

# Thompson Sampling for Bandit Convex Optimisation

**Alireza Bakhtiari**

*University of Alberta*

**Tor Lattimore**

*Google DeepMind*

**Csaba Szepesvári**

*Google DeepMind and University of Alberta*

SBAKHTIA@UALBERTA.CA

LATTIMORE@GOOGLE.COM

SZEPI@GOOGLE.COM

**Editors:** Nika Haghtalab and Ankur Moitra

## Abstract

We show that Thompson sampling has a Bayesian regret of at most  $\tilde{O}(\sqrt{n})$  for 1-dimensional bandit convex optimisation where  $n$  is the time horizon and no assumptions are made on the loss function beyond convexity, boundedness and a mild Lipschitz assumption. For general high-dimensional problems we show that Thompson sampling can fail catastrophically. More positively, we show that Thompson sampling has Bayesian regret of  $\tilde{O}(d^{2.5}\sqrt{n})$  for generalised linear bandits with an unknown convex monotone link function. Lastly, we prove that the standard information-theoretic machinery can never give a bound on the regret in the general case that improves on the best known bound of  $\tilde{O}(d^{1.5}\sqrt{n})$ .

**Keywords:** Bandit convex optimisation; Thompson sampling.

## 1. Introduction

Thompson sampling (TS) is a simple and often practical algorithm for interactive decision-making with a long history (Thompson, 1933; Russo et al., 2018). Our interest is in its application to (Bayesian) bandit convex optimisation (Lattimore, 2024). Let  $K$  be a convex body in  $\mathbb{R}^d$  and  $\mathcal{F}$  be a set of convex functions from  $K$  to  $[0, 1]$ . We assume there is a known (prior) probability measure  $\xi$  on  $\mathcal{F}$ . The interaction between the learner and environment lasts for  $n$  rounds. At the beginning the environment secretly samples  $f$  from the prior  $\xi$ . Subsequently the learner and environment interact sequentially. In round  $t$  the learner plays an action  $X_t \in K$  and observes  $Y_t \in \{0, 1\}$  for which  $\mathbb{E}[Y_t | X_1, Y_1, \dots, X_t, f] = f(X_t)$ . The assumption that the noise is Bernoulli is for convenience only. Our analysis would be unchanged with any bounded noise model and would continue to hold for subgaussian noise with minor modifications. A learner  $\mathcal{A}$  is a (possibly random) mapping from sequences of action/loss pairs to actions and its Bayesian regret with respect to prior  $\xi$  is

$$\text{BReg}_n(\mathcal{A}, \xi) = \mathbb{E} \left[ \sup_{x \in K} \sum_{t=1}^n (f(X_t) - f(x)) \right].$$

Note that both  $f$  and the iterates  $(X_t)$  are random elements. Moreover, in the Bayesian setting the learner  $\mathcal{A}$  is allowed to depend on the prior  $\xi$ . The main quantity of interest is

$$\sup_{\xi \in \mathcal{P}(\mathcal{F})} \text{BReg}_n(\text{TS}, \xi), \tag{1}$$

where  $\mathcal{P}(\mathcal{F})$  is the space of probability measures on  $\mathcal{F}$  (with a suitable  $\sigma$ -algebra) and TS is Thompson sampling (Algorithm 1) with prior  $\xi$  (the dependence on the prior is always omitted from the notation). The quantity in Eq. (1) depends on the function class  $\mathcal{F}$ . Our analysis explores this dependence for various natural classes of convex functions. TS (Algorithm 1) is theoretically near-trivial. In every round it samples  $f_t$  from the posterior and plays  $X_t$  as the minimizer of  $f_t$ .

**Contributions** Our main contribution is showing the following:

- When  $d = 1$ ,  $\text{BReg}_n(\text{TS}, \xi) = \tilde{O}(\sqrt{n})$  for all priors (Theorem 4).
- A convex function  $f$  is called a monotone ridge function if there exists a convex monotone function  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  and  $\theta \in \mathbb{R}^d$  such that  $f(x) = \ell(\langle x, \theta \rangle)$ . Theorem 6 shows when  $\xi$  is supported on monotone ridge functions, then  $\text{BReg}_n(\text{TS}, \xi) = \tilde{O}(d^{2.5}\sqrt{n})$ .
- In general, the Bayesian regret of TS can be exponential in the dimension (Theorem 11).
- The classical information-theoretic machinery used by [Bubeck and Eldan \(2018\)](#) and [Lattimore \(2020\)](#) cannot improve the regret for bandit convex optimisation beyond the best known upper bound of  $\tilde{O}(d^{1.5}\sqrt{n})$ .

Although the regret bounds are known already in the frequentist setting for different algorithms, there is still value in studying Bayesian algorithms and especially TS. Most notably, none of the frequentist algorithms can make use of prior information about the loss functions and adapting them to exploit such information is often painstaking and ad-hoc. TS, on the other hand, automatically exploits prior information. Our bounds for ridge functions can be viewed as a Bayesian regret bound for a kind of generalised linear bandit where the link function is unknown and assumed to be convex and monotone increasing. This choice is a natural fit for a variety of problems such as resource allocation, which we discuss in (a little) more detail in the discussion.

Our lower bounds show that TS does not behave well in the general bandit convex optimisation unless possibly the dimension is quite small. Perhaps more importantly, we show that the classical information-theoretic machinery used by [Bubeck and Eldan \(2018\)](#) and [Lattimore \(2020\)](#) cannot be used to improve the current best dimension dependence of the regret for bandit convex optimisation. Combining this with the duality between exploration-by-optimisation and information-directed sampling shows that exploration-by-optimisation (with negentropy potential) also cannot naively improve on the best known  $\tilde{O}(d^{1.5}\sqrt{n})$  upper bound ([Zimmert and Lattimore, 2019](#); [Lattimore and György, 2023](#)). We note that this does not imply a lower bound for bandit convex optimisation. The construction in the lower bound is likely amenable to methods for learning a direction based on the power method ([Lattimore and Hao, 2021](#); [Huang et al., 2021](#)). The point is that the information ratio bound characterises the signal-to-noise ratio for the prior, but does not prove the signal-to-noise ratio does not increase as the learner gains information.

**Related work** Bandit convex optimisation in the regret setting was first studied by [Flaxman et al. \(2005\)](#) and [Kleinberg \(2005\)](#). Since then the field has grown considerably as summarised in the recent monograph by [Lattimore \(2024\)](#). Our focus is on the Bayesian version of the problem, which has seen only limited attention. [Bubeck et al. \(2015\)](#) consider the adversarial version of the Bayesian regret and show that a (heavy) modification of TS enjoys a Bayesian regret of  $\tilde{O}(\sqrt{n})$  when  $d = 1$ . Interestingly, they argue that TS without modification is not amenable to analysis via the information-theoretic machinery, but this argument only holds in the adversarial setting as our analysis shows. [Bubeck and Eldan \(2018\)](#) and [Lattimore \(2020\)](#) generalised the information-theoretic machinery used by [Bubeck et al. \(2015\)](#) to higher dimensions, also in the adversarial setting. These works focus on proving bounds for information-directed sampling (IDS), which is a conceptually simple but computationally more complicated algorithm introduced by [Russo and Van Roy \(2014\)](#). Nevertheless, we borrow certain techniques from these papers. Convex ridge functions have been studied before by [Lattimore \(2021\)](#), who showed that IDS has a Bayesian regret in the

stochastic setting of  $\tilde{O}(d\sqrt{n})$ , which matches the lower bound provided by linear bandits (Dani et al., 2008). Regrettably, however, this algorithm is not practically implementable, even under the assumption that you can sample efficiently from the posterior. Saha et al. (2021) also study a variation on the problem where the losses have the form  $f(g(x))$  with  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  a *known* function and  $f : \mathbb{R} \rightarrow \mathbb{R}$  an unknown convex function. When  $g$  is linear, then  $f \circ g$  is a convex ridge function. The assumption that  $g$  is known dramatically changes the setting, however. The best known bound for an efficient algorithm in the monotone convex ridge function setting is  $\tilde{O}(d^{1.5}\sqrt{n})$ , which also holds for general convex functions, even in the frequentist setting (Fokkema et al., 2024). Convex ridge functions can also be viewed as a special case of the generalised linear model, which has been studied extensively as a reward model for stochastic bandits (Filippi et al., 2010, and many more). TS and other randomised algorithms have been studied with generalised linear models in Bayesian and frequentist settings (Abeille and Lazaric, 2017; Dong et al., 2019; Kveton et al., 2020). None of these papers assume convexity (concavity for rewards) and consequently suffer a regret that depends on other properties of the link function that can be arbitrarily large. Moreover, in generalised linear bandits it is standard to assume the link function is known.

**Notation** Let  $\|\cdot\|$  be the standard euclidean norm on  $\mathbb{R}^d$ . For natural number  $k$  let  $[k] = \{1, \dots, k\}$ . Define  $\|x\|_\Sigma = \sqrt{x^\top \Sigma x}$  for positive definite  $\Sigma$ . Given a function  $f : K \rightarrow \mathbb{R}$ , let  $\|f\|_\infty = \sup_{x \in K} f(x)$ . The centered euclidean ball of radius  $r > 0$  is  $\mathbb{B}_r = \{x \in \mathbb{R}^d : \|x\| \leq r\}$  and the sphere is  $\mathbb{S}_r = \{x \in \mathbb{R}^d : \|x\| = r\}$ . We also let  $\mathbb{B}_r(x) = \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$ . We let  $H(x, \eta) = \{y : \langle y, \eta \rangle \geq \langle x, \eta \rangle\}$ , which is a half-space with inward-facing normal  $\eta$ . Given a finite set  $\mathcal{C}$  let  $\text{PAIR}(\mathcal{C}) = \{(x, y) \in \mathcal{C} : x \neq y\}$  be the set of all distinct ordered pairs and abbreviate  $\text{PAIR}(k) = \text{PAIR}([k])$ . The convex hull of a subset  $A$  of a linear space is  $\text{conv}(A)$ . The space of probability measures on  $K$  with respect to the Borel  $\sigma$ -algebra is  $\mathcal{P}(K)$ . Similarly,  $\mathcal{P}(\mathcal{F})$  is a space of probability measures on  $\mathcal{F}$  with some unspecified  $\sigma$ -algebra ensuring that  $f \mapsto f(x)$  is measurable for all  $x \in K$ . Given a convex function  $f : K \rightarrow \mathbb{R}$  we define  $\text{Lip}_K(f) = \sup_{x \neq y \in K} (f(x) - f(y)) / \|x - y\|$  and  $f_\star = \inf_{x \in K} f(x)$  and  $x_f = \arg \min_{x \in K} f(x)$  where ties are broken in an arbitrary measurable fashion. Such a mapping exists and  $f \mapsto f_\star$  is also measurable as we explain in Appendix B.  $\mathbb{P}_t = \mathbb{P}(\cdot | X_1, Y_1, \dots, X_t, Y_t)$  and  $\mathbb{E}_t$  be the expectation operator with respect to  $\mathbb{P}_t$ . The following assumption on  $K$  is considered global throughout:

**Assumption 1**  $K$  is a convex body (compact, convex with non-empty interior) and  $\mathbf{0} \in K$ .

**Spaces of convex functions** A function  $f : K \rightarrow \mathbb{R}$  is called a convex ridge function if there exists a convex  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  and  $\theta \in \mathbb{R}^d$  such that  $f(x) = \ell(\langle x, \theta \rangle)$ . Moreover,  $f$  is called a monotone convex ridge function if it is a convex ridge function and  $\ell$  is monotone increasing. We are interested in the following classes of convex functions: (a)  $\mathcal{F}_b$  is the space of all bounded convex functions  $f : K \rightarrow [0, 1]$ . (b)  $\mathcal{F}_1$  is the space of convex functions  $f : K \rightarrow \mathbb{R}$  with  $\text{Lip}(f) \leq 1$ . (c)  $\mathcal{F}_r$  is the space of all convex ridge functions. (d)  $\mathcal{F}_{rm}$  is the space of all monotone convex ridge functions. Intersections are represented as you might expect:  $\mathcal{F}_{b1} = \mathcal{F}_b \cap \mathcal{F}_1$  and similarly for other combinations. The set  $\mathcal{F}$  refers to a class of convex functions, which will always be either  $\mathcal{F}_{b1}$  or  $\mathcal{F}_{b1rm}$ .

## 2. Information Ratio and Thompson Sampling

TS is formally straightforward (Algorithm 1). You might wonder whether or not the algorithm is robust to approximation errors in sampling and finding the minimiser of the sampled function. In

Appendix F we introduce and analyse an approximate version of TS showing that indeed a certain kind of robustness does exist.

```

1  args: prior  $\xi$ 
2  for  $t = 1$  to  $\infty$ :
3      sample  $f_t$  from  $\mathbb{P}(f = \cdot | X_1, Y_1, \dots, X_{t-1}, Y_{t-1})$ 
4      play  $X_t = x_{f_t}$  and observe  $Y_t$ 
    
```

**Algorithm 1:** Thompson sampling

**Information ratio** The main theoretical tool is a version of the information ratio, which was introduced by Russo and Van Roy (2016) as a means to bound the Bayesian regret of TS for finite-armed and linear bandits. Given a distribution  $\xi \in \mathcal{P}(\mathcal{F})$  and policy  $\pi \in \mathcal{P}(K)$ , let  $(X, f)$  have law  $\pi \otimes \xi$  and with  $\bar{f} = \mathbb{E}[f]$  define

$$\Delta(\pi, \xi) = \mathbb{E} [\bar{f}(X) - f_*] \quad \text{and} \quad \mathcal{I}(\pi, \xi) = \mathbb{E} [(f(X) - \bar{f}(X))^2],$$

which are both non-negative. The quantity  $\Delta(\pi, \xi)$  is the regret suffered by  $\pi$  when the loss function is sampled from  $\xi$ , while  $\mathcal{I}(\pi, \xi)$  is a measure of the observed variation of the loss function. Intuitively, if  $\mathcal{I}(\pi, \xi)$  is large, then the learner is gaining information. A classical version of the information ratio is  $\Delta(\pi, \xi)^2 / \mathcal{I}(\pi, \xi)$ , though the most standard version replaces  $\mathcal{I}(\pi, \xi)$  with an expected relative entropy term that is never smaller than  $\mathcal{I}(\pi, \xi)$  (Russo and Van Roy, 2016). Given a distribution  $\xi$  and a random function  $f$  with law  $\xi$ , we let  $\pi_{\text{TS}}^\xi \in \mathcal{P}(K)$  be the law of  $x_f$ , which is the minimiser of  $f$ . The minimax generalised information ratio associated with TS on class of loss functions  $\mathcal{F}$  is

$$\text{IR}(\mathcal{F}) = \left\{ (\alpha, \beta) \in \mathbb{R}_+^2 : \sup_{\xi \in \mathcal{P}(\mathcal{F})} \left[ \Delta(\pi_{\text{TS}}^\xi, \xi) - \alpha - \sqrt{\beta \mathcal{I}(\pi_{\text{TS}}^\xi, \xi)} \right] \leq 0 \right\}.$$

Note that  $(0, \beta) \in \text{IR}(\mathcal{F})$  is equivalent to  $\Delta(\pi_{\text{TS}}^\xi, \xi)^2 / \mathcal{I}(\pi_{\text{TS}}^\xi, \xi) \leq \beta$  for all  $\xi \in \mathcal{P}(\mathcal{F})$ . The  $\alpha$  term is used to allow a small amount of slack that eases analysis and may even be essential in non-parametric and/or infinite-action settings.

**Theorem 2** Suppose that  $\mathcal{F} \in \{\mathcal{F}_{\text{bl}}, \mathcal{F}_{\text{blrm}}\}$  and  $(\alpha, \beta) \in \text{IR}(\mathcal{F})$ . Then, for any prior  $\xi \in \mathcal{P}(\mathcal{F})$ , the regret of TS (Algorithm 1) is at most

$$\text{BReg}_n(\text{TS}, \xi) \leq n\alpha + O\left(\sqrt{\beta n d \log(n \text{diam}(K))}\right),$$

where the Big-O hides only a universal constant.

The proof of Theorem 2 is deferred to the Appendix F. At a high level the argument is based on similar results by Bubeck et al. (2015) and Bubeck and Eldan (2018). Also note that the space of ridge functions is not closed under convex combinations, which introduces certain challenges also noticed by Lattimore (2021).

