Thompson Sampling for Bandit Convex Optimization

by

Alireza Bakhtiari

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

University of Alberta

Abstract

Thompson sampling (TS) is a popular and empirically successful algorithm for online decision-making problems. This thesis advances our understanding of TS when applied to bandit convex optimization (BCO) problems, by providing new theoretical guarantees and characterizing its limitations.

First, we analyze 1-dimensional BCO and show that TS achieves a near-optimal Bayesian regret of at most $\tilde{O}(\sqrt{n})$, where n is the time horizon. This result holds without strong assumptions on the loss functions, requiring only convexity, boundedness, and a mild Lipschitz condition. In sharp contrast, we demonstrate that for general high-dimensional problems, TS can fail catastrophically.

More positively, we establish a Bayesian regret bound of $\tilde{O}(d^{2.5}\sqrt{n})$ for TS in generalized linear bandits, even when the convex monotone link function is unknown. Finally, we prove a fundamental limitation of current analysis techniques: we show that the standard information-theoretic machinery can never yield a regret bound better than the existing $\tilde{O}(d^{1.5}\sqrt{n})$ in the general case.

Preface

TODO.

Acknowledgements

TODO.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

Convexity is a common assumption in optimization problems [BV04]. Bandit convex optimization (BCO) addresses the fundamental problem of minimizing a convex function over a convex set when only noisy evaluations of the function are available at selected points. This setting naturally arises in scenarios where:

(i) Limited Access: The algorithm can only observe noisy evaluations of the objective function.

(ii) Cumulative Cost: The goal is to minimize the cumulative cost of these evaluations over time, rather than to simply identify the function's minimizer.

In classical optimization, it is typically assumed that the optimizer has access to the full function, or at least to its value and gradient at any point in the domain. However, this assumption often fails in practice. A representative example is dynamic pricing, where a seller selects a price (the input) and observes the resulting profit (the output), which is often modeled as a concave function of the price. The seller cannot directly observe the profit function itself, but only noisy feedback through customer purchases at chosen prices. Each evaluation corresponds to a real transaction and thus incurs a potentially significant cost. In such cases, the objective is not to eventually find the best price at any expense, but rather to make pricing decisions that yield high cumulative profit over time.

This thesis focuses on this *zeroth-order* or *bandit feedback* setting, which is prevalent in domains where gradients are unavailable or costly to compute, and where each function evaluation carries a tangible cost. While convexity is an idealization, it is a broadly applicable assumption that facilitates principled algorithm design and analysis. Unlike traditional optimization set-

tings—where function evaluations are often treated as free abstractions—many real-world problems demand an approach that explicitly accounts for the cumulative cost incurred during learning.

Formally, the goal is to approximately solve the optimization problem

$$\arg\min_{x \in \mathcal{K}} f(x)\,, \tag{1.1}$$

where $\mathcal{K} \subset \mathbb{R}^d$ is a convex set (typically compact with non-empty interior) and $f : \mathcal{K} \to \mathbb{R}$ is a convex function. In the BCO setting, the learner does not observe gradients or even function values at arbitrary points; instead, in each round of interaction it selects a point $x_t \in \mathcal{K}$ and receives noisy feedback $y_t = f(x_t) + \varepsilon_t$, where ε_t models the noise. While the convexity assumption may seem restrictive, it captures a rich class of problems and provides a tractable framework to develop principled algorithms with provable guarantees.

The learner aims to minimize the cumulative loss over T rounds, which leads to the online variant of the optimization problem, where the learner's goal is to minimize cumulative loss compared to the best action in hindsight, which is

$$\text{Regret}(T) = \sum_{t=1}^{T} f(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f(x)\,, \tag{1.2}$$

often referred to as *bandit regret*. This perspective connects BCO to the broader literature on online learning and multi-armed bandits. While BCO is challenging due to the simultaneous absence of gradient information and the need for exploration, it offers a powerful abstraction for decision-making under uncertainty with limited feedback. The convexity assumption, though idealized, provides a tractable and theoretically rich framework for developing and analyzing algorithms that strike a balance between exploration and exploitation.

1.1 Examples of BCO

This setting naturally arises in a variety of applications where gradient information is unavailable, unreliable, or too expensive to compute, such as

Manufacturing: Consider a cheese factory that aims to optimize its recipe by adjusting the temperature and humidity of its warehouse. The quality of the final product can be modeled as a convex function of these parameters. However, each measurement becomes available only after the batch

is complete, and producing a batch incurs a financial cost. The factory's goal is to iteratively improve product quality based on customer feedback, but it cannot afford to ruin too many batches in the process. This situation is even more pronounced for expensive or custom products—like cars, airplanes, or musical instruments—where each failed experiment is prohibitively costly and feedback may only be available post-sale.

Dynamic Pricing: In dynamic pricing, a retailer interacts sequentially with an uncertain market. At each round, they select a price $X_t \in \mathcal{K} \subset \mathbb{R}$, and the associated loss $f(X_t)$ represents the negative of expected profit. Prices that are too high may deter purchases, while prices that are too low leave revenue on the table. The profit function f is unknown in advance and customer behavior introduces noise into observations. The goal is to adjust prices over time to maximize cumulative profit, not just to identify an optimal price after at any cost.

Service Personalization: Large Language Models (LLMs) often personalize their responses based on user preferences. At each interaction, the system selects a response style $X_t \in \mathcal{K} \subset \mathbb{R}^d$ — e.g., controlling tone, formality, or humor. The user's satisfaction is captured by a loss function $f(X_t)$, which reflects poor alignment with their preferences. This function is unknown, subjective, and observed only through noisy feedback such as click-through rates or engagement metrics. The system must learn and adapt over time to minimize dissatisfaction, making this a natural fit for the bandit convex optimization setting.

Resource Allocation: Many decision-making problems involve allocating limited resources—like budget or bandwidth—across competing options. For example, a company might distribute marketing funds across various channels. The return on investment typically exhibits diminishing returns and can be modeled by a convex utility function. The utility is not directly observable but can be estimated after committing to a specific allocation. Since each evaluation carries cost, the company seeks to minimize cumulative regret over time by carefully balancing exploration and exploitation.

Online Advertising—$\mathbb{R}$-valued parameters: Online advertisement is a classical application of bandit algorithms. Often, advertisers can choose among nearly-continuous design parameters—like font size and color. These decisions affect user engagement in a way that can be modeled by a convex function over the parameter space. The feedback (e.g., click-through rates) is noisy and delayed, and testing each variation has opportunity cost. Bandit convex optimization provides a principled framework to navigate this space efficiently and improve ad performance over time.

Efficiency Tuning: In commercial aviation, dispatchers must decide the cruise altitude and Mach number for each flight, represented as $X_t = (\text{Mach}, \text{Altitude}) \in \mathcal{K} \subset \mathbb{R}^2$. The loss $f(X_t)$ is

the fuel burned per seat-kilometre—a quantity well-approximated by a convex function due to the trade-offs between speed, altitude, and drag. Crucially, the true fuel burn is revealed only after the flight, and is confounded by weather, routing, and payload variability. Since each evaluation corresponds to a real flight with significant operational cost, the airline cannot afford extensive trial-and-error. Instead, it must adapt its cruise settings sequentially, improving fuel efficiency over time while minimizing total cost—an ideal use case for bandit convex optimization.

1.2 Thompson Sampling and Bayesian Bandits

Thompson sampling (TS) is a simple and often practical algorithm for interactive decision-making with a long history [Tho33, RVK+18]. Our interest is in its application to Bayesian bandit convex optimization [Lat24].

A Bayesian bandit problem is a sequential decision-making problem where the learner has access to a prior distribution over the unknown objective function. The learner interacts with the environment by selecting actions and observing noisy feedback, which is modeled as a realization of a random variable whose distribution depends on the unknown objective function. This prior distribution captures the learner or the domain expert's beliefs about the objective function before any interaction.

At its core, Thompson Sampling is a Bayesian algorithm that maintains a posterior distribution over the unknown objective (or cost) function. In each round, it samples a function from this posterior and selects the action that minimizes the sampled function. The elegance of TS lies in its simplicity and flexibility—it requires no explicit exploration bonus or confidence bounds and can be implemented in a wide range of settings, provided posterior sampling is computationally feasible.

Chapter 2

Related work

BCO in the regret setting was first studied by [FKM05] and [Kle05]. Since then the field has grown considerably as summarized in the recent monograph by [Lat24]. Our focus is on the Bayesian version of the problem, which has seen only limited attention. [BDKP15] 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. [BE18] and [Lat20] generalized the information-theoretic machinery used by [BDKP15] 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 [RV14]. Nevertheless, we borrow certain techniques from these papers. Convex ridge functions have been studied before by [Lat21], 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 [DHK08]. Regrettably, however, this algorithm is not practically implementable, even under the assumption that you can sample efficiently from the posterior. [SNNJ21] also study a variation on the problem where the losses have the form $f(g(x))$ with $g : \mathbb{R}^d \to \mathbb{R}$ a *known* function and $f : \mathbb{R} \to \mathbb{R}$ 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 [FvdHLM24]. Convex ridge functions can also be viewed as a special case of the generalized linear model, which has been studied extensively as a reward model for stochastic bandits [FCGS10, and many more]. TS and other randomized algorithms have been studied with

generalized linear models in Bayesian and frequentist settings [AL17, DMR19, KZS+20]. None of these papers assume convexity (concavity for rewards) and consequentially suffer a regret that depends on other properties of the link function that can be arbitrarily large. Moreover, in generalized linear bandits it is standard to assume the link function is known.

