# Thompson Sampling for Bandit Convex Optimization
# (Version 4.2)

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

The first part of this thesis has been published as Alireza Bakhtiari, Tor Lattimore, and Csaba Szepesvári. Thompson Sampling for Bandit Convex Optimization. In International Conference on Learning Theory, 2025.

# Acknowledgements

I would like to sincerely thank my supervisor, Csaba, for his time, mentorship, guidance, and patience throughout my studies. Csaba gave me the opportunity to pursue what I love most and provided me with unconditional intellectual freedom. His passion for research and honest approach to science have been a true inspiration. I also extend my gratitude to Tor Lattimore for his time and for teaching me the ways of research; his insights have profoundly shaped my thinking. Special thanks go to Dave Janz and Johannes Krishner for their close guidance and constructive feedback in shaping my research. Moreover, I am deeply grateful to my labmates—listed in alphabetical order as Alex prefers—Alex Ayoub, Kushagra Chandak, Shuai Liu, Samuel Robertson, and Vlad Tkachuk, for patiently listening to my ideas, providing valuable feedback, and assisting me in my research. This journey would have been far less enjoyable without their presence. I am also thankful to my friends for always being there when needed, and to my partner, Zahra, for her unwavering patience, support, and encouragement. Lastly, I feel incredibly fortunate to have my parents, Pooneh and Majid, whose although been far away, have always kept their love and care present in my life, and my brother, Amin, whose presence makes my life much more colorful and vibrant.

# Contents

# List of Tables

# List of Figures

# List of Algorithms

# Chapter 1

# Introduction

In this chapter we introduce the bandit convex optimization (BCO) problem, and its relation to other related settings. Convexity is a common assumption in optimization problems [Boyd and Vandenberghe, 2004]. In convex optimization, the goal is to find the minimum of a convex function over a convex set, typically with access to the function's value and gradient at any point in the domain. Formally, the objective 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. Broadly, convex optimization algorithms use function evaluations $f(x)$ as well as gradient and higher order information of $f$ to find the minimizer.

A related setting is the *zeroth-order* optimization, where the learner (optimization algorithm) just gets access to function's evaluations for points $x \in \mathcal{K}$, and no other information about $f$ like its gradient. Precisely, for $n$ many rounds, the algorithm selects a point $x_t \in \mathcal{K}$ and receives feedback $y_t = f(x_t)$ for $t \in \{1, \dots, n\}$, and use these evaluations to find (approximate) the minimizer of $f$.

Compared to classical convex optimization and *zeroth-order* optimization, the BCO problem is characterized by two key features:

(i) Limited Noisy Access: The algorithm can only observe *noisy* point evaluations of the objective function, i.e. the feedback is of the form $f(x_t) + \varepsilon_t$, where $\varepsilon_t$ is the noise.

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

A representative example of BCO is dynamic pricing, where a seller selects a price (the input) and observes the resulting negative-profit (the output), which is often modeled as a convex 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.

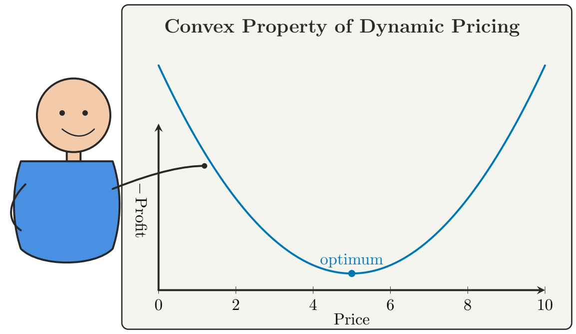



**Figure 1.1:** Convex property of dynamic pricing ($x =$ price, $y = -$profit).

Precisely, the learner in the BCO setting aims to minimize the cumulative loss over $n$ rounds, which leads to the online variant of the optimization problem. The goal after $n$ rounds of interaction is to minimize the cumulative loss compared to the best fixed point in hindsight, which is

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

referred to as *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.

## 1.1 Examples of BCO

BCO naturally arises in a variety of applications where gradient information is unavailable, unreliable, or too expensive to compute, and where each $f$ evaluation comes with a cost. Here are some examples:

*Hyperparameter Tuning:* Optimizing hyperparameters is a challenge for systems that can only get evaluations of their performance through live deployment. For instance, consider an automated trading algorithm with tunable hyperparameters. Its performance—measured by profit—can often be modeled as a convex function of these parameters. Evaluating performance requires running the algorithm live, which comes with financial risk and opportunity cost. Since the system is deployed in production, the goal is not to identify the best hyperparameters in hindsight, but to continually adapt and improve performance over time with minimal cumulative loss.

*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.

# Chapter 2

# Related work

BCO in the regret setting was first studied by Flaxman et al. [2005] and Kleinberg [2005]. Since then the field has grown considerably as summarized in the recent monograph by Lattimore [2024]. Our focus is on the Bayesian version of the problem, which has seen only limited attention. Bubeck et al. [2015] consider the adversarial version of the Bayesian regret and show that a (heavy) modification of TS enjoys a Bayesian regret of $\tilde{O}(\sqrt{n})$ when $d = 1$. Interestingly, they argue that TS without modification is not amenable to analysis via the information-theoretic machinery, but this argument only holds in the adversarial setting as our analysis shows. Bubeck and Eldan [2018] and Lattimore [2020] generalized the information-theoretic machinery used by Bubeck et al. [2015] to higher dimensions, also in the adversarial setting. These works focus on proving bounds for information-directed sampling (IDS), which is a conceptually simple but computationally more complicated algorithm introduced by Russo and Van Roy [2014]. Nevertheless, we borrow certain techniques from these papers. Convex ridge functions have been studied before by Lattimore [2021], who showed that IDS has a Bayesian regret in the stochastic setting of $\tilde{O}(d\sqrt{n})$, which matches the lower bound provided by linear bandits [Dani et al., 2008]. Regrettably, however, this algorithm is not practically implementable, even under the assumption that you can sample efficiently from the posterior. Saha et al. [2021] also study a variation on the problem where the losses have the form $f(g(x))$ with $g : \mathbb{R}^d \to \mathbb{R}$ 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 [Fokkema et al., 2024a]. 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 [Filippi et al., 2010, and many more]. TS and other randomized algorithms have been studied with generalized linear models in Bayesian and frequentist settings [Abeille and Lazaric, 2017, Dong et al., 2019, Kveton et al., 2020]. None of these papers assume convexity (concavity for rewards) and 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

# The Bayesian BCO Problem and the TS Algorithm

In this chapter we introduce the Bayesian version of the BCO problem and the Thompson sampling (TS) algorithm. The Bayesian version of the BCO problem is a sequential decision-making problem where the learner has access to a prior distribution over the unknown loss function $f$. This prior distribution captures the learner or the domain expert's beliefs about the objective function before any interaction. Thompson sampling (TS) is a simple and often practical algorithm for interactive decision-making with a long history [Thompson, 1933, Russo et al., 2018]. Our interest is in its application to Bayesian bandit convex optimization [Lattimore, 2024]. 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.

## 3.1 The Bayesian BCO Problem

Let $\mathcal{K}$ be a convex body in $\mathbb{R}^d$ and $\mathscr{F}$ be a set of convex functions from $\mathcal{K}$ to $[0, 1]$. We assume there is a known (prior) probability measure $\xi$ on $\mathscr{F}$ (equipped with a $\sigma$-algebra). The interaction between the learner and environment lasts for $n$ rounds. At the beginning the environment secretly samples $f$ from the prior $\xi$. Subsequently, the learner and environment interact sequentially. In round $t$ the learner chooses an action $X_t \in \mathcal{K}$ and observes $Y_t \in \{0, 1\}$ for which $\mathbb{E}[Y_t | X_1, Y_1, \dots, X_t, f] = f(X_t)$. The assumption that the observations are Binary 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 $\xi$ is