**Decomposition lemma** We introduce a new mechanism for bounding  $\text{IR}(\mathcal{F})$ .

**Lemma 3** Suppose there exist natural numbers  $k$  and  $m$  such that for all  $\bar{f} \in \text{conv}(\mathcal{F})$  there exists a disjoint union  $\mathcal{F} = \cup_{i=1}^m \mathcal{F}_i$  of measurable sets for which

$$\max_{i \in [m]} \left[ \sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f_\star) - \alpha - \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2} \right] \leq 0.$$

Then  $(\alpha, k(k-1)m\beta) \in \text{IR}(\mathcal{F})$ .

Let us pause for a moment to provide some intuition. The supremum term is the worst possible regret within  $\mathcal{F}_i$  while the infimum represents a kind of bound on the minimum amount of information obtained by TS. In particular, TS plays the optimal action for some sampled loss and gains information when there is variation of the losses at that point. The appearance of  $m$  in the information ratio bound arises from a Cauchy-Schwarz (what else?) that is somehow the ‘same’ Cauchy-Schwarz used in the analysis of the information ratio for finite-armed bandits (Russo and Van Roy, 2014) and in the information ratio decomposition by Lattimore (2020). The proof can be found in Appendix D.

### 3. Thompson Sampling in One Dimension

Our first main theorem shows that TS is statistically efficient when the loss is bounded and Lipschitz and  $d = 1$ .

**Theorem 4** When  $d = 1$ ,  $\sup_{\xi \in \mathcal{P}(\mathcal{F}_{b1})} B\text{Reg}_n(\text{TS}, \xi) = O\left(\sqrt{n \log(n) \log(n \text{diam}(K))}\right)$ .

Theorem 4 is established by combining the following bound on the information ratio and  $\alpha = 1/n$  with Theorem 2.

**Theorem 5** Suppose that  $d = 1$  and  $\alpha \in (0, 1)$ . Then  $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \text{IR}(\mathcal{F}_{b1})$ .

**Proof** Let  $\bar{f} \in \text{conv}(\mathcal{F}_{b1})$  and for integer  $i$  define

$$\mathcal{F}_i = \begin{cases} \{f \in \mathcal{F}_{b1} : \bar{f}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}], x_f \geq x_{\bar{f}}\} & \text{if } i > 0 \\ \{f \in \mathcal{F}_{b1} : \bar{f}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}], x_f < x_{\bar{f}}\} & \text{if } i < 0 \\ \{f \in \mathcal{F}_{b1} : \bar{f}(x_f) - f_\star \leq \alpha\} & \text{if } i = 0. \end{cases} \quad (2)$$

Since the losses in  $\mathcal{F}_{b1}$  are bounded by assumption, for  $|i| > m = \lceil \log_2(1/\alpha) \rceil$ ,  $\mathcal{F}_i = \emptyset$  so that  $\mathcal{F}_{b1} = \cup_{i=-m}^m \mathcal{F}_i$ . In a moment we will show that with  $k = 4$  and  $-m \leq i \leq m$  and  $\varepsilon = \alpha 2^{|i|}$  that

$$\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f_\star) \leq \varepsilon \leq \alpha + \sqrt{230 \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}. \quad (3)$$

Hence, by Lemma 3 and naive simplification of constants  $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \text{IR}(\mathcal{F}_{b1})$  as desired. The first inequality in Eq. (3) is an immediate consequence of the definition of  $\mathcal{F}_i$  and  $\varepsilon$ . The second is also immediate when  $i = 0$ . The situation when  $i < 0$  and  $i > 0$  is symmetric, so for the

remainder we prove that the second inequality in Eq. (3) holds for any  $i > 0$ . Let  $f_1, \dots, f_k \in \mathcal{F}_i$  and for  $j \in [4]$  let  $x_j = x_{f_j}$  and assume without loss of generality that  $x_1 \leq x_2 \leq x_3 \leq x_4$ . Suppose that

$$\sum_{j,l \in \text{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2 < c^2 \varepsilon^2 \quad \text{with} \quad c = \frac{\sqrt{65} - 7}{16}. \quad (4)$$

Let us now establish a contradiction, which we do in three steps. The main argument in each step is illustrated in Figure 1.

**STEP 1** We start by showing that  $x_2$  must be somewhat closer to  $x_3$  than to  $x_1$ .

$$f_1(x_3) \stackrel{(a)}{\leq} \bar{f}(x_3) + \varepsilon c \stackrel{(b)}{\leq} f_3(x_3) + \varepsilon[c+1] \leq f_3(x_1) + \varepsilon[c+1] \stackrel{(c)}{\leq} \bar{f}(x_1) + \varepsilon[2c+1],$$

where (a) follows from Eq. (4), (b) since for  $f \in \mathcal{F}_i$ ,  $f(x_f) \geq \bar{f}(x_f) - \varepsilon$  and (c) from Eq. (4) again. Hence, with  $p \in [0, 1]$  such that  $x_2 = (1-p)x_1 + px_3$ ,

$$\begin{aligned} \bar{f}(x_1) &\stackrel{(a)}{\leq} \bar{f}(x_2) \stackrel{(b)}{\leq} f_1(x_2) + c\varepsilon \stackrel{(c)}{\leq} (1-p)f_1(x_1) + pf_1(x_3) + c\varepsilon \\ &\stackrel{(d)}{\leq} (1-p)(\bar{f}_1(x_1) - \varepsilon/2) + pf_1(x_3) + c\varepsilon \\ &\stackrel{(e)}{\leq} \bar{f}(x_1) + \varepsilon[c + p[2c + 3/2] - 1/2], \end{aligned}$$

where (a) follows because  $\bar{f}$  is non-decreasing on  $[x_1, x_4]$  by the definition of  $\mathcal{F}_i$  and  $i > 0$ , (b) from Eq. (4), (c) by convexity and the definition of  $p$ , (d) since  $f_1(x_1) \leq \bar{f}_1(x_1) - \varepsilon/2$  by the definition of  $\mathcal{F}_i$  and (e) is true by the previous display. Therefore  $p \geq (1/2 - c)/(2c + 3/2) \approx 0.27$ .

**STEP 2** Having shown that  $x_2$  is close to  $x_3$ , we now show that  $f_3(x_3)$  is not much smaller than  $\bar{f}(x_1)$ . Indeed,

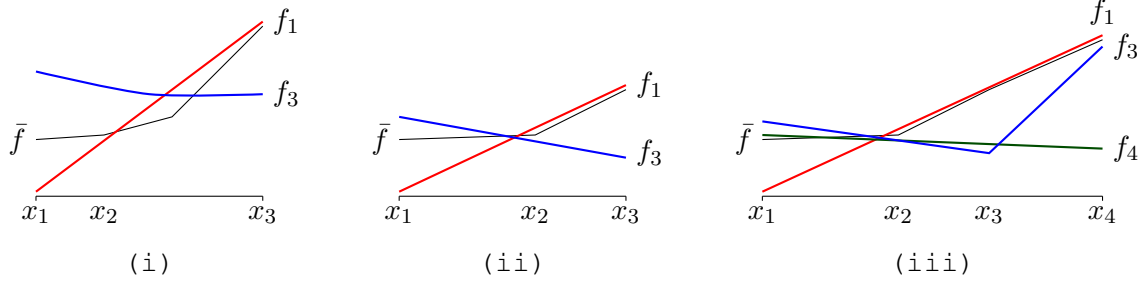
$$\begin{aligned} \bar{f}(x_1) &\stackrel{(a)}{\leq} \bar{f}(x_2) \stackrel{(b)}{\leq} f_3(x_2) + c\varepsilon \\ &\stackrel{(c)}{\leq} (1-p)f_3(x_1) + pf_3(x_3) + c\varepsilon \\ &\stackrel{(d)}{\leq} (1-p)\bar{f}(x_1) + pf_3(x_3) + 2c\varepsilon, \end{aligned}$$

(a) – (c) follows as above in **STEP 1** and (d) from Eq. (4). Rearranging shows that  $f_3(x_3) \geq \bar{f}(x_1) - \frac{2c\varepsilon}{p}$ .

**STEP 3** Lastly we derive a contradiction using **STEP 1** and **STEP 2** since

$$\begin{aligned} f_4(x_3) &\stackrel{(a)}{\leq} f_4(x_1) \stackrel{(b)}{\leq} \bar{f}(x_1) + c\varepsilon \\ &\stackrel{(c)}{\leq} f_3(x_3) + \varepsilon \left[ c + \frac{2c}{p} \right] \\ &\stackrel{(d)}{\leq} \bar{f}(x_3) + \varepsilon \left[ c + \frac{2c}{p} - \frac{1}{2} \right] \\ &\stackrel{(e)}{\leq} \bar{f}(x_3) - c\varepsilon, \end{aligned}$$

where (a) follows by convexity and because  $f_4$  is minimised at  $x_4$ , (b) from Eq. (4), (c) from **STEP 2**, (d) since  $f_3(x_3) \leq \bar{f}(x_3) - \varepsilon/2$  by the definition of  $\mathcal{F}_i$  and (e) from the bound on  $p$  in **STEP 1** and the definition of  $c$ . But this contradicts Eq. (4). Hence Eq. (4) does not hold. And since  $c^2 \geq \frac{1}{230}$  it follows that Eq. (3) holds.  $\blacksquare$



**Figure 1:** (i) shows that if  $x_2$  is too close to  $x_1$ , then  $f_1(x_3)$  must be large, which implies that  $f_3(x_3)$  must be large and so too must  $f_3(x_1)$ , which shows that  $f_3(x_1) - \bar{f}(x_1)$  is large. (ii) shows what happens if  $f_3(x_3)$  is too far below  $\bar{f}(x_1)$ , which is that  $f_3(x_1)$  must be much larger than  $\bar{f}(x_1)$ . (iii) shows that  $f_4(x_3)$  cannot be much larger than  $f_3(x_3)$  and therefore  $\bar{f}(x_3) - f_4(x_3)$  must be large.

#### 4. Thompson Sampling for Ridge Functions

We now consider the multi-dimensional convex monotone ridge function setting where  $\mathcal{F} = \mathcal{F}_{\text{blrm}}$

**Theorem 6**  $\sup_{\xi \in \mathcal{P}(\mathcal{F}_{\text{blrm}})} B\text{Reg}_n(\text{TS}, \xi) = O(d^{2.5} \sqrt{n} \log(nd \text{diam}(K))^2).$

Russo and Van Roy (2016) used information-theoretic means to show that for linear bandits the regret is at most  $\tilde{O}(d\sqrt{n})$ . Lattimore (2021) showed that for (possibly non-monotone) convex ridge functions a version of IDS has Bayesian regret at most  $\tilde{O}(d\sqrt{n})$ . The downside is that IDS is barely implementable in practice, even given efficient access to posterior samples. Like Theorem 4, Theorem 6 is established by combining a bound on the information ratio with Theorem 2.

**Theorem 7**  $(\alpha, \beta \lceil \log(1/\alpha) \rceil) \in \text{IR}(\mathcal{F}_{\text{blrm}})$  whenever  $\alpha \in (0, 1)$  and

$$\beta = \Omega \left( d^4 \log \left( \frac{d \text{diam}(K)}{\alpha} \right)^2 \right).$$

with the Big-O hiding only a universal constant.

**Proof** Abbreviate  $\mathcal{F} = \mathcal{F}_{\text{blmr}}$ . The high-level argument follows the proof of Theorem 5. The main challenge is lower bounding the quadratic (information gain) term that appears in Lemma 3, which uses an argument based on the method of inscribed ellipsoid for optimisation (Tarasov et al., 1988). Let  $\bar{f} \in \text{conv}(\mathcal{F})$ . Given a nonempty finite set  $\mathcal{C} \subset \mathcal{F}$  and  $\delta > 0$ , let  $J_\delta(\mathcal{C}) = \text{conv}(\cup_{g \in \mathcal{C}} \mathbb{B}_\delta(x_g))$ . Moreover, let  $E_\delta(\mathcal{C})$  be the ellipsoid of maximum volume enclosed in  $J_\delta(\mathcal{C})$ , which is called John's



ellipsoid. We now need two lemmas. The first shows that for a suitable subset  $\mathcal{C}$  of loss functions either the information gain is reasonably large or some function  $f$  can be removed from  $\mathcal{C}$  in such a way that  $E_\delta(\mathcal{C} \setminus \{f\})$  is considerably smaller than  $E_\delta(\mathcal{C})$ .

**Lemma 8** *Let  $\varepsilon > 0$ ,  $\delta = \frac{\varepsilon}{12(d+1)}$  and  $\mathcal{C} \subset \mathcal{F}$  be a nonempty finite set such that for all  $f, g \in \mathcal{C}$ ,  $\bar{f}(x_f) - f_\star \in [\varepsilon/2, \varepsilon]$  and  $|\bar{f}(x_f) - \bar{f}(x_g)| \leq \delta$ . Then at least one of the following holds:*

(i) *There exists an  $f \in \mathcal{C}$  such that  $\text{vol}(E_\delta(\mathcal{C} \setminus \{f\})) \leq 0.85 \text{vol}(E_\delta(\mathcal{C}))$ .*

(ii) *There exists a pair  $f, g \in \text{PAIR}(\mathcal{C})$  such that  $(f(x_g) - \bar{f}(x_g))^2 \geq \delta^2$ .*

**Proof** Let  $\mu \in \mathbb{R}^d$  and  $\Sigma$  be positive definite such that  $E_\delta(\mathcal{C}) = \{x : \|x - \mu\|_{\Sigma^{-1}} \leq 1\}$  and for  $r > 0$ , let  $E_{\delta,r} = \{x : \|x - \mu\|_{\Sigma^{-1}} \leq r\}$ . By John's theorem (Artstein-Avidan et al., 2015, Remark 2.1.17),  $E_\delta(\mathcal{C}) \subset J_\delta(\mathcal{C}) \subset E_{\delta,d}(\mathcal{C})$ . Let  $f \in \mathcal{C}$  be arbitrary. By assumption there exists a convex monotone increasing  $\ell$  and  $\theta \in \mathbb{S}_1$  such that  $f = \ell(\langle \cdot, \theta \rangle)$ . Let  $H = H(\mu, \theta)$ , which is the half-space passing through the center of John's ellipsoid  $E_\delta(\mathcal{C})$  with inward-facing normal  $\theta$ . Consider the following cases, illustrated in Figure 2:

**CASE 1**  $\langle x_g, \theta \rangle \geq \langle \mu, \theta \rangle + \delta$  for all  $g \in \mathcal{C} \setminus \{f\}$ . In this case  $J_\delta(\mathcal{C} \setminus \{f\}) \subset H \cap J_\delta(\mathcal{C})$  and therefore the inequality of Khachiyan (1990) shows that  $\text{vol}(E_\delta(\mathcal{C} \setminus \{f\})) \leq 0.85 \text{vol}(E_\delta(\mathcal{C}))$ .