Chapter 3

Bayesian BCO Problem and the TS Algorithm

Let K be a convex body in $\mathbb{R}^d$ and $\mathscr{F}$ be a set of convex functions from K to $[0,1]$. We assume there is a known (prior) probability measure ξ on $\mathscr{F}$. The interaction between the learner and environment lasts for n rounds. At the beginning the environment secretly samples f from the prior ξ. 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 $\mathbb{E}[Y_t | X_1, Y_1, \ldots, X_t, f] = f(X_t)$. 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 sub-gaussian noise with minor modifications. A learner $\mathscr{A}$ is a (possibly random) mapping from sequences of action/loss pairs to actions and its Bayesian regret with respect to prior ξ is

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

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

$$\sup_{\xi \in \mathscr{P}(\mathscr{F})} \mathrm{BReg}_n(\mathrm{TS}, \xi) \,, \tag{3.1}$$

where $\mathscr{P}(\mathscr{F})$ is the space of probability measures on $\mathscr{F}$ (with a suitable σ-algebra) and TS is Thompson sampling (Algorithm 1) with prior ξ (the dependence on the prior is always omitted from the notation). The quantity in Eq. (3.1) depends on the function class $\mathscr{F}$. 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.

```
args: prior ξ
for t = 1 to ∞:
    sample f_t from ℙ(f = ·|X_1, Y_1, ..., X_{t-1}, Y_{t-1})
    play X_t = x_{f_t} and observe Y_t
```

Algorithm 1: Thompson sampling

3.1 Thompson Sampling for Bandit Convex Optimization

With these definitions in place, we can now summarize our results:

- When $d = 1$, $\mathrm{BReg}_n(\mathrm{TS}, \xi) = \tilde{O}(\sqrt{n})$ for all priors (Theorem 15).
- A convex function f is called a monotone ridge function if there exists a convex monotone function $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{R}^d$ such that $f(x) = \ell(\langle x, \theta \rangle)$. Theorem 17 shows when ξ is supported on monotone ridge functions, then $\mathrm{BReg}_n(\mathrm{TS}, \xi) = \tilde{O}(d^{2.5}\sqrt{n})$.
- In general, the Bayesian regret of TS can be exponential in the dimension (Theorem 22).
- The classical information-theoretic machinery used by [BE18] an [Lat20] cannot improve the regret for BCO 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 generalized linear bandit where the link function is unknown and assumed to be convex and monotone increasing.

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 $K = \{x \geq \mathbf{0} : \|x\|_1 \leq 1\}$.

Our lower bounds show that TS does not behave well in the general BCO unless possibly the dimension is quite small. Perhaps more importantly, we show that the classical information-theoretic machinery used by [BE18] and [Lat20] cannot be used to improve the current best dimension dependence of the regret for BCO. 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 [ZL19, LG23]. We note that this does not imply a lower bound for BCO. The construction in the lower bound is likely amenable to methods for learning a direction based on the power method [LH21, HHK+21]. 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.

3.2 Notation

Let $\|\cdot\|$ be the standard euclidean norm on $\mathbb{R}^d$. For natural number k let $[k] = \{1, \ldots, k\}$. Define $\|x\|_\Sigma = \sqrt{x^\top \Sigma x}$ for positive definite Σ. Given a function $f : K \to \mathbb{R}$, 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 η. 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 σ-algebra is $\mathscr{P}(K)$. Similarly, $\mathscr{P}(\mathscr{F})$ is a space of probability measures on $\mathscr{F}$ with some unspecified σ-algebra ensuring that $f \mapsto f(x)$ is measurable for all $x \in K$. Given a convex function $f : K \to \mathbb{R}$ 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; [Nie92] showed that such a mapping always exists. Of course it follows that $f \mapsto f_\star = f(x_f)$ is also measurable. $\mathbb{P}_t = \mathbb{P}(\cdot | X_1, Y_1, \ldots, X_t, Y_t)$ and $\mathbb{E}_t$ be the expectation operator with respect to $\mathbb{P}_t$. The following assumption on $\mathcal{K}$ is considered global throughout:

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

3.2.1 Spaces of convex functions

A function $f : K \to \mathbb{R}$ is called a convex ridge function if there exists a convex $\ell : \mathbb{R} \to \mathbb{R}$ 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 ℓ is monotone increasing. We are interested in the following classes of convex functions: (a) $\mathscr{F}_{\mathtt{b}}$ is the space of all bounded convex functions $f : K \to [0, 1]$. (b) $\mathscr{F}_{\mathtt{l}}$ is the space of convex functions $f : K \to \mathbb{R}$ with $\mathrm{Lip}(f) \leq 1$. (c) $\mathscr{F}_{\mathtt{r}}$ is the space of all convex ridge functions. (d) $\mathscr{F}_{\mathtt{rm}}$ is the space of all monotone convex ridge functions. Intersections are represented as you might expect: $\mathscr{F}_{\mathtt{bl}} = \mathscr{F}_{\mathtt{b}} \cap \mathscr{F}_{\mathtt{l}}$ and similarly for other combinations. The set $\mathscr{F}$ refers to a class of convex functions, which will always be either $\mathscr{F}_{\mathtt{bl}}$ or $\mathscr{F}_{\mathtt{blrm}}$.

The representation of f as a ridge convex function is not unique, meaning that there could be (are) two sets (θ_1, ℓ_1) and (θ_2, ℓ_2) such that $f = \ell_1(\langle x, \theta_1 \rangle)$ and $f = \ell_2(\langle x, \theta_2 \rangle)$. The following lemma ensures that the *link function* ℓ can be chosen in a way that the Lipschitzness of the original function f is preserved.

Lemma 2. *Suppose that K is a convex body and $f \in \mathscr{F}_{\mathtt{lr}}$ is a Lipschitz convex ridge function. Then there exists a $\theta \in \mathbb{S}_1$ and a convex $\ell : \mathbb{R} \to \mathbb{R}$ such that $f(x) = \ell(\langle x, \theta \rangle)$ and $\mathrm{Lip}(\ell) \leq \mathrm{Lip}_K(f)$.*

Proof. By assumption there exists a convex function $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{S}_1$ such that $f(x) = \ell(\langle x, \theta \rangle)$. It remains to show that ℓ can be chosen so that $\mathrm{Lip}(\ell) \leq \mathrm{Lip}_K(f)$. Let h_K be the support function associated with K, given by $h_K(v) = \sup_{x \in K} \langle v, x \rangle$. Therefore ℓ 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

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

Then define ℓ on all of $\mathbb{R}$ via the classical extension [Lat24, Proposition 3.18, for example]. $\square$

Chapter 4

Generalized Information Ratio

The main theoretical tool is a version of the information ratio, which was introduced by [RV16] as a means to bound the Bayesian regret of TS for finite-armed and linear bandits. Given a distribution $\xi \in \mathscr{P}(\mathscr{F})$ and policy $\pi \in \mathscr{P}(K)$, let (X, f) have law $\pi \otimes \xi$ and with $\bar{f} = \mathbb{E}[f]$ define

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

which are both non-negative. The quantity $\Delta(\pi, \xi)$ is the regret suffered by π when the loss function is sampled from ξ, 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 $\mathcal{I}(\pi, \xi)$ [RV16]. Given a distribution ξ and a random function f with law ξ, we let $\pi^\xi_{\text{TS}} \in \mathscr{P}(K)$ 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 $\mathscr{F}$ is

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

Note that $(0, \beta) \in \text{IR}(\mathscr{F})$ is equivalent to $\Delta(\pi^\xi_{\text{TS}}, \xi)^2/\mathcal{I}(\pi^\xi_{\text{TS}}, \xi) \leq \beta$ for all $\xi \in \mathscr{P}(\mathscr{F})$. The α 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 3. *Suppose that* $\mathscr{F} \in \{\mathscr{F}_{\texttt{bl}}, \mathscr{F}_{\texttt{blrm}}\}$ *and* $(\alpha, \beta) \in \text{IR}(\mathscr{F})$. *Then, for any prior* $\xi \in \mathscr{P}(\mathscr{F})$,

the regret of TS *(Algorithm 1) is at most*

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

where the Big-O hides only a universal constant.

This theorem is a direct consequence of Theorem 13, so we defer the proof to Section 5.3. At a high level the argument is based on similar results by [BDKP15] and [BE18]. Also note that the space of ridge functions is not closed under convex combinations, which introduces certain challenges also noticed by [Lat21]. To address this last issue, we introduce a cover in Section 5.1 that let us work with subsets of $\mathscr{F}$ that are closed under convex combinations, and also satisfy some other properties.

