \sum_{k=1}^d x_k^2 + \frac{2C}{1+\alpha} \sqrt{d\alpha(d+\alpha) \sum_{k=1}^d x_k^2} + dC$$

Moreover, it holds that

$$(d+\alpha)\sum_{k=1}^{d} x_k^2 \le d\sum_{k=1}^{d}\frac{x_k^2}{1-x_k^2} + \alpha\sum_{k=1}^{d} 1 = 2\,\alpha d\,,$$

and hence

$$\|\hat{y}\|_t^2 \le \frac{2C\alpha^2 d}{(1+\alpha)^2} + \frac{2\sqrt{2}C\alpha d}{1+\alpha} + dC \le (3+2\sqrt{2})dC \tag{22}$$

Finally, we bound the value of $C$. Let

$$g(t) = \frac{L'(t)}{(1-L^2(t))^2}$$

for all $t \in \mathbb{R}$. This function is even so it suffices to analyze its behavior on $\mathbb{R}^+$. We want to show that for all $t > 0$, $g(t) \le 1/2$. Direct computation gives

$$L'(t) = \frac{1}{t^2} - \frac{1}{\sinh^2 t}, \qquad 1 - L(t)^2 = \frac{2t\cosh t\,\sinh t - t^2 - \sinh^2 t}{t^2\sinh^2 t}.$$

Hence, the desired inequality is equivalent to

$$(1-L^2)^2 - 2L' \ge 0.$$

Clearing the denominator $t^4\sinh^4 t$ and setting

$$N(t) := \big(2t\cosh t\,\sinh t - t^2 - \sinh^2 t\big)^2 - 2t^2\sinh^2 t\,(\sinh^2 t - t^2),$$

one obtains after a short algebraic rearrangement the identity

$$N(t) = \Big(t^2 - t\sinh(2t) + \tfrac{1}{2}(\cosh(2t)-1)\Big)^2 + 2t^2\sinh^2 t\,(\sinh t - t)^2.$$

Both terms on the right-hand side are nonnegative for $t > 0$, so $N(t) \ge 0$ and therefore $(1-L^2)^2 - 2L' \ge 0$. This yields $g(t) \le 1/2$, as claimed, and thus $C \le 1/2$. Hence we have,

$$\|\hat{y}\|_t^2 \le (3+2\sqrt{2})\cdot\frac{1}{2}\cdot d \le 3\,d\,,$$

which concludes the proof. ■

### E.3. Proof of Theorem 2

**Proof** Almost surely, it holds that for all $t \in [n]$,

$$4\,\eta^2\,\|\hat{y}_t\|_t^2 \le 12\,d\,\eta^2 = 36\,d\,\frac{\ln n}{n} \le 1\,,$$

where the first inequality follows from the almost sure bound in Proposition 8, the equality from the definition of $\eta$ and the last inequality from the theorem's assumption. Thus, the requirements of Theorem 5 are satisfied, yielding the following regret bound:

$$R_n \le \frac{\vartheta\ln n}{\eta} + \eta\sum_{t=1}^{n}\mathbb{E}[\|\hat{y}_t\|_t^2] + 2\,,$$

By Proposition 8, we have $\mathbb{E}[\|\hat{y}_t\|_t^2] \le d/3$ for all $t$. Furthermore, since the self-concordance parameter $\vartheta$ of $\mathcal{D}$ equals $d$, it follows that

$$R_n \le \frac{d \ln n}{\eta} + \frac{1}{3}\eta n d + 2\,.$$

This bound is optimal for $\eta = \sqrt{\frac{3 \ln n}{n}}$, which yields the claimed result. ■

### E.4. Complexity Analysis of SC-FTPL on the Hypercube

At each round $t \in [n]$, SC-FTPL performs three main steps: sampling $\xi_t$ from $\mathcal{D}$, solving the linear program $a_t = \arg\min_{a \in K} \langle a, \eta \hat{Y}_{t-1} - \xi_t \rangle$, and computing the estimator $\hat{y}_t = Q_t^{-1} a_t \langle y_t, a_t \rangle$.

Sampling from $\mathcal{D}$ given in Proposition 7 reduces to drawing $d$ i.i.d. random variables with probability density function $f$. Using the fact that $f(t) \le 1/(1+t^2)$ for all $t \in \mathbb{R}$, we can use rejection sampling with a Cauchy proposal to sample each coordinate in expected constant time. So the overall sampling step requires $\mathcal{O}(d)$ operations.

The linear program over the hypercube admits the closed-form solution

$$a_{t,i} = \operatorname{sgn}(-\eta \hat{Y}_{t-1,i} + \xi_{t,i})$$

for each coordinate $i \in [d]$, which can be computed in constant time per coordinate, yielding $\mathcal{O}(d)$ complexity.

To compute the estimator, we first compute $x_t = \nabla \mathcal{R}^*(-\eta \hat{Y}_{t-1})$, which is separable across coordinates and therefore requires $\mathcal{O}(d)$ time. From the proof of Proposition 8, $Q_t^{-1} a_t$ admits the closed form

$$Q_t^{-1} a_t = \Big( \frac{-1}{1+\alpha} \frac{x_{t,i}}{1 - x_{t,i}^2} \sum_{j=1}^{d} \frac{x_{t,j} a_{t,j}}{1 - x_{t,j}^2} + \frac{a_{t,i}}{1 - x_{t,i}^2} \Big)_{1 \le i \le d}$$

where $\alpha = \sum_{i=1}^{d} x_{t,i}^2 / (1 - x_{t,i}^2)$. This allows computing $Q_t^{-1} a_t$ efficiently: we first evaluate $\alpha$ and

$$\beta = \sum_{j=1}^{d} \frac{x_{t,j} a_{t,j}}{1 - x_{t,j}^2}$$

in $\mathcal{O}(d)$, and then each coordinate of $Q_t^{-1} a_t$ can be computed in constant time as

$$[Q_t^{-1} a_t]_i = \frac{-\beta}{1+\alpha} \frac{x_{t,i}}{1 - x_{t,i}^2} + \frac{a_{t,i}}{1 - x_{t,i}^2}.$$

Combining all steps, SC-FTPL on the hypercube has a total per-round complexity of $\mathcal{O}(d)$. For comparison, both SCRIBLE and OSMD enjoy a per-round complexity in $\mathcal{O}(d)$ when the action set of the hypercube. Indeed, for SCRIBLE, when the action set is the hypercube, we have access to closed forms for the eigenvalues and eigenvectors of $\nabla^2 \mathcal{R}(x_t)$, which allows to attain a per-round complexity in $\mathcal{O}(d)$.