**CASE 2** There exists a  $g \in \mathcal{C} \setminus \{f\}$  such that  $\langle x_g, \theta \rangle < \langle \mu, \theta \rangle + \delta$ . Since  $E_\delta(\mathcal{C}) \subset J_\delta(\mathcal{C})$ , there exists an  $x \in J_\delta(\mathcal{C})$  such that  $\langle x, \theta \rangle \geq \langle \mu, \theta \rangle + \|\theta\|_\Sigma$ . By the definition of  $J_\delta(\mathcal{C})$  it follows that there exists a  $h \in \mathcal{C}$  such that  $\langle x_h, \theta \rangle \geq \langle \mu, \theta \rangle + \|\theta\|_\Sigma - \delta$ . Let  $x'_h = x_h + \delta\theta$  and  $x'_g = x_g - \delta\theta$  and  $x'_f = x_f - \delta\theta$ . Collecting the above facts, we have

$$\langle x'_f, \theta \rangle \geq \langle \mu, \theta \rangle - d\|\theta\|_\Sigma \quad \langle x'_g, \theta \rangle \leq \langle \mu, \theta \rangle \quad \langle x'_h, \theta \rangle \geq \langle \mu, \theta \rangle + \|\theta\|_\Sigma. \quad (5)$$

Since  $\ell$  is nondecreasing and  $f(x_f) \leq f(x_g)$  it follows that  $\langle x'_f, \theta \rangle \leq \langle x'_g, \theta \rangle \leq \langle x'_h, \theta \rangle$ . Therefore,

$$\begin{aligned} f(x_g) &\stackrel{(a)}{\leq} \delta + f(x'_g) \stackrel{(b)}{=} \delta + f\left(\frac{\langle x'_g - x'_f, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_h + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_f\right) \\ &\stackrel{(c)}{\leq} \delta + f(x'_h) + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} (f(x'_f) - f(x'_h)) \\ &\stackrel{(d)}{\leq} \delta + f(x'_h) + \frac{1}{d+1} (f(x'_f) - f(x'_h)) \stackrel{(e)}{\leq} 3\delta + f(x_h) + \frac{1}{d+1} (f(x_f) - f(x_h)), \end{aligned} \quad (6)$$

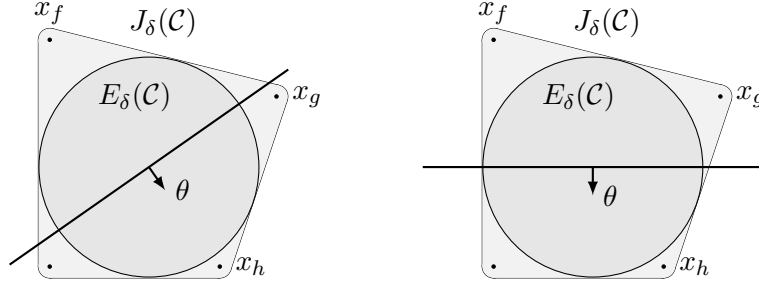
where (a) follows because  $f$  is a Lipschitz ridge function and using Lemma 13 and the definition of  $x'_g$  so that  $|\langle x_g - x'_g, \theta \rangle| = \delta$ . (b) by definitions and the fact that  $f(\cdot) = \ell(\langle \cdot, \theta \rangle)$ , (c) by convexity of  $f$ , (d) from Eq. (5) and because  $f(x'_f) \leq f(x'_h)$ . (e) uses again that  $f$  is a Lipschitz ridge function and Lemma 13 and that  $|\langle x'_h - x_h, \theta \rangle| = \delta$ . Suppose that  $f(x_h) \geq \bar{f}(x_h) + \delta$ , then  $(f(x_h) - \bar{f}(x_h))^2 \geq \delta^2$  and (ii) holds. Otherwise  $f(x_h) < \bar{f}(x_h) + \delta$  and so

$$\begin{aligned} f(x_g) &\stackrel{(a)}{\leq} 3\delta + \frac{d}{d+1} f(x_h) + \frac{1}{d+1} f(x_f) \stackrel{(b)}{\leq} 3\delta + \frac{d}{d+1} (\bar{f}(x_h) + \delta) + \frac{1}{d+1} (\bar{f}(x_h) + \delta - \varepsilon/2) \\ &\stackrel{(c)}{\leq} 4\delta + \bar{f}(x_h) - \frac{\varepsilon}{2(d+1)} \stackrel{(d)}{=} \bar{f}(x_h) - 2\delta \stackrel{(e)}{\leq} \bar{f}(x_g) - \delta, \end{aligned}$$



where (a) follows from Eq. (6), (b) follows from the assumption in the lemma statement that  $f(x_f) \leq \bar{f}(x_f) - \varepsilon/2 \leq \bar{f}(x_h) + \delta - \varepsilon/2$  and  $f(x_h) < \bar{f}(x_h) + \delta$ . (c) by naive simplification, (d) by the definition of  $\gamma$  and  $\delta$  and (e) by the assumptions in the lemma. Therefore  $f(x_g) \leq \bar{f}(x_g) - \delta$ , which implies that  $(f(x_g) - \bar{f}(x_g))^2 \geq \delta^2$  and again (ii) holds.

Summarising, in **CASE 1**, (i) holds while in **CASE 2**, (ii) holds. ■



**Figure 2:** The two cases considered in the proof of Lemma 8. In the left figure, the situation is such that  $E_\delta(\mathcal{C} \setminus \{f\})$  is a constant fraction less volume than  $E_\delta(\mathcal{C})$ . On the other hand, in the figure on the right one of  $(f(x_h) - \bar{f}(x_h))^2$  or  $(f(x_g) - \bar{f}(x_g))^2$  must be reasonably large.

The next lemma uses an inductive argument to show that any suitably large set  $\mathcal{C}$  satisfying the conditions of the previous lemma necessarily yields a large information gain.

**Lemma 9** *Suppose that  $\mathcal{C}$  satisfies the conditions of Lemma 8 for some  $\varepsilon \geq \alpha$  and  $|\mathcal{C}| = 1 + 2d + 8d \left\lceil \log \left( \frac{24d(d+1) \text{diam}(K)}{\alpha} \right) \right\rceil$ . Then*

$$\sum_{f,g \in \text{PAIR}(\mathcal{C})} (f(x_g) - \bar{f}(x_g))^2 \geq d\delta^2.$$

**Proof** Define a sequence  $(\mathcal{C}_k)$  of sets as follows. Let  $\mathcal{C}_1 = \mathcal{C}$  and  $2m - 1 \triangleq |\mathcal{C}|$ . Then, given  $\mathcal{C}_k$ , define  $\mathcal{C}_{k+1} \subset \mathcal{C}_k$  as a set such that one of two properties hold:

- (i)  $|\mathcal{C}_{k+1}| = |\mathcal{C}_k| - 1$  and  $\text{vol}(E_\delta(\mathcal{C}_{k+1})) \leq 0.85 \text{vol}(E_\delta(\mathcal{C}_k))$ ; or
- (ii)  $\mathcal{C}_{k+1} = \mathcal{C}_k \setminus \{f, g\}$  for some  $f, g \in \text{PAIR}(\mathcal{C}_k)$  and  $(f(x_g) - \bar{f}(x_g))^2 \geq \delta^2$ .

Such a sequence exists by Lemma 8. By definition  $|\mathcal{C}_1| = 2m - 1$  and since  $|\mathcal{C}_{k+1}| \geq |\mathcal{C}_k| - 2$ ,  $|\mathcal{C}_m| \geq |\mathcal{C}_1| - 2(m - 1) = 1$ . Recall that by John's theorem  $E_\delta(\mathcal{C}_m) \subset J_\delta(\mathcal{C}_m) \subset E_{\delta,d}(\mathcal{C}_m)$ , which means that

$$\text{vol}(E_\delta(\mathcal{C}_m)) = \left(\frac{1}{d}\right)^d \text{vol}(E_{\delta,d}(\mathcal{C}_m)) \geq \left(\frac{1}{d}\right)^d \text{vol}(J_\delta(\mathcal{C}_m)) \geq \left(\frac{1}{d}\right)^d \text{vol}(\mathbb{B}_\delta).$$

Furthermore,  $E_\delta(\mathcal{C}_1) \subset K + \mathbb{B}_\delta \subset \mathbb{B}_{\text{diam}(K) + \delta}$ . Let  $\tau$  be the number of times (i) occurs. Then

$$\left(\frac{1}{d}\right)^d \text{vol}(\mathbb{B}_\delta) \leq \text{vol}(E_\delta(\mathcal{C}_m)) \leq (0.85)^\tau \text{vol}(E_\delta(\mathcal{C}_1)) \leq (0.85)^\tau \text{vol}(\mathbb{B}_{\text{diam}(K) + \delta}).$$

Therefore  $(0.85)^\tau \geq \left(\frac{\delta}{d(\text{diam}(K)+\delta)}\right)^d$ , which shows that

$$\tau \leq \frac{d \log \left( \frac{\delta}{d(\text{diam}(K)+\delta)} \right)}{\log(0.85)} \leq 8d \log \left( \frac{2d \text{diam}(K)}{\delta} \right) \leq |\mathcal{C}| - 2d - 1.$$

Therefore (iii) happens at least  $d$  times and the claim follows by the definition of (iii).  $\blacksquare$

The last step is to introduce a decomposition of the space of loss functions and show how to obtain finite sets satisfying the conditions of Lemma 9. Let

$$k = (25d + 24) \left[ 1 + 2d + 8d \left\lceil \log \left( \frac{24d(d+1) \text{diam}(K)}{\alpha} \right) \right\rceil \right] = O \left( d^2 \log \left( \frac{d \text{diam}(K)}{\alpha} \right) \right).$$

For  $0 \leq i \leq \lceil \log_2(1/\alpha) \rceil$  let

$$\mathcal{F}_i = \begin{cases} \{f \in \mathcal{F} : \bar{f}(x_f) - f_\star \in [\alpha 2^{|i|-1}, \alpha 2^{|i|}]\} & \text{if } i > 0 \\ \{f \in \mathcal{F} : \bar{f}(x_f) - f_\star < \alpha\} & \text{if } i = 0. \end{cases}$$

In order to apply Lemma 3 we will show that for all  $0 \leq i \leq \lceil \log_2(1/\alpha) \rceil$  and with  $\varepsilon = \alpha 2^i$  that for all  $f_1, \dots, f_k \in \mathcal{F}_i$ ,

$$\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f_\star) \leq \varepsilon \leq \alpha + \sqrt{512 \sum_{j,l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}. \quad (7)$$

This implies that  $(\alpha, 512k(k-1)(1 + \lceil \log(1/\alpha) \rceil)) \in \text{IR}(\mathcal{F})$  as required. The first inequality in Eq. (7) follows immediately from the definition of  $\varepsilon$  and  $\mathcal{F}_i$ . The second is also immediate when  $i = 0$  by the definition of  $\mathcal{F}_i$ . Suppose now that  $i > 0$ . Let  $f_1, \dots, f_k \in \mathcal{F}_i$  and assume without loss of generality that  $j \mapsto \bar{f}(x_{f_j})$  is nonincreasing. The second inequality in Eq. (7) holds immediately if  $f_j = f_l$  for some  $j, l \in \text{PAIR}(k)$  since  $f_j(x_{f_l}) = f_j(x_{f_j}) \leq \bar{f}(x_{f_j}) - \varepsilon/2$ . Suppose this is not the case and consider two cases:

**CASE 1**  $\bar{f}(x_{f_1}) \geq \bar{f}(x_{f_k}) + 2\varepsilon$ . In this case  $f_1(x_{f_k}) \geq f_1(x_{f_1}) \geq \bar{f}(x_{f_1}) - \varepsilon \geq \bar{f}(x_{f_k}) + \varepsilon$ , which shows that  $(f_1(x_{f_k}) - \bar{f}(x_{f_k}))^2 \geq \varepsilon^2$  and the second inequality Eq. (7) holds.

**CASE 2**  $\bar{f}(x_1) < \bar{f}(x_k) + 2\varepsilon$ . Let  $\delta = \frac{\varepsilon}{12(d+1)}$  as in Lemma 9. Let  $b = 25d + 24$  and  $\mathcal{C}_1, \dots, \mathcal{C}_b$  be formed by dividing  $\{f_1, \dots, f_k\}$  in order into  $b$  blocks of equal size. Let  $s_a = \max_{f,g \in \mathcal{C}_a} |\bar{f}(x_f) - \bar{f}(x_g)|$ . Given the conditions of the case we have  $2\varepsilon > \sum_{a=1}^b s_a \geq \sum_{a=1}^b \delta \mathbf{1}(s_a > \delta)$ , which means that  $\sum_{a=1}^b \mathbf{1}(s_a > \delta) \leq 2\varepsilon/\delta \leq 24(d+1) = b - d$ . Hence there exist at least  $d$  blocks  $\mathcal{C}_a$  for which  $s_a \leq \delta$  and these blocks satisfy the conditions of Lemma 9 so that  $\sum_{j,l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2 \geq d^2 \delta^2 \geq \varepsilon^2/512$  and again the second inequality in Eq. (7) holds.

Hence Eq. (7) holds and the claim follows from Lemma 3 and the definitions of  $k$  and  $m$ .  $\blacksquare$

## 5. Lower Bounds

We now prove that Thompson sampling has poor behaviour for general multi-dimensional convex functions and that the classical information-theoretic techniques cannot improve on the best known bound for general bandit convex optimisation of  $\tilde{O}(d^{1.5}\sqrt{n})$ . While these seem like quite different results, they are based on the same construction, which is based on finding a family of functions and prior that makes learning challenging.

**Lemma 10** *Let  $\varepsilon \in [0, 1/2]$  and  $\theta \in \mathbb{S}_1$  and define functions  $f$  and  $f_\theta$  by*

$$f(x) = \varepsilon + \frac{1}{2} \|x\|^2 \quad f_\theta(x) = \begin{cases} f(x) & \text{if } \|\theta - x\|^2 \geq 1 + 2\varepsilon \\ \langle \theta, x \rangle - 1 + \sqrt{1 + 2\varepsilon} \|\theta - x\| & \text{otherwise.} \end{cases}$$

*Then  $f_\theta$  is convex and minimised at  $\theta$  and  $\text{Lip}_{\mathbb{B}_1}(f_\theta) \leq \sqrt{2 + 2\varepsilon}$ .*

The function  $f_\theta$  arises naturally as the largest convex function for which both  $f_\theta(\theta) = 0$  and  $f_\theta(x) \leq f(x)$  for all  $x \in \mathbb{R}^d$ . Equivalently, its epigraph is the convex hull of the epigraphs of  $f$  and the convex indicator function:  $\infty \mathbf{1}_\theta(\cdot)$ . The proof of Lemma 10 as well as detailed proofs of all the remaining results in this section are given in Appendix E.

**Theorem 11** *When  $K = \mathbb{B}_1$  is the standard euclidean ball. There exists a prior  $\xi$  on  $\mathcal{F}_{\mathbb{B}_1}$  such that*

$$\text{BReg}_n(\text{TS}, \xi) \geq \frac{1}{2} \min \left( n, \left\lfloor \frac{1}{4} \exp(d/32) \right\rfloor \right).$$

**Proof** [sketch] The idea is to construct a prior such that with high probability TS obtains limited information while suffering high regret. We assume there is no noise and let  $f$  and  $f_\theta$  be defined as in Lemma 10 with  $\varepsilon = 1/4$ . Let  $\sigma$  be the uniform probability measure on  $\mathbb{S}_1$  and the prior  $\xi$  be the law of  $f_\theta$  when  $\theta$  is sampled from  $\sigma$ . By the definition of  $f_\theta$  and the fact that  $\varepsilon = 1/4$ , for any  $x \in \mathbb{S}_1$  ( $f(x) = f_\theta(x)$ )  $\Leftrightarrow \langle x, \theta \rangle \leq \frac{1}{4}$ . Since TS plays the minimiser of some  $f_\theta$  in every round, it follows that TS always plays in  $\mathbb{S}_1$ . Let  $\mathcal{C}_\theta = \{x \in \mathbb{S}_1 : \langle x, \theta \rangle > \frac{1}{4}\}$  and  $\delta = \sigma(\mathcal{C}_\theta)$ . Let  $f_{\theta_*}$  be the true loss function sampled from  $\xi$ . Suppose that  $X_1, \dots, X_t \in \mathbb{S}_1 \setminus \mathcal{C}_{\theta_*}$ , which means that  $Y_s = \frac{3}{4}$  for  $1 \leq s \leq t$  and the posterior is the uniform distribution on  $\Theta_{t+1} = \mathbb{S}_1 \setminus \bigcup_{s=1}^t \mathcal{C}_{X_s}$ . Provided that  $t\delta \leq \frac{1}{2}$ ,

$$\mathbb{P}(X_{t+1} \in \mathcal{C}_{\theta_*} | X_1, \dots, X_t \notin \mathcal{C}_{\theta_*}) = \frac{\sigma(\mathcal{C}_{\theta_*} \cap \Theta_{t+1})}{\sigma(\Theta_{t+1})} \leq \frac{\delta}{1 - t\delta} \leq 2\delta.$$

Hence, with  $n_0 = \min(n, \lfloor 1/(4\delta) \rfloor)$ ,

$$\text{BReg}_n(\text{TS}, \xi) \geq \text{BReg}_{n_0}(\text{TS}, \xi) \geq \frac{3n_0}{4} \mathbb{P}(X_1, \dots, X_{n_0} \notin \mathcal{C}_{\theta_*}) \geq \frac{3n_0}{4} (1 - 2n_0\delta) \geq \frac{3n_0}{8}.$$

The result is completed since  $\delta \leq \exp(-d/32)$  follows from concentration of measure on the sphere (Tkocz, 2018, Theorem B.1).  $\blacksquare$

Theorem 11 shows that Thompson sampling has large regret for general bandit convex optimisation. The next theorem shows there exist priors for which the information ratio for any policy is at least  $\Omega(d^2)$ . At least naively, this means that the information-theoretic machinery will not yield a bound on the regret for general bandit convex optimisation that is better than  $\tilde{O}(d^{1.5}\sqrt{n})$

**Theorem 12** *Suppose that  $K = \mathbb{B}_1$  and  $d > 256$ . Then there exists a prior  $\xi$  on  $\mathcal{F}_{b1}$  such that for all probability measures  $\pi$  on  $K$ ,  $\Delta(\pi, \xi) \geq 2^{-19} \frac{d}{\log(d)} \sqrt{\mathcal{I}(\pi, \xi)}$ .*

The prior  $\xi$  is the same as used in the proof of Theorem 11 but with  $\varepsilon = \tilde{\Theta}(1/d)$ . The argument is based on proving that for any policy the regret is  $\Omega(\varepsilon)$  while the information gain is  $\tilde{O}(\varepsilon^2)$ .