4.1 Decomposition lemma

We also introduce a new mechanism for deriving information ratio bounds specially for TS. The rough idea is to partition the function class $\mathscr{F}$ into disjoint subsets $\mathscr{F}_i$ such that an inequality similar to that of general information ratio holds for each partition.

Lemma 4. *Suppose there exist natural numbers k and m such that for all $\bar{f} \in \operatorname{conv}(\mathscr{F})$ there exists a disjoint union $\mathscr{F} = \cup_{i=1}^{m} \mathscr{F}_i$ of measurable sets for which*

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

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

Let us pause for a moment to provide some intuition. The supremum term is the worst possible regret within $\mathscr{F}_i$ 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 [RV14] and in the information ratio decomposition by [Lat20].

Proof of Lemma 4. Let $\xi \in \mathscr{P}(\mathscr{F})$ and $\bar{f} = \mathbb{E}[f]$ and $\mathscr{F}_1, \dots, \mathscr{F}_m$ be disjoint subsets of $\mathscr{F}$ such that $\mathscr{F} = \cup_{i=1}^m \mathscr{F}_i$ and

$$\max_{i \in [m]} \left[\sup_{f \in \mathscr{F}_i} \left(\bar{f}(x_f) - f_\star \right) - \alpha - \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathscr{F}_i} \sum_{j,l \in \text{PAIR}(k)} (\bar{f}_j(x_{f_l}) - f(x_{f_l}))^2} \right] \leq 0 \,, \tag{4.1}$$

which exists by the assumptions in the lemma. When $\xi(\mathscr{F}_i) = 0$ define ν_i as an arbitrary probability measure on $\mathscr{F}$ and otherwise let $\nu_i(\cdot) = \xi(\cdot \cap \mathscr{F}_i)/\xi(\mathscr{F}_i)$ and $w_i = \xi(\mathscr{F}_i)$. Therefore

$$\begin{aligned}
\Delta(\pi, \xi) &= \int_{\mathscr{F}} \left(\bar{f}(x_f) - f_\star \right) \mathrm{d}\xi(f) \\
&= \sum_{i=1}^m w_i \int_{\mathscr{F}} \left(\bar{f}(x_f) - f_\star \right) \mathrm{d}\nu_i(f) \\
&\overset{\text{(a)}}{\leq} \sum_{i=1}^m w_i \sup_{f \in \mathscr{F}_i} (\bar{f}(x_f) - f_\star) \\
&\overset{\text{(b)}}{\leq} \alpha + \sum_{i=1}^m w_i \sqrt{\beta \inf_{f_1, \dots, f_k \subset \mathscr{F}_i} \sum_{j,l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2} \\
&\overset{\text{(c)}}{\leq} \alpha + \sum_{i=1}^m w_i \sqrt{\beta k(k-1) \int_{\mathscr{F}} \int_{\mathscr{F}} (\bar{f}(x_g) - f(x_g))^2 \, \mathrm{d}\nu_i(f) \, \mathrm{d}\nu_i(g)} \\
&\overset{\text{(d)}}{\leq} \alpha + \sqrt{\beta m k(k-1) \sum_{i=1}^m w_i^2 \int_{\mathscr{F}} \int_{\mathscr{F}} (\bar{f}(x_g) - f(x_g))^2 \, \mathrm{d}\nu_i(f) \, \mathrm{d}\nu_i(g)} \\
&\overset{\text{(e)}}{\leq} \alpha + \sqrt{\beta m k(k-1) \sum_{i=1}^m \sum_{j=1}^m w_i w_j \int_{\mathscr{F}} \int_{\mathscr{F}} (\bar{f}(x_g) - f(x_g))^2 \, \mathrm{d}\nu_i(f) \, \mathrm{d}\nu_j(g)} \\
&= \alpha + \sqrt{\beta m k(k-1) \int_{\mathscr{F}} \int_{\mathscr{F}} (\bar{f}(x_g) - f(x_g))^2 \, \mathrm{d}\xi(f) \, \mathrm{d}\xi(g)} \\
&= \alpha + \sqrt{\beta m k(k-1) \mathcal{I}(\pi, \xi)} \,,
\end{aligned} \tag{4.2}$$

where (a) is immediate from the definition of the integral. (b) follows from Eq. (4.1). (c) is

true because if $f_1, \ldots, f_k$ are sampled independently from ν_i, then

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

`(d)` follows from Cauchy-Schwarz and `(e)` by introducing additional non-negative terms. Since Eq. (4.2) holds for all $\xi \in \mathscr{P}(\mathscr{F})$ it follows that

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

and therefore $(\alpha, \beta m k(k-1)) \in \textsc{ir}(\xi)$. □

Chapter 5

Approximate Thompson Sampling

Often *exact* minimization a convex function is computationally expensive. Therefore, it is natural to consider approximate minimization. In fact, we analyze this approximate version of TS, which is a strict generalization of the exact version. Later, we specialize the analysis of this approximate version, which we call approximate Thompson sampling (ATS), to get regret bounds for the exact version of TS. ATS is defined in Algorithm 2 and is similar to TS except that it only approxi-

```
args: prior ξ
for t = 1 to ∞:
  sample f_t from ℙ(f = ·|X_1, Y_1, ..., X_{t-1}, Y_{t-1})
  play X_t ∈ x̄_{f_t}
  observe Y_t
```

Algorithm 2: Approximate Thompson sampling

mately minimizes the sampled loss function, i.e. X_t only needs to approximately minimize f_t. The analysis of this algorithm is surprisingly subtle, and indeed, we were only able to analyze an approximate version of TS that uses a small amount of regularization.

Definition 5. Let $\epsilon_O \leq \epsilon_R$ be non-negative constants called the optimization accuracy and regularization parameter, respectively. Given $f \in \mathscr{F}_1$ let $\tilde{f}(x) = f(x) + \frac{\epsilon_R}{2}\|x\|^2$ when $\epsilon_R > 0$ define

$$\tilde{x}_f = \operatorname*{arg\,min}_{x \in K} \tilde{f}(x) \qquad\qquad \bar{x}_f = \left\{ x : \tilde{f}(x) \leq \min_{y \in K} \tilde{f}(y) + \epsilon_O \right\}.$$

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

When $\epsilon_R = \epsilon_O = 0$, 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 regularization 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 regularization and define $\bar{x}_f$ by $\{x : f_t(x) \leq f_{t\star} + \epsilon\}$ for suitably small $\epsilon \geq 0$. We suspect the answer is yes but do not currently have a proof. The regularization ensures that $\bar{x}_f$ has small diameter, which need not be true in general for $\{x : f(x) \leq f_\star + \epsilon\}$, even if ϵ is arbitrarily small.

Remark 6. It's also important to note that x_f is not $\bar{x}_f$ necessarily.

5.1 A Convex Cover

We start by defining a kind of cover of a set of convex functions $\mathscr{F}$. In the standard analysis introduced by [BDKP15] and [BE18], this cover was defined purely in terms of the optimal action. As noticed by [Lat21], this argument relies on $\mathscr{F}$ being closed under convex combinations, which is not true for the space of ridge functions. Here we introduce a new notion of cover for function classes $\mathscr{F}$ that are not closed under convex combinations.

Definition 7. Let $\mathscr{F}$ be a set of convex functions from $\mathcal{K}$ to $\mathbb{R}$ and $\epsilon > 0$. Define $N(\mathscr{F}, \epsilon)$ to be the smallest number N such that there exists $\{\mathscr{F}_1, \ldots, \mathscr{F}_N\}$ such that for all $k \in [N]$:

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

Moreover:

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

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

Proposition 8. *Suppose that* $\mathscr{F} = \mathscr{F}_{\texttt{bl}}$. *Then* $\log N(\mathscr{F}, \epsilon) = O\left(d \log\left(\frac{\mathrm{diam}(K)}{\epsilon}\right)\right)$.

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 $\|x - y\| \leq \epsilon$. Standard bounds on covering numbers [AAGM15, §4] show that $\mathcal{C}_K$ can be chosen so that

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

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

Proposition 9. *Suppose that* $\mathscr{F} = \mathscr{F}_{\texttt{blrm}}$. *Then* $\log N(\mathscr{F}, \epsilon) = O\left(d \log\left(\frac{\operatorname{diam}(K)}{\epsilon}\right)\right)$.