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

Note that both $f$ and the actions $(X_t)$ are random elements. In addition the learner $\mathscr{A}$ is allowed to use the prior $\xi$. 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 $\sigma$-algebra) and TS is Thompson sampling (Algorithm 1) with prior $\xi$. In what follows, the dependence on the prior will be omitted from the notation to minimize clutter. 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—assuming efficient sampling from the posterior, and efficient minimization of the sampled function, it does not require any additional machinery (e.g., exploration bonuses or confidence bounds). 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 ∈ argmin_{x∈K} F_t(x) and observe Y_t
```

**Algorithm 1:** Thompson sampling

## 3.2 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 13).
- We call a convex function $f$ a monotone ridge function if there exists a convex monotone (non-decreasing or non-increasing) function $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{R}^d$ such that $f(x) = \ell(\langle x, \theta \rangle)$. Theorem 15 shows that when $\xi$ 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 for general multidimensional convex losses (Theorem 22).
- The classical information-theoretic machinery used by Bubeck and Eldan [2018] an Lattimore [2020] 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. It's worth mentioning that prior dependence is also a limitation for Bayesian algorithms like TS; in the case of prior mismatch, the performance of these algorithms can degrade significantly. 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 $\mathcal{K}$ 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 $d$ projects. 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 $\mathcal{K} = \{x \geq \mathbf{0} : \|x\|_1 \leq 1\}$.

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

## 3.3 Notation

Let $\|\cdot\|$ be the standard euclidean norm on $\mathbb{R}^d$. Let $\mathbb{R}_+$ be the set of non-negative real numbers. For natural number $k$ let $[k] = \{1, \dots, k\}$. Define $\|x\|_\Sigma = \sqrt{x^\top \Sigma x}$ for positive definite $\Sigma \in \mathbb{R}^{d\times d}$ and $x \in \mathbb{R}^d$. Given a function $f : \mathcal{K} \to \mathbb{R}$, let $\|f\|_\infty = \sup_{x\in\mathcal{K}} |f(x)|$. The centered euclidean ball of radius $r > 0$ is $\mathbb{B}_r = \{x \in \mathbb{R}^d : \|x\| \leq r\}$ and the sphere is $\mathbb{S}_r = \{x \in \mathbb{R}^d : \|x\| = r\}$. We also let $\mathbb{B}_r(x) = \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$. We let $H(x, \eta) = \{y : \langle y, \eta\rangle \geq \langle x, \eta\rangle\}$, which is a half-space with inward-facing normal $\eta$. Given a finite set $\mathcal{C}$ let $\textsc{pair}(\mathcal{C}) = \{(x, y) \in \mathcal{C} : x \neq y\}$ be the set of all distinct ordered pairs in $\mathcal{C}$ and abbreviate $\textsc{pair}(k) = \textsc{pair}([k] \times [k])$. For nonempty $A, B \subset \mathbb{R}^d$, the Minkowski sum is $A + B = \{a + b : a \in A, b \in B\}$. Similarly, $\mathscr{P}(\mathscr{F})$ is a space of probability measures on $\mathscr{F}$ with some unspecified $\sigma$-algebra ensuring that $f \mapsto f(x)$ is measurable for all $x \in \mathcal{K}$. Given a convex function $f : \mathcal{K} \to \mathbb{R}$ we define $\operatorname{Lip}_K(f) = \sup_{x\neq y\in\mathcal{K}}(f(x) - f(y))/\|x - y\|$. Further, let $f_\star = \inf_{x\in\mathcal{K}} f(x)$ and $x_f = \arg\min_{x\in\mathcal{K}} f(x)$ where ties are broken in an arbitrary measurable fashion; [Niemiro, 1992] showed that such a mapping exists and $f \mapsto f_\star$ is also measurable. Of course it follows that $f \mapsto f_\star = f(x_f)$ is also measurable. Let $\mathbb{P}_t = \mathbb{P}(\cdot|X_1, Y_1, \dots, X_t, Y_t)$ and $\mathbb{E}_t$ be the expectation operator with respect to $\mathbb{P}_t$. The following assumption on $\mathcal{K}$ is assumed globally:

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

### 3.3.1 Spaces of Convex Functions

Recall that we define the function $f : \mathcal{K} \to \mathbb{R}$ a convex ridge function if there exists a convex $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{R}^d$ such that $f(x) = \ell(\langle x, \theta\rangle)$ (hence, $f$ is convex). Moreover, $f$ is called a monotone convex ridge function if it is a convex ridge function and $\ell$ is monotone. We are interested in the following classes of convex functions:

(a) $\mathscr{F}_{\text{b}}$ is the space of all bounded convex functions $f : \mathcal{K} \to [0, 1]$.

(b) $\mathscr{F}_{\mathtt{l}}$ is the space of convex functions $f : \mathcal{K} \to \mathbb{R}$ 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 pairs $(\theta_1, \ell_1)$ and $(\theta_2, \ell_2)$ 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* $\ell$ can be chosen in a way that the Lipschitzness of the original function $f$ is preserved.

**Lemma 2.** *Suppose that $\mathcal{K}$ 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 $f(x) = \ell(\langle x, \theta \rangle)$, for all $x \in \mathcal{K}$. It remains to show that $\ell$ can be chosen so that $\mathrm{Lip}(\ell) \leq \mathrm{Lip}_K(f)$. Let $h_K$ be the support function associated with $\mathcal{K}$: $h_K(v) = \sup_{x \in \mathcal{K}} \langle v, x \rangle$. Therefore $\ell$ is uniquely defined on $I = [-h_K(-\theta), h_K(\theta)]$ and can be defined in any way that preserves convexity outside. Let $Dg(x)[v]$ be the directional derivative of $g$ at $x$ in direction $v$, which for convex $g$ exists for all $x$ in the interior of the domain of $g$. Then

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

Then define $\ell$ on all of $\mathbb{R}$ via the smallest convex extension [Lattimore, 2024, Proposition 3.18, for example]. $\square$

# Chapter 4

# Information Ratio

Information ratio is a now classical tool in the analysis of Bayesian regret for bandit problems [Russo and Van Roy, 2016]. It can be thought of as a complexity measure for Bayesian decision-making problems—the regret of any Bayesian algorithm can be bounded in terms of its worst-case information ratio. In this chapter we introduce our notion of information ratio which is a generalization of the information ratio introduced by Russo and Van Roy [2016]. Next, we move on to one of our main results, which provides a way to bound the generalized information ratio for TS, by partitioning the function class $\mathscr{F}$ into disjoint subsets.

## 4.1 Generalized Information Ratio

Recall that in the Bayesian setting, the true function $F$ is unknown, but the learner has access to a prior distribution $\xi \in \mathscr{P}(\mathscr{F})$ over the class of functions $\mathscr{F}$. The prior $\xi$ naturally induces a prior over the optimal action $X_F = \arg\min_{x \in \mathcal{K}} F(x)$, where the ties are broken in a deterministic measurable way. Therefore, from the information-theoretic perspective, there is only finite amount of information that the learner can gain about the optimal action $X_F$, which is equal to the entropy of $\Pr(X_f = \cdot)$. By taking actions $X \in \mathcal{K}$ and observing the loss $F(X)$, the learner can gain information about the function $F$ and the optimal action $X_F$. The main idea of the information ratio is to quantify the amount of information gained by the learner compared to the regret suffered by the learner in each round. More accurately, if the learner can upper bound its per-round regret by its information gain (either multiplicatively or additively), then its total regret can be bounded in terms of this upper bound and the entropy of the prior. Nevertheless, the actual measure of information gain needs to be carefully defined.

A policy $\pi \in \mathscr{P}(\mathcal{K})$ is a distribution over actions $X \in \mathcal{K}$. Given a distribution $\xi \in \mathscr{P}(\mathscr{F})$ and a policy $\pi \in \mathscr{P}(\mathcal{K})$, let $F \sim \xi$ and $X \sim \pi$, then with $\bar{F} = \mathbb{E}[F]$ define the gap (expected instantaneous regret) as

$$\Delta(\pi, \xi) = \mathbb{E}\left[F(X) - F_\star\right] = \mathbb{E}\left[\bar{F}(X) - F_\star\right]$$

where $F_\star = \inf_{x \in \mathcal{K}} F(x)$ is the optimal value of $F$. The classical information ratio [Russo and Van Roy, 2016] uses the mutual information as the measure of information gain,

$$\mathcal{I}_{KL}(\pi, \xi) = I\left(X_F; (X, F(X))\right) ,$$

where $I\left(X_F; (X, F(X))\right) = H(X_F) - H(X_F | (X, F(X)))$ is the mutual information between the optimal action $X_F$ and the pair $(X, F(X))$. So to summaries, the quantity $\Delta(\pi, \xi)$ is the regret suffered by $\pi$ when the loss function is sampled from $\xi$, while $\mathcal{I}(\pi, \xi)$ is a measure of the observed variation of the loss function. The classical version of the information ratio is multiplicative and defined as

$$\frac{\Delta(\pi, \xi)^2}{\mathcal{I}_{KL}(\pi, \xi)} .$$

An upper bound on this ratio for all $\xi \in \mathscr{P}(\mathscr{F})$, often implies an upper bound on the Bayesian regret of the learner Russo and Van Roy [2016]. Intuitively, this ratio stays the same if the learner suffers a large regret but also gains a lot of information, or if the learner suffers a small regret but gains little information.

Compared to the classical information ratio, our generalization is additive and uses a different measure of information gain. Firstly, we use variance based measure of information, which lower bounds the mutual information (up to constant factors), but is easier to work with, defined as

$$\mathcal{I}(\pi, \xi) = \mathbb{E}\left[(F(X) - \bar{F}(X))^2\right] .$$

Secondly, given a distribution $\xi$ and a random function $F$ with law $\xi$, we let $\pi_{\mathrm{TS}}^{\xi} \in \mathscr{P}(\mathcal{K})$ be the law of $x_f$, which is the minimizer of $F$. The generalized information ratio associated with TS on class of loss functions $\mathscr{F}$ is the set

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

To see why this is a generalization of the classical information ratio, note that $(0, \beta) \in \text{IR}(\mathscr{F})$ is equivalent to $\Delta(\pi_{\text{TS}}^\xi, \xi)^2 / \mathcal{I}(\pi_{\text{TS}}^\xi, \xi) \leq \beta$ for all $\xi \in \mathscr{P}(\mathscr{F})$, which is the classical information ratio of TS. The $\alpha$ term is used to allow a small amount of slack that eases analysis and may even be essential in non-parametric and/or infinite-action settings—it allows the learner to suffer a small amount of regret ($\alpha = O(1/\sqrt{n})$) without gaining any information.

The following theorem shows that finding a pair $(\alpha, \beta) \in \text{IR}(\mathscr{F})$ is sufficient to bound the Bayesian regret of TS.

**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*

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

*where the Big-O hides only a universal constant.*

This theorem is a direct consequence of Theorem 12, so to avoid repetition we kindly refer the reader to Section 5.4. At a high level the argument is based on similar results by Bubeck et al. [2015] and Bubeck and Eldan [2018].

## 4.2 Decomposition Lemma

Theorem 3 proves that if $(\alpha, \beta) \in \text{IR}_{\text{TS}}(\mathscr{F})$, then the Bayesian regret of TS is bounded in terms of $\alpha$ and $\beta$. Note that the generalized information ratio associated with TS, i.e. $\text{IR}_{\text{TS}}(\mathscr{F})$, is only a function of the class of functions $\mathscr{F}$. We introduce a mechanism for deriving information ratio bounds for TS through partitioning the function class $\mathscr{F}$ into disjoint subsets, and in later chapters we will use this mechanism to derive information ratio bounds for various classes of convex functions. Precisely, if we 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, then we can find a pair $(\alpha, \beta) \in \text{IR}_{\text{TS}}(\mathscr{F})$.

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

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

*Then* $(\alpha, k(k-1)m\beta) \in \textsc{ir}(\mathscr{F})$.

To build intution, note that 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.

*Proof of Lemma 4.* Let $\xi \in \mathscr{P}(\mathscr{F})$ and $\tilde{f} = \bar{F} = \mathbb{E}[F]$ be the expected loss function sampled from $\xi$. Then from the assumption of the lemma, there exist disjoint $\mathscr{F}_1, \dots, \mathscr{F}_m$ subsets of $\mathscr{F}$ such that $\mathscr{F} = \cup_{i=1}^m \mathscr{F}_i$ and

$$\sup_{f \in \mathscr{F}_i} \left( \bar{F}(x_f) - f_\star \right) \leq \alpha + \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathscr{F}_i} \sum_{j,l \in \textsc{pair}(k)} (f_j(x_{f_l}) - \bar{F}(x_{f_l}))^2} \,, \ \forall i \in [m] \,. \tag{4.1}$$

When $\xi(\mathscr{F}_i) = 0$ define $\nu_i$ 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, we have

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

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 $\nu_i$, then

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

`(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 $(\alpha, \beta mk(k-1)) \in \textsc{ir}(\xi)$. $\square$

# Chapter 5

# Approximate Thompson Sampling

The goal of this chapter is to prove Theorem 12, which shows that the regret of TS can be bounded using the generalized information ratio $\text{IR}_{\text{TS}}(\mathscr{F})$. Along the way, we introduce approximate Thompson sampling (ATS), which is a generalization of TS, and show that our results hold for ATS.

## 5.1 Approximate Minimization

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 approximately