## 6. Discussion

*Adversarial setup:* In the Bayesian adversarial setting a sequence of loss functions  $f_1, \dots, f_n$  are sampled from a joint distribution on  $\mathcal{F}^n$ . The learner plays  $X_t$  and observes  $Y_t = f_t(X_t)$  and the Bayesian regret is  $\text{BReg}(\mathcal{A}, \xi) = \mathbb{E}[\sup_{x \in K} \sum_{t=1}^n (f_t(X_t) - f_t(x))]$ . One can envisage two possible definitions of Thompson sampling in this setting. One samples  $g_t$  from the marginal of the posterior and plays  $X_t = x_{g_t}$ . The second samples  $g_1, \dots, g_n$  from the posterior and plays  $X_t$  as the minimiser of  $\sum_{t=1}^n g_t$ . The former has linear regret, while [Bubeck et al. \(2015\)](#) notes that the latter has an unbounded information ratio. More details are in Appendix C.

*Tightness of bounds:* At present we are uncertain whether or not the monotonicity assumption is needed in the ridge setting. Our best guess is that it is not. One may also wonder if the bound on the information ratio in Theorem 7 can be improved. We cautiously believe that when the loss has the form  $f(x) = \ell(\langle x, \theta \rangle)$  for *known* convex link function  $\ell : \mathbb{R} \rightarrow \mathbb{R}$ , then the information ratio is at most  $d$ . This would mean that convex generalised linear bandits are no harder than linear bandits.

*TS vs IDS:* Theorem 11 shows that TS can have more-or-less linear regret in high-dimensional problems. On the other hand, [Bubeck and Eldan \(2018\)](#) and [Lattimore \(2020\)](#) show that IDS has a well-controlled information ratio, but is much harder to compute. An obvious question is whether some simple adaptation of Thompson sampling has a well-controlled information ratio.

*Applications:* Many problems are reasonably modelled as 1-dimensional convex bandits, with the classical example being dynamic pricing where  $K$  is a set of prices and convexity is a reasonable assumption based on the response of demand to price. The monotone ridge function class is a natural model for resource allocation problems where a single resource (e.g., money) is allocated to  $d$  locations. The success of some global task increases as more resources are allocated, but with diminishing returns. Problems like this can reasonably be modelled by convex monotone ridge functions with  $K = \{x \geq \mathbf{0} : \|x\|_1 \leq 1\}$ .

*Lipschitz assumption:* Our bounds depend logarithmically on the Lipschitz constant associated with the class of loss functions. There is a standard trick to relax this assumption based on the observation that bounded convex functions must be Lipschitz on a suitably defined interior of the constraint set  $K$ . Concretely, suppose that  $K$  is a convex body and  $f : K \rightarrow [0, 1]$  is convex and  $\mathbb{B}_r \subset K$  and  $K_\varepsilon = (1 - \varepsilon)K$ . Then  $\min_{x \in K_\varepsilon} f(x) \leq \inf_{x \in K} f(x) + \varepsilon$  and  $f$  is  $1/(r\varepsilon)$ -Lipschitz on  $K_\varepsilon$  ([Lattimore, 2024](#), Chapter 3). Hence, you can run TS on  $K_\varepsilon$  with  $\varepsilon = 1/n$  and the Lipschitz constant is at most  $n/r$ . Moreover, if  $K$  is in (approximate) isotropic or John’s position, then  $\mathbb{B}_1 \subset K \subset \mathbb{B}_{2d}$  by [Kannan et al. \(1995\)](#) and John’s theorem, respectively.

*Frequentist regret:* An ambitious goal would be to prove a bound on the frequentist regret of TS for some well-chosen prior. This is already quite a difficult problem in multi-armed ([Kaufmann et al., 2012](#); [Agrawal and Goyal, 2012](#)) and linear bandits ([Agrawal and Goyal, 2013](#)) and is out of reach of the techniques developed here. On the other hand, the Bayesian algorithm has the advantage of

being able to specify a prior that makes use of background knowledge and the theoretical guarantees for TS provide a degree of comfort.

*Choice of prior:* The choice of the prior depends on the application. A variety of authors of constructed priors supported on non-parametric classes of 1-dimensional convex functions using a variety of methods (Ramgopal et al., 1993; Chang et al., 2007; Shively et al., 2011). In many cases you may know the loss belongs to a simple parametric class, in which case the prior and posterior computations may simplify dramatically.

## References

- M. Abeille and A. Lazaric. Linear Thompson sampling revisited. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pages 176–184, Fort Lauderdale, FL, USA, 2017. JMLR.org.
- S. Agrawal and N. Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In Proceedings of the 25th Conference on Learning Theory, 2012.
- S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In Proceedings of the 30th International Conference on Machine Learning, pages 127–135, Atlanta, GA, USA, 2013. JMLR.org.
- S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. Asymptotic geometric analysis, Part I, volume 202. American Mathematical Soc., 2015.
- S. Bubeck and R. Eldan. Exploratory distributions for convex functions. Mathematical Statistics and Learning, 1(1):73–100, 2018.
- S. Bubeck, O. Dekel, T. Koren, and Y. Peres. Bandit convex optimization:  $\sqrt{T}$  regret in one dimension. In Proceedings of the 28th Conference on Learning Theory, pages 266–278, Paris, France, 2015. JMLR.org.
- I-S. Chang, L-C. Chien, C. Hsiung, C-C. Wen, and Y-J. Wu. Shape restricted regression with random bernstein polynomials. Lecture Notes-Monograph Series, pages 187–202, 2007.
- V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In Proceedings of the 21st Conference on Learning Theory, pages 355–366, 2008.
- S. Dong, T. Ma, and B. Van Roy. On the performance of thompson sampling on logistic bandits. In Conference on Learning Theory, pages 1158–1160. PMLR, 2019.
- S. Filippi, O. Cappé, A. Garivier, and Cs. Szepesvári. Parametric bandits: The generalized linear case. In Advances in Neural Information Processing Systems, pages 586–594. Curran Associates, Inc., 2010.
- A Flaxman, A Kalai, and HB McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In SODA’05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, pages 385–394, 2005.

- H. Fokkema, D. van der Hoeven, T. Lattimore, and J. Mayo. Online Newton method for bandit convex optimisation. In Conference on Learning Theory, 2024.
- B. Huang, K. Huang, S. Kakade, J. D. Lee, Q. Lei, R. Wang, and J. Yang. Optimal gradient-based algorithms for non-concave bandit optimization. Advances in Neural Information Processing Systems, 34:29101–29115, 2021.
- R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. Discrete & Computational Geometry, 13:541–559, 1995.
- E. Kaufmann, N. Korda, and R. Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In Proceedings of the 23rd International Conference on Algorithmic Learning Theory, volume 7568 of Lecture Notes in Computer Science, pages 199–213. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-34105-2.
- L. G. Khachiyan. An inequality for the volume of inscribed ellipsoids. Discrete & computational geometry, 5:219–222, 1990.
- R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In Advances in Neural Information Processing Systems, pages 697–704. MIT Press, 2005.
- B. Kveton, M. Zaheer, Cs. Szepesvári, L. Li, M. Ghavamzadeh, and C. Boutilier. Randomized exploration in generalized linear bandits. In International Conference on Artificial Intelligence and Statistics, pages 2066–2076. PMLR, 2020.
- T. Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. Mathematical Statistics and Learning, 2(3/4):311–334, 2020.
- T. Lattimore. Minimax regret for bandit convex optimisation of ridge functions. arXiv preprint arXiv:2106.00444, 2021.
- T. Lattimore. Bandit Convex Optimisation. 2024.
- T. Lattimore and A. György. A second-order method for stochastic bandit convex optimisation. arXiv preprint arXiv:2302.05371, 2023.
- T. Lattimore and B. Hao. Bandit phase retrieval. Advances in Neural Information Processing Systems, 34:18801–18811, 2021.
- W. Niemi. Asymptotics for m-estimators defined by convex minimization. The Annals of Statistics, pages 1514–1533, 1992.
- P. Ramgopal, P. Laud, and A. Smith. Nonparametric bayesian bioassay with prior constraints on the shape of the potency curve. Biometrika, 80(3):489–498, 1993.
- D. Russo and B. Van Roy. Learning to optimize via information-directed sampling. In Advances in Neural Information Processing Systems, pages 1583–1591. Curran Associates, Inc., 2014.
- D. Russo and B. Van Roy. An information-theoretic analysis of Thompson sampling. Journal of Machine Learning Research, 17(1):2442–2471, 2016. ISSN 1532-4435.

- D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen. A tutorial on Thompson sampling. Foundations and Trends in Machine Learning, 11(1):1–96, 2018.
- A. Saha, N. Natarajan, P. Netrapalli, and P. Jain. Optimal regret algorithm for pseudo-1d bandit convex optimization. In Marina Meila and Tong Zhang, editors, Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 9255–9264. PMLR, 2021.
- T. Shively, S. Walker, and P. Damien. Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. Journal of Econometrics, 161(2):166–181, 2011.
- S. Tarasov, L. G. Khachiyan, and I. I. Erlich. The method of inscribed ellipsoids. In Soviet Mathematics-Doklady, volume 37, pages 226–230, 1988.
- W. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3/4):285–294, 1933.
- T. Tkocz. Asymptotic convex geometry lecture notes. 2018.
- J. Zimmert and T. Lattimore. Connections between mirror descent, thompson sampling and the information ratio. In Advances in Neural Information Processing Systems, pages 11973–11982. Curran Associates, Inc., 2019.



## Appendix A. Technical inequalities

**Lemma 13** *Suppose that  $K$  is a convex body and  $f \in \mathcal{F}_{1_K}$  is a Lipschitz ridge function. Then there exists a  $\theta \in \mathbb{S}_1$  and convex  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  such that  $f(x) = \ell(\langle x, \theta \rangle)$  and  $\text{Lip}(\ell) \leq \text{Lip}(f)$ .*

**Proof** By assumption there exists a convex function  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  and  $\theta \in \mathbb{S}_1$  such that  $f(x) = \ell(\langle x, \theta \rangle)$ . What remains to show is that  $\ell$  can be chosen so that  $\text{Lip}(\ell) \leq \text{Lip}_K(f)$ . Let  $h_K$  be the support function associated with  $K$ , given by  $h_K(v) = \sup_{x \in K} \langle v, x \rangle$ . Therefore  $\ell$  is uniquely defined on  $I = [-h_K(-\theta), h_K(\theta)]$  and can be defined in any way that preserves convexity outside. Let  $Dg(x)[v]$  be the directional derivative of  $g$  at  $x$  in direction  $v$ , which for convex  $g$  exists for all  $x$  in the interior of the domain of  $g$ . Then

$$\begin{aligned} \text{Lip}_K(f) &\geq \sup_{x \in \text{int}(K)} \max(Df(x)[\theta], Df(x)[- \theta]) \\ &= \sup_{x \in \text{int}(K)} \max(D\ell(\langle x, \theta \rangle)[1], D\ell(\langle x, \theta \rangle)[-1]) \\ &= \sup_{x \in \text{int}(I)} \max(|D\ell(x)[1]|, |D\ell(x)[-1]|) \\ &= \text{Lip}_{\text{int}(I)}(\ell) \\ &= \text{Lip}_I(\ell). \end{aligned}$$

Then define  $\ell$  on all of  $\mathbb{R}$  via the classical extension (Lattimore, 2024, Proposition 3.17, for example). ■

## Appendix B. Comments on Measurability

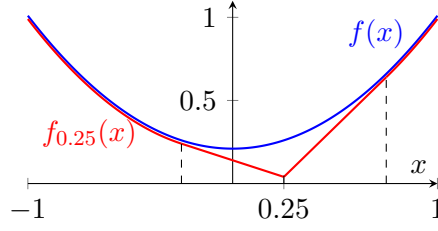
We assumed that probability measures on  $\mathcal{F}$  are defined with respect to some  $\sigma$ -algebra such that  $x \mapsto f(x)$  is measurable for all  $f \in \mathcal{F}$ . The mapping  $f \mapsto x_f$  was also assumed to be measurable. [Niemiro \(1992\)](#) showed that such a mapping always exists. Of course it follows that  $f \mapsto f_\star = f(x_f)$  is also measurable.

## Appendix C. Thompson Sampling for Adversarial Problems

In the Bayesian adversarial setting the prior is a probability measure on  $\mathcal{F}^n$  and a whole sequence of loss functions is sampled in secret by the environment. The natural generalisations of TS in this setting are the following:

- (a) Sample  $(f_s)_{s=1}^n$  from the posterior and play  $X_t = \arg \min_{x \in K} f_t(x)$ .
- (b) Sample  $(f_s)_{s=1}^n$  from the posterior and play  $X_t = \arg \min_{x \in K} \sum_{s=1}^t f_s(x)$ .

The version in (a) suffers linear regret as the following example shows. Let  $d = 1$  and  $K = [-1, 1]$  and  $f(x) = \varepsilon + \max(\varepsilon x, (1 - \varepsilon)x)$  and  $g(x) = f(-x)$ . Note that  $f \in \mathcal{F}_{b_1}$  and is piecewise linear and minimised at  $-1$  with  $f(-1) = 0$  and  $f(0) = \varepsilon$  and  $f(1) = 1$ . The function  $g$  is the mirror image of  $f$ . Now let  $\nu$  be the uniform distribution on  $\{f, g\}$  and  $\xi = \nu^n$  be the product measure. TS as defined in (a) plays uniformly on  $\{1, -1\}$  and an elementary calculation shows that the regret is  $\Omega(n)$ .



**Figure 3:** The function  $f(x) = 0.2 + 0.8x^2$  and the function  $f_{0.25}$  with  $\varepsilon = 0.2$ . The function  $f_{0.25}$  is the largest convex function that is smaller than  $f$  and has  $f_{0.25}(0.25) = f(0.25) - 0.2$ .