Proof. To begin, define $\epsilon_{\mathbb{S}} = \epsilon / \operatorname{diam}(K)$. Given a ridge function $f \in \mathscr{F}$, 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

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

Note that $\{f \in \mathscr{F} : f_\theta = \theta\}$ is convex and hence $\operatorname{conv}(\mathscr{F}_{x,\theta}) \subset \mathscr{F}$ 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 $\|\theta - \eta\| \leq \epsilon_{\mathbb{S}}$. 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 $\|x - y\| \leq \epsilon$. Classical covering number results [AAGM15, §4] show that $\mathcal{C}_{\mathbb{S}}$ and $\mathcal{C}_K$ can be chosen so that

$$|\mathcal{C}_{\mathbb{S}}| \leq \left(1 + \frac{4}{\epsilon_{\mathbb{S}}}\right)^d \qquad\qquad |\mathcal{C}_K| \leq \left(1 + \frac{2\operatorname{diam}(K)}{\epsilon}\right)^d .$$

Consider the collection $\{\mathscr{F}_{x,\theta} : x \in \mathcal{C}_K, \theta \in \mathcal{C}_{\mathbb{S}}\}$, which has size $N = |\mathcal{C}_K||\mathcal{C}_{\mathbb{S}}|$. Let $f \in \mathscr{F}$ 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 $\|x - \tilde{x}_f\| \leq \epsilon$. Then define $g = u_f(\langle \cdot, \theta \rangle) \in \mathscr{F}$, which satisfies

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

Therefore the approximation property holds. □

5.2 Continuity of Regret and Information Gain

In order to find a pair (α, β) in $\mathrm{IR}(\mathscr{F})$, we need to bound the regret $\Delta(\pi, \xi)$ in terms of the information gain $I(\pi, \xi)$ for a prior $\xi \in \mathscr{P}(\mathscr{F})$ and policy $\pi \in \mathscr{P}(K)$. It turns out useful to prove continuity (Lipschitzness) properties of these quantities in terms of the distance between two different priors $\xi, \nu \in \mathscr{P}(\mathscr{F})$ and the distance between two different policies $\pi, \rho \in \mathscr{P}(K)$. Of course, the proper *distance* metric on $\mathscr{P}(\mathscr{F})$ and $\mathscr{P}(K)$ needs to be specified to make this precise.

Lemma 10. *Suppose f and g are random elements in $\mathscr{F}$ with laws ξ and ν and that $X, Y \in K$ are independent of f and g and have laws π and ρ. Suppose that $\|f - g\|_\infty \leq \epsilon$ almost surely and $\|X - Y\| \leq \epsilon$ almost surely. Then*

(a) $I(\pi, \nu)^{1/2} \leq I(\pi, \xi)^{1/2} + \epsilon$.

(b) $I(\pi, \nu)^{1/2} \leq I(\rho, \nu)^{1/2} + \epsilon$.

Proof. For random variable X let $\|X\|_{\mathrm{L2}} = \mathbb{E}[X^2]^{1/2}$, 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 $I(\pi, \xi)^{1/2} = \|f(X) - \bar{f}(X)\|_{\mathrm{L2}}$ and $I(\pi, \nu)^{1/2} = \|g(X) - \bar{g}(X)\|_{\mathrm{L2}}$. The first claim follows since $\|\cdot\|_{\mathrm{L2}}$ is a norm. The second claim follows in the same manner and using the fact that $f, g, \bar{f}, \bar{g}$ are Lipschitz. □

Lemma 11. *Suppose that $\alpha, \beta \in \mathrm{IR}(\mathscr{F})$. Suppose that X and f are (possibly dependent) random elements with laws $\pi \in \mathscr{P}(K)$ and $\nu \in \mathscr{P}(\mathscr{F})$ and such that $f(X) \leq f_\star + \epsilon$ almost surely. Then*

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

Proof. Let $g(x) = \max(f(x), f(X))$ and ξ be the law of g, which means that $\pi \in \mathrm{TS}(\xi)$. As usual, let $\bar{f} = \mathbb{E}[f]$ and $\bar{g} = \mathbb{E}[g]$. By construction

$$\begin{aligned}
\|f - g\|_\infty &= \sup_{x \in \mathcal{K}} |\max(f(x), f(X)) - f(x)| \\
&= \sup_{x \in \mathcal{K}} \max(f(x), f(X)) - f(x) \\
&\leq \sup_{x \in \mathcal{K}} \max(f(x), f(x) + \epsilon) - f(x) \\
&\leq \epsilon ,
\end{aligned}$$

almost surely. Therefore

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

The result now follows from Lemma 10. □

The next lemma establishes basic properties of the regularised minimisers $\tilde{x}_f$ and $\bar{x}_f$, which are defined in Definition 5. Remember that ϵ_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 ϵ_O, which can be chosen extremely small.

Lemma 12. *Suppose that $f, g \in \mathscr{F}$. Then*

(a) $\sup\{\|\tilde{x}_f - y\| : y \in \bar{x}_f\} \leq \sqrt{2\epsilon_O/\epsilon_R}$ *with* $0/0 \triangleq 0$.

(b) $f(\tilde{x}_f) \leq f_\star + \frac{\epsilon_R}{2}\operatorname{diam}(K)^2$.

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

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

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

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

And the result follows since $\|y\|^2 - \|x\|^2 \leq \operatorname{diam}(K)^2$. □

5.3 A Regret Bound in Terms of the Information Ratio

We can now state a general theorem from which Theorem 3 follows.

Theorem 13. *Suppose that $\epsilon \in (0,1)$ and $\frac{1}{2}\epsilon_R \operatorname{diam}(K)^2 \leq \epsilon$ and $2\epsilon_O/\epsilon_R \leq \epsilon^2$ and let $\mathscr{F}$ be a set of convex functions from K to $[0,1]$ and $(\alpha,\beta) \in \text{IR}(\mathscr{F})$. Then the Bayesian regret of* ATS *for any prior ξ is at most*

$$\mathit{BReg}_n(\text{ATS},\xi) \leq n\alpha + 3n\epsilon[1+\sqrt{\beta}] + \sqrt{\frac{\beta n}{2}\log\left(N\left(\mathscr{F},1/\epsilon\right)\right)}\,.$$

Theorem 3 follows by choosing $\epsilon = 1/n$ and $\epsilon_R = \epsilon_O = 0$ and by Proposition 8 and Proposition 9 to bound the covering numbers for the relevant classes.

Proof. Let $N = N(\mathscr{F}, \epsilon)$ and $\mathscr{F}_1, \ldots, \mathscr{F}_N$ be a collection of subset of $\mathscr{F}$ satisfying the conditions of Definition 7. Hence, there exists a sequence $x_1, \ldots, x_N$ such that for all $k \in [N]$ and $f \in \mathscr{F}_k$,

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

Let $\xi \in \mathscr{P}(\mathscr{F})$ be any prior. Suppose that f is sampled from ξ 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 7 there exists an $[N]$-valued random variable κ such that:

(i) There exists a random function $f_\kappa \in \mathscr{F}_\kappa$ with $\|f - f_\kappa\|_\infty \leq \epsilon$; and

(ii) $\|\tilde{x}_f - x_\kappa\| \leq \epsilon$.

Define π as the law of X, which is a (approximate) Thompson sampling policy for ξ.

Lemma 14. *The following holds:*

$$\Delta(\pi, \xi) \leq \alpha + \sqrt{\beta I(\kappa; X, Y)} + 3\epsilon[1 + \sqrt{\beta}] \,,$$

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

Proof. Let ν be the law of $\mathbb{E}[f|\kappa]$ and ν_κ be the law of $\mathbb{E}[f_\kappa|\kappa]$ and π_κ be the law of x_κ. Let

$\bar{f}_\kappa = \mathbb{E}[f_\kappa]$. Then

$$\begin{aligned}
\Delta(\pi, \xi) &= \mathbb{E}[\bar{f}(X) - f_\star] \\
&\overset{(a)}{=} \mathbb{E}[\bar{f}(X) - \mathbb{E}[f_\star|\kappa]] \\
&\overset{(b)}{\leq} \mathbb{E}[\bar{f}_\kappa(X) - \mathbb{E}[f_\star|\kappa]] + \epsilon \\
&\overset{(c)}{\leq} \mathbb{E}[\bar{f}_\kappa(x_\kappa) - \mathbb{E}[f_\kappa|\kappa]_\star] + 3\epsilon \\
&\overset{(d)}{=} \Delta(\pi_\kappa, \nu_\kappa) + 3\epsilon \\
&\overset{(e)}{\leq} \alpha + \sqrt{\beta I(\pi_\kappa, \nu_\kappa)} + 3\epsilon \\
&\overset{(f)}{\leq} \alpha + \sqrt{\beta I(\pi_\kappa, \nu)} + \epsilon[3 + \sqrt{\beta}] \\
&\overset{(g)}{\leq} \alpha + \sqrt{\beta I(\pi, \nu)} + 3\epsilon[1 + \sqrt{\beta}] \\
&\overset{(h)}{\leq} \alpha + \sqrt{\frac{\beta I(\kappa; X, Y)}{2}} + 3\epsilon[1 + \sqrt{\beta}]\,,
\end{aligned}$$

where

(a) by the tower rule.