```
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

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

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

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

When $\varepsilon_R = \varepsilon_O = 0$, then ATS and TS are equivalent, though we note the importance in our analysis that the ties in TS are broken in a consistent fashion—the optimization algorithm used to compute $\tilde{x}_f$ from $f$ must be deterministic. 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 $\{x : f_t(x) \leq f_{t\star} + \varepsilon\}$ for suitably small $\varepsilon \geq 0$. 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 $\{x : f(x) \leq f_\star + \varepsilon\}$, even if $\varepsilon$ is arbitrarily small.

## 5.2 A Convex Cover

The need for the cover in this section is subtle, and would be clear later in the proof of Theorem 12. We start by defining a kind of cover of a set of convex functions $\mathscr{F}$. In the standard analysis introduced by Bubeck et al. [2015] and Bubeck and Eldan [2018], this cover was defined purely in terms of the optimal action. As noticed by Lattimore [2021], this argument relies on $\mathscr{F}$ being closed under convex combinations, which is not true for some subsets of convex functions such as the space of convex ridge functions. Here we introduce a new notion of cover for function classes $\mathscr{F}$ that are not closed under convex combinations.

**Definition 6.** Let $\mathscr{F}$ be a set of convex functions from $\mathcal{K}$ to $\mathbb{R}$ and $\varepsilon > 0$. Define $N(\mathscr{F}, \varepsilon)$ to be the smallest number $N$ such that there exists $\{\mathscr{F}_1, \ldots, \mathscr{F}_N\}$ and points $\{x_1, \ldots, x_N\} \subset \mathcal{K}$, such that the following properties hold:

- *Closure:* For all $k \in [N]$, $\operatorname{conv}(\mathscr{F}_k) \subset \mathscr{F}$.
- *Approximation:* For all $f \in \mathscr{F}$, there exists a $k \in [N]$ such that $\|\tilde{x}_f - x_k\| \leq \varepsilon$, and $\inf_{g \in \mathscr{F}_k} \|f - g\|_\infty \leq \varepsilon$.

We now bound the covering number $N(\mathscr{F}, \varepsilon)$ 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 7.** *Suppose that* $\mathscr{F} = \mathscr{F}_{\texttt{bl}}$. *Then* $\log N(\mathscr{F}, \varepsilon) = O\left(d \log\left(\frac{\operatorname{diam}(\mathcal{K})}{\varepsilon}\right)\right)$.

*Proof.* Let $\mathcal{C}_K = x_1, \ldots, x_N$ be a finite subset of $\mathcal{K}$ such that for all $x \in \mathcal{K}$ there exists a $y \in \mathcal{C}_K$ with $\|x - y\| \leq \varepsilon$. Standard bounds on covering numbers [Artstein-Avidan et al., 2015, §4] show that $\mathcal{C}_K$ can be chosen so that

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

Given $x \in \mathcal{C}_K$ define $\mathscr{F}_x = \{f \in \mathscr{F} : \|\tilde{x}_f - x\| \leq \varepsilon\}$. We let $\mathscr{F}_i = \mathscr{F}_{x_i}, i \in [N]$, and show that the properties of Definition 6 hold for $\{\mathscr{F}_1, \ldots, \mathscr{F}_N\}$ and $\{x_1, \ldots, x_N\}$. Since $\operatorname{conv}(\mathscr{F}) = \mathscr{F}$ it follows trivially that $\operatorname{conv}(\mathscr{F}_x) \subset \operatorname{conv}(\mathscr{F}) = \mathscr{F}$. Suppose that $f \in \mathscr{F}$ is arbitrary and let $x_i \in \mathcal{C}_K$ be such that $\|x_i - \tilde{x}_f\| \leq \varepsilon$, where $i \in [N]$ exists by construction. Therefore, $f \in \mathscr{F}_i$ and the approximation property also holds. □

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

*Proof.* To begin, define $\varepsilon_{\mathbb{S}} = \varepsilon / \operatorname{diam}(\mathcal{K})$. 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 $x \in \mathcal{K}$ and $\theta \in \mathbb{S}_1$ let

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

Note that $\{f \in \mathscr{F} : \theta_f = \theta\}$ 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 $\|\theta - \eta\| \leq \varepsilon_{\mathbb{S}}$. Similarly, let $\mathcal{C}_K$ be a finite subset of $\mathcal{K}$ such that for all $x \in \mathcal{K}$ there exists a $y \in \mathcal{C}_K$ with $\|x - y\| \leq \varepsilon$. Classical covering number results [Artstein-Avidan et al., 2015, §4] show that $\mathcal{C}_{\mathbb{S}}$ and $\mathcal{C}_K$ can be chosen so that

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

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

$g = u_f(\langle \cdot, \theta \rangle) \in \mathscr{F}$, which satisfies

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

Therefore the approximation property holds. $\square$

## 5.3 Continuity of Regret and Information Gain

In order to find a pair $(\alpha, \beta)$ in $\mathrm{IR}(\mathscr{F})$, we need to bound the regret $\Delta(\pi, \xi)$ in terms of the information measure $\mathcal{I}(\pi, \xi)$ for a prior $\xi \in \mathscr{P}(\mathscr{F})$ and policy $\pi \in \mathscr{P}(\mathcal{K})$. It turns out useful to prove the Lipschitzness properties of these quantities with respect to both arguments. To make this precise, we choose an appropriate *distance* metric both for $\mathscr{P}(\mathscr{F})$ and $\mathscr{P}(\mathcal{K})$.

**Lemma 9.** *Suppose $F, G$ are random elements in $\mathscr{F}$ with laws $\xi$ and $\nu$, respectively, and that $X, Y \in \mathcal{K}$ are independent of $F$ and $G$ and have laws $\pi$ and $\rho$, respectively, all jointly distributed. Further, suppose that $\|F - G\|_\infty \leq \varepsilon$ and $\|X - Y\| \leq \varepsilon$ hold almost surely. Then*

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

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

*Proof.* For random variable $U$ let $\|U\|_{\mathrm{L}^2} = \mathbb{E}[U^2]^{1/2}$, and $\|U\|_{L^\infty} = \operatorname{ess\,sup}|U|$, which are norms on the space of square integrable random variables. Further, it holds that $\|U\|_{L^2} \leq \|U\|_{L^\infty}$.