We do not know if the version of TS defined in (b) has  $\tilde{O}(\sqrt{n})$  Bayesian regret. However, the following example shows that in general the adversarial version of the information ratio is not bounded. Because the loss function changes from round to round, the action  $X_t$  may not minimise  $f_t$ . This must be reflected in the definition of the information ratio. Let  $\xi$  be a probability measure on  $\mathcal{F} \times K$  and  $\pi$  be a probability measure on  $K$  and let

$$\Delta(\pi, \xi) = \mathbb{E}[f(X) - f(X_\star)] \quad \text{and} \quad \mathcal{I}(\pi, \xi) = \mathbb{E}[(f(X) - \mathbb{E}[f(X)|X])^2],$$

where  $(f, X_\star, X)$  has law  $\xi \otimes \pi$ . Thompson sampling as in Item (b) is the policy  $\pi$  with the same law as  $X_\star$ . The claim is that in general it does not hold that

$$\Delta(\pi, \xi) \leq \alpha + \sqrt{\beta \mathcal{I}(\pi, \xi)},$$

unless  $\alpha$  is unreasonably large. Let  $d = 1, \varepsilon \in (0, 2^{-7})$ ,  $K = [-1, 1]$ , and  $f(x) = \varepsilon + (1 - \varepsilon)x^2$ . Given  $\theta \in [-1, 1]$  let  $f_\theta(x)$  be defined as

$$f_\theta(x) = \begin{cases} (1 - \varepsilon)(\theta^2 + 2(x - \theta)(\theta + \sqrt{\frac{\varepsilon}{1 - \varepsilon}})) & \theta \leq x \leq \theta + \sqrt{\frac{\varepsilon}{1 - \varepsilon}} \\ (1 - \varepsilon)(\theta^2 + 2(x - \theta)(\theta - \sqrt{\frac{\varepsilon}{1 - \varepsilon}})) & \theta - \sqrt{\frac{\varepsilon}{1 - \varepsilon}} \leq x < \theta \\ f(x) & \text{otherwise,} \end{cases}$$

which is convex and smaller than  $f$  for all  $x \in K$ . Essentially,  $f_\theta$  should be thought of as the largest convex function that is smaller than  $f$  and has  $f_\theta(\theta) = f(\theta) - \varepsilon$  (see Fig. 3). Moreover, an elementary calculation shows that  $\max_{x \in K} |f(x) - f_\theta(x)| = \varepsilon$  for all  $\theta \in [-1, 1]$ . Let  $\xi$  be the law of  $(f_\theta, \theta)$  when  $\theta$  is sampled uniformly from  $[-1, 1]$  and  $\pi$  be uniform on  $[-1, 1]$  which is the TS policy as defined in (b). Then, by letting  $\theta'$  be an i.i.d. copy of  $\theta$  we have

$$\Delta(\pi, \xi) = \mathbb{E}[f_\theta(\theta') - f_\theta(\theta)] = \mathbb{E}[f_\theta(\theta') - f(\theta)] + \varepsilon = \mathbb{E}[f_{\theta'}(\theta) - f(\theta)] + \varepsilon$$

where the second equality follows from the definition of  $f_\theta$  and the third equality follows from  $f_\theta(\theta) = f(\theta) - \varepsilon$ . Next, we have

$$\begin{aligned} \mathbb{E}[f_{\theta'}(\theta) - f(\theta)] &= \mathbb{E}\left[\mathbf{1}_{\{|\theta - \theta'| \leq \sqrt{\frac{\varepsilon}{1-\varepsilon}}\}} (f_{\theta'}(\theta) - f(\theta)) + \mathbf{1}_{\{|\theta - \theta'| \geq \sqrt{\frac{\varepsilon}{1-\varepsilon}}\}} (f_{\theta'}(\theta) - f(\theta))\right] \\ &\stackrel{(a)}{=} \mathbb{E}\left[\mathbf{1}_{\{|\theta - \theta'| \leq \sqrt{\frac{\varepsilon}{1-\varepsilon}}\}} (f_{\theta'}(\theta) - f(\theta))\right] \\ &\stackrel{(b)}{\geq} -\mathbb{P}\left(|\theta - \theta'| \leq \sqrt{\frac{\varepsilon}{1-\varepsilon}}\right) \varepsilon \\ &\stackrel{(c)}{\geq} -2\sqrt{\frac{\varepsilon}{1-\varepsilon}} \varepsilon, \end{aligned}$$

where (a) follows from the fact that  $f_{\theta'}(\theta) = f(\theta)$  if  $|\theta - \theta'| \geq \sqrt{\frac{\varepsilon}{1-\varepsilon}}$ ; (b) follows from the fact that  $f_{\theta'} \leq f(\theta)$  and the fact that  $\max_{x \in K} |f(x) - f_\theta(x)| = \varepsilon$ ; and (c) follows from the fact that  $\theta$  and  $\theta'$  are i.i.d. on  $[-1, 1]$ . Therefore,

$$\Delta(\pi, \xi) \geq \varepsilon \left(1 - 2\sqrt{\frac{\varepsilon}{1-\varepsilon}}\right).$$

Next, we turn our attention to  $\mathcal{I}(\pi, \xi)$ , which can be upper bounded as

$$\begin{aligned} \mathcal{I}(\pi, \xi) &= \mathbb{E}\left[(f_{\theta'}(\theta) - \mathbb{E}[f_{\theta'}(\theta)|\theta])^2\right] \\ &\stackrel{(a)}{\leq} \mathbb{E}\left[(f_{\theta'}(\theta) - f(\theta))^2\right] \\ &\stackrel{(b)}{\leq} \mathbb{P}\left(|\theta - \theta'| \leq \sqrt{\frac{\varepsilon}{1-\varepsilon}}\right) \varepsilon^2 \\ &\stackrel{(c)}{\leq} 2\varepsilon^2 \sqrt{\frac{\varepsilon}{1-\varepsilon}}, \end{aligned}$$

where (a) follows from the fact that the mean minimizes the squared deviation; (b) follows from the fact that  $f_{\theta'}(\theta) = f(\theta)$  if  $|\theta - \theta'| \geq \sqrt{\frac{\varepsilon}{1-\varepsilon}}$ ; and (c) follows from the fact that  $\theta$  and  $\theta'$  are i.i.d. on  $[-1, 1]$ . Therefore, by putting the two inequalities together we have

$$\frac{\Delta(\pi, \xi)}{\sqrt{\mathcal{I}(\pi, \xi)}} \geq \frac{\varepsilon \left(1 - 2\sqrt{\frac{\varepsilon}{1-\varepsilon}}\right)}{\sqrt{2\varepsilon^2 \sqrt{\frac{\varepsilon}{1-\varepsilon}}}} = \frac{1 - \sqrt{\frac{4\varepsilon}{1-\varepsilon}}}{\sqrt[4]{\frac{4\varepsilon}{1-\varepsilon}}} = \sqrt[4]{\frac{1-\varepsilon}{4\varepsilon}} - \sqrt[4]{\frac{4\varepsilon}{1-\varepsilon}},$$

which can be further lower bounded by

$$\frac{\Delta(\pi, \xi)}{\sqrt{\mathcal{I}(\pi, \xi)}} \geq \sqrt[4]{\frac{1-\varepsilon}{4\varepsilon}} - \sqrt[4]{\frac{4\varepsilon}{1-\varepsilon}} \geq \sqrt[4]{\frac{1}{4\varepsilon} - \frac{1}{4}} - \sqrt[4]{8\varepsilon} \geq \sqrt[4]{\frac{1}{8\varepsilon}} - \sqrt[4]{8\varepsilon} \geq \frac{1}{4}\varepsilon^{-\frac{1}{4}},$$

where the all inequalities follow from  $\varepsilon \in (0, 2^{-7})$ . Therefore, the information ratio is unbounded as  $\varepsilon \rightarrow 0$ .

### Appendix D. Proof of Lemma 3

Let  $\xi \in \mathcal{P}(\mathcal{F})$  and  $\bar{f} = \mathbb{E}[f]$  and  $\mathcal{F}_1, \dots, \mathcal{F}_m$  be disjoint subsets of  $\mathcal{F}$  such that  $\mathcal{F} = \cup_{i=1}^m \mathcal{F}_i$  and

$$\max_{i \in [m]} \left[ \sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f_\star) - \alpha - \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(k)} (\bar{f}_j(x_{f_l}) - f(x_{f_l}))^2} \right] \leq 0, \quad (8)$$

which exists by the assumptions in the lemma. When  $\xi(\mathcal{F}_i)$  define  $\nu_i$  as an arbitrary probability measure on  $\mathcal{F}$  and otherwise let  $\nu_i(\cdot) = \xi(\cdot \cap \mathcal{F}_i) / \xi(\mathcal{F}_i)$  and  $w_i = \xi(\mathcal{F}_i)$ . Therefore

$$\begin{aligned} \Delta(\pi, \xi) &= \int_{\mathcal{F}} (\bar{f}(x_f) - f_\star) d\xi(f) \\ &= \sum_{i=1}^m w_i \int_{\mathcal{F}} (\bar{f}(x_f) - f_\star) d\nu_i(f) \\ &\stackrel{(a)}{\leq} \sum_{i=1}^m w_i \sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f_\star) \\ &\stackrel{(b)}{\leq} \alpha + \sum_{i=1}^m w_i \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2} \\ &\stackrel{(c)}{\leq} \alpha + \sum_{i=1}^m w_i \sqrt{\beta k(k-1) \int_{\mathcal{F}} \int_{\mathcal{F}} (\bar{f}(x_g) - f(x_g))^2 d\nu_i(f) d\nu_i(g)} \\ &\stackrel{(d)}{\leq} \alpha + \sqrt{\beta m k(k-1) \sum_{i=1}^m w_i^2 \int_{\mathcal{F}} \int_{\mathcal{F}} (\bar{f}(x_g) - f(x_g))^2 d\nu_i(f) d\nu_i(g)} \\ &\stackrel{(e)}{\leq} \alpha + \sqrt{\beta m k(k-1) \sum_{i=1}^m \sum_{j=1}^m w_i w_j \int_{\mathcal{F}} \int_{\mathcal{F}} (\bar{f}(x_g) - f(x_g))^2 d\nu_i(f) d\nu_j(g)} \\ &= \alpha + \sqrt{\beta m k(k-1) \int_{\mathcal{F}} \int_{\mathcal{F}} (\bar{f}(x_g) - f(x_g))^2 d\xi(f) d\xi(g)} \\ &= \alpha + \sqrt{\beta m k(k-1) \mathcal{I}(\pi, \xi)}, \end{aligned} \quad (9)$$

where (a) is immediate from the definition of the integral. (b) follows from Eq. (8). (c) is true because if  $f_1, \dots, f_k$  are sampled independently from  $\nu_i$ , then

$$\begin{aligned} \int_{\mathcal{F}} \int_{\mathcal{F}} (\bar{f}(x_g) - f(x_g)) d\nu_i(f) d\nu_i(g) &= \frac{1}{k(k-1)} \mathbb{E} \left[ \sum_{j, l \in \text{PAIR}(k)} (\bar{f}(x_{f_l}) - f_j(x_{f_l}))^2 \right] \\ &\geq \frac{1}{k(k-1)} \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(k)} (\bar{f}(x_{f_l}) - f_j(x_{f_l}))^2. \end{aligned}$$

(d) follows from Cauchy-Schwarz and (e) by introducing additional non-negative terms. Since Eq. (9) holds for all  $\xi \in \mathcal{P}(\mathcal{F})$  it follows that

$$\sup_{\xi \in \mathcal{P}(\mathcal{F})} \left[ \Delta(\pi, \xi) - \alpha - \sqrt{\beta m k(k-1) \mathcal{I}(\pi, \xi)} \right] \leq 0$$

and therefore  $(\alpha, \beta m k(k-1)) \in \text{IR}(\xi)$ .

## Appendix E. Lower Bounds

**Proof** [Proof of Lemma 10] Recall that  $\varepsilon \in [0, 1/2]$  and

$$f(x) = \varepsilon + \frac{1}{2} \|x\|^2 \quad f_\theta(x) = \begin{cases} f(x) & \text{if } \|x - \theta\|^2 \geq 1 + 2\varepsilon \\ \langle \theta, x \rangle - 1 + \sqrt{1 + 2\varepsilon} \|\theta - x\| & \text{otherwise.} \end{cases}$$

We need to prove that  $f_\theta$  on  $\mathbb{B}_1$  is convex, Lipschitz and minimised at  $\theta$ . Convexity follows because

$$\begin{aligned} g_\theta(x) &\triangleq \sup_{y \in \mathbb{R}^d} \{ f(y) + \langle f'(y), x - y \rangle : f(y) + \langle f'(y), \theta - y \rangle \leq 0 \} \\ &= \sup_{\substack{y \in \mathbb{R}^d \\ r \leq 0}} \{ f(y) + \langle f'(y), x - y \rangle : f(y) + \langle f'(y), \theta - y \rangle = r \} \\ &= \sup_{\substack{y \in \mathbb{R}^d \\ r \leq 0}} \{ \langle f'(y), x - \theta \rangle + r : f(y) + \langle f'(y), \theta - y \rangle = r \} \\ &= \sup_{\substack{y \in \mathbb{R}^d \\ r \leq 0}} \{ \langle y, x - \theta \rangle + r : \|y - \theta\|^2 = 1 + 2\varepsilon - 2r \} \\ &= \sup_{r \leq 0} \{ \langle \theta, x - \theta \rangle + r + \sqrt{1 + 2\varepsilon - 2r} \|x - \theta\| \} \\ &= \sup_{r \leq 0} \{ \langle \theta, x \rangle - 1 + r + \sqrt{1 + 2\varepsilon - 2r} \|x - \theta\| \} \\ &= f_\theta(x), \end{aligned} \tag{10}$$

where in the final inequality we note that the maximising  $r$  is

$$r = \begin{cases} \frac{1}{2} + \varepsilon - \frac{1}{2} \|x - \theta\|^2 & \text{if } \|x - \theta\|^2 \geq 1 + 2\varepsilon \\ 0 & \text{otherwise.} \end{cases}$$

Therefore  $f_\theta$  is the supremum over a set of linear functions and hence convex. That  $f_\theta$  is minimised at  $\theta$  follows directly from the first-order optimality conditions. Let  $\eta \in \mathbb{S}_1$ . Then

$$Df_\theta(\theta)[\eta] = \langle \theta, \eta \rangle + \sqrt{1 + 2\varepsilon} \|\eta\| > 0,$$

where  $Df_\theta(\theta)[\cdot]$  is the directional derivative operator (noting that  $f_\theta$  is not differentiable at  $\theta$ ). Lastly, for Lipschitzness. Since  $f_\theta$  is continuous, it suffices to bound  $\|f'_\theta(\cdot)\|$  on  $\text{int}(\mathbb{B}_1)$  where  $f_\theta$

is differentiable. When  $\|x - \theta\|^2 \geq 1 + 2\varepsilon$ , then  $\|f'_\theta(x)\| = \|f'(x)\| = \|x\| \leq 1$ . On the other hand, if  $\|x - \theta\|^2 < 1 + 2\varepsilon$ , then

$$\begin{aligned} \|f'_\theta(x)\|^2 &= \left\| \theta + \sqrt{1 + 2\varepsilon} \frac{x - \theta}{\|x - \theta\|} \right\|^2 \\ &= 2 + 2\varepsilon + \sqrt{1 + 2\varepsilon} \frac{\langle \theta, x - \theta \rangle}{\|x - \theta\|} \\ &\leq 2 + 2\varepsilon. \end{aligned}$$

Therefore  $\text{Lip}_{\mathbb{B}_1}(f) \leq \sqrt{2 + 2\varepsilon}$ . ■

**Definition 14** For  $\theta \in \mathbb{S}_1$  and  $\varepsilon \in [0, 1/2]$ , define

$$\mathcal{C}_{\theta, \varepsilon} = \left\{ x \in \mathbb{S}_1 : \|\theta - x\|^2 < 1 + 2\varepsilon \right\}$$

and  $\bar{\mathcal{C}}_{\theta, \varepsilon} = \mathbb{S}_1 \setminus \mathcal{C}_{\theta, \varepsilon}$ .

**Definition 15** Let  $\sigma(\cdot)$  be the uniform distribution over the unit sphere  $\mathbb{S}_1$ . Moreover, let  $\sigma_S(\cdot)$  be the uniform distribution over a set  $S \subseteq \mathbb{S}_1$  defined as  $\sigma_S(\cdot) = \frac{\sigma(\cdot \cap S)}{\sigma(S)}$ .

We use the following theorem from (Tkocz, 2018) to bound the surface area of spherical caps.

**Theorem 16** [Tkocz, 2018, Theorem B.1] For all  $\varepsilon \in [0, 1]$  and  $\theta \sim \sigma$  we have

$$\mathbb{P}(\langle \theta, e_1 \rangle \geq \varepsilon) \leq \exp\left(-\frac{d\varepsilon^2}{2}\right).$$

### E.1. Proof of Theorem 11

**Proof** [Proof of Theorem 11] Define  $f$  and  $f_\theta$  as in Lemma 10 with  $\varepsilon = 1/4$ , which then using Definition 14 can be written as

$$f_\theta(x) = \begin{cases} f(x) & \text{if } x \in \bar{\mathcal{C}}_\theta, \\ \langle \theta, x \rangle - 1 + \sqrt{\frac{3}{2}} \|\theta - x\| & \text{if } x \in \mathcal{C}_\theta, \end{cases}$$

where we drop the  $\varepsilon$  from  $\mathcal{C}_{\theta, \varepsilon}$  and  $\bar{\mathcal{C}}_{\theta, \varepsilon}$  in the notation for simplicity. We define the bandit instance by setting the prior  $\xi_1$  to be the law of  $f_\theta$  when  $\theta$  has law  $\sigma$ , and letting the observation noise to be zero, meaning that

$$Y_t = f_{\theta_*}(X_t),$$

where  $X_t$  is the action played at round  $t$ ,  $Y_t$  is the loss observed at round  $t$ , and  $f_{\theta_*}$  is the true function that is secretly sampled from the prior  $\xi_1$ . Also, define the random sets  $\Theta_t \subseteq \mathbb{S}_1$  as

$$\Theta_t = \left\{ \theta \in \mathbb{S}_1 : f_\theta(X_s) = \frac{3}{4}, \forall s \in [t-1] \right\}.$$

We also make extensive use of the fact that for any two  $\theta_1, \theta_2 \in \mathbb{S}_1$ , we have

$$f_{\theta_1}(\theta_2) = f_{\theta_2}(\theta_1) = \frac{3}{4}, \quad \text{if and only if } \theta_1 \in \bar{\mathcal{C}}_{\theta_2} \Leftrightarrow \theta_2 \in \bar{\mathcal{C}}_{\theta_1},$$

which follows from the definition of  $f_\theta$ .