# Appendix F. Proofs of Results on the $\ell_2$ Ball

## F.1. Proof of Proposition 9

The proof of Proposition 9 follows from the following lemma, which gives the expression of the expected value of a multivariate $t$-distribution projected over the unit sphere.

**Lemma 14** *Let $\mathbf{T}$ be a random vector in $\mathbb{R}^d$ following a multivariate $t$-distribution with location $\mu \in \mathbb{R}^d$, scale matrix $I_d$ and $d+1$ degrees of freedom. Then,*

$$\mathbb{E}\left[\frac{\mathbf{T}}{\|\mathbf{T}\|}\right] = \frac{\mu}{\sqrt{d+1+\|\mu\|^2}}\,.$$

**Proof of Lemma 14.** Let $\mathbf{T}$ be a random vector following a multivariate $t$-distribution with location $\mu \in \mathbb{R}^d$, scale matrix $I_d$, and $d+1$ degrees of freedom. We use the stochastic representation of the multivariate $t$-distribution (Kotz and Nadarajah, 2004): for independent random variables $\mathbf{N} \sim \mathcal{N}(0, I_d)$ and $Z \sim \chi_{d+1}$, the random vector

$$\mu + \frac{\sqrt{d+1}}{Z}\mathbf{N}$$

has the same distribution as $\mathbf{T}$. Consequently, the expected projection can be expressed as

$$\mathbb{E}\left[\frac{\mathbf{T}}{\|\mathbf{T}\|}\right] = \mathbb{E}\left[\frac{Z\mu + \sqrt{d+1}\mathbf{N}}{\|Z\mu + \sqrt{d+1}\mathbf{N}\|}\right].$$

Conditioning on $Z$, the vector $Z\mu + \sqrt{d+1}\mathbf{N}$ follows a multivariate normal distribution $\mathcal{N}(Z\mu, (d+1)I_d)$. Applying the known characterization for the mean of a projected isotropic normal distribution (Herrera-Esposito and Burge, 2025), we have:

$$\mathbb{E}\left[\frac{Z\mu + \sqrt{d+1}\mathbf{N}}{\|Z\mu + \sqrt{d+1}\mathbf{N}\|}\,\middle|\, Z\right] = \frac{\Gamma(\frac{d+1}{2})}{\sqrt{2(d+1)}\,\Gamma(\frac{d}{2}+1)}\ {}_1F_1\left(\frac{1}{2}; \frac{d}{2}+1; \frac{-Z^2\|\mu\|^2}{2(d+1)}\right) Z\mu\,,$$

where ${}_1F_1(a; b; z)$ denotes the confluent hypergeometric function. By the tower rule, it follows that

$$\mathbb{E}\left[\frac{\mathbf{T}}{\|\mathbf{T}\|}\right] = \mu\,\frac{\Gamma(\frac{d+1}{2})}{\sqrt{2(d+1)}\,\Gamma(\frac{d}{2}+1)}\,\mathbb{E}\left[{}_1F_1\left(\frac{1}{2}; \frac{d}{2}+1; \frac{-Z^2\|\mu\|^2}{2(d+1)}\right) Z\right].$$

Given that $Z^2 \sim \chi^2_{d+1}$, we evaluate the remaining expectation in integral form

$$\mathbb{E}\left[{}_1F_1\left(\frac{1}{2}; \frac{d}{2}+1; \frac{-Z^2\|\mu\|^2}{2(d+1)}\right) Z\right]$$
$$= \frac{1}{2^{\frac{d+1}{2}}\,\Gamma(\frac{d+1}{2})}\int_0^{+\infty} {}_1F_1\left(\frac{1}{2}; \frac{d}{2}+1; -t\frac{\|\mu\|^2}{2(d+1)}\right) t^{d/2}\, e^{-t/2}\, dz.$$

Using the standard identity for integrals of confluent hypergeometric functions (Gradshteyn and Ryzhik, 2007), the integral evaluates to

$$\int_0^{+\infty} {}_1F_1\left(\frac{1}{2}; \frac{d}{2}+1; -t\frac{\|\mu\|^2}{2(d+1)}\right) t^{d/2}\, e^{-t/2}\, dz = \Gamma\left(\frac{d}{2}+1\right) 2^{\frac{d}{2}+1}\left(1+\frac{\|\mu\|^2}{d+1}\right)^{-1/2}.$$

Substituting this result back into the main expression and simplifying the constants, we arrive at the final form

$$\mathbb{E}\left[\frac{\mathbf{T}}{\|\mathbf{T}\|}\right] = \frac{\mu}{\sqrt{n+1+\|\mu\|^2}}\,.$$

■

**Proof of Proposition 9.** Let $\mathcal{D}$ be the distribution of $\xi$. By construction, $\mathcal{D}$ is an absolutely continuous distribution on $\mathbb{R}^d$ with full support. We now show that $\mathcal{D}$ replicates the 1-self-concordant barrier $x \in \operatorname{int}\mathbb{B}^d \mapsto -\ln(1-\|x\|^2)$. It suffices to show that for all $\theta \in \mathbb{R}^d$,

$$\mathbb{E}_\xi[\nabla\phi_{\mathbb{B}^d}(\theta+\xi)] = \nabla\mathcal{R}^*(\theta) = \theta\cdot\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}\,.$$

For all $x \in \mathbb{R}^d \setminus \{0\}$, the differential of the support function of $\mathbb{B}^d$ in $x$ is

$$\nabla\phi_{\mathbb{B}^d}(x) = \underset{\|a\|\le 1}{\arg\max}\langle a, x\rangle = \frac{x}{\|x\|}\,.$$