Let $\bar{F} = \mathbb{E}[F]$ and $\bar{G} = \mathbb{E}[G]$, and note that

$$\left\|\bar{F} - \bar{G}\right\|_\infty = \|\mathbb{E}[F] - \mathbb{E}[G]\|_\infty = \|\mathbb{E}[F - G]\|_\infty \leq \mathbb{E}[\|F - G\|_\infty] \leq \varepsilon\,. \tag{5.1}$$

By definition $\mathcal{I}(\pi, \xi)^{1/2} = \|F(X) - \bar{F}(X)\|_{\mathrm{L}^2}$ and $\mathcal{I}(\pi, \nu)^{1/2} = \|G(X) - \bar{G}(X)\|_{\mathrm{L}^2}$. The first claim follows since

$$\begin{aligned}
\left|\mathcal{I}(\pi, \xi)^{1/2} - \mathcal{I}(\pi, \nu)^{1/2}\right| &= \left|\|F(X) - \bar{F}(X)\|_{\mathrm{L}^2} - \|G(X) - \bar{G}(X)\|_{\mathrm{L}^2}\right| \\
&\overset{\text{(a)}}{\leq} \|(F(X) - \bar{F}(X)) - (G(X) - \bar{G}(X))\|_{\mathrm{L}^2} \\
&\overset{\text{(b)}}{\leq} \left\|(F(X) - \bar{F}(X)) - (G(X) - \bar{G}(X))\right\|_{L^\infty} \\
&\overset{\text{(c)}}{\leq} \|F(X) - G(X)\|_{L^\infty} + \left\|\bar{F}(X) - \bar{G}(X)\right\|_{L^\infty} \\
&\overset{\text{(d)}}{\leq} \varepsilon + \varepsilon = 2\varepsilon\,,
\end{aligned}$$

where `(a)` follows from the reverse triangle inequality, `(b)` follows from the definition of $\|\cdot\|_{\mathrm{L}^2} \leq \|\cdot\|_{L^\infty}$, `(c)` follows from the triangle inequality, and `(d)` follows from Eq. (5.1) and the fact that

$$\|F(X) - G(X)\|_{L^\infty} \leq \left\|\sup_{x\in\mathcal{K}} |F(x) - G(x)|\right\|_{L^\infty} = \|\||F(x) - G(x)|\|_\infty\|_{L^\infty} \leq \varepsilon\,.$$

The second claim follows similarly, since

$$\begin{aligned}\left|\mathcal{I}(\pi,\xi)^{1/2} - \mathcal{I}(\rho,\xi)^{1/2}\right| &= \left|\|F(X) - \bar{F}(X)\|_{\mathrm{L}^2} - \|F(Y) - \bar{F}(Y)\|_{\mathrm{L}^2}\right| \\ &\overset{\text{(a)}}{\leq} \|(F(X) - \bar{F}(X)) - (F(Y) - \bar{F}(Y))\|_{\mathrm{L}^2} \\ &\overset{\text{(b)}}{\leq} \left\|(F(X) - \bar{F}(X)) - (F(Y) - \bar{F}(Y))\right\|_{L^\infty} \\ &\overset{\text{(c)}}{\leq} \|F(X) - F(Y)\|_{L^\infty} + \left\|\bar{F}(X) - \bar{F}(Y)\right\|_{L^\infty} \\ &\leq \varepsilon + \varepsilon = 2\varepsilon\,,\end{aligned}$$

where `(a)` follows from the reverse triangle inequality, `(b)` follows from the definition of $\|\cdot\|_{\mathrm{L}^2} \leq \|\cdot\|_{L^\infty}$, `(c)` follows from the triangle inequality, and `(d)` follows from the same argument as in `(d)` of the first claim and the Lipschitzness of $F$ and $\bar{F}$. $\square$

The point of the next lemma is to show that approximate minimization policies benefit from the same information ratio as exact minimization policies, plus a small additive term that depends on how well they approximate the minimization.

**Lemma 10.** *Suppose that $(\alpha,\beta) \in \mathrm{IR}_{\mathrm{TS}}(\mathscr{F})$. Suppose that $X$ and $F$ are jointly distributed (possibly dependent) random elements with laws $\pi \in \mathscr{P}(\mathcal{K})$ and $\nu \in \mathscr{P}(\mathscr{F})$, respectively, such that $F(X) \leq F_\star + \varepsilon$ almost surely. Then*

$$\Delta(\pi,\nu) \leq \alpha + \sqrt{\beta\mathcal{I}(\pi,\nu)} + \varepsilon\left[1 + 2\sqrt{\beta}\right]\,.$$

*Proof.* Let $G(x) = \max(F(x), F(X))$ and $\xi$ be the law of $G$, which means that $\pi$ is a TS policy for $\xi$, meaning that for any measurable $A \subseteq \mathcal{F}$,

$$\mathbb{P}(G \in A) = \mathbb{P}\left(X \in \bigcup_{f\in A} \operatorname*{arg\,min}_{x\in\mathcal{K}} f(x)\right)\,.$$

Now, from the assumption of the theorem that $(\alpha,\beta) \in \mathrm{IR}_{\mathrm{TS}}(\mathscr{F})$, we have $\Delta(\pi,\xi) \leq \alpha +$

$\sqrt{\beta \mathcal{I}(\pi, \xi)}$. By construction

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

almost surely. As usual, let $\bar{F} = \mathbb{E}[F]$ and $\bar{G} = \mathbb{E}[G]$. Putting these together we have

$$
\begin{aligned}
\Delta(\pi, \nu) &= \mathbb{E}[\bar{F}(X) - F_\star] \\
&\leq \mathbb{E}[\bar{G}(X) - G_\star] + \varepsilon \\
&= \Delta(\pi, \xi) + \varepsilon \\
&\leq \alpha + \sqrt{\beta I(\pi, \xi)} + \varepsilon \\
&\leq \alpha + \sqrt{\beta}(\sqrt{I(\pi, \nu)} + 2\varepsilon) + \varepsilon \,,
\end{aligned}
$$

where the last inequality follows from Lemma 9 (a) and the fact that $\|F - G\|_\infty \leq \varepsilon$. $\square$

The next lemma establishes basic properties of the regularized minimizers $\tilde{x}_f$ and $\bar{x}_f$, which are defined in Definition 5. Remember that $\varepsilon_R$ is the amount of regularization. Larger values make $\tilde{x}_f$ more stable but also a worse approximation to $x_f$. The optimization error in the definition of $\bar{x}_f$ is $\varepsilon_O$.

**Lemma 11.** *Suppose that $f \in \mathscr{F}$. Then*

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

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

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

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

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

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

and the result follows by rearranging, as

$$f(\tilde{x}_f) \leq f(y) + \frac{\varepsilon_R}{2}(\|y\|^2 - \|\tilde{x}_f\|^2) \leq f(y) + \frac{\varepsilon_R}{2} \operatorname{diam}(\mathcal{K})^2 ,$$

where the last inequality follows from the assumption that $0 \in \mathcal{K}$. $\square$

## 5.4 A Regret Bound in Terms of the Information Ratio

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

**Theorem 12.** *Suppose that $\varepsilon \in (0,1)$, $\frac{1}{2}\varepsilon_R \operatorname{diam}(\mathcal{K})^2 \leq \varepsilon$, $2\varepsilon_O/\varepsilon_R \leq \varepsilon^2$, and $(\alpha, \beta) \in \mathrm{IR}_{\mathrm{TS}}(\mathscr{F})$ with $\mathscr{F}$ be a set of convex functions from $\mathcal{K}$ to $[0,1]$. Then the Bayesian regret of* ATS *for any prior $\xi$ is at most*

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

Theorem 3 follows by choosing $\varepsilon = 1/n$ and $\varepsilon_R = \varepsilon_O = 0$ and by Proposition 7 and Proposition 8 to bound the covering numbers for the relevant classes.

*Proof.* Let $N = N(\mathscr{F}, \varepsilon)$ and $\mathscr{F}_1, \ldots, \mathscr{F}_N \subset \mathscr{F}$ together with $x_1, x_2, \ldots, x_N \in \mathcal{K}$ satisfy the conditions of Definition 6. Let $\xi \in \mathscr{P}(\mathscr{F})$ be any prior, and $F$ is sampled from $\xi$, then by Definition 6 there exists an $[N]$-valued random variable $\kappa$ such that:

(i) There exists a projection $\Pi : \mathscr{F} \to \mathscr{F}$ such that $\Pi(F) \in \mathscr{F}_\kappa$ and $\|F - \Pi(F)\|_\infty \leq \varepsilon$; and

(ii) $\|\tilde{x}_F - x_\kappa\| \leq \varepsilon$.

Further, suppose that $X$ is a random element in $\mathcal{K}$ such that $X \in \bar{x}_F$ and let $(X', F', \kappa)$ be an independent copy of $(X, F, \kappa)$. Also let $Y$ the observed cost by the learner, i.e., $Y = F(X)$ and $Y \in [0,1]$ almost surely. Define $\pi$ as the law of $X'$, which is an approximate Thompson sampling policy for $\xi$.