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

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

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

$$\begin{aligned}
\mathbb{E}[f_\star|\kappa] &= \mathbb{E}[f(x_f)|\kappa] \\
&\geq \mathbb{E}[f(\tilde{x}_f)|\kappa] - \frac{\epsilon_R}{2}\operatorname{diam}(K)^2 && \text{by Lemma 12 (b)} \\
&\geq \mathbb{E}[f(x_\kappa)|\kappa] - 2\epsilon && \text{by the assumption on } \epsilon_R \text{ and (ii)} \\
&\geq \mathbb{E}[f_\kappa(x_\kappa)|\kappa] - 3\epsilon && \text{by (i)} \\
&= \mathbb{E}[f_\kappa|\kappa]_\star - 3\epsilon\,.
\end{aligned}$$

And because by the triangle inequality, the definition of $X_\star$ and Lemma 12 (a),

$$\|X_\star - x_\kappa\| \leq \|X_\star - \tilde{x}_f\| + \|\tilde{x}_f - x_\kappa\| \leq \sqrt{2\epsilon_O/\epsilon_R} + \epsilon \leq 2\epsilon\,, \tag{5.1}$$

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

(d) follows by definition.

(e) follows by $(\alpha, \beta) \in \text{IR}(\mathcal{F})$.

(f) follows from Lemma 10 and because

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

(g) follows from Lemma 10 and Eq. (5.1).

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

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

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

We are now in a position to prove Theorem 13. Let π_t be the law of X_t under $\mathbb{P}_{t-1}$ and ξ_t be the law of f under $\mathbb{P}_{t-1}$. By Lemma 14,

$$\Delta(\pi_t, \xi_t) \leq \alpha + \sqrt{\beta I_t(\kappa; X_t, Y_t)} + 3\epsilon[1 + \sqrt{\beta}]\,.$$

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

$$\begin{aligned}
\text{BReg}_n(\text{ATS}, \xi) &= \mathbb{E}\left[\sum_{t=1}^n \Delta(\pi_t, \xi_t)\right] \\
&\leq n\alpha + \mathbb{E}\left[\sum_{t=1}^n \sqrt{\beta I_{t-1}(\kappa; X_t, Y_t)}\right] + 3n\epsilon[1 + \sqrt{\beta}] \\
&\leq n\alpha + \sqrt{\beta n \mathbb{E}\left[\sum_{t=1}^n I_{t-1}(\kappa; X_t, Y_t)\right]} + 3n\epsilon[1 + \sqrt{\beta}] \\
&\leq n\alpha + \sqrt{\beta n \log(N)} + 3n\epsilon[1 + \sqrt{\beta}]\,,
\end{aligned}$$

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 $\log(N)$. □

Chapter 6

Thompson Sampling in 1-dimension

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

Theorem 15. *When* $d = 1$, $\sup_{\xi \in \mathscr{P}(\mathscr{F}_{\texttt{bl}})} \mathit{BReg}_n(\mathrm{TS}, \xi) = O\left(\sqrt{n \log(n) \log(n \operatorname{diam}(K))}\right)$.

Theorem 15 is established by combining the following bound on the information ratio and $\alpha = 1/n$ with Theorem 3.

Theorem 16. *Suppose that* $d = 1$ *and* $\alpha \in (0, 1)$. *Then* $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \textsc{ir}(\mathscr{F}_{\texttt{bl}})$.

Proof. Let $\bar{f} \in \operatorname{conv}(\mathscr{F}_{\texttt{bl}})$ and for integer i define

$$\mathscr{F}_i = \begin{cases} \{f \in \mathscr{F}_{\texttt{bl}} : \bar{f}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f \geq x_{\bar{f}}\} & \text{if } i > 0 \\ \{f \in \mathscr{F}_{\texttt{bl}} : \bar{f}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f < x_{\bar{f}}\} & \text{if } i < 0 \\ \{f \in \mathscr{F}_{\texttt{bl}} : \bar{f}(x_f) - f_\star \leq \alpha\} & \text{if } i = 0\,. \end{cases} \tag{6.1}$$

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

$$\sup_{f \in \mathscr{F}_i} \left(\bar{f}(x_f) - f_\star\right) \leq \epsilon \leq \alpha + \sqrt{230 \inf_{f_1, \ldots, f_k \in \mathscr{F}_i} \sum_{j,l \in \textsc{pair}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,. \tag{6.2}$$

Hence, by Lemma 4 and naive simplification of constants $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \textsc{ir}(\mathscr{F}_{\texttt{bl}})$ as desired. The first inequality in Eq. (6.2) is an immediate consequence of the definition of $\mathscr{F}_i$ and

ϵ. 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. (6.2) holds for any $i > 0$. Let $f_1, \dots, f_k \in \mathscr{F}_i$ and for $j \in [4]$ let $x_j = x_{f_j}$ and assume without loss of generality that $x_1 \le x_2 \le x_3 \le x_4$. Suppose that

$$\sum_{j,l \in \text{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2 < c^2\epsilon^2 \qquad \text{with} \qquad c = \frac{\sqrt{65} - 7}{16}\,. \tag{6.3}$$

Let us now establish a contradiction, which we do in three steps. The main argument in each step is illustrated in Figure 6.1.

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

$$f_1(x_3) \overset{\text{(a)}}{\le} \bar{f}(x_3) + \epsilon c \overset{\text{(b)}}{\le} f_3(x_3) + \epsilon[c+1] \le f_3(x_1) + \epsilon[c+1] \overset{\text{(c)}}{\le} \bar{f}(x_1) + \epsilon[2c+1]\,,$$

where (a) follows from Eq. (6.3), (b) since for $f \in \mathscr{F}_i$, $f(x_f) \ge \bar{f}(x_f) - \epsilon$ and (c) from Eq. (6.3) again. Hence, with $p \in [0,1]$ such that $x_2 = (1-p)x_1 + px_3$,

$$\begin{aligned}
\bar{f}(x_1) \overset{\text{(a)}}{\le} \bar{f}(x_2) \overset{\text{(b)}}{\le} f_1(x_2) + c\epsilon &\overset{\text{(c)}}{\le} (1-p)f_1(x_1) + pf_1(x_3) + c\epsilon \\
&\overset{\text{(d)}}{\le} (1-p)(\bar{f}_1(x_1) - \epsilon/2) + pf_1(x_3) + c\epsilon \\
&\overset{\text{(e)}}{\le} \bar{f}(x_1) + \epsilon\left[c + p[2c + 3/2] - 1/2\right]\,,
\end{aligned}$$

where (a) follows because $\bar{f}$ is non-decreasing on $[x_1, x_4]$ by the definition of $\mathscr{F}_i$ and $i > 0$, (b) from Eq. (6.3), (c) by convexity and the definition of p, (d) since $f_1(x_1) \le \bar{f}_1(x_1) - \epsilon/2$ by the definition of $\mathscr{F}_i$ and (e) is true by the previous display. Therefore $p \ge (1/2 - c)/(2c + 3/2) \approx 0.27$.

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,

$$\begin{aligned}
\bar{f}(x_1) &\overset{\text{(a)}}{\le} \bar{f}(x_2) \overset{\text{(b)}}{\le} f_3(x_2) + c\epsilon \\
&\overset{\text{(c)}}{\le} (1-p)f_3(x_1) + pf_3(x_3) + c\epsilon \\
&\overset{\text{(d)}}{\le} (1-p)\bar{f}(x_1) + pf_3(x_3) + 2c\epsilon\,,
\end{aligned}$$

(a)–(c) follows as above in **STEP 1** and (d) from Eq. (6.3). Rearranging shows that $f_3(x_3) \ge \bar{f}(x_1) - \frac{2c\epsilon}{p}$.

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

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

where `(a)` follows by convexity and because f_4 is minimised at x_4, `(b)` from Eq. (6.3), `(c)` from **STEP 2**, `(d)` since $f_3(x_3) \leq \bar{f}(x_3) - \epsilon/2$ by the definition of $\mathscr{F}_i$ and `(e)` from the bound on p in **STEP 1** and the definition of c. But this contradicts Eq. (6.3). Hence Eq. (6.3) does not hold. And since $c^2 \geq \frac{1}{230}$ it follows that Eq. (6.2) holds. □

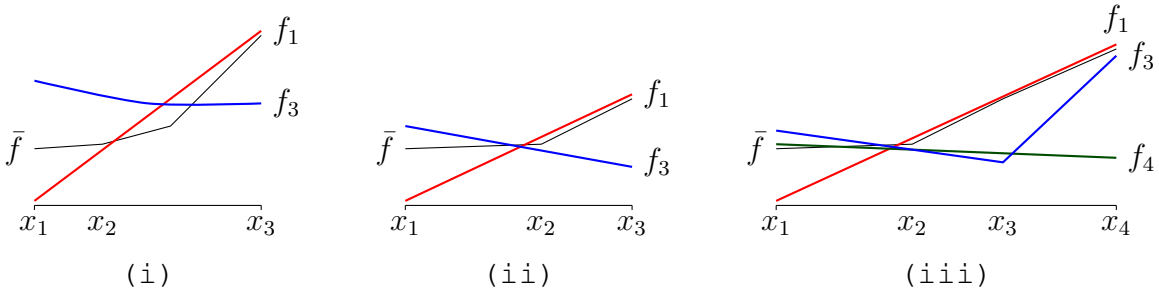

Figure 6.1: `(i)` shows that if x_2 is too close to x_1, then $f_1(x_3)$ must be large, which implies that $f_3(x_3)$ must be large and so too must $f_3(x_1)$, which shows that $f_3(x_1) - \bar{f}(x_1)$ is large. `(ii)` shows what happens if $f_3(x_3)$ is too far below $\bar{f}(x_1)$, which is that $f_3(x_1)$ 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.