Thus,

$$\mathbb{E}[\nabla\phi_{\mathbb{B}^d}(\theta+\xi)] = \mathbb{E}\left[\frac{\theta+\xi}{\|\theta+\xi\|}\right] = \mathbb{E}\left[\frac{\sqrt{d+1}\,U\theta+\mathbf{T}}{\|\sqrt{d+1}\,U\theta+\mathbf{T}\|}\right]\,.$$

Conditionally to $U$, $\sqrt{d+1}\,U\theta+\mathbf{T}$ follows a multivariate $t$-distribution with location $\sqrt{d+1}\,U\theta$, scale matrix $I_d$ and $d+1$ degrees of freedom. Hence, by Lemma 14,

$$\mathbb{E}\left[\frac{\sqrt{d+1}\,U\theta+\mathbf{T}}{\|\sqrt{d+1}\,U\theta+\mathbf{T}\|}\,\middle|\, U\right] = \frac{\sqrt{d+1}\,U\,\theta}{\sqrt{d+1+(d+1)U^2\|\theta\|^2}} = \frac{U\,\theta}{\sqrt{1+U^2\|\theta\|^2}}$$

Finally, by tower rule,

$$\begin{aligned}\mathbb{E}[\nabla\phi_{\mathbb{B}^d}(\theta+\xi)] &= \mathbb{E}\left[\frac{U\,\theta}{\sqrt{1+U^2\|\theta\|^2}}\right]\\ &= \theta\int_0^1\frac{u}{\sqrt{1+\|\theta\|^2u^2}}du\\ &= \theta\,\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}\,,\end{aligned}$$

which concludes the proof. ■

### F.2. Proof of Proposition 10

**Proof** Let $t \in [n]$. For the sake of brevity, we denote $\theta = -\eta\,\hat{Y}_{t-1}$ and $\xi = \xi_t$. The action chosen at time $t$ by SC-FTPL is then $a = \nabla\phi_{\mathbb{B}^d}(\theta+\xi)$, and its covariance matrix is

$$Q = \mathbb{E}_{t-1}[aa^\top] = \mathbb{E}\left[\frac{(\theta+\xi)(\theta+\xi)^\top}{\|\theta+\xi\|^2}\right]\,.$$

The expectation is with respect to $\xi$, conditionally on the $t-1$ first rounds. We also denote $y = y_t$ the loss vector at time $t$ and $\hat{y} = Q^{-1}aa^\top y$ the estimator. Finally, we recall that the norm $\|\cdot\|_t$ is the energy norm relatively to the matrix $\nabla^2\mathcal{R}^*(\theta)$.

**Proof Outline.** In Step 1, we use the rotational symmetry of the distribution of $\xi$ to show that $Q$ belongs to the linear span of $(P_\theta, P_{\theta^\perp})$, where $P_\theta$ and $P_{\theta^\perp}$ are the orthogonal projection matrices onto $\theta$ and $\theta^\perp$, respectively. In Step 2, we develop the expression of the eigenvalue $q_\perp$ corresponding to $P_{\theta^\perp}$ and show that it can be reduced to a one-dimensional integral expression. Subsequently, in Step 3 we derive lower bounds on both $q$ and $q_\perp$, the eigenvalues corresponding to $P_\theta$ and $P_{\theta^\perp}$, respectively, using the integral expression of $q_\perp$ and an affine relation between $q$ and $q\ \perp$. Step 4 provides the explicit expression for $Q^{-1}$ and for the Hessian $\nabla^2\mathcal{R}^*(\theta)$. In Steps 5, 6, and 7, we combine the explicit expressions for $Q^{-1}$ and $\nabla^2\mathcal{R}^*(\theta)$ with the bounds on $q$ and $q_\perp$ to derive the claimed bounds on the estimator $\hat{y}$. Finally, in Step 8, we analyze the degenerate case $\theta = 0$, where $P_\theta$ is not well-defined, and show that the inequalities still hold.

**Step 1: Proving that $Q \in \operatorname{span}(P_\theta, P_{\theta^\perp})$.** By the stochastic representation of the multivariate $t$-distribution, for independent random variables $\mathbf{N} \sim \mathcal{N}(0, I_d)$, $Z \sim \chi_{d+1}$, and $U \sim \mathcal{U}([0,1])$, the random vector $\mathbf{N}/(Z\,U)$ follows the same distribution as $\xi$. So

$$Q = \mathbb{E}\left[\frac{(ZU\theta + \mathbf{N})(ZU\theta + \mathbf{N})^\top}{\|ZU\theta + \mathbf{N}\|^2}\right]. \tag{23}$$

Conditioning on $Z$ and $U$, $ZU\theta + \mathbf{N}$ follows a multivariate normal distribution with mean $ZU\theta$ and isotropic covariance. Due to the rotational symmetry of this normal distribution around the axis defined by $ZU\theta$, it is a known result (Herrera-Esposito and Burge, 2025) that

$$\mathbb{E}\left[\left.\frac{(ZU\theta + \mathbf{N})(ZU\theta + \mathbf{N})^\top}{\|ZU\theta + \mathbf{N}\|^2}\right| Z, U\right] \in \operatorname{span}((ZU\theta)(ZU\theta)^\top, I_d) = \operatorname{span}(\theta\theta^\top, I_d)\,.$$

Then, by tower rule and linearity of the expectation,

$$Q = \mathbb{E}\left[\mathbb{E}\left[\left.\frac{(ZU\theta + \mathbf{N})(ZU\theta + \mathbf{N})^\top}{\|ZU\theta + \mathbf{N}\|^2}\right| Z, U\right]\right] \in \operatorname{span}(\theta\theta^\top, I_d).$$