**Step 1:** First we show that if the posterior distribution  $\xi_t$  at round  $t \in [T]$  is uniform over  $\Theta_t$ , and the algorithm observes the loss  $Y_t = \frac{3}{4}$  as a result of playing  $X_t$ , then the posterior distribution  $\xi_{t+1}$  at round  $t+1$  is uniform over  $\Theta_{t+1}$ , i.e.,  $\xi_{t+1} = \sigma_{\Theta_{t+1}}$ . To this end, observe that if  $Y_t = \frac{3}{4}$  then for any set  $B \subseteq \Theta_t$ ,

$$\begin{aligned} \xi_{t+1}(B) &= \mathbb{P}_{t-1} \left( \theta_\star \in B \mid Y_t = \frac{3}{4}, X_t \right) = \frac{\mathbb{P}_{t-1}(\theta_\star \in B, Y_t = \frac{3}{4} \mid X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4} \mid X_t)} \\ &= \frac{\mathbb{P}_{t-1}(Y_t = \frac{3}{4} \mid X_t, \theta_\star \in B) \mathbb{P}_{t-1}(\theta_\star \in B \mid X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4} \mid X_t)} \\ &= \frac{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4} \mid X_t, \theta_\star \in B) \mathbb{P}_{t-1}(\theta_\star \in B)}{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4} \mid X_t)}. \quad (11) \end{aligned}$$

Note that TS samples  $f_{\theta_t}$  from  $\xi_t$ , and then plays the minimizer of  $f_{\theta_t}$ , which from Lemma 10 is  $\theta_t$ , i.e.  $X_t = \theta_t$ . Consequently, continuing from Eq. (11) with the assumption of this step that  $\theta_t, \theta_\star \sim \sigma_{\Theta_t}$  and the fact that  $X_t = \theta_t$ , we have

$$\begin{aligned} \xi_{t+1}(B) &= \frac{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4} \mid X_t, \theta_\star \in B) \mathbb{P}_{t-1}(\theta_\star \in B)}{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4} \mid X_t)} \\ &= \frac{\mathbb{P}_{t-1}(\theta_\star \in \bar{\mathcal{C}}_{\theta_t} \mid \theta_t, \theta_\star \in B) \mathbb{P}_{t-1}(\theta_\star \in B)}{\mathbb{P}_{t-1}(\theta_\star \in \bar{\mathcal{C}}_{\theta_t} \mid \theta_t)} \\ &= \frac{\frac{\sigma(B \cap \bar{\mathcal{C}}_{\theta_t})}{\sigma(B)} \cdot \frac{\sigma(B)}{\sigma(\Theta_t)}}{\frac{\sigma(\bar{\mathcal{C}}_{\theta_t} \cap \Theta_t)}{\sigma(\Theta_t)}} \\ &= \frac{\sigma(B \cap \bar{\mathcal{C}}_{\theta_t})}{\sigma(\Theta_t \cap \bar{\mathcal{C}}_{\theta_t})} \end{aligned}$$

which implies that  $\xi_{t+1}$  is uniform over  $\Theta_t \cap \bar{\mathcal{C}}_{\theta_t}$ . Lastly, note that

$$\Theta_{t+1} = \left\{ \theta \in \mathbb{S}_1 : f_\theta(X_s) = \frac{3}{4}, \forall s \in [t] \right\} = \Theta_t \cap \left\{ \theta \in \mathbb{S}_1 : f_\theta(\theta_t) = \frac{3}{4} \right\} = \Theta_t \cap \bar{\mathcal{C}}_{\theta_t},$$

which means that  $\xi_{t+1}$  is uniform over  $\Theta_{t+1}$ .

**Step 2:** Let  $\delta = \sigma(\mathcal{C}_{\theta_\star})$ , and note that  $\sigma(\mathcal{C}_\theta) = \delta$  for all  $\theta \in \mathbb{S}_1$  due to the shape of the  $\mathcal{C}_\theta$  which is a spherical cap with a fixed radius. Consider the event  $\mathcal{E}_t$  where  $X_1, \dots, X_t \in \bar{\mathcal{C}}_{\theta_\star}$ , which implies both that  $Y_1, \dots, Y_t = \frac{3}{4}$ , and that  $\xi_{t+1}$  is uniform over  $\Theta_{t+1}$ . Conditioned on  $\mathcal{E}_t$ , we have

$$\begin{aligned} \sigma(\Theta_{t+1}) &= \sigma \left( \left\{ \theta \in \mathbb{S}_1 : f_\theta(X_s) = \frac{3}{4}, \forall s \in [t] \right\} \right) \\ &= \sigma \left( \left\{ \theta \in \mathbb{S}_1 : f_\theta(\theta_s) = \frac{3}{4}, \forall s \in [t] \right\} \right) \\ &= \sigma \left( \left\{ \theta \in \mathbb{S}_1 : \theta \in \bar{\mathcal{C}}_{\theta_s}, \forall s \in [t] \right\} \right) \\ &= \sigma \left( \left\{ \theta \in \mathbb{S}_1 : \theta \notin \bigcup_{s=1}^t \mathcal{C}_{\theta_s} \right\} \right) \\ &\geq 1 - t\delta. \end{aligned}$$



Therefore, the probability of TS playing  $X_{t+1} \in \mathcal{C}_{\theta_*}$  is upper bounded by

$$\mathbb{P}(X_{t+1} \in \mathcal{C}_{\theta_*} | \theta_*, \mathcal{E}_t) = \frac{\sigma(\mathcal{C}_{\theta_*} \cap \Theta_{t+1})}{\sigma(\Theta_{t+1})} \leq \frac{\delta}{1 - t\delta},$$

which further implies that

$$\mathbb{P}(\mathcal{E}_{t+1} | \mathcal{E}_t) = \mathbb{P}\left(Y_{t+1} = \frac{3}{4} | \mathcal{E}_t\right) = \mathbb{P}(X_{t+1} \in \bar{\mathcal{C}}_{\theta_*} | \mathcal{E}_t) \geq 1 - \frac{\delta}{1 - t\delta},$$

and therefore

$$\mathbb{P}(\mathcal{E}_{t+1}) = \mathbb{P}(\mathcal{E}_t) \mathbb{P}(\mathcal{E}_{t+1} | \mathcal{E}_t) \geq \mathbb{P}(\mathcal{E}_t) \left(1 - \frac{\delta}{1 - t\delta}\right) \geq \mathbb{P}(\mathcal{E}_t) - \frac{\delta}{1 - t\delta}.$$

Let  $n_0 = \min(\lfloor \frac{1}{4\delta} \rfloor, n)$ , then

$$\mathbb{P}(\mathcal{E}_{n_0}) \geq \mathbb{P}(\mathcal{E}_{n_0-1}) - \frac{\delta}{1 - (n_0 - 1)\delta} \geq \mathbb{P}(\mathcal{E}_1) - \sum_{t=1}^{n_0-1} \frac{\delta}{1 - t\delta} = 1 - \sum_{t=0}^{n_0-1} \frac{\delta}{1 - t\delta}$$

where the last equality follows from  $\mathbb{P}(\mathcal{E}_1) = \mathbb{P}(\theta_1 \in \bar{\mathcal{C}}_{\theta_*}) = 1 - \delta$ . Since  $t\delta \leq n_0\delta \leq 1/4$  for all  $t < n_0$ , we have

$$\mathbb{P}(\mathcal{E}_{n_0}) \geq 1 - \sum_{t=0}^{n_0-1} \frac{\delta}{1 - 1/4} = 1 - \frac{4}{3}n_0\delta \geq \frac{2}{3}.$$

Therefore, the expected regret of TS is lower bounded by

$$\text{BReg}_n(\text{TS}, \xi_1) \geq \text{BReg}_{n_0}(\text{TS}, \xi_1) \geq \frac{3}{4}n_0\mathbb{P}(\mathcal{E}_{n_0}) \geq \frac{1}{2}n_0,$$

since the algorithm incurs maximum regret of  $\frac{3}{4}$  in every round  $s \in [n_0]$  given the event  $\mathcal{E}_{n_0}$ . Finally, using Theorem 16, we have

$$\delta \leq \exp(-d/32),$$

which implies that

$$\text{BReg}_n(\text{TS}, \xi_1) \geq \frac{1}{2} \min\left(n, \left\lfloor \frac{1}{4} \exp(d/32) \right\rfloor\right).$$

■

## E.2. Proof of Theorem 12

Throughout this section, we use the same construction as the one used in Appendix E.1 except with  $\varepsilon = \frac{8 \log(d)}{d}$ . Therefore, we have

$$f(x) = \frac{1}{2} \|x\| + \frac{8 \log(d)}{d}, \quad \text{and} \quad f_\theta(x) = \begin{cases} f(x) & \text{if } x \in \bar{\mathcal{C}}_\theta, \\ \langle \theta, x \rangle - 1 + \sqrt{1 + \frac{16 \log(d)}{d}} \|\theta - x\| & \text{if } x \in \mathcal{C}_\theta. \end{cases}$$

Further, let  $\xi$  be the law of  $f_\theta$  where  $\theta \sim \sigma$ , and for  $x \in \mathbb{B}_1$  define

$$\Delta_x = \mathbb{E}[f_\theta(x) - f_\theta(\theta)] = \mathbb{E}[f_\theta(x)], \quad \text{and} \quad \mathcal{I}_x = \mathbb{E}\left[(f_\theta(x) - \mathbb{E}[f_\theta(x)])^2\right],$$

which should be thought of as the expected loss and the expected information gain at  $x$ . Therefore, for any policy  $\pi$  we have

$$\Delta(\pi, \xi) = \mathbb{E}[\Delta_X] \quad \text{and} \quad \mathcal{I}(\pi, \xi) = \mathbb{E}[\mathcal{I}_X],$$

where  $X \sim \pi$ . The basic idea is to prove that  $\Delta_x = \Omega(\log(d)/d)$  for all  $x \in \mathbb{B}_1$  and  $\mathcal{I}_x = O(\log(d)^4 d^{-4})$  for all  $x \in \mathbb{B}_1$ , which implies that  $\Delta(\pi, \xi) = \Omega(\log(d)/d)$  and  $\mathcal{I}(\pi, \xi) = \tilde{O}(\log(d)^4 d^{-4})$  for any policy  $\pi$ , and hence the claimed lower bound.

Additional to  $\varepsilon = 8 \log(d)/d$ , we fix  $\tau = \sqrt{8\varepsilon} = 8\sqrt{\log(d)/d}$  in the rest of this section.

**Lemma 17** For  $\tau \leq r$ , and  $x \in \mathbb{B}_1$  with  $\|x\| = r$ ,  $\mathbb{P}(f_\theta(x) = f(x)) \geq 1 - \frac{1}{d^4}$ .

**Proof** From Lemma 10,  $f_\theta(x) = f(x)$  if  $\|\theta - x\|^2 \geq 1 + 2\varepsilon$ . Moreover, for  $x \in \mathbb{B}_1$  with  $\|x\| = r$ ,

$$\|\theta - x\|^2 = \|\theta\|^2 + r^2 - 2\langle \theta, x \rangle = 1 + r^2 - 2\langle \theta, x \rangle,$$

which implies that  $f(x) = f_\theta(x)$  if  $\langle \theta, x \rangle \leq \frac{r^2}{2} - \varepsilon$ . Since  $\theta \sim \sigma$ ,

$$\begin{aligned} \mathbb{P}\left(\langle \theta, x \rangle \leq \frac{r^2}{2} - \varepsilon\right) &= \mathbb{P}\left(\langle \theta, e_1 \rangle \leq \left(\frac{r^2}{2} - \varepsilon\right) \|x\|^{-1}\right) \\ &= 1 - \mathbb{P}\left(\langle \theta, e_1 \rangle > \left(\frac{r}{2} - \frac{\varepsilon}{r}\right)\right) \\ &\geq 1 - \exp\left(-\left(\frac{r}{2} - \frac{\varepsilon}{r}\right)^2 \frac{d}{2}\right) \\ &\geq 1 - \exp\left(-\left(4\sqrt{\frac{\log(d)}{d}} - \frac{\sqrt{d}\log(d)}{d\sqrt{\log(d)}}\right)^2 \frac{d}{2}\right) \\ &= 1 - \exp\left(-\frac{9\log(d)}{2}\right) \\ &\geq 1 - \exp(-\log(d^4)) \\ &\geq 1 - \frac{1}{d^4}, \end{aligned}$$

where the first inequality follows from Theorem 16, and the second inequality follows from the fact that  $r \geq \tau \geq \sqrt{2\varepsilon}$ .  $\blacksquare$

**Lemma 18** For all  $x \in \mathbb{B}_1$  and  $\theta \in \mathbb{S}_1$ , we have

$$f_\theta(x) \geq \langle \theta, x \rangle - 1 + \sqrt{1 + 2\varepsilon} \|\theta - x\|.$$

**Proof** The proof follows by setting  $r = 0$  in Equation (10).  $\blacksquare$

**Lemma 19** For  $d \geq 2^8$  and  $x \in \mathbb{B}_1$ , we have  $\Delta_x \geq 2 \frac{\log(d)}{d}$ .

**Proof** Let  $r = \|x\|$ . We prove this result by considering two cases.

**CASE 1** If  $r \geq \tau$ , then

$$\mathbb{E}[f_\theta(x)] \geq \mathbb{P}(f_\theta(x) = f(x)) f(x) \stackrel{(a)}{\geq} \left(1 - \frac{1}{d^4}\right) \left(\frac{1}{2}r^2 + \varepsilon\right) \geq \frac{1}{2} \left(\frac{32 \log(d)}{d} + \frac{8 \log(d)}{d}\right) \geq \frac{20 \log(d)}{d},$$

where (a) follows from Lemma 17.

**CASE 2** If  $r < \tau$ , using the lower bound on  $f_\theta(x)$  from Lemma 18,

$$\begin{aligned} \mathbb{E}[f_\theta(x)] &\geq \mathbb{E}[\langle \theta, x \rangle - 1 + \sqrt{1 + 2\varepsilon} \|\theta - x\|] \\ &\stackrel{(a)}{=} \sqrt{1 + 2\varepsilon} \mathbb{E}[\sqrt{1 + r^2 - 2\langle \theta, x \rangle}] - 1 \\ &\stackrel{(b)}{\geq} \left(1 + \frac{2\varepsilon - 4\varepsilon^2}{2}\right) \mathbb{E}\left[1 + \frac{r^2 - 2\langle \theta, x \rangle - (r^2 - 2\langle \theta, x \rangle)^2}{2}\right] - 1 \\ &\stackrel{(c)}{=} (1 + \varepsilon - 2\varepsilon^2) \left(1 + \frac{r^2 - r^4}{2} - \mathbb{E}[2\langle \theta, x \rangle^2]\right) - 1 \\ &\stackrel{(d)}{=} (1 + \varepsilon - 2\varepsilon^2) \left(1 + \frac{r^2 - r^4}{2} - \frac{2r^2}{16}\right) - 1 \\ &= (1 + \varepsilon - 2\varepsilon^2) \left(1 + \frac{3r^2}{8} - \frac{r^4}{2}\right) - 1, \end{aligned}$$

where (a) follows from  $\mathbb{E}[\langle \theta, x \rangle] = 0$ , (b) follows from the inequality  $\sqrt{1+a} \geq 1 + \frac{a-a^2}{2}$  for  $a \geq -1$ , (c) follows from  $\mathbb{E}[\langle \theta, x \rangle] = 0$ , (d) follows from  $\mathbb{E}[\langle \theta, x \rangle^2] = \frac{r^2}{d}$  and  $d > 16$ . Note that  $(1 + 3r^2/8 - r^4/2) \geq 1$  for  $0 \leq r \leq \sqrt{3/4}$ , and is decreasing in  $r$  for  $r \in [\sqrt{3/4}, 1]$ . Therefore,

$$1 + \frac{3r^2}{8} - \frac{r^4}{2} \geq \min(1, 1 + \frac{3\tau^2}{8} - \frac{\tau^4}{2}) = 1 + \min(0, 3\varepsilon - 4\varepsilon^2).$$

This let us further lower bound  $\mathbb{E}[f_\theta(x)]$  as

$$\mathbb{E}[f_\theta(x)] \geq (1 + \varepsilon - 2\varepsilon^2) (1 + \min(0, 3\varepsilon - 4\varepsilon^2)) - 1.$$

Now, since we have  $\varepsilon \leq 1/3$ , we have  $3\varepsilon - 4\varepsilon^2 \geq 0$ , which gives

$$\mathbb{E}[f_\theta(x)] \geq 1 + \varepsilon - 2\varepsilon^2 - 1 = \varepsilon - 2\varepsilon^2 \geq \varepsilon - \frac{2}{3}\varepsilon \geq \frac{8 \log(d)}{3d}.$$

■

**Lemma 20** For all  $x \in \mathbb{B}_1$  and  $d > 256$ ,  $\mathcal{I}_x \leq 2^{40} \frac{\log(d)^4}{d^4}$ .

**Proof** Let  $x \in \mathbb{B}_1$  and  $r = \|x\|$ . We prove this result by considering two cases.

**CASE 1** If  $r \geq \tau$ , then we have

$$\begin{aligned}
 \mathbb{E} \left[ (f_\theta(x) - \mathbb{E}[f_\theta(x)])^2 \right] &\leq \mathbb{E} \left[ (f_\theta(x) - f(x))^2 \right] \\
 &\leq \mathbb{P}(f_\theta(x) = f(x)) (f(x) - f(x))^2 + \mathbb{P}(f_\theta(x) \neq f(x)) \left( \frac{1}{2}r^2 + \varepsilon \right)^2 \\
 &\leq \frac{1}{d^4} \left( \frac{1}{2}r^2 + \varepsilon \right)^2 \leq \frac{1}{d^4} \left( \frac{1}{2} + \frac{1}{3} \right)^2 \leq \frac{25}{36d^4},
 \end{aligned} \tag{12}$$

where the first inequality holds since the mean minimizes the squared deviation, and the second inequality holds since  $0 \leq f(x) - f_\theta(x) \leq \frac{1}{2}r^2 + \varepsilon$ .