**STEP 1** We start by showing that

$$\Delta(\pi, \xi) \leq \Delta(\pi_\kappa, \xi_\kappa) + 5\varepsilon\,, \tag{5.2}$$

where $\pi_\kappa$ is the law of $x_\kappa$ and $\xi_\kappa$ is the law of $\mathbb{E}[\Pi(F)|\kappa]$. To see this, we can write

$$\begin{aligned}
\Delta(\pi, \xi) &= \mathbb{E}\left[F(X') - F(X_F)\right] \\
&\overset{\text{(a)}}{\leq} \mathbb{E}\left[F(\tilde{x}_{F'}) - F(\tilde{x}_F)\right] + 2\varepsilon \\
&\overset{\text{(b)}}{\leq} \mathbb{E}\left[F(x_{\kappa'}) - F(x_\kappa)\right] + 4\varepsilon \\
&\overset{\text{(c)}}{\leq} \mathbb{E}\left[\Pi(F)(x_{\kappa'}) - \Pi(F)(x_\kappa)\right] + 6\varepsilon \\
&\overset{\text{(d)}}{=} \mathbb{E}\left[\mathbb{E}[\Pi(F)(x_{\kappa'})|\kappa] - \mathbb{E}[\Pi(F)(x_\kappa)|\kappa]\right] + 6\varepsilon \\
&\overset{\text{(e)}}{=} \mathbb{E}\left[\mathbb{E}[\Pi(F)|\kappa](x_{\kappa'}) - \mathbb{E}[\Pi(F)|\kappa](x_\kappa)\right] + 6\varepsilon \\
&= \Delta(\pi_\kappa, \xi_\kappa) + 6\varepsilon\,,
\end{aligned}$$

where `(a)` follows from the fact that $F(\tilde{x}_F) \leq F(X_F) + \frac{\varepsilon_R}{2}\operatorname{diam}(\mathcal{K})^2$ by Lemma 11, Lipschitzness of $F$ and $\|X' - \tilde{x}_{F'}\| \leq \varepsilon$; `(b)` follows from the fact that $\|\tilde{x}_{F'} - x_{\kappa'}\| \leq \varepsilon$ and $\|\tilde{x}_F - x_\kappa\| \leq \varepsilon$; `(c)` follows from the fact that $\|F - \Pi(F)\|_\infty \leq \varepsilon$; `(d)` follows from the tower rule; and `(e)` follows from the independence of $(X', F', \kappa)$ and $(X, F, \kappa)$.

**STEP 2** Next, we use that $(\alpha, \beta) \in \mathrm{IR}_{\mathrm{TS}}(\mathscr{F})$ to bound $\Delta(\pi_\kappa, \xi_\kappa)$, as

$$\Delta(\pi_\kappa, \xi_\kappa) \leq \alpha + \sqrt{\beta \mathcal{I}(\pi_\kappa, \xi_\kappa)}\,, \tag{5.3}$$

Note that this step is true since $\mathbb{E}[\pi(F)|\kappa] \in \mathscr{F}$, which is a direct consequence of Definition 6`(i)`.

**STEP 3** We now bound the $\mathcal{I}(\pi_\kappa, \xi_\kappa)$ term. Let $\bar{\xi}_\kappa$ be the law of $\mathbb{E}[F|\kappa]$. We have

$$\begin{aligned}
\sqrt{\mathcal{I}(\pi_\kappa, \xi_\kappa)} &\overset{\text{(a)}}{\leq} \sqrt{\mathcal{I}(\pi_\kappa, \bar{\xi}_\kappa)} + 2\varepsilon \\
&\overset{\text{(b)}}{\leq} \sqrt{\mathcal{I}(\pi, \bar{\xi}_\kappa)} + 6\varepsilon \\
&\overset{\text{(c)}}{\leq} \sqrt{\tfrac{1}{2} I(\kappa; X', Y)} + 6\varepsilon,
\end{aligned} \tag{5.4}$$

where `(a)` follows from Lemma 9`(a)` and since

$$\|\mathbb{E}[F|\kappa] - \mathbb{E}[\Pi(F)|\kappa]\|_\infty = \|\mathbb{E}[F - \Pi(F)|\kappa]\|_\infty \leq \|F - \Pi(F)\|_\infty \leq \varepsilon\,;$$

(b) follows from Lemma 9 (b) and the fact that

$$\|X' - X_{\kappa'}\| \leq \|X' - \tilde{x}_{F'}\| + \|\tilde{x}_{F'} - x_{\kappa'}\| \leq 2\varepsilon\,;$$

and (c) follows from

$$\begin{aligned}
\mathcal{I}(\pi, \bar{\xi}_\kappa) &= \mathbb{E}\left[(\mathbb{E}[F|\kappa](X') - \mathbb{E}[F](X'))^2\right] \\
&= \mathbb{E}\left[(\mathbb{E}[F(X')|\kappa] - \mathbb{E}[F(X')|X'])^2\right] \\
&= \mathbb{E}\left[(\mathbb{E}[Y|\kappa, X'] - \mathbb{E}[Y|X'])^2\right] \\
&\leq \tfrac{1}{2}\mathbb{E}\left[D_{KL}\left(\mathbb{P}_{Y|X'}, \mathbb{P}_{Y|X',\kappa}\right)\right] \\
&= \tfrac{1}{2} I(\kappa; X', Y)\,,
\end{aligned}$$

where the first inequality follows from the Pinsker's inequality.

**STEP 4** By putting Eqs. (5.2) to (5.4) together we have

$$\begin{aligned}
\Delta(\pi, \xi) &\leq \Delta(\pi_\kappa, \xi_\kappa) + 6\varepsilon && \text{(Eq. (5.2))} \\
&\leq \alpha + \sqrt{\beta \mathcal{I}(\pi_\kappa, \xi_\kappa)} + 6\varepsilon && \text{(Eq. (5.3))} \\
&\leq \alpha + \sqrt{\beta}\left(\sqrt{\tfrac{1}{2} I(\kappa; X', Y)} + 6\varepsilon\right) + 6\varepsilon && \text{(Eq. (5.4))} \\
&\leq \alpha + \sqrt{\tfrac{\beta}{2} I(\kappa; X', Y)} + 6\varepsilon\left(1 + \sqrt{\beta}\right)\,, && (5.5)
\end{aligned}$$

We are now ready to prove Theorem 12. Let $\pi_t$ be the law of $X_t$ under $\mathbb{P}_{t-1}$ and $\xi_t$ be the law of $F$ under $\mathbb{P}_{t-1}$, and $I_t$ be the mutual information with respect to probability measure $\mathbb{P}_t$. Note that $\kappa$ is jointly distributed with $F$. Then we have

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

where `(a)` follows from Eq. (5.5); `(b)` follows from Cauchy-Schwarz inequality; and `(c)` holds by the chain rule for the mutual information and because $\kappa \in [N]$ and hence its entropy is at most $\log(N)$. $\square$

# 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 13.** *When* $d = 1$, $\sup_{\xi \in \mathscr{P}(\mathscr{F}_{b1})} \mathit{BReg}_n(\mathrm{TS}, \xi) = O\left(\sqrt{n \log(n) \log(n \operatorname{diam}(\mathcal{K}))}\right)$.

Theorem 13 follows by combining a bound on the information ratio established next with Theorem 3. Interestingly, the following theorem doesn't need the Lipschitzness property.

**Theorem 14.** *Suppose that* $d = 1$ *and* $\alpha \in (0, 1)$. *Then* $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \mathrm{IR}(\mathscr{F}_{b1})$.

*Proof.* Fix $\xi \in \mathcal{P}(\mathscr{F}_{b1})$ and let $\pi$ be the TS policy for $\xi$. As usual, let $\bar{F} = \mathbb{E}[F]$. Our plan is to use Lemma 4. For this we define a partitioning of $\mathscr{F}_{b1}$. In particular, for $f \in \mathscr{F}_{b1}$, $f$ is sorted into one of $\mathscr{F}_i$ defined below based on $\bar{F}(x_f) - f_\star$ and $x_f$: If $\bar{F}(x_f) - f_\star \leq \alpha$, the $f \in \mathscr{F}_0$, otherwise $f \in \mathscr{F}_i$ where $i \in \mathbb{Z}$ is so that $\bar{F}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}]$, and $i < 0$ when $x_f < x_{\bar{F}}$ and $i > 0$ otherwise. Precisely, define

$$\mathscr{F}_i = \begin{cases} \{f \in \mathscr{F}_{b1} : \bar{F}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\, x_f \geq x_{\bar{F}}\}\,, & \text{if } i > 0\,; \\ \{f \in \mathscr{F}_{b1} : \bar{F}(x_f) - f_\star \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\, x_f < x_{\bar{F}}\}\,, & \text{if } i < 0\,; \\ \{f \in \mathscr{F}_{b1} : \bar{F}(x_f) - f_\star \leq \alpha\}\,, & \text{if } i = 0\,. \end{cases} \tag{6.1}$$