Let

$$P_\theta \coloneqq \frac{\theta\theta^\top}{\|\theta\|^2}\,, \quad \text{and} \quad P_{\theta^\perp} \coloneqq I_d - P_\theta$$

be the orthogonal projection matrices onto $\theta$ and $\theta^\perp$, respectively. Then, $\operatorname{span}(\theta\theta^\top, I_d) = \operatorname{span}(P_\theta, P_{\theta^\perp})$. Hence, there exists $q, q_\perp \in \mathbb{R}$ such that

$$Q = q\,P_\theta + q_\perp\,P_{\theta^\perp}\,. \tag{24}$$

**Step 2: Developing $q_\perp$.** We will now develop $q_\perp$ in order to derive upper and lower bounds. First notice that we only need to study one among $q$ an $q_\perp$ as it holds that

$$\operatorname{tr} Q = q\operatorname{tr}(P_\theta) + q_\perp\operatorname{tr}(P_{\theta^\perp}) = q + (d-1)q_\perp$$

and

$$\operatorname{tr} Q = \operatorname{tr}\mathbb{E}_{t-1}[aa^\top] = \mathbb{E}_{t-1}[\operatorname{tr}(aa^\top)] = \mathbb{E}_{t-1}[\|a\|^2] = 1\,.$$

Hence,

$$q = 1 - (d-1)q_\perp\,. \tag{25}$$

Now let $y \in \theta^\perp$ be such that $\|y\| = 1$. Multiplying (23) by $y$ on the left and on the right, it holds that

$$q_\perp = y^\top Q y = \mathbb{E}\left[\frac{(y^\top \mathbf{N})^2}{\|ZU\theta + \mathbf{N}\|^2}\right]\,.$$

To evaluate this expectation, we can always rotate our coordinate system without loss of generality so that $\theta = \|\theta\| e_1$ and $y = e_2$. Then, since $\mathbf{N} \sim \mathcal{N}(0, I_d)$ is isotropically distributed, the rotation preserves its distribution. We thus have that that

$$y^\top \mathbf{N} = \mathbf{N}_2 \quad \text{and} \quad ZU\theta + \mathbf{N} = (ZU\|\theta\| + \mathbf{N}_1)e_1 + \sum_{k=2}^{d} \mathbf{N}_k e_k$$

The expectation thus becomes

$$q_\perp = \mathbb{E}\Big[\frac{\mathbf{N}_2^2}{(ZU\|\theta\| + \mathbf{N}_1)^2 + \sum_{k=2}^d \mathbf{N}_k^2}\Big]\,.$$

Denote $D$ the denominator in the previous expression. To handle it, we use the integral identity

$$\frac{1}{D} = \int_0^{+\infty} e^{-vD} dv\,.$$

Conditioned on $Z, U$, we have that

$$\begin{aligned}
\mathbb{E}\Big[\frac{\mathbf{N}_2^2}{D}\,\Big|\, Z, U\Big] &= \mathbb{E}\Big[\int_0^{+\infty} e^{-v(ZU\|\theta\|+\mathbf{N}_1)^2}\, \mathbf{N}_2^2 e^{-v\mathbf{N}_2^2} \prod_{k=3}^{d} e^{-v\mathbf{N}_k^2}\, dv\Big] \\
&= \int_0^{+\infty} \mathbb{E}[e^{-v(ZU\|\theta\|+\mathbf{N}_1)^2}]\, \mathbb{E}[N_2^2 e^{-v\mathbf{N}_2^2}] \prod_{k=3}^{d} \mathbb{E}[e^{-v\mathbf{N}_k^2}]\, dv
\end{aligned}$$

where the second line comes from the independence of the coordinates of $\mathbf{N}$. Evaluating each of these standard Gaussian expectation yields

$$\mathbb{E}\Big[\frac{\mathbf{N}_2^2}{D}\,\Big|\, Z, U\Big] = \int_0^{+\infty} (1+2v)^{-d/2-1} \exp\Big(-\frac{vZ^2U^2\|\theta\|^2}{1+2v}\Big) dv\,.$$

Making the change of variable $s = \frac{2v}{1+2v}$, we can simplify the integral to

$$\mathbb{E}\Big[\frac{\mathbf{N}_2^2}{D}\,\Big|\, Z, U\Big] = \frac{1}{2}\int_0^1 (1-s)^{d/2-1} e^{-\frac{1}{2}Z^2U^2\|\theta\|^2 s} ds$$

Then, by tower rule,

$$q_\perp = \frac{1}{2}\int_0^1 (1-s)^{d/2-1} \mathbb{E}[e^{-\frac{1}{2}Z^2U^2\|\theta\|^2 s}]\, ds$$

Conditioning on $U$, $Z \sim \chi^2_{d+1}$ so by expression of its moment-generating function,

$$\mathbb{E}[e^{-\frac{1}{2}Z^2U^2\|\theta\|^2 s} | U] = (1 + U^2\|\theta\|^2 s)^{-(d+1)/2}$$

which gives

$$q_\perp = \frac{1}{2}\int_0^1 (1-s)^{d/2-1}\int_0^1 (1+u^2\|\theta\|^2 s)^{-(d+1)/2}\,du\,ds$$

These integrals are converging and using properties of generalized hypergeometric functions (Gradshteyn and Ryzhik, 2007), one can show that

$$\begin{aligned}
q_\perp &= \frac{1}{2}\int_0^1 (1-s)^{d/2-1}\,{}_2F_1\Big(\frac{1}{2},\frac{d+1}{2};\frac{3}{2};-\|\theta\|^2 s\Big)\,ds\\
&= \frac{1}{d}\,{}_3F_2\Big(\frac{1}{2},1,\frac{d+1}{2};\frac{3}{2},\frac{d}{2}+1;-\|\theta\|^2\Big)\\
&= \frac{\Gamma(\frac{d}{2}+1)}{d\sqrt{\pi}\Gamma(\frac{d}{2})}\int_0^1 s^{\frac{d+3}{2}}(1-s)^{-\frac{1}{2}}\,{}_2F_1\Big(\frac{1}{2},1;\frac{3}{2};-\|\theta\|^2 s\Big)\,ds\,.
\end{aligned}$$