Chapter 7

Thompson Sampling for Ridge Functions

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

Theorem 17. $\sup_{\xi \in \mathscr{P}(\mathscr{F}_{\texttt{blrm}})} \textit{BReg}_n(\mathrm{TS}, \xi) = O\left(d^{2.5}\sqrt{n}\log(nd\,\mathrm{diam}(K))^2\right)$.

[RV16] used information-theoretic means to show that for linear bandits the regret is at most $\tilde{O}(d\sqrt{n})$. [Lat21] 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 15, Theorem 17 is established by combining a bound on the information ratio with Theorem 3.

Theorem 18. $(\alpha, \beta\lceil\log(1/\alpha)\rceil) \in \textsc{ir}(\mathscr{F}_{\texttt{blrm}})$ *whenever* $\alpha \in (0,1)$ *and*

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

with the Big-O hiding only a universal constant.

Proof. Abbreviate $\mathscr{F} = \mathscr{F}_{\texttt{blmr}}$. The high-level argument follows the proof of Theorem 16. The main challenge is lower bounding the quadratic (information gain) term that appears in Lemma 4, which uses an argument based on the method of inscribed ellipsoid for optimisation [TKE88]. Let $\bar{f} \in \mathrm{conv}(\mathscr{F})$. Given a nonempty finite set $\mathcal{C} \subset \mathscr{F}$ and $\delta > 0$, let $J_\delta(\mathcal{C}) = \mathrm{conv}\left(\cup_{g\in\mathcal{C}} \mathbb{B}_\delta(x_g)\right)$. 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 19. *Let $\epsilon > 0$, $\delta = \frac{\epsilon}{12(d+1)}$ and $\mathcal{C} \subset \mathscr{F}$ be a nonempty finite set such that for all $f, g \in \mathcal{C}$, $\bar{f}(x_f) - f_\star \in [\epsilon/2, \epsilon]$ 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 $\operatorname{vol}(E_\delta(\mathcal{C} \setminus \{f\})) \leq 0.85 \operatorname{vol}(E_\delta(\mathcal{C}))$.*

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

Proof. Let $\mu \in \mathbb{R}^d$ and Σ be positive definite such that $E_\delta(\mathcal{C}) = \{x : \|x - \mu\|_{\Sigma^{-1}} \leq 1\}$ and for $r > 0$, let $E_{\delta,r}(\mathcal{C}) = \{x : \|x - \mu\|_{\Sigma^{-1}} \leq r\}$. By John's theorem [AAGM15, 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 ℓ 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 θ. Consider the following cases, illustrated in Figure 7.1:

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 [Kha90] shows that $\operatorname{vol}(E_\delta(\mathcal{C} \setminus \{f\}))) \leq 0.85 \operatorname{vol}(E_\delta(\mathcal{C}))$.

CASE 2 There exists a $g \in \mathcal{C} \setminus \{f\}$ such that 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 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

$$\langle x'_f, \theta \rangle \geq \langle \mu, \theta \rangle - d\|\theta\|_\Sigma \qquad (\text{since } J_\delta(\mathcal{C}) \subset E_{\delta,d}(\mathcal{C})) \tag{7.1}$$

$$\langle x'_g, \theta \rangle \leq \langle \mu, \theta \rangle \tag{7.2}$$

$$\langle x'_h, \theta \rangle \geq \langle \mu, \theta \rangle + \|\theta\|_\Sigma \,. \qquad (\text{since } E_\delta(\mathcal{C}) \subset J_\delta(\mathcal{C})) \tag{7.3}$$

Since ℓ 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,

$$\begin{aligned} f(x_g) &\overset{(a)}{\leq} \delta + f(x'_g) \overset{(b)}{=} \delta + f\left(\frac{\langle x'_g - x'_f, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_h + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_f\right) \\ &\overset{(c)}{\leq} \delta + f(x'_h) + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle}(f(x'_f) - f(x'_h)) \\ &\overset{(d)}{\leq} \delta + f(x'_h) + \frac{1}{d+1}(f(x'_f) - f(x'_h)) \overset{(e)}{\leq} 3\delta + f(x_h) + \frac{1}{d+1}(f(x_f) - f(x_h))\,, \end{aligned} \tag{7.4}$$

where (a) follows because f is a Lipschitz ridge function and using Lemma 2 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. (7.3) and because $f(x'_f) \leq f(x'_h)$. (e) uses again that f is a

Lipschitz ridge function and Lemma 2 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

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

where (a) follows from Eq. (7.4), (b) follows from the assumption in the lemma statement that $f(x_f) \leq \bar{f}(x_f) - \epsilon/2 \leq \bar{f}(x_h) + \delta - \epsilon/2$ and $f(x_h) < \bar{f}(x_h) + \delta$. (c) by naive simplification, (d) by the definition of ϵ and δ 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. □

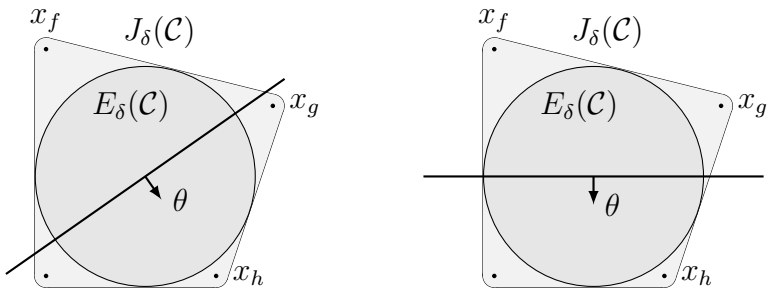

Figure 7.1: The two cases considered in the proof of Lemma 19. In the left figure, the situation is such that $E_\delta(\mathcal{C} \setminus \{f\})$ is a constant fraction less volume than $E_\delta(\mathcal{C})$. On the other hand, in the figure on the right one of $(f(x_h) - \bar{f}(x_h))^2$ or $(f(x_g) - \bar{f}(x_g))^2$ 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 20. *Suppose that $\mathcal{C}$ satisfies the conditions of Lemma 19 for some $\epsilon \geq \alpha$ and $|\mathcal{C}| = 1 + 2d + 8d\left\lceil \log\left(\frac{24d(d+1)\,\mathrm{diam}(K)}{\alpha}\right)\right\rceil$. Then*

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

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 $\mathrm{vol}(E_\delta(\mathcal{C}_{k+1})) \leq 0.85\,\mathrm{vol}(E_\delta(\mathcal{C}_k))$; or

(ii) $\mathcal{C}_{k+1} = \mathcal{C}_k \setminus \{f, g\}$ for some $f, g \in \textsc{pair}(\mathcal{C}_k)$ and $(f(x_g) - \bar{f}(x_g))^2 \geq \delta^2$.

Such a sequence exists by Lemma 19. 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

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

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

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

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

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

Therefore (ii) happens at least d times and the claim follows by the definition of (ii). □

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 20. Let

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

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

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

In order to apply Lemma 4 we will show that for all $0 \leq i \leq \lceil \log_2(1/\alpha) \rceil$ and with $\epsilon = \alpha 2^i$ that for all $f_1, \ldots, f_k \in \mathscr{F}_i$,

$$\sup_{f \in \mathscr{F}_i} \left(\bar{f}(x_f) - f_\star\right) \leq \epsilon \leq \alpha + \sqrt{512 \sum_{j,l \in \textsc{pair}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,. \tag{7.5}$$

This implies that $(\alpha, 512k(k-1)(1+\lceil\log(1/\alpha)\rceil)) \in \text{IR}(\mathscr{F})$ as required. The first inequality in Eq. (7.5) follows immediately from the definition of ϵ and $\mathscr{F}_i$. The second is also immediate when $i = 0$ by the definition of $\mathscr{F}_i$. Suppose now that $i > 0$. Let $f_1, \ldots, f_k \in \mathscr{F}_i$ and assume without loss of generality that $j \mapsto \bar{f}(x_{f_j})$ is nonincreasing. The second inequality in Eq. (7.5) holds immediately if $f_j = f_l$ for some $j, l \in \text{PAIR}(k)$ since $f_j(x_{f_l}) = f_j(x_{f_j}) \leq \bar{f}(x_{f_j}) - \epsilon/2$. Suppose this is not the case and consider two cases:

CASE 1 $\bar{f}(x_{f_1}) \geq \bar{f}(x_{f_k}) + 2\epsilon$. In this case $f_1(x_{f_k}) \geq f_1(x_{f_1}) \geq \bar{f}(x_{f_1}) - \epsilon \geq \bar{f}(x_{f_k}) + \epsilon$, which shows that $(f_1(x_{f_k}) - \bar{f}(x_{f_k}))^2 \geq \epsilon^2$ and the second inequality Eq. (7.5) holds.