Clearly, since by assumption $\bar{F}(x_f), f_\star \in [0, 1]$ for any $f \in \mathcal{F}_{b1}$, for $|i| > m = \lceil \log_2(1/\alpha) \rceil$, $\mathscr{F}_i = \emptyset$. Hence, $\mathscr{F}_{b1} = \cup_{i=-m}^{m} \mathscr{F}_i$. In a moment we will show that with $k = 4$ and $-m \leq i \leq m$

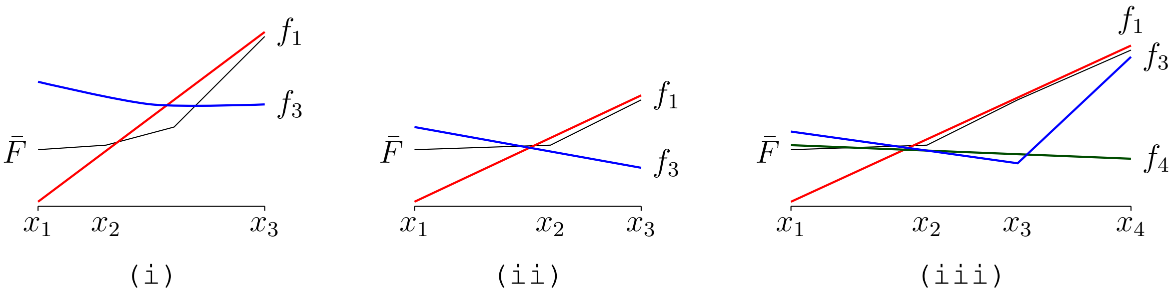



**Figure 6.1:** `(i)` shows that if $x_2$ is too close to $x_1$, then $f_1(x_3)$ must be large, which implies that $f_3(x_3)$ must be large and so too must $f_3(x_1)$, which shows that $f_3(x_1) - \bar{F}(x_1)$ is large. `(ii)` shows what happens if $f_3(x_3)$ is too far below $\bar{F}(x_1)$, which is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

and $\varepsilon_i = \alpha 2^{|i|}$,

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

holds. Hence, by Lemma 4 and naive simplification of constants $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \text{IR}(\mathscr{F}_{\text{bl}})$ as desired. The first inequality in Eq. (6.2) is an immediate consequence of the definition of $\mathscr{F}_i$ and $\varepsilon_i$. 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 $i > 0$. Fix such an $i > 0$ and let $\varepsilon = \varepsilon_i = \alpha 2^i$. Let $k = 4$, $f_1, \ldots, f_4 \in \mathscr{F}_i$, and $x_j = x_{f_j}$ and assume without loss of generality that $x_1 \le x_2 \le x_3 \le x_4$. Note that $x_{\bar{F}} \le x_1$ also hods because $i > 0$. It suffices to show that $\sum_{j,l \in \text{PAIR}(4)} (f_j(x_{f_l}) - \bar{F}(x_{f_l}))^2 < c^2 \varepsilon_i^2$ for a suitable universal constant $c$. We prove this by contradiction. That is, assume

$$\sum_{j,l \in \text{PAIR}(4)} (f_j(x_{f_l}) - \bar{F}(x_{f_l}))^2 < c^2 \varepsilon^2 \qquad \text{where we take} \qquad c = \frac{\sqrt{65} - 7}{16}, \tag{6.3}$$

for reasons that become clear later. We establish a contradiction in three steps. The main argument in each step is illustrated in Figure 6.1.

**STEP 1** We start by showing that $x_2$ cannot be too far from $x_3$. In particular, writing $x_2 = (1 - p)x_1 + px_3$, we show that $p \ge 0.27$. First, we have

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

where (a) follows from Eq. (6.3); (b) holds since for $f \in \mathscr{F}_i$, $\bar{F}(x_f) - f(x_f) \leq \varepsilon$ and $f_3 \in \mathscr{F}_i$; and (c) follows from Eq. (6.3) again. Hence,

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

where (a) follows because $\bar{F}$ is non-decreasing on $[x_1, x_4]$ because $x_f \leq x_1$ as discussed before; (b) follows from Eq. (6.3); (c) holds by convexity and the definition of $p$; (d) follows since $f_1(x_1) < \bar{f}_1(x_1) - \varepsilon/2$ by the definition of $\mathscr{F}_i$ and (e) is true by the previous display. Therefore $p \geq (1/2 - c)/(2c + 3/2) \approx 0.27$.

**STEP 2** Having shown that $x_2$ cannot be far from $x_3$, we now show that $\bar{F}(x_1)$ is not much larger than $f_3(x_3)$. Indeed,

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

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

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

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

where (a) follows by convexity and because $f_4$ is minimized at $x_4$, (b) from Eq. (6.3), (c) from **STEP 2**, (d) since $f_3(x_3) < \bar{F}(x_3) - \varepsilon/2$ 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. $\square$

*Proof of Theorem 13.* Use Theorem 14 with $\alpha = 1/n$ and then Theorem 3. $\square$

# 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 15.** *The following holds:* $\sup_{\xi \in \mathscr{P}(\mathscr{F}_{blrm})} \mathit{BReg}_n(\mathrm{TS}, \xi) = O\left(d^{2.5}\sqrt{n}\log(nd\,\mathrm{diam}(\mathcal{K}))^2\right)$.

The class of convex ridge functions is an extension of the class of linear functions, which have been studied extensively in the literature. Russo and Van Roy [2016] used information-theoretic means to show that for linear bandits the regret is at most $\tilde{O}(d\sqrt{n})$. Lattimore [2021] showed that for (possibly non-monotone) convex ridge functions a version of IDS has Bayesian regret at most $\tilde{O}(d\sqrt{n})$. The downside is that IDS is often not implementable in practice, even given efficient access to posterior samples, whereas Thompson Sampling only needs to minimize a convex function given an efficient sampling oracle from the posterior. Like Theorem 13, Theorem 15 is established by combining a bound on the information ratio with Theorem 3.

We first provide two lemmas that are used in the proof.

**Theorem 16** (John's Theorem, Todd [2016] Theorem 1.1)**.** *Let $X \subset \mathbb{R}^d$ be a convex body, and $E$ be the ellipsoid of maximal volume contained in $X$. Then $X \subseteq dE$.*

**Theorem 17** (Khachiyan volume bound Khachiyan [1990])**.** *Let $E = \{x \in \mathbb{R}^d : (x-\mu)^\top \Sigma^{-1}(x-\mu) \leq 1\}$ and let $H = \{x : a^\top(x-\mu) \leq 0\}$ be a halfspace that contains the center $\mu$ of $E$ and splits $E$ into two parts $E^+ = E \cap H$ and $E^- = E \cap H^c$. Let $V^+$ and $V^-$ be the volumes of the maximal volume ellipsoids contained in $E^+$ and $E^-$, respectively. Then it holds that* $\max(V^+, V^-) \leq 0.85\,\mathrm{vol}(E)$.

**Theorem 18.** *For any* $\alpha \in (0,1)$ *and*

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

*it holds that* $(\alpha, \beta\lceil\log(1/\alpha)\rceil) \in \mathrm{IR}(\mathscr{F}_{\texttt{blrm}})$ *with the Big-O hiding only a universal constant.*

*Proof.* Abbreviate $\mathscr{F} = \mathscr{F}_{\texttt{blrm}}$. The high-level argument follows the proof of Theorem 14—first using Lemma 4 to partition the class $\mathscr{F}_{\texttt{blrm}}$ and then using convexity of the functions to lower bound the information gain. Let $\bar{f} \in \mathrm{conv}(\mathscr{F})$.

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

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

To be able to apply Lemma 4, the decomposition lemma, we will show that for all $0 \le i \le \lceil\log_2(1/\alpha)\rceil$, for $\varepsilon_i = \alpha 2^i$ and $f_1, \ldots, f_k \in \mathscr{F}_i$,

$$\sup_{f \in \mathscr{F}_i} \left(\bar{f}(x_f) - f_\star\right) \le \varepsilon_i \le \alpha + \sqrt{512 \sum_{(j,l) \in \mathrm{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,, \tag{7.1}$$

with

$$k = 9d\left[1 + 2d + 8d\left\lceil\log\left(\frac{2d\,\mathrm{diam}(\mathcal{K})}{\alpha}\right)\right\rceil\right] = O\left(d^2 \log\left(\frac{d\,\mathrm{diam}(\mathcal{K})}{\alpha}\right)\right).$$

Then Lemma 4 will imply that $(\alpha, 512k(k-1)(1 + \lceil\log(1/\alpha)\rceil)) \in \mathrm{IR}(\mathscr{F})$, as required.