Finally, using the closed-form expression of ${}_2F_1(\frac{1}{2},1;\frac{3}{2};\cdot)$ and the change of variable $s=\sin^2\phi$, we get

$$q_\perp = \frac{\Gamma(\frac{d}{2})}{\|\theta\|\sqrt{\pi}\Gamma(\frac{d+1}{2})}\int_0^{\pi/2}\sin^{d-1}(\phi)\arctan(\|\theta\|\sin\phi)\,d\phi\,. \tag{26}$$

**Step 3: Bounding $q_\perp$.** We now want to lower-bound both $q$ and $q_\perp$, using the integral expression of $q_\perp$ in (26). First, notice that for all $x\geq 0$, $\arctan x\leq x$. Thus, we have that

$$q_\perp \leq \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d+1}{2})}\int_0^{\pi/2}\sin^{d}(\phi)\,d\phi\,.$$

Evaluating the integral, we get

$$q_\perp \leq \frac{1}{d}\,.$$

Now, using Equation (25) linking $q$ and $q_\perp$,

$$q \geq 1-\frac{d-1}{d} = \frac{1}{d}\,. \tag{27}$$

Second, to lower-bound $q_\perp$, note that for all $x\geq 0$, $\arctan x \geq \frac{x}{1+x}$. Thus,

$$q_\perp \geq \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d+1}{2})}\int_0^{\pi/2}\frac{\sin^{d}(\phi)}{1+\|\theta\|\sin\phi}\,d\phi\,.$$

The denominator of the integrand is smaller than $1+\|\theta\|$. Hence,

$$q_\perp \geq \frac{1}{d(\|\theta\|+1)}\,. \tag{28}$$

**Step 4: Expression of $Q^{-1}$ and $\nabla^2\mathcal{R}^*(\theta)$.** We recall that we have

$$Q = q\,P_\theta + q_\perp\,P_{\theta^\perp}.$$

In Steps 2 and 3, we have shown that $q$ and $q_\perp$ are positive. Moreover, because the matrices $P_\theta$ and $P_{\theta^\perp}$ satisfy the following identities

$$P_\theta^2 = P_\theta, \qquad P_{\theta^\perp}^2 = P_{\theta^\perp}, \qquad P_\theta P_{\theta^\perp} = P_{\theta^\perp} P_\theta = 0, \qquad P_\theta + P_{\theta^\perp} = I_d,$$

we have that

$$Q^{-1} = q^{-1}\,P_\theta + q_\perp^{-1}\,P_{\theta^\perp}.$$

For the Hessian of $\mathcal{R}^*$ in $\theta$, we recall that $\nabla\mathcal{R}^*(\theta)$ is a radial function with magnitude $\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|}$. Thus,

$$\nabla^2\mathcal{R}^*(\theta) = \frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2\sqrt{1+\|\theta\|^2}}\,P_\theta + \frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}\,P_{\theta^\perp}.$$

**Step 5: Bounding $\mathbb{E}[\|\hat{y}\|_t^2]$.** By definition of $\hat{y}$,

$$\hat{y} = Q^{-1}a\langle y, a\rangle.$$

Because the loss $\langle y, a\rangle$ is bounded by 1 almost surely, we have that

$$\begin{aligned}
\mathbb{E}[\|\hat{y}\|_t^2] &= \mathbb{E}[\langle y, a\rangle a^\top Q^{-1}\nabla^2\mathcal{R}^*(\theta)Q^{-1}a\langle y, a\rangle] \\
&\le \mathbb{E}[a^\top Q^{-1}\nabla^2\mathcal{R}^*(\theta)Q^{-1}a] \\
&= \mathrm{tr}(Q^{-1}\nabla^2\mathcal{R}^*(\theta)).
\end{aligned}$$

Plugging the expressions of $Q^{-1}$ and $\nabla^2\mathcal{R}^*(\theta)$ into this last inequality, we get

$$\begin{aligned}
\mathbb{E}[\|\hat{y}\|_t^2] &\le q^{-1}\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2\sqrt{1+\|\theta\|^2}}\ \mathrm{tr}(P_\theta) + q_\perp^{-1}\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}\ \mathrm{tr}(P_{\theta^\perp}) \\
&\le d\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2\sqrt{1+\|\theta\|^2}} + d(d-1)(\|\theta\|+1)\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}
\end{aligned}$$

We now need to bound the functions

$$f(t) = \frac{\sqrt{1+t^2}-1}{t^2\sqrt{1+t^2}} \quad \text{and} \quad g(t) = \frac{(t+1)(\sqrt{1+t^2}-1)}{t^2},$$

for all $t \ge 0$. Define

$$h(t) = t^2\sqrt{1+t^2} - 2(\sqrt{1+t^2}-1), \quad t \ge 0.$$

The derivative of $h$ is

$$h'(t) = \frac{3t^3}{\sqrt{1+t^2}} \ge 0.$$

Thus, for all $t \ge 0$, $h(t) \ge h(0) = 0$. Thus, by rearranging the terms, for all $t \ge 0$,

$$f(t) \le \frac{1}{2}.$$

To bound $g(t)$, a routine calculation shows

$$g'(t) = \frac{t^3(t+1-\sqrt{1+t^2})}{t^4\sqrt{1+t^2}(\sqrt{1+t^2}+1)}\,.$$

The denominator is always non-negative. For the numerator, we split into two cases based on the sign of $t$. If $t \geq 0$, $t^3 \geq 0$ and $\sqrt{1+t^2} \leq 1+t$, so the numerator is non-negative. Else, if $t \leq 0$, $\sqrt{1+t^2} \geq 1$ so $t+1-\sqrt{1+t^2} \leq t$, and $t^3 \leq 0$, so the numerator is greater than $t^4$. Hence it is non-negative. So $g'(t) \geq 0$ for all $t \geq 0$. So $g$ is non-decreasing on $\mathbb{R}$. Moreover, $\lim_{t\to+\infty} g(t) = 1$, so for all $t \in \mathbb{R}$, $g(t) \leq 1$. Thus, we conclude

$$\mathbb{E}[\|\hat{y}\|_t^2] \leq \frac{1}{2}d + d(d-1) \leq d^2\,.$$

This inequality implies the stated inequality in $\frac{5}{4}d^2$.