CASE 2 $\bar{f}(x_1) < \bar{f}(x_k) + 2\epsilon$. Let $\delta = \frac{\epsilon}{12(d+1)}$ as in Lemma 20. Let $b = 25d + 24$ and $\mathcal{C}_1, \ldots, \mathcal{C}_b$ be formed by dividing $\{f_1, \ldots, f_k\}$ 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 $2\epsilon > \sum_{a=1}^{b} s_a \geq \sum_{a=1}^{b} \delta \mathbf{1}(s_a > \delta)$, which means that $\sum_{a=1}^{b} \mathbf{1}(s_a > \delta) \leq 2\epsilon/\delta \leq 24(d+1) = b - d$. Hence there exist at least d blocks $\mathcal{C}_a$ for which $s_a \leq \delta$ and these blocks satisfy the conditions of Lemma 20 so that $\sum_{j,l \in \text{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2 \geq d^2\delta^2 \geq \epsilon^2/512$ and again the second inequality in Eq. (7.5) holds.

Hence Eq. (7.5) holds and the claim follows from Lemma 4 and the definitions of k and m. □

Chapter 8

TS Lower-Bound for general Convex Functions

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 21. *Let $\epsilon \in [0, 1/2]$ and $\theta \in \mathbb{S}_1$ and define functions f and f_θ by*

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

Then f_θ is convex and minimised at θ and $\operatorname{Lip}_{\mathbb{B}_1}(f_\theta) \leq \sqrt{2+2\epsilon}$.

The function f_θ 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: $\infty \mathbf{1}_\theta(\cdot)$.

Proof of Lemma 21. Recall that $\epsilon \in [0, 1/2]$ and

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

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

$$\begin{aligned}
g_\theta(x) &\triangleq \sup_{y\in\mathbb{R}^d} \{f(y) + \langle f'(y), x-y\rangle : f(y) + \langle f'(y), \theta - y\rangle \le 0\} \\
&= \sup_{\substack{y\in\mathbb{R}^d \\ r\le 0}} \{f(y) + \langle f'(y), x-y\rangle : f(y) + \langle f'(y), \theta - y\rangle = r\} \\
&= \sup_{\substack{y\in\mathbb{R}^d \\ r\le 0}} \{\langle f'(y), x-\theta\rangle + r : f(y) + \langle f'(y), \theta - y\rangle = r\} \\
&= \sup_{\substack{y\in\mathbb{R}^d \\ r\le 0}} \left\{\langle y, x-\theta\rangle + r : \|y-\theta\|^2 = 1 + 2\epsilon - 2r\right\} \\
&= \sup_{r\le 0} \left\{\langle \theta, x-\theta\rangle + r + \sqrt{1+2\epsilon-2r}\,\|x-\theta\|\right\} \\
&= \sup_{r\le 0} \left\{\langle \theta, x\rangle - 1 + r + \sqrt{1+2\epsilon-2r}\,\|x-\theta\|\right\} \qquad (8.1) \\
&= f_\theta(x)\,,
\end{aligned}$$

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

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

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

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

where $Df_\theta(\theta)[\cdot]$ is the directional derivative operator (noting that f_θ is not differentiable at θ). Lastly, for Lipschitzness. Since f_θ is continuous, it suffices to bound $\|f'_\theta(\cdot)\|$ on $\mathrm{int}(\mathbb{B}_1)$ where f_θ is differentiable. When $\|x-\theta\|^2 \ge 1 + 2\epsilon$, then $\|f'_\theta(x)\| = \|f'(x)\| = \|x\| \le 1$. On the other hand, if $\|x-\theta\|^2 < 1 + 2\epsilon$, then

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

Therefore $\mathrm{Lip}_{\mathbb{B}_1}(f) \leq \sqrt{2+2\epsilon}$. □

Theorem 22. *When $K = \mathbb{B}_1$ is the standard euclidean ball. There exists a prior ξ on $\mathscr{F}_{b1}$ such that*

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

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_θ be defined as in Lemma 21 with $\epsilon = 1/4$. Let σ be the uniform probability measure on $\mathbb{S}_1$ and the prior ξ be the law of f_θ when θ is sampled from σ. By the definition of f_θ and the fact that $\epsilon = 1/4$, 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_θ 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 $f_{\theta_\star}$ be the true loss function sampled from ξ. Suppose that $X_1, \ldots, X_t \in \mathbb{S}_1 \setminus \mathcal{C}_{\theta_\star}$, which means that $Y_s = \frac{3}{4}$ for $1 \leq s \leq t$ and the posterior is the uniform distribution on $\Theta_{t+1} = \mathbb{S}_1 \setminus \cup_{s=1}^t \mathcal{C}_{X_s}$. Provided that $t\delta \leq \frac{1}{2}$,

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

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

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

The result is completed since $\delta \leq \exp(-d/32)$ follows from concentration of measure on the sphere [Tko18, Theorem B.1]. □

Definition 23. For $\theta \in \mathbb{S}_1$ and $\epsilon \in [0, 1/2]$, define

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

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

Definition 24. 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 [Tko18] to bound the surface area of spherical caps.

Theorem 25. *[Tko18, Theorem B.1] For all $\epsilon \in [0,1]$ and $\theta \sim \sigma$ we have*

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

Proof of Theorem 22. Define f and f_θ as in Lemma 21 with $\epsilon = 1/4$, which then using Definition 23 can be written as

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

where we drop the ϵ from $\mathcal{C}_{\theta,\epsilon}$ and $\bar{\mathcal{C}}_{\theta,\epsilon}$ in the notation for simplicity. We define the bandit instance by setting the prior ξ_1 to be the law of f_θ when θ has law σ, 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_{θ_*} is the true function that is secretly sampled from the prior ξ_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

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

which follows from the definition of f_θ.

Step 1: First we show that if the posterior distribution ξ_t at round $t \in [T]$ is uniform over Θ_t, and the algorithm observes the loss $Y_t = \frac{3}{4}$ as a result of playing X_t, then the posterior distribution ξ_{t+1} at round $t+1$ is uniform over Θ_{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 $B \subseteq \Theta_t$,

$$\xi_{t+1}(B) = \mathbb{P}_{t-1}\left(\theta_\star \in B | Y_t = \frac{3}{4}, X_t\right) = \frac{\mathbb{P}_{t-1}(\theta_\star \in B, Y_t = \frac{3}{4}|X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t)}$$
$$= \frac{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t, \theta_\star \in B)\mathbb{P}_{t-1}(\theta_\star \in B|X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t)}$$
$$= \frac{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4}|X_t, \theta_\star \in B)\mathbb{P}_{t-1}(\theta_\star \in B)}{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4}|X_t)}. \quad (8.2)$$

Note that TS samples f_{θ_t} from ξ_t, and then plays the minimizer of f_{θ_t}, which from Lemma 21 is θ_t, i.e. $X_t = \theta_t$. Consequently, continuing from Eq. (8.2) 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

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

which implies that ξ_{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 ξ_{t+1} is uniform over Θ_{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 $X_1, \ldots, X_t \in \bar{\mathcal{C}}_{\theta_\star}$, which implies

both that $Y_1, \ldots, Y_t = \frac{3}{4}$, and that ξ_{t+1} is uniform over Θ_{t+1}. Conditioned on $\mathcal{E}_t$, we have

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

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

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

which further implies that

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

and therefore

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

Let $n_0 = \min(\left\lfloor \frac{1}{4\delta} \right\rfloor, n)$, 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 $\mathbb{P}(\mathcal{E}_1) = \mathbb{P}(\theta_1 \in \bar{\mathcal{C}}_{\theta_\star}) = 1 - \delta$. 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 25, we have

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

which implies that

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

□

Chapter 9

IR Lower-Bound for General Convex Functions

Theorem 22 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 26. *Suppose that $K = \mathbb{B}_1$ and $d > 256$. Then there exists a prior ξ on $\mathscr{F}_{b1}$ such that for all probability measures π on K, $\Delta(\pi, \xi) \geq 2^{-19}\frac{d}{\log(d)}\sqrt{\mathcal{I}(\pi, \xi)}$.*

The prior ξ is the same as used in the proof of Theorem 22 but with $\epsilon = \tilde{\Theta}(1/d)$. The argument is based on proving that for any policy the regret is $\Omega(\epsilon)$ while the information gain is $\tilde{O}(\epsilon^2)$.

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

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

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

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

which should be thought of as the expected loss and the expected information gain at x. Therefore,

for any policy π we have

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

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 π, and hence the claimed lower bound.

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

Lemma 27. *For $\tau \le r$, and $x \in \mathbb{B}_1$ with $\|x\| = r$, $\mathbb{P}\left(f_\theta(x) = f(x)\right) \ge 1 - \dfrac{1}{d^4}$.*

Proof. From Lemma 21, $f_\theta(x) = f(x)$ if $\|\theta - x\|^2 \ge 1 + 2\epsilon$. 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 $\langle\theta, x\rangle \le \frac{r^2}{2} - \epsilon$. Since $\theta \sim \sigma$,

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

where the first inequality follows from Theorem 25, and the second inequality follows from the fact that $r \ge \tau \ge \sqrt{2\epsilon}$. □

Lemma 28. *For all $x \in \mathbb{B}_1$ and $\theta \in \mathbb{S}_1$, we have*

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

Proof. The proof follows by setting $r = 0$ in Equation (8.1). □

Lemma 29. *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

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

where (a) follows from Lemma 27.