It remains to verify Eq. (7.1). The first inequality in Eq. (7.1) follows immediately from the definition of $\varepsilon_i$ and $\mathscr{F}_i$. The second is also immediate when $i = 0$ by the definition of $\mathscr{F}_i$. Suppose now that $i > 0$, and to reduce the clutter let $\varepsilon = \varepsilon_i$. 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 non-increasing. The second inequality in Eq. (7.1) holds immediately if $f_j = f_l$ for some $(j,l) \in \mathrm{PAIR}(k)$ since then $f_j(x_{f_l}) = f_l(x_{f_l}) \le \bar{f}(x_{f_l}) - \varepsilon/2$. Therefore, we can focus on the case where $f_j \ne f_l$ for all $(j,l) \in \mathrm{PAIR}(k)$. We consider two cases:

**CASE 1** Large drop of $\bar{f}$: $\bar{f}(x_{f_1}) \ge \bar{f}(x_{f_k}) + 2\varepsilon$. In this case $f_1(x_{f_k}) \ge f_1(x_{f_1}) \ge \bar{f}(x_{f_1}) - \varepsilon \ge \bar{f}(x_{f_k}) + \varepsilon$, which shows that $(f_1(x_{f_k}) - \bar{f}(x_{f_k}))^2 \ge \varepsilon^2$ and the second inequality in Eq. (7.1) holds.

**CASE 2** Small drop of $\bar{f}$: $\bar{f}(x_1) < \bar{f}(x_k) + 2\varepsilon$. Let $\delta = \frac{\varepsilon}{4d}$. Let $b = 9d$ and $\mathcal{C}_1, \ldots, \mathcal{C}_b$ be formed by dividing $\{f_1, \dots, f_k\}$ in order into $b$ blocks of equal size. Let $s_a = \max_{f,g \in \mathcal{C}_a} |\bar{f}(x_f) - \bar{f}(x_g)|$. Given the conditions of the case we have $2\varepsilon > \sum_{a=1}^{b} s_a \geq \sum_{a=1}^{b} \delta \mathbf{1}(s_a > \delta)$, which means that $\sum_{a=1}^{b} \mathbf{1}(s_a > \delta) < 2\varepsilon/\delta \leq 8d = b - d$. Hence, there exist at least $d$ blocks $\mathcal{C}_a$ for which $s_a \leq \delta$. For each of these blocks, we have

$$s_a = \max_{f,g \in \mathcal{C}_a} |\bar{f}(x_f) - \bar{f}(x_g)| \leq \delta \,,$$

which intuitively means that the value of $\bar{f}$ is approximately constant within each block $\mathcal{C}_a$. Next, we exploit this property to lower bound the quadratic (information gain) term within each of these blocks. To this end, we use an iterative process based on the method of inscribed ellipsoid for optimization [Tarasov et al., 1988], where we use shrinking ellipsoids to show that the information gain is sufficiently large (for a pair $f_j, f_l \in \mathcal{C}_a$) unless the volume of certain ellipsoid (defined later) shrinks significantly.

Given a nonempty finite set $\mathcal{C} \subset \mathscr{F}$ and $\delta > 0$, let $J_\delta(\mathcal{C}) = \operatorname{conv}(\cup_{g \in \mathcal{C}} \mathbb{B}_\delta(x_g)) \subseteq \mathbb{R}^d$. Moreover, let $E_\delta(\mathcal{C})$ be the ellipsoid of maximum volume enclosed in $J_\delta(\mathcal{C})$ (this is known as John's ellipsoid John [1948]). We now state two lemmas. The first shows that for a suitable subset $\mathcal{C}$ of loss functions either the information gain is reasonably large or one can find a function $f$ in $\mathcal{C}$ such that the ellipsoid $E_\delta(\mathcal{C} \setminus \{f\})$ has a considerably smaller volume than $E_\delta(\mathcal{C})$.

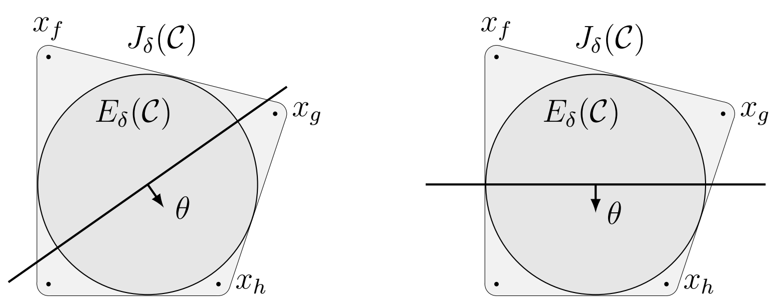



**Figure 7.1:** The two cases considered in the proof of Lemma 19. In the left figure, the situation is such that $E_\delta(\mathcal{C} \setminus \{f\})$ is a constant fraction less volume than $E_\delta(\mathcal{C})$. On the other hand, in the figure on the right, one of $(f(x_h) - \bar{f}(x_h))^2$ or $(f(x_g) - \bar{f}(x_g))^2$ must be reasonably large.

**Lemma 19.** *Let $\bar{f} \in \operatorname{conv}(\mathscr{F})$. Let $\varepsilon > 0$, $\delta = \frac{\varepsilon}{4d}$ and $\mathcal{C} \subset \mathscr{F}$ be a nonempty finite set such that for all $f, g \in \mathcal{C}$, $\bar{f}(x_f) - f_\star \in [\varepsilon/2, \varepsilon]$ and $|\bar{f}(x_f) - \bar{f}(x_g)| \leq \delta$. Pick any $f \in \mathcal{C}$. Then at least one of the following holds:*

(i) *We have* $\operatorname{vol}(E_\delta(\mathcal{C} \setminus \{f\})) \leq 0.85 \operatorname{vol}(E_\delta(\mathcal{C}))$.

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

*Proof.* Let $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ 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 assumption there exists a convex non-decreasing function $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{S}_1$ such that $f = \ell(\langle \cdot, \theta \rangle)$. Let $H = H(\mu, \theta)$, which is the half-space passing through the center of John's ellipsoid $E_\delta(\mathcal{C})$ with inward-facing normal $\theta$. Consider the following cases, illustrated in Figure 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 Theorem 17 shows that $\mathrm{vol}(E_\delta(\mathcal{C} \setminus \{f\})) \leq 0.85 \, \mathrm{vol}(E_\delta(\mathcal{C}))$.

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

$$\langle x_g, \theta \rangle < \langle \mu, \theta \rangle + \delta \, . \tag{7.2}$$

We now show that for some $h \in \mathcal{C}$, $\langle x_h, \theta \rangle \geq \langle \mu, \theta \rangle + \|\theta\|_{\Sigma^{-1}}^{-1} - \delta$. Let $\overline{x} = \arg\max_{z \in E_\delta(\mathcal{C})} \langle z, \theta \rangle$, which has the closed form of $\overline{x} = \mu + \frac{\Sigma\theta}{\|\theta\|_\Sigma}$. Since $J_\delta(\mathcal{C}) = \mathrm{conv}(\cup_{g \in \mathcal{C}} \mathbb{B}_\delta(x_g))$, $x = \sum_{g \in \mathcal{C}} \lambda_g x'_g$ for some $x'_g \in \mathbb{B}_\delta(x_g)$, with $\lambda_g \geq 0$ and $\sum_{g \in \mathcal{C}} \lambda_g = 1$. Since $\langle \cdot, \theta \rangle$ is linear, there exists $h \in \mathcal{C}$ such that $\langle x'_h, \theta \rangle = \langle \overline{x}, \theta \rangle$. Further, since $x'_h \in \mathbb{B}_\delta(x_h)$ and $\|\theta\| = 1$, we have $\langle x'_h, \theta \rangle \geq \langle x_h, \theta \rangle - \delta$. Therefore,

$$\langle x_h, \theta \rangle \geq \langle x'_h, \theta \rangle - \delta \geq \langle \overline{x}, \theta \rangle - \delta = \langle \mu, \theta \rangle + \|\theta\|_\Sigma - \delta \, . \tag{7.3}$$

Next, we show that $\langle x'_f, \theta \rangle \geq \langle \mu, \theta \rangle - d \, \|\theta\|_\Sigma$. By Theorem 16, we have $J_\delta(\mathcal{C}) \subseteq E_{\delta,d}(\mathcal{C})$. Let $\underline{x} = \arg\min_{z \in E_{\delta,d}(\mathcal{C})} \langle z, \theta \rangle$, which has the closed form of $\underline{x} = \mu - \frac{\Sigma\theta}{\|\theta\|_\Sigma}$. Note that $J_\delta(\mathcal{C}) \subseteq E_{\delta,d}(\mathcal{C})$, so $x'_f \in E_{\delta,d}(\mathcal{C})$ and $\langle \underline{x}, \theta \rangle \leq \langle x'_f, \theta \rangle$. This together with $x'_f \in \mathbb{B}_\delta(x_f)$ implies that

$$\langle x_f, \theta \rangle \geq \langle x'_f, \theta \rangle - \delta \geq \langle \underline{x}, \theta \rangle - \delta = \langle \mu, \theta \rangle - d \, \|\theta\|_\Sigma - \delta \, . \tag{7.4}$$