**Step 6: Bounding $\|\hat{y}\|_t^2$ almost surely.** Because the loss $\langle y, a\rangle$ is bounded by 1 almost surely, we have that

$$\begin{aligned}
\|\hat{y}\|_t^2 &\leq a^\top Q^{-1}\nabla^2\mathcal{R}^*(\theta)Q^{-1}a \\
&= a^\top\left(q^{-2}\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2\sqrt{1+\|\theta\|^2}}\,P_\theta + q_\perp^{-2}\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2}\,P_{\theta^\perp}\right)a.
\end{aligned}$$

For any projection $P$, we can bound $a^\top Pa$ with $a^\top Pa = \|Pa\|^2 \leq \|a\|^2 = 1$. Thus, applying the bounds on $q$ and $q_\perp$,

$$\begin{aligned}
\|\hat{y}\|_t^2 &\leq d^2\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2\sqrt{1+\|\theta\|^2}} + d^2(\|\theta\|+1)^2\frac{\sqrt{1+\|\theta\|^2}-1}{\|\theta\|^2} \\
&\leq d^2 f(\|\theta\|) + d^2(\|\theta\|+1)g(\|\theta\|)
\end{aligned}$$

where $f$ and $g$ have been defined above. Using the bounds on $f$ and $g$ obtained previously, we get

$$\|\hat{y}\|_t^2 \leq \frac{1}{2}d^2 + d^2\left(\|\theta\|+1\right) \leq d^2\|\theta\| + \frac{3}{2}d^2\,,$$

which implies the stated bound.

**Step 7: Bounding $\mathbb{E}[\|\hat{y}\|_2^2]$.** We bound the variance of the $\ell_2$ norm of $\hat{y}$ by

$$\begin{aligned}
\mathbb{E}[\|\hat{y}\|_2^2] &= \mathbb{E}[\langle y, a\rangle a^\top Q^{-1}Q^{-1}a\langle y, a\rangle] \\
&\leq \mathbb{E}[a^\top Q^{-1}Q^{-1}a] \\
&= \mathrm{tr}(Q^{-1}).
\end{aligned}$$

Using the expression of $Q^{-1}$ and the bounds on $q$ and $q_\perp$, we bound the trace of $Q^{-1}$ by

$$\begin{aligned}
\mathrm{tr}(Q^{-1}) &= q^{-1}\ \mathrm{tr}(P_\theta) + q_\perp^{-1}\ \mathrm{tr}(P_{\theta^\perp}) \\
&\leq d + d(\|\theta\|+1)(d-1).
\end{aligned}$$

Hence,

$$\mathbb{E}[\|\hat{y}\|_2^2] \leq d^2\|\theta\|^2 + 2d^2.$$

**Step 8:** $\theta = 0$ **Case.** Until now we have assumed that $\theta \neq 0$. Indeed, we cannot define $P_\theta = \theta\theta^\top/\|\theta\|^2$ for $\theta = 0$. We now show that the required bounds continue to hold in the degenerate case $\theta = 0$. Since $\xi$ is spherically symmetric, $a = \xi/\|\xi\|$ follows a uniform distribution on $\mathbb{S}^{d-1}$. Thus, its covariance matrix is

$$Q = \frac{1}{d} I_d.$$

A direct computation shows that the Hessian of $\mathcal{R}^*$ in 0 is

$$\nabla^2 \mathcal{R}^*(0) = \frac{1}{2} I_d.$$

Then, $\mathbb{E}[\|\hat{y}_t\|_t^2]$ simplifies to

$$\mathbb{E}[\|\hat{y}\|_t^2] = \frac{d^2}{2}\mathbb{E}[y^\top a a^\top a a^\top y].$$

Since $a^\top a = 1$, $\mathbb{E}[aa^\top aa^\top] = Q$ and we have

$$\mathbb{E}[\|\hat{y}\|_t^2] = \frac{d}{2}\|y\|^2 \leq \frac{d}{2}.$$

For the almost-sure bound, we have similarly

$$\|\hat{y}\|_t^2 = \frac{d^2}{2}(y^\top a)^2 \leq \frac{d^2}{2}.$$

And for the $\ell_2$ norm bound,

$$\mathbb{E}[\|\hat{y}\|_2^2] = d^2\, \mathbb{E}[y^\top a a^\top a a^\top y] \leq d.$$

Since $\theta = 0$, the above bounds are consistent with the general inequalities. ■

### F.3. Proof of Theorem 3

**Proof**

**Step 1: Establishing that** $\eta\|\hat{y}_t\|_t \leq 1/2$ **holds with high-probability.** Define the event

$$E \coloneqq \{\forall t \in [n], \eta\, \|\hat{y}_t\|_t \leq 1/2\}\,.$$

From Proposition 10, we can bound the local norm of the estimator almost-surely by $\|\hat{y}_t\|_t^2 \leq d^2\,\eta\,\|\hat{Y}_{t-1}\| + 4d^2$ almost surely. It follows that

$$E \supseteq \left\{\forall t \in [n],\, \|\hat{Y}_{t-1}\| \leq \frac{1}{4\,d^2\,\eta^3} - \frac{4}{\eta}\right\}.$$

Using the expression of $\eta$ and the assumption that $\frac{n}{\ln n} \geq 64$,

$$\begin{aligned} E &\supseteq \left\{\sup_{0\leq t\leq n-1} \|\hat{Y}_t\|^2 \leq \frac{5}{4} d^2 \frac{n}{\ln n}\Big(\frac{5}{16}\frac{n}{\ln n} - 4\Big)^2\right\} \\ &\supseteq \left\{\sup_{0\leq t\leq n-1} \|\hat{Y}_t\|^2 \leq \frac{5}{64} d^2 \Big(\frac{n}{\ln n}\Big)^3\right\}. \end{aligned} \tag{29}$$