**CASE 2** Now suppose that  $r < \tau$ . We have

$$\begin{aligned}
 0 &\stackrel{(a)}{\leq} f(x) - f_\theta(x) \\
 &\stackrel{(b)}{\leq} f(x) - \langle \theta, x \rangle + 1 - \sqrt{1 + 2\varepsilon} \|\theta - x\| \\
 &= f(x) - \langle \theta, x \rangle + 1 - \sqrt{1 + 2\varepsilon} \sqrt{1 + r^2 - 2\langle \theta, x \rangle} \\
 &\stackrel{(c)}{\leq} \varepsilon + \frac{r^2}{2} - \langle \theta, x \rangle + 1 - (1 + \varepsilon - 2\varepsilon^2) \left( 1 + \frac{r^2}{2} - \langle \theta, x \rangle - \frac{(r^2 - 2\langle \theta, x \rangle)^2}{2} \right) \\
 &= \frac{(r^2 - 2\langle \theta, x \rangle)^2}{2} + 2\varepsilon^2 - (\varepsilon - 2\varepsilon^2) \left( \frac{r^2}{2} - \langle \theta, x \rangle - \frac{(r^2 - \langle \theta, x \rangle)^2}{2} \right) \\
 &= \frac{r^4}{2} + 2\langle \theta, x \rangle^2 - 2\langle \theta, x \rangle r^2 + 2\varepsilon^2 - (\varepsilon - 2\varepsilon^2) \left( \frac{r^2}{2} - \langle \theta, x \rangle - \frac{r^4}{2} - 2\langle \theta, x \rangle^2 + 2\langle \theta, x \rangle r^2 \right) \\
 &\stackrel{(d)}{\leq} 30(r^2 + |\langle \theta, x \rangle| + \varepsilon)^2,
 \end{aligned}$$

where (a) follows from  $f(x) \geq f_\theta(x)$ , the (b) follows from Lemma 18, (c) follows from  $\sqrt{1+x} \geq 1 + \frac{x}{2} - \frac{x^2}{2}$  for all  $x \geq -1$ , and (d) follows from the fact that  $r, \varepsilon, |\langle \theta, x \rangle| \leq 1$ , and the expression in the previous line a polynomial of these terms with degree at least 2 and sum of coefficients at most 30.

Next, starting from the right-hand side of Eq. (12), we have

$$\begin{aligned}
 \mathbb{E} \left[ (f(x) - f_\theta(x))^2 \right] &\leq \mathbb{E} \left[ 30^2 (r^2 + |\langle \theta, x \rangle| + \varepsilon)^4 \right] \\
 &\stackrel{(a)}{\leq} 30^2 \mathbb{E} \left[ 27 (r^8 + \langle \theta, x \rangle^4 + \varepsilon^4) \right] \\
 &\leq 30^3 \left( r^8 + \mathbb{E} \left[ \langle \theta, x \rangle^4 \right] + \varepsilon^4 \right) \\
 &\stackrel{(b)}{\leq} 30^3 \left( r^8 + \frac{3r^4}{d^2} + \varepsilon^4 \right) \\
 &\stackrel{(c)}{\leq} 30^3 \left( 8^4 \varepsilon^4 + \frac{3 \cdot 8^2 \varepsilon^2}{d^2} + \varepsilon^4 \right) \\
 &\leq 30^3 (8^4 \varepsilon^4 + 3\varepsilon^4 + \varepsilon^4) \\
 &\leq 2^{28} \varepsilon^4 = 2^{40} \frac{\log(d)^4}{d^4},
 \end{aligned}$$

where (a) follows from  $(a+b+c)^4 \leq 27(a^4+b^4+c^4)$ , (b) follows from the fact that  $\mathbb{E}[\langle \theta, x \rangle^4] = \frac{3r^4}{d(d+2)} < \frac{3r^4}{d^2}$ , and (c) follows from  $r < \tau = \sqrt{8\varepsilon}$ .  $\blacksquare$

**Proof** [Proof of Theorem 12] Let  $\xi$  be the law of  $f_\theta$  when  $\theta$  has law  $\sigma$ . Then for any policy  $\pi$ , and  $X \sim \pi$ , from Lemma 19 we have

$$\Delta(\pi, \xi) = \mathbb{E}[\Delta_X] \geq \frac{\log(d)}{2d},$$

and from Lemma 20 we have

$$\mathcal{I}(\pi, \xi) = \mathbb{E}[\mathcal{I}_X] \leq \frac{2^{40} \log(d)^4}{d^4},$$

which together imply

$$\frac{\Delta(\pi, \xi)}{\sqrt{\mathcal{I}(\pi, \xi)}} \geq \frac{d}{2^{19} \log(d)}.$$

$\blacksquare$

## Appendix F. Approximate Thompson sampling and Proof of Theorem 2

Before proving Theorem 2 we explain how to generalise TS so that  $X_t$  only needs to approximately minimise  $f_t$ . This is surprisingly subtle. And indeed, we were only able to analyse an approximate version of TS that uses a small amount of regularisation.

**Definition 21** Let  $\varepsilon_O \leq \varepsilon_R$  be non-negative constants called the optimisation accuracy and regularisation parameter, respectively. Given  $f \in \mathcal{F}_1$  let  $\tilde{f}(x) = f(x) + \frac{\varepsilon_R}{2} \|x\|^2$  when  $\varepsilon_R > 0$  define

$$\tilde{x}_f = \arg \min_{x \in K} \tilde{f}(x) \qquad \bar{x}_f = \left\{ x : \tilde{f}(x) \leq \min_{y \in K} \tilde{f}(y) + \varepsilon_O \right\}.$$

When the regularisation parameter  $\varepsilon_R = 0$ , define  $\tilde{x}_f = x_f$  and  $\bar{x}_f = \{x_f\}$ .

Approximate Thompson Sampling (ATS) samples  $f_t$  from the posterior and plays any action  $X_t \in \bar{x}_{f_t}$ .

```

1  args: prior  $\xi$ 
2  for  $t = 1$  to  $\infty$ :
3      sample  $f_t$  from  $\mathbb{P}(f = \cdot | X_1, Y_1, \dots, X_{t-1}, Y_{t-1})$ 
4      play  $X_t \in \bar{x}_{f_t}$ 
5      observe  $Y_t$ 
    
```

**Algorithm 2:** Approximate Thompson sampling

When  $\varepsilon_R = \varepsilon_O = 0$ , then ATS and TS are equivalent, though we note the importance in our analysis that the ties in TS are broken in a consistent fashion. The regularisation in the definition of  $\tilde{f}$  ensures that all points in  $\bar{x}_f$  are reasonably close to  $\tilde{x}_f$  and introduces a degree of stability into ATS. An obvious question is whether or not you could do away with the regularisation and define  $\bar{x}_f$  by  $\{x : f_t(x) \leq f_{t*} + \varepsilon\}$  for suitably small  $\varepsilon \geq 0$ . We suspect the answer is yes but do not currently have a proof. The regularisation ensures that  $\bar{x}_f$  has small diameter, which need not be true in general for  $\{x : f(x) \leq f_* + \varepsilon\}$ , even if  $\varepsilon$  is arbitrarily small.

**Analysis** We start by defining a kind of cover of a set of convex functions  $\mathcal{F}$ . In the standard analysis introduced by [Bubeck et al. \(2015\)](#) and [Bubeck and Eldan \(2018\)](#), this cover was defined purely in terms of the optimal action. As noticed by [Lattimore \(2021\)](#), this argument relies on  $\mathcal{F}$  being closed under convex combinations, which is not true for the space of ridge functions. Here we introduce a new definition notion of cover for function classes  $\mathcal{F}$  that are not closed under convex combinations.

**Definition 22** Let  $\mathcal{F}$  be a set of convex functions from  $K$  to  $\mathbb{R}$  and  $\varepsilon > 0$ . Define  $N(\mathcal{F}, \varepsilon)$  to be the smallest number  $N$  such that there exists  $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$  such that for all  $k \in [N]$ :

- *Closure:*  $\mathcal{F}_k$  is a subset of  $\mathcal{F}$  and  $\text{conv}(\mathcal{F}_k) \subset \mathcal{F}$ .
- *Common near-minimiser:* There exists an  $x_k \in K$  such that  $\|\tilde{x}_f - x_k\| \leq \varepsilon$  for all  $f \in \mathcal{F}_k$ .

Moreover:

- *Approximation:* For all  $f \in \mathcal{F}$  there exists a  $k \in [N]$  and  $g \in \mathcal{F}_k$  such that  $\|f - g\|_\infty \leq \varepsilon$  and  $\|\tilde{x}_f - x_k\| \leq \varepsilon$ .

We now bound the covering number  $N(\mathcal{F}, \varepsilon)$  for function classes  $\mathcal{F}_{\text{bl}}$  and  $\mathcal{F}_{\text{blrm}}$ . The former class is closed under convex combinations, which somewhat simplifies the situation.

**Proposition 23** Suppose that  $\mathcal{F} = \mathcal{F}_{\text{bl}}$ . Then  $\log N(\mathcal{F}, \varepsilon) = O\left(d \log\left(\frac{\text{diam}(K)}{\varepsilon}\right)\right)$ .

**Proof** Let  $\mathcal{C}_K$  be a finite subset of  $K$  such that for all  $x \in K$  there exists a  $y \in \mathcal{C}_K$  with  $\|x - y\| \leq \varepsilon$ . Standard bounds on covering numbers ([Artstein-Avidan et al., 2015](#), §4) show that  $\mathcal{C}_K$  can be chosen so that

$$|\mathcal{C}_K| \leq \left(1 + \frac{2 \text{diam}(K)}{\varepsilon}\right)^d.$$

Given  $x \in \mathcal{C}_K$  define  $\mathcal{F}_x = \{f \in \mathcal{F} : \|\tilde{x}_f - x\| \leq \varepsilon\}$ . Since  $\text{conv}(\mathcal{F}) = \mathcal{F}$  it follows trivially that  $\text{conv}(\mathcal{F}_x) \subset \text{conv}(\mathcal{F}) = \mathcal{F}$ . The common near minimiser property is satisfied automatically by definition. Suppose that  $f \in \mathcal{F}$  is arbitrary and let  $x \in \mathcal{C}_K$  be such that  $\|x - \tilde{x}_f\| \leq \varepsilon$ , which exists by construction. Therefore  $f \in \mathcal{F}_x$  and the approximation property also holds. ■

**Proposition 24** Suppose that  $\mathcal{F} = \mathcal{F}_{\text{blrm}}$ . Then  $\log N(\mathcal{F}, \varepsilon) = O\left(d \log\left(\frac{\text{diam}(K)}{\varepsilon}\right)\right)$ .

**Proof** To begin, define  $\varepsilon_{\mathbb{S}} = \varepsilon / \text{diam}(K)$ . Given a ridge function  $f \in \mathcal{F}$ , let  $\theta_f \in \mathbb{S}_1$  be a direction such that  $f(\cdot) = u(\langle \theta, \cdot \rangle)$  for some convex function  $u$ . Given  $x \in K$  and  $\theta \in \mathbb{S}_1$  let

$$\mathcal{F}_{x,\theta} = \{f \in \mathcal{F} : \|\tilde{x}_f - x\| \leq \varepsilon \text{ and } \theta_f = \theta\}.$$

Note that  $\{f \in \mathcal{F} : \theta_f = \theta\}$  is convex and hence  $\text{conv}(\mathcal{F}_{x,\theta}) \subset \mathcal{F}$  holds. Let  $\mathcal{C}_{\mathbb{S}}$  be a finite subset of  $\mathbb{S}_1$  such that for all  $\theta \in \mathbb{S}_1$  there exists an  $\eta \in \mathcal{C}_{\mathbb{S}}$  for which  $\|\theta - \eta\| \leq \varepsilon_{\mathbb{S}}$ . Similarly, let  $\mathcal{C}_K$  be a finite subset of  $K$  such that for all  $x \in K$  there exists a  $y \in \mathcal{C}_K$  with  $\|x - y\| \leq \varepsilon$ . Classical covering number results (Artstein-Avidan et al., 2015, §4) show that  $\mathcal{C}_{\mathbb{S}}$  and  $\mathcal{C}_K$  can be chosen so that

$$|\mathcal{C}_{\mathbb{S}}| \leq \left(1 + \frac{4}{\varepsilon_{\mathbb{S}}}\right)^d \quad |\mathcal{C}_K| \leq \left(1 + \frac{2 \text{diam}(K)}{\varepsilon}\right)^d.$$

Consider the collection  $\{\mathcal{F}_{x,\theta} : x \in \mathcal{C}_K, \theta \in \mathcal{C}_{\mathbb{S}}\}$ , which has size  $N = |\mathcal{C}_K| |\mathcal{C}_{\mathbb{S}}|$ . Let  $f \in \mathcal{F}$  be arbitrary and let  $\theta \in \mathcal{C}_{\mathbb{S}}$  and  $x \in \mathcal{C}_K$  be such that  $\|\theta - \theta_f\| \leq \delta$  and  $\|x - \tilde{x}_f\| \leq \varepsilon$ . Then define  $g = u_f(\langle \cdot, \theta \rangle) \in \mathcal{F}$ , which satisfies

$$\|f - g\|_{\infty} = \sup_{x \in K} |u_f(\langle x, \theta \rangle) - u_f(\langle x, \theta_f \rangle)| \leq \sup_{x \in K} |\langle x, \theta - \theta_f \rangle| \leq \varepsilon_{\mathbb{S}} \text{diam}(K) \leq \varepsilon.$$

Therefore the approximation property holds.  $\blacksquare$

Next we need some elementary continuity properties of the regret and information gain functions.