CASE 2 If $r < \tau$, using the lower bound on $f_\theta(x)$ from Lemma 28,

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

where (a) follows from $\mathbb{E}\left[\langle\theta, x\rangle\right] = 0$, (b) follows from the inequality $\sqrt{1+a} \geq 1 + \frac{a - a^2}{2}$ for $a \geq -1$, (c) follows from $\mathbb{E}\left[\langle\theta, x\rangle\right] = 0$, (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,

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

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

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

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

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

□

Lemma 30. *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

$$\begin{aligned}\mathbb{E}\left[(f_\theta(x) - \mathbb{E}\left[f_\theta(x)\right])^2\right] &\leq \mathbb{E}\left[(f_\theta(x) - f(x))^2\right] \\ &\leq \mathbb{P}\left(f_\theta(x) = f(x)\right)\left(f(x) - f(x)\right)^2 + \mathbb{P}\left(f_\theta(x) \neq f(x)\right)\left(\frac{1}{2}r^2 + \epsilon\right)^2 \\ &\leq \frac{1}{d^4}\left(\frac{1}{2}r^2 + \epsilon\right)^2 \leq \frac{1}{d^4}\left(\frac{1}{2} + \frac{1}{3}\right)^2 \leq \frac{25}{36d^4}\,,\end{aligned} \tag{9.1}$$

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

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

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

where (a) follows from $f(x) \geq f_\theta(x)$, the (b) follows from Lemma 28, (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 $r, \epsilon, |\langle\theta, x\rangle| \leq 1$, 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. (9.1), we have

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

where (a) follows from $(a+b+c)^4 \le 27(a^4+b^4+c^4)$, (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 $r < \tau = \sqrt{8\epsilon}$. □

Proof of Theorem 26. Let ξ be the law of f_θ when θ has law σ. Then for any policy π, and $X \sim \pi$, from Lemma 29 we have

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

and from Lemma 30 we have

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

which together imply

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

□

Chapter 10

Thompson Sampling for Adversarial Problems

In the Bayesian adversarial setting the prior is a probability measure on $\mathscr{F}^n$ 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 $f(x) = \epsilon + \max(\epsilon x, (1-\epsilon)x)$ and $g(x) = f(-x)$. Note that $f \in \mathscr{F}_{\text{b1}}$ and is piecewise linear and minimised at -1 with $f(-1) = 0$ and $f(0) = \epsilon$ and $f(1) = 1$. The function g is the mirror image of f. Now let ν 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 $\Omega(n)$.

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 ξ be a probability measure on $\mathscr{F} \times K$ and π be a probability measure on K and let

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

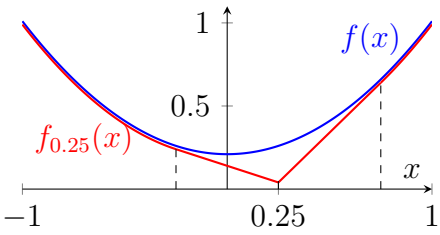

Figure 10.1: The function $f(x) = 0.2 + 0.8x^2$ and the function $f_{0.25}$ with $\epsilon = 0.2$. The function $f_{0.25}$ is the largest convex function that is smaller than f and has $f_{0.25}(0.25) = f(0.25) - 0.2$.

where $(f, X_\star, X)$ has law $\xi \otimes \pi$. Thompson sampling as in Item (b) is the policy π 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 α is unreasonably large. Let $d = 1, \epsilon \in (0, 2^{-7}), K = [-1, 1]$, and $f(x) = \epsilon + (1 - \epsilon)x^2$. Given $\theta \in [-1, 1]$ let $f_\theta(x)$ be defined as

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

which is convex and smaller than f for all $x \in K$. Essentially, f_θ should be thought of as the largest convex function that is smaller than f and has $f_\theta(\theta) = f(\theta) - \epsilon$ (see Fig. 10.1). Moreover, an elementary calculation shows that $\max_{x \in K} |f(x) - f_\theta(x)| = \epsilon$ for all $\theta \in [-1, 1]$. Let ξ be the law of (f_θ, θ) when θ is sampled uniformly from $[-1, 1]$ and π be uniform on $[-1, 1]$ which is the TS policy as defined in (b). Then, by letting θ' be an i.i.d. copy of θ we have

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

where the second equality follows from the definition of f_θ and the third equality follows from

$f_\theta(\theta) = f(\theta) - \epsilon$. Next, we have

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

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

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

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

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

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 $|\theta - \theta'| \geq \sqrt{\frac{\epsilon}{1-\epsilon}}$; and `(c)` follows from the fact that θ and θ' are i.i.d. on $[-1, 1]$. Therefore, by putting the two inequalities together we have

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

which can be further lower bounded by

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

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

Chapter 11

Discussion

11.1 Adversarial setup

In the Bayesian adverarial setting a sequence of loss functions $f_1, \dots, f_n$ are sampled from a joint distribution on $\mathscr{F}^n$. The learner plays X_t and observes $Y_t = f_t(X_t)$ and the Bayesian regret is $\mathrm{BReg}(\mathscr{A}, \xi) = \mathbb{E}[\sup_{x \in K} \sum_{t=1}^n (f_t(X_t) - f_t(x))]$. 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 [BDKP15] notes that the latter has an unbounded information ratio. More details are in **??**.

11.2 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 18 can be improved. We are cautiously believe that when the loss has the form $f(x) = \ell(\langle x, \theta \rangle)$ for *known* convex link function $\ell : \mathbb{R} \to \mathbb{R}$, then the information ratio is at most d. This would mean that convex generalised linear bandits are no harder than linear bandits.

11.3 TS vs IDS

Theorem 22 shows that TS can have more-or-less linear regret in high-dimensional problems. On the other hand, [BE18] and [Lat20] 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.

11.4 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 $K = \{x \geq \mathbf{0} : \|x\|_1 \leq 1\}$.

11.5 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 $f : K \to [0, 1]$ is convex and $\mathbb{B}_r \subset K$ and $K_\epsilon = (1 - \epsilon)K$. Then $\min_{x \in K_\epsilon} f(x) \leq \inf_{x \in K} f(x) + \epsilon$ and f is $1/(r\epsilon)$-Lipschitz on K_ϵ [Lat24, Chapter 3]. Hence, you can run TS on K_ϵ with $\epsilon = 1/n$ 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 [KLS95] and John's theorem, respectively.

11.6 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 [KKM12, AG12] and linear

bandits [AG13] 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.

11.7 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 [RLS93, CCH+07, SWD11]. 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

[AAGM15] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. *Asymptotic geometric analysis, Part I*, volume 202. American Mathematical Soc., 2015.

[AG12] 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.

[AG13] 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.

[AL17] 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.

[BDKP15] 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.

[BE18] S. Bubeck and R. Eldan. Exploratory distributions for convex functions. *Mathematical Statistics and Learning*, 1(1):73–100, 2018.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CCH+07] 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.

[DHK08] 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.

[DMR19] 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.

[FCGS10] 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.

[FKM05] 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.

[FvdHLM24] H. Fokkema, D. van der Hoeven, T. Lattimore, and J. Mayo. Online Newton method for bandit convex optimisation. In *Conference on Learning Theory*, 2024.

[HHK+21] 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.

[Kha90] L. G. Khachiyan. An inequality for the volume of inscribed ellipsoids. *Discrete & computational geometry*, 5:219–222, 1990.

[KKM12] 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.

[Kle05] R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704. MIT Press, 2005.

[KLS95] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

[KZS+20] 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.

[Lat20] T. Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2(3/4):311–334, 2020.

[Lat21] T. Lattimore. Minimax regret for bandit convex optimisation of ridge functions. *arXiv preprint arXiv:2106.00444*, 2021.

[Lat24] T. Lattimore. *Bandit Convex Optimisation*. 2024.

[LG23] T. Lattimore and A. György. A second-order method for stochastic bandit convex optimisation. *arXiv preprint arXiv:2302.05371*, 2023.

[LH21] T. Lattimore and B. Hao. Bandit phase retrieval. *Advances in Neural Information Processing Systems*, 34:18801–18811, 2021.

[Nie92] W. Niemiro. Asymptotics for m-estimators defined by convex minimization. *The Annals of Statistics*, pages 1514–1533, 1992.

[RLS93] 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.

[RV14] 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.

[RV16] D. Russo and B. Van Roy. An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17(1):2442–2471, 2016.

[RVK+18] 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.

[SNNJ21] 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.

[SWD11] 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.

[Tho33] 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.

[TKE88] S. Tarasov, L. G. Khachiyan, and I. I. Erlich. The method of inscribed ellipsoids. In *Soviet Mathematics-Doklady*, volume 37, pages 226–230, 1988.

[Tko18] T. Tkocz. Asymptotic convex geometry lecture notes. 2018.

[ZL19] 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.