For all $t \in [n]$, we can bound $\|\hat{Y}_t\|^2$ in expectation using the third inequality of Proposition 10, which yields

$$\begin{aligned}\mathbb{E}_{t-1}[\|\hat{Y}_t\|_2] &= \|\hat{Y}_{t-1}\|^2 + \mathbb{E}_{t-1}[\|\hat{y}_t\|^2] + 2\langle y_t, \hat{Y}_{t-1}\rangle \\ &\le \|\hat{Y}_{t-1}\|^2 + d^2\eta\,\|\hat{Y}_{t-1}\| + 2d^2 + 2\|\hat{Y}_{t-1}\|.\end{aligned}$$

Using the assumption that $n/\ln n \ge 2d^2$, we have $d^2\eta \le \sqrt{2/5}$ and $2d^2 \le n$, so the bound becomes

$$\mathbb{E}_{t-1}[\|\hat{Y}_t\|_2] \le \|\hat{Y}_{t-1}\|^2 + \left(2 + \sqrt{\frac{2}{5}}\right)^2 \|\hat{Y}_{t-1}\| + n.$$

By the AM-GM inequality, this trinomial can be bounded by

$$\begin{aligned}\mathbb{E}_{t-1}[\|\hat{Y}_t\|_2] &\le \left(1 + \frac{1}{n}\right)\|\hat{Y}_{t-1}\|^2 + \frac{1}{4}\left(2 + \sqrt{\frac{2}{5}}\right)^2 n + n \\ &\le \left(1 + \frac{1}{n}\right)\|\hat{Y}_{t-1}\|^2 + 3n. \end{aligned} \tag{30}$$

If we now define the positive, adapted, stochastic process

$$M_t \coloneqq \left(1 + \frac{1}{n}\right)^{-t}\left(\|\hat{Y}_t\|^2 + 3n^2\right),$$

then (30) bounds the conditional expectancy of $M_t$ by

$$\mathbb{E}_{t-1}[M_t] \le \left(1 + \frac{1}{n}\right)^{-t}\left(\left(1 + \frac{1}{n}\right)\|\hat{Y}_{t-1}\|^2 + 3n + 3n^2\right) = M_{t-1}\,.$$

Hence, $(M_t)_t$ is a supermartingale, adapted to the filtration $\{\mathcal{F}_t\}_t$. Therefore, by Ville's inequality, for all $\varepsilon \ge 0$,

$$\mathbb{P}\Big[\sup_{0\le t\le n-1} M_t \ge \varepsilon\Big] \le \frac{\mathbb{E}[M_0]}{\varepsilon} = \frac{3n^2}{\varepsilon}\,.$$

Moreover, for all $t \in [n]$, it holds that

$$\|\hat{Y}_t\|^2 = \left(1 + \frac{1}{n}\right)^t M_t - 3n^2 \le e\,M_t\,.$$

Combining this last inequality with (29), we get

$$E \supseteq \left\{\sup_{0\le t\le n-1} M_t \le \frac{5}{64\,e} d^2\Big(\frac{n}{\ln n}\Big)^3\right\}.$$

Hence,

$$\mathbb{P}(E^c) \le \mathbb{P}\Big[\sup_{0\le t<n} \|\hat{Y}_t\|^2 > \frac{5}{64} d^2\Big(\frac{n}{\ln n}\Big)^3\Big] \le \frac{192\,e}{5}\frac{(\ln n)^3}{d^2\,n}\,. \tag{31}$$

**Step 2: Bounding the regret $R_n$.** In order to bound the regret using similar arguments as in Theorem 5, we introduce an appropriate stopping time and invoke Doob's optional stopping theorem to bound the regret conditionally on $E$. Define the random variable

$$\tau \coloneqq \inf\Big\{t < n : \|\hat{Y}_t\|^2 > \frac{5}{64}d^2\Big(\frac{n}{\ln n}\Big)^3\Big\} \wedge n\,.$$

Because $(\hat{Y}_t)_t$ is a stochastic process adapted to the filtration $(\mathcal{F}_t)_t$, $\tau$ is a stopping time with respect to the same filtration. For all $u \in K$, we can decompose the regret of SC-FTPL w.r.t $u$ as

$$R_n(u) = \mathbb{E}\Big[\sum_{t=1}^{\tau}\langle y_t, a_t - u\rangle\Big] + \mathbb{E}\Big[\sum_{t=\tau+1}^{n}\langle y_t, a_t - u\rangle\Big]\,. \tag{32}$$

We now proceed to bound each term of the right-hand side of (32) separately. For the second term, we have

$$\begin{aligned}
\mathbb{E}\Big[\sum_{t=\tau+1}^{n}\langle y_t, a_t - u\rangle\Big] &\le \mathbb{P}(\tau < n)\,\mathbb{E}\Big[\sum_{t=1}^{n}|\langle y_t, a_t - u\rangle|\Big|\tau < n\Big] \\
&\le \mathbb{P}\Big[\sup_{0\le t<n}\|\hat{Y}_t\|^2 > \frac{5}{64}d^2\Big(\frac{n}{\ln n}\Big)^3\Big]\;2n \\
&\le \frac{384\,e}{5\,d^2}\;\ln^3 n\,,
\end{aligned} \tag{33}$$

where the second inequality follows from the boundedness of the loss vectors $y_t$ and the last inequality follows from (31). For the first term, let

$$B_t \coloneqq \sum_{s=1}^{t}\langle y_s, a_s - u\rangle - \langle \hat{y}_s, x_s - u\rangle\,.$$

The process $(B_t)_t$ is adapted to the filtration $(\mathcal{F}_t)_t$ and for all $t \in [n]$,

$$\mathbb{E}_{t-1}[B_t] = B_{t-1} + \langle y_t, \mathbb{E}_{t-1}[a_t - u]\rangle - \langle \mathbb{E}_{t-1}[\hat{y}_t], x_t - u\rangle = B_{t-1}\,.$$

Therefore, $(B_t)_t$ is a martingale with $B_0 = 0$. Thus, by Doob's optional stopping theorem, $\mathbb{E}[B_\tau] = 0$, i.e.,

$$\mathbb{E}\Big[\sum_{t=1}^{\tau}\langle y_t, a_t - u\rangle\Big] = \mathbb{E}\Big[\sum_{t=1}^{\tau}\langle \hat{y}_t, x_t - u\rangle\Big]\,.$$