**Lemma 25** Suppose  $f$  and  $g$  are random elements in  $\mathcal{F}$  with laws  $\xi$  and  $\nu$  and that  $X, Y \in K$  are independent of  $f$  and  $g$  and have laws  $\pi$  and  $\rho$ . Suppose that  $\|f - g\|_{\infty} \leq \varepsilon$  almost surely and  $\|X - Y\| \leq \varepsilon$  almost surely. Then

- (a)  $I(\pi, \nu)^{1/2} \leq I(\pi, \xi)^{1/2} + \varepsilon.$
- (b)  $I(\pi, \nu)^{1/2} \leq I(\rho, \nu)^{1/2} + \varepsilon.$

**Proof** For random variable  $X$  let  $\|X\|_{L_2} = \mathbb{E}[X^2]^{1/2}$ , which is a norm on the space of square integrable random variables on some probability space with suitable a.s. identification. Let  $\bar{f} = \mathbb{E}[f]$  and  $\bar{g} = \mathbb{E}[g]$ . By definition  $I(\pi, \xi)^{1/2} = \|f(X) - \bar{f}(X)\|_{L_2}$  and  $I(\pi, \nu)^{1/2} = \|g(X) - \bar{g}(X)\|_{L_2}$ . The first claim follows since  $\|\cdot\|_{L_2}$  is a norm. The second claim follows in the same manner and using the fact that  $f, g, \bar{f}, \bar{g}$  are Lipschitz.  $\blacksquare$

**Lemma 26** Suppose that  $\alpha, \beta \in \text{IR}(\mathcal{F})$ . Suppose that  $X$  and  $f$  are (possibly dependent) random elements with laws  $\pi \in \mathcal{P}(K)$  and  $\nu \in \mathcal{P}(\mathcal{F})$  and such that  $f(X) \leq f_{\star} + \varepsilon$  almost surely. Then

$$\Delta(\pi, \nu) \leq \alpha + \sqrt{\beta I(\pi, \nu)} + \varepsilon \left[1 + \sqrt{\beta}\right].$$

**Proof** Let  $g(x) = \max(f(x), f(X))$  and  $\xi$  be the law of  $g$ , which means that  $\pi \in \text{TS}$ . As usual, let  $\bar{f} = \mathbb{E}[f]$  and  $\bar{g} = \mathbb{E}[g]$ . By construction  $\|f - g\|_{\infty} \leq \varepsilon$  and therefore

$$\Delta(\pi, \nu) = \mathbb{E}[\bar{f}(X) - f_{\star}] \leq \mathbb{E}[\bar{g}(X) - g_{\star}] + \varepsilon = \Delta(\pi, \xi) + \varepsilon \leq \alpha + \sqrt{\beta I(\pi, \xi)} + \varepsilon.$$



The result now follows from Lemma 25. ■

The next lemma establishes basic properties of the regularised minimisers  $\tilde{x}_f$  and  $\bar{x}_f$ , which are defined in Definition 21. Remember that  $\varepsilon_R$  is the amount of regularisation. Larger values make  $\tilde{x}_f$  more stable but also a worse approximation of  $x_f$ . The approximation error in the definition of  $\bar{x}_f$  is  $\varepsilon_O$ , which can be chosen extremely small.

**Lemma 27** *Suppose that  $f, g \in \mathcal{F}$ . Then*

- (a)  $\sup\{\|\tilde{x}_f - y\| : y \in \bar{x}_f\} \leq \sqrt{2\varepsilon_O/\varepsilon_R}$  with  $0/0 \triangleq 0$ .
- (b)  $f(\tilde{x}_f) \leq f_\star + \frac{\varepsilon_R}{2} \text{diam}(K)^2$ .

**Proof** Note the special case that  $\varepsilon_R = \varepsilon_O = 0$ , then  $\bar{x}_f = \{\tilde{x}_f\}$  by definition and (a) is immediate. Otherwise let  $\tilde{f}(x) = f(x) + \frac{\varepsilon_R}{2} \|x\|^2$  and  $x = \tilde{x}_f$  and  $y \in \bar{x}_f$ . Then

$$\tilde{f}(\tilde{x}_f) + \varepsilon_O \geq \tilde{f}(y) \geq \tilde{f}(\tilde{x}_f) + D\tilde{f}(\tilde{x}_f)[y - \tilde{x}_f] + \frac{\varepsilon_R}{2} \|\tilde{x}_f - y\|^2 \geq \tilde{f}(\tilde{x}_f) + \frac{\varepsilon_R}{2} \|\tilde{x}_f - y\|^2.$$

Rearranging completes the proof of the first part. For (b), let  $y \in K$  be arbitrary

$$f(\tilde{x}_f) + \frac{\varepsilon_R}{2} \|\tilde{x}_f\|^2 \leq f(y) + \frac{\varepsilon_R}{2} \|y\|^2.$$

And the result follows since  $\|y\|^2 - \|\tilde{x}_f\|^2 \leq \text{diam}(K)^2$ . ■

We can now state a general theorem from which Theorem 2 follows.

**Theorem 28** *Suppose that  $\varepsilon \in (0, 1)$  and  $\frac{1}{2}\varepsilon_R \text{diam}(K)^2 \leq \varepsilon$  and  $2\varepsilon_O/\varepsilon_R \leq \varepsilon^2$  and let  $\mathcal{F}$  be a set of convex functions from  $K$  to  $[0, 1]$  and  $(\alpha, \beta) \in \text{IR}(\mathcal{F})$ . Then the Bayesian regret of ATS for any prior  $\xi$  is at most*

$$B\text{Reg}_n(\text{ATS}, \xi) \leq n\alpha + n\varepsilon[6 + 2\sqrt{\beta}] + \sqrt{\frac{\beta n}{2} \log(N(\mathcal{F}, 1/\varepsilon))}.$$

Theorem 2 follows by choosing  $\varepsilon = 1/n$  and  $\varepsilon_R = \varepsilon_O = 0$  and by Proposition 23 and Proposition 24 to bound the covering numbers for the relevant classes.

**Proof** Let  $N = N(\mathcal{F}, \varepsilon)$  and  $\mathcal{F}_1, \dots, \mathcal{F}_N$  be a collection of subset of  $\mathcal{F}$  satisfying the conditions of Definition 22. Hence, there exists a sequence  $x_1, \dots, x_N$  such that for all  $k \in [N]$  and  $f \in \mathcal{F}_k$ ,

$$\|\tilde{x}_f - x_k\| \leq \varepsilon.$$

Let  $\xi \in \mathcal{P}(\mathcal{F})$  be any prior. Suppose that  $f$  is sampled from  $\xi$  and  $X_\star$  be a random element in  $K$  such that  $X_\star \in \bar{x}_f$  and let  $X$  be an independent copy of  $X_\star$  and  $Y$  be a random variable such that  $\mathbb{E}[Y|f, X] = f(X)$  and  $Y \in [0, 1]$  almost surely. By Definition 22 there exists an  $[N]$ -valued random variable  $\kappa$  such that:

- (i) There exists a random function  $f_\kappa \in \mathcal{F}_\kappa$  with  $\|f - f_\kappa\| \leq \varepsilon$ ; and
- (ii)  $\|\tilde{x}_f - x_\kappa\| \leq \varepsilon$ .
- (iii) There exists a random function  $f_\kappa \in \mathcal{F}_\kappa$  with  $\|f - f_\kappa\|_\infty \leq \varepsilon$ ; and

$$(iv) \quad \|\tilde{x}_f - x_\kappa\| \leq \varepsilon.$$

Define  $\pi$  as the law of  $X$ , which is a (approximate) Thompson sampling policy for  $\xi$ .

**Lemma 29** *The following holds:*

$$\Delta(\pi, \xi) \leq \alpha + \sqrt{\beta I(\kappa; X, Y)} + \varepsilon \text{ poly}(\text{diam}(K), \beta), \Delta(\pi, \xi) \leq \alpha + \sqrt{\beta I(\kappa; X, Y)} + \varepsilon[6 + 2\sqrt{\beta}],$$

where  $I(\kappa; X, Y)$  is the mutual information between  $\kappa$  and the pair  $(X, Y)$ . where  $I(\kappa; X, Y)$  is the mutual information between  $\kappa$  and the pair  $(X, Y)$ .

**Proof** Let  $\nu$  be the law of  $\mathbb{E}[f|\kappa]$  and  $\nu_\kappa$  be the law of  $\mathbb{E}[f_\kappa|\kappa]$  and  $\pi_\kappa$  be the law of  $x_\kappa$ . Let  $\bar{f}_\kappa = \mathbb{E}[f_\kappa]$ . Then

$$\begin{aligned} \Delta(\pi, \xi) &= \mathbb{E}[\bar{f}(X) - f_\star] \\ &\stackrel{(a)}{=} \mathbb{E}[\bar{f}(X) - \mathbb{E}[f_\star|\kappa]] \\ &\stackrel{(b)}{\leq} \mathbb{E}[\bar{f}_\kappa(X) - \mathbb{E}[f_\star|\kappa]] + \varepsilon \\ &\stackrel{(c)}{\leq} \mathbb{E}[\bar{f}_\kappa(x_\kappa) - \mathbb{E}[f_\kappa|\kappa]_\star] + 6\varepsilon \\ &\stackrel{(d)}{=} \Delta(\pi_\kappa, \nu_\kappa) + 6\varepsilon \\ &\stackrel{(e)}{\leq} \alpha + \sqrt{\beta I(\pi_\kappa, \nu_\kappa)} + 6\varepsilon \\ &\stackrel{(f)}{\leq} \alpha + \sqrt{\beta I(\pi_\kappa, \nu)} + \varepsilon[6 + \sqrt{\beta}] \\ &\stackrel{(g)}{\leq} \alpha + \sqrt{\beta I(\pi, \nu)} + \varepsilon[6 + 2\sqrt{\beta}] \\ &\stackrel{(h)}{\leq} \alpha + \sqrt{\frac{\beta I(\kappa; X, Y)}{2}} + \varepsilon[6 + 2\sqrt{\beta}], \end{aligned}$$

where

(a) by the tower rule.

(b) follows since  $\|f_\kappa - f\|_\infty \leq \varepsilon$  by definition and by convexity of  $\|\cdot\|_\infty$ ,

$$\|\bar{f} - \bar{f}_\kappa\|_\infty = \|\mathbb{E}[f] - \mathbb{E}[f_\kappa]\|_\infty \leq \mathbb{E}[\|f - f_\kappa\|_\infty] \leq \varepsilon.$$

(c) follows because  $\|f - f_\kappa\|_\infty \leq \varepsilon$  and  $\|\tilde{x}_f - x_\kappa\| \leq \varepsilon$  so that

$$\begin{aligned} \mathbb{E}[f_\star|\kappa] &= \mathbb{E}[f(x_f)|\kappa] \\ &\geq \mathbb{E}[f(\tilde{x}_f)|\kappa] - \frac{\varepsilon R}{2} \text{diam}(K)^2 \\ &\geq \mathbb{E}[f(x_\kappa)|\kappa] - 2\varepsilon \\ &\geq \mathbb{E}[f_\kappa(x_\kappa)|\kappa] - 3\varepsilon \\ &\geq \mathbb{E}[f_\kappa|\kappa]_\star - 3\varepsilon \cdot \mathbb{E}[f_\star|\kappa] &= \mathbb{E}[f(x_f)|\kappa] \\ &\geq \mathbb{E}[f(\tilde{x}_f)|\kappa] - \frac{\varepsilon R}{2} \text{diam}(K)^2 &\text{by Lemma 27 (b)} \\ &\geq \mathbb{E}[f(x_\kappa)|\kappa] - 2\varepsilon &\text{by the assumption on } \varepsilon_R \text{ and (iv)} \\ &\geq \mathbb{E}[f_\kappa(x_\kappa)|\kappa] - 3\varepsilon &\text{by (iii)} \\ &\geq \mathbb{E}[f_\kappa|\kappa]_\star - 3\varepsilon. \end{aligned}$$

And because  $\|X_\star - x_\kappa\| \leq \|X_\star - \tilde{x}_f\| + \|\tilde{x}_f - x_\kappa\| \leq 2\varepsilon$  so that

$$\mathbb{E}[\bar{f}_\kappa(X)] = \mathbb{E}[\bar{f}_\kappa(X_\star)] \leq \mathbb{E}[\bar{f}_\kappa(x_\kappa)] + 2\varepsilon.$$

And because by the triangle inequality, the definition of  $X_\star$  and Lemma 27 (a),

$$\|X_\star - x_\kappa\| \leq \|X_\star - \tilde{x}_f\| + \|\tilde{x}_f - x_\kappa\| \leq \sqrt{2\varepsilon_O/\varepsilon_R} + \varepsilon \leq 2\varepsilon, \quad (13)$$

which implies that  $\mathbb{E}[\bar{f}_\kappa(X)] = \mathbb{E}[\bar{f}_\kappa(X_\star)] \leq \mathbb{E}[\bar{f}_\kappa(x_\kappa)] + 2\varepsilon$ .

(d) follows by definition.

(e) follows because

$$\mathbb{E}[f_\kappa|\kappa](x_\kappa) = \mathbb{E}[f_\kappa(x_\kappa)|\kappa] \leq \mathbb{E}[f_{\kappa\star}|\kappa] + \varepsilon \leq \mathbb{E}[f_\kappa|\kappa]_\star + \varepsilon$$

Then use Lemma 26.  $\mathbb{E}[f_\kappa|\kappa](x_\kappa) = \mathbb{E}[f_\kappa(x_\kappa)|\kappa] \leq \mathbb{E}[f_{\kappa\star}|\kappa] + \varepsilon \leq \mathbb{E}[f_\kappa|\kappa]_\star + \varepsilon$  and Lemma 26.

(f) follows from Lemma 25 and because

$$\|\mathbb{E}[f_\kappa|\kappa] - \mathbb{E}[f|\kappa]\|_\infty \leq \mathbb{E}[\|f_\kappa - f\|_\infty] \leq \varepsilon.$$

(g) follows from Lemma 25 and because

$$\|X_\star - x_\kappa\| \leq \|X_\star - \tilde{x}_f\| + \|\tilde{x}_f - x_\kappa\| \leq \sqrt{2\varepsilon_O/\varepsilon_R} + \varepsilon \leq 2\varepsilon.$$

$\|\mathbb{E}[f_\kappa|\kappa] - \mathbb{E}[f|\kappa]\|_\infty \leq \mathbb{E}[\|f_\kappa - f\|_\infty] \leq \varepsilon$ .

(h) follows from Lemma 25 and Eq. (13).

(i) follows from Pinsker's inequality. Let  $\text{KL}(\cdot, \cdot)$  be the relative entropy. Then

$$\begin{aligned} \mathcal{I}(\pi, \nu) &= \mathbb{E}[(\mathbb{E}[f(X)|X] - \mathbb{E}[f(X)|\kappa, X])^2] \\ &= \mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y|\kappa, X])^2] \\ &\leq \frac{1}{2} \mathbb{E}[\text{KL}(\mathbb{P}_{Y|X}, \mathbb{P}_{Y|\kappa, X})] \\ &= \frac{1}{2} I(\kappa|X, Y) \end{aligned}$$

$$\begin{aligned} \mathcal{I}(\pi, \nu) &= \mathbb{E}[(\mathbb{E}[f(X)|X] - \mathbb{E}[f(X)|\kappa, X])^2] \\ &= \mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y|\kappa, X])^2] \\ &\leq \frac{1}{2} \mathbb{E}[\text{KL}(\mathbb{P}_{Y|X}, \mathbb{P}_{Y|\kappa, X})] \\ &= \frac{1}{2} I(\kappa|X, Y). \end{aligned}$$

This concludes the explanation of the steps and so the proof of the lemma. ■

We are now in a position to prove Theorem 28. Let  $\pi_t$  be the law of  $X_t$  under  $\mathbb{P}_{t-1}$  and  $\xi_t$  be the law of  $f$  under  $\mathbb{P}_{t-1}$ . By Lemma 29,

$$\Delta(\pi_t, \xi_t) \leq \alpha + \sqrt{\beta I_t(\kappa; X_t, Y_t)} + \varepsilon[6 + 2\sqrt{\beta}].$$

Hence, letting  $I_t$  be the mutual information with respect to probability measure  $\mathbb{P}_t$ ,

$$\begin{aligned}
 \text{BReg}_n(\text{TS}, \xi) &= \mathbb{E} \left[ \sum_{t=1}^n \Delta(\pi_t, \xi_t) \right] \\
 &\leq n\alpha + \mathbb{E} \left[ \sum_{t=1}^n \sqrt{\beta I_{t-1}(\kappa; X_t, f(X_t))} \right] + n\varepsilon[6 + 2\sqrt{\beta}] \\
 &\leq n\alpha + \sqrt{\beta n \mathbb{E} \left[ \sum_{t=1}^n I_{t-1}(\kappa; X_t, f(X_t)) \right]} + n\varepsilon[6 + 2\sqrt{\beta}] \\
 \text{BReg}_n(\text{ATS}, \xi) &= \mathbb{E} \left[ \sum_{t=1}^n \Delta(\pi_t, \xi_t) \right] \\
 &\leq n\alpha + \mathbb{E} \left[ \sum_{t=1}^n \sqrt{\beta I_{t-1}(\kappa; X_t, Y_t)} \right] + n\varepsilon[6 + 2\sqrt{\beta}] \\
 &\leq n\alpha + \sqrt{\beta n \mathbb{E} \left[ \sum_{t=1}^n I_{t-1}(\kappa; X_t, Y_t) \right]} + n\varepsilon[6 + 2\sqrt{\beta}] \\
 &\leq n\alpha + \sqrt{\beta n \log(N)} + n\varepsilon[6 + 2\sqrt{\beta}],
 \end{aligned}$$

where the final inequality holds by the chain rule for the mutual information and because  $\kappa \in [N]$  and hence its entropy is at most  $\log(N)$ .  $\blacksquare$