Moreover, for all $t \le \tau$, we have that $\|\hat{Y}_{t-1}\|^2 \le \frac{d^2}{8}(\frac{n}{\ln n})^3$, and thus $\eta\|\hat{y}_t\|_t \le \frac{1}{2}$. Thus, we can bound $\sum_{t=1}^{\tau}\mathbb{E}[\langle \hat{y}_t, x_t - u\rangle]$ following the arguments from the proofs of Lemma 13 and of Theorem 5, which yield

$$\mathbb{E}\Big[\sum_{t=1}^{\tau}\langle \hat{y}_t, x_t - u\rangle\Big] \le \frac{\vartheta \ln n}{\eta} + \eta\sum_{t=1}^{n}\mathbb{E}[\|\hat{y}_t\|_t^2] + 2\,. \tag{34}$$

Note that Theorem 5 could not have been applied directly, as we needed the bound $\eta\|\hat{y}_t\|_t \le \frac{1}{2}$ to hold almost-surely. Also, conditioning the expectation in the regret definition by the event $E$ would not have enabled to use the unbiasedness of $a_t$ and $\hat{y}_t$, which justifies the need for the stopping time argument we have developed. Combining (33) and (34), we have that

$$\begin{aligned}R_n(u) &\leq \frac{\vartheta \ln n}{\eta} + \eta \sum_{t=1}^{n} \mathbb{E}[\|\hat{y}_t\|_t^2] + 2 + \frac{384\, e}{5} \frac{\ln^3 n}{d^2} \\ &\leq \frac{\ln n}{\eta} + \frac{5}{4} \eta\, n\, d^2 + 2 + \frac{384\, e}{5} \frac{\ln^3 n}{d^2}\,,\end{aligned}$$

where the last inequality follows from the value of the self-concordance parameter $\vartheta = 1$ and from the bound on $\mathbb{E}[\|\hat{y}_t\|_t^2]$ of Proposition 10. Finally, plugging the value of $\eta$ and taking the supremum over all $u \in K$, we have

$$R_n \leq d\sqrt{5\, n \ln n} + 2 + \frac{384\, e}{5} \frac{\ln^3 n}{d^2}\,.$$

■

### F.4. Complexity Analysis of SC-FTPL on the $\ell_2$ Ball

At each round $t \in [n]$, SC-FTPL performs theses three main steps:

1. Sampling $\xi_t \sim \mathcal{D}$,
2. Solving the linear program $a_t = \arg\min_{a \in \mathbb{B}^d} \langle a, \eta \hat{Y}_{t-1} - \xi_t \rangle$, and
3. Computing the estimator $\hat{y}_t = Q_t^{-1} a_t \langle y_t, a_t \rangle$.

Sampling from $\mathcal{D}$ given in Proposition 9 reduces to drawing $\mathbf{T}$ from a multivariate $t$-distribution with location 0, scale matrix $I_d$ and $d+1$ degrees of freedom, and $U$ from a uniform distribution on $[0, 1]$ independently of $\mathbf{T}$, and to compute

$$\xi_t = \frac{1}{\sqrt{d+1}\, U} \mathbf{T}.$$

By the stochastic representation of the multivariate $t$-distribution (Kotz and Nadarajah, 2004), sampling $\mathbf{T}$ amounts to sampling $\mathbf{N} \sim \mathcal{N}(0, I_d)$ and $Z \sim \chi_{d+1}$ independently and computing

$$\mathbf{T} = \frac{\sqrt{d+1}}{Z} \mathbf{N}.$$

Sampling from a standard multivariate normal distribution can be done in $\mathcal{O}(d)$ operations, while sampling from a $\chi_{d+1}$ distribution or from an uniform distribution on $[0, 1]$ can be done in constant times. Hence, the overall sampling step requires $\mathcal{O}(d)$ time.

The linear program over the Euclidean ball admits the closed-form solution

$$a_t = \left[ \frac{-\eta \hat{Y}_{t-1,i} + \xi_{t,i}}{\| - \eta \hat{Y}_{t-1} + \xi_t \|} \right]_{i=1}^{d}.$$

The norm of $-\eta \hat{Y}_{t-1} + \xi_t$ can be computed in $\mathcal{O}(d)$ operations, after which each coordinate of $a_t$ can be computed in constant time, yielding overall $\mathcal{O}(d)$ complexity.

For the estimator, from the proof of Proposition 10, we have that

$$Q_t^{-1} = q^{-1}\, P_\theta + q_\perp^{-1}\, P_{\theta^\perp},$$

where

$$\theta = -\eta \hat{Y}_{t-1}, \qquad P_\theta = \frac{\theta\theta^\top}{\|\theta\|^2}, \qquad P_{\theta^\perp} = I_d - P_\theta$$

and $q$ and $q_\perp$ are given by

$$q = 1 - (d-1)q_\perp, \qquad q_\perp = \frac{\Gamma(\frac{d}{2})}{\|\theta\|\sqrt{\pi}\Gamma(\frac{d+1}{2})} \int_0^{\pi/2} \sin^{d-1}(\phi) \arctan(\|\theta\| \sin\phi)\, d\phi\,.$$

A closed-form expression for $q_\perp$ would involve the generalized hypergeometric function ${}_3F_2$, for which there exists no general numerical evaluation method. Instead, this integral expression can be computed using Gauss-Legendre quadrature. Because the integrand in analytic, this method yields exponential convergence.

Once $q_\perp$ has been numerically computed, $q$ follows and

$$Q_t^{-1} a_t = q_\perp^{-1} a_t + (q^{-1} - q_\perp^{-1})\, \frac{\langle a_t, \theta\rangle}{\|\theta\|^2}\theta.$$

So $Q_t^{-1} a_t$ can be computed in $\mathcal{O}(d)$ time knowing $q$ and $q_\perp$.

Combining all steps, the per-round computational complexity of SC-FTPL on the $\ell_2$-ball is $\mathcal{O}(d)$.