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

$$\begin{aligned}
f(x_g) &\overset{\text{(a)}}{\leq} \delta + f(x'_g) \overset{\text{(b)}}{=} \delta + f\left( \frac{\langle x'_g - x'_f, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_h + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} x'_f \right) \\
&\overset{\text{(c)}}{\leq} \delta + f(x'_h) + \frac{\langle x'_h - x'_g, \theta \rangle}{\langle x'_h - x'_f, \theta \rangle} (f(x'_f) - f(x'_h)) \\
&\overset{\text{(d)}}{\leq} \delta + f(x'_h) + \frac{1}{d}(f(x'_f) - f(x'_h)) \overset{\text{(e)}}{\leq} 2\delta + f(x_h) + \frac{1}{d}(f(x_f) - f(x_h)) \, ,
\end{aligned} \tag{7.5}$$

where (a) follows because $f$ is a 1-Lipschitz ridge function, and because $|\langle x_g - x'_g, \theta\rangle| = \delta$; (b) holds by definitions and the fact that $f(\cdot) = \ell(\langle \cdot, \theta\rangle)$; (c) by convexity of $f$; (d) follows from Eqs. (7.2) to (7.4) and because $f(x'_f) \leq f(x'_h)$; and (e) uses again that $f$ is a 1-Lipschitz ridge function, Lemma 2, and that $|\langle x'_h - x_h, \theta\rangle| = \delta$. If $f(x_h) \geq \bar{f}(x_h) + \delta$, then $(f(x_h) - \bar{f}(x_h))^2 \geq \delta^2$ and (ii) holds and we are done. So it remains to consider the case when $f(x_h) < \bar{f}(x_h) + \delta$. Then

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

where (a) follows from Eq. (7.5), (b) follows from on the one hand from our assumption that $f(x_f) \leq \bar{f}(x_f) - \varepsilon/2 \leq \bar{f}(x_h) + \delta - \varepsilon/2$, and on the other hand, that $f(x_h) < \bar{f}(x_h) + \delta$ holds thanks to the case we are considering; (c) follows by calculations, (d) by the definition of $\varepsilon$ and $\delta$ and (e) by the assumptions of 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, finishing the proof.

□

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 $\mathcal{C}$, $f$, and $\varepsilon$ satisfy the conditions of Lemma 19 for some $\varepsilon \geq \alpha \geq 0, \delta = \frac{\varepsilon}{4d}$ and $|\mathcal{C}| \geq 2 + 2d + 14d \left\lceil \log\left(\frac{2d\,\mathrm{diam}(\mathcal{K})}{\alpha}\right)\right\rceil$. Then*

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

*Proof.* Define a sequence $(\mathcal{C}_k)_{k\geq 1}$ of sets as follows. Let $\mathcal{C}_1 = \mathcal{C}$ and $m$ be such that $2m - 1 \leq |\mathcal{C}| \leq 2m$. Note that $\frac{|\mathcal{C}|}{2} \leq m \leq \frac{|\mathcal{C}|+1}{2}$. 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. It suffices to show that in this process (ii) happens at least $d$ times. By definition $|\mathcal{C}_1| \geq 2m - 1$ and since $|\mathcal{C}_{k+1}| \geq |\mathcal{C}_k| - 2$, $|\mathcal{C}_m| \geq |\mathcal{C}_1| - 2(m-1) \geq 1$.

Recall that by Theorem 16 $E_\delta(\mathcal{C}_m) \subset J_\delta(\mathcal{C}_m) \subset E_{\delta,d}(\mathcal{C}_m)$, which means that

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

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

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

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

$$\tau \leq \frac{d\log\left(\frac{-\delta}{d(\mathrm{diam}(\mathcal{K})+\delta)}\right)}{\log(0.85)} \leq 7d\log\left(\frac{2d\,\mathrm{diam}(\mathcal{K})}{\delta}\right) \leq \frac{|\mathcal{C}|}{2} - d - 1\,.$$

Note that `(ii)` happens $s = m - \tau - 1$ times in the process produces $\mathcal{C}_m$ from $\mathcal{C}_1$. Using the upper bound on $\tau$, we get

$$s \geq m - 1 - \left(\tfrac{|\mathcal{C}|}{2} - d - 1\right) \geq m - 1 - \tfrac{2m}{2} + d + 1 = d\,.$$

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

To summarize, with $k = 9d\left[1 + 2d + 8d\left\lceil\log\left(\frac{2d\,\mathrm{diam}(\mathcal{K})}{\alpha}\right)\right\rceil\right]$, at least $d$ many of the blocks $\mathcal{C}_a$ satisfy the conditions of Lemma 20 and therefore

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

where the last inequality holds by Lemma 20, finishing the proof. □

# 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 optimization 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. This family is based on constructing a convex function $f_\theta$ for each $\theta \in \mathbb{S}_1$ such that $f_\theta$ is minimized at $\theta$ with minimum value of $0$, and $f_\theta$ exactly agrees with the upward-shifted quadratic function outside of a ball of radius $\sqrt{1+2\varepsilon}$ centered at $\theta$, where $\varepsilon$ is the minimum value the upward-shifted quadratic function takes, which is realized at $x = 0$.

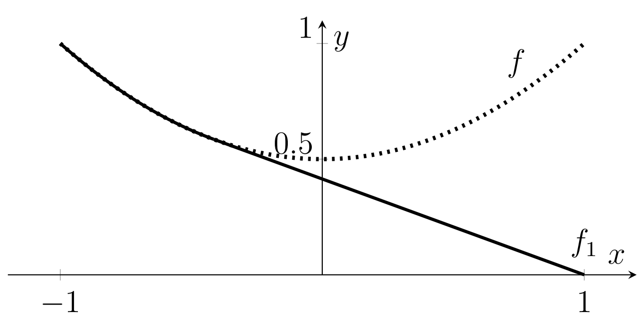



**Figure 8.1:** Illustration of the function $f_\theta$ for $\varepsilon = 0.5$ and $\theta = 1$ in 1 dimension. It can be seen that $f_\theta$ is convex, minimized at $\theta$ and lies below $f$.

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

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

*Then* $f_\theta$ *is convex, is minimized at* $\theta$*,* $\operatorname{Lip}_{\mathbb{B}_1}(f_\theta) \leq \sqrt{2+2\varepsilon}$*, and* $f_\theta(x) \leq f(x)$ *for all* $x \in \mathbb{R}^d$*.*

The function $f_\theta$ arises naturally as the largest convex function for which both $f_\theta(\theta) = 0$ and $f_\theta(x) \leq f(x)$ for all $x \in \mathbb{R}^d$. The 1 dimensional version of this function is illustrated in Figure 8.1. While for our later argument it suffices to consider $f_\theta$ only on $\mathbb{B}_1$, in Lemma 21 we consider the $\mathbb{R}^d$ as the domain for generality.

*Proof of Lemma 21.* We need to prove that $f_\theta$ is convex, minimized at $\theta$, $f_\theta(x) \leq f(x)$ for all $x \in \mathbb{R}^d$, and when restricted to $\mathbb{B}_1$, $f_\theta$ is $\sqrt{2+2\varepsilon}$-Lipschitz. That $f_\theta(x) \leq f(x)$ for all $x \in \mathbb{R}^d$ and that $f_\theta$ is convex follows because

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

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

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

Therefore $f_\theta$ is the supremum over a set of linear functions and hence convex.

That $f_\theta$ is minimized at $\theta$ follows directly from the first-order optimality conditions. To see this, let $\eta \in \mathbb{S}_1$. Then

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

where $Df_\theta(\theta)[\cdot]$ is the directional derivative operator (noting that $f_\theta$ is not differentiable at $\theta$). Hence, only 0 is a subgradient of $f_\theta$ at $\theta$.

Lastly, for the Lipschitzness of $f_\theta$ on $\mathbb{B}_1$: since $f_\theta$ is continuous, it suffices to bound $\|f'_\theta(\cdot)\|$ on $\operatorname{int}(\mathbb{B}_1)$ where $f_\theta$ is differentiable. When $\|x-\theta\|^2 \geq 1+2\varepsilon$, then $\|f'_\theta(x)\| = \|f'(x)\| = \|x\| \leq 1$. On the other hand, if $\|x-\theta\|^2 < 1+2\varepsilon$, then

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

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

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

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

We first give a quick sketch of the proof. We will construct a prior such that with high probability TS obtains limited information while suffering high regret. We assume there is no noise and let $f$ and $f_\theta$ be defined as in Lemma 21 with $\varepsilon = 1/4$. Let $\sigma$ be the uniform probability measure on $\mathbb{S}_1$ and the prior $\xi$ be the law of $f_\theta$ when $\theta$ is sampled from $\sigma$. By the definition of $f_\theta$ and the fact that $\varepsilon = 1/4$, for any $x \in \mathbb{S}_1$, $f(x) = f_\theta(x)$ holds if and only if $\langle x, \theta \rangle \leq \frac{1}{4}$. Since TS plays the minimizer of some $f_\theta$ in every round, it follows that TS always plays in $\mathbb{S}_1$. Let $\mathcal{C}_\theta = \{x \in \mathbb{S}_1 : \langle x, \theta \rangle > \frac{1}{4}\}$ and $\delta = \sigma(\mathcal{C}_\theta)$. Let $f_{\theta_\star}$ be the true loss function sampled from $\xi$. Suppose that $X_1, \ldots, X_t \in \mathbb{S}_1 \setminus \mathcal{C}_{\theta_\star}$, which means that $Y_s = f(X_s) = \frac{3}{4}$ 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 $\delta \leq \exp(-d/32)$, which follows from the upper bound on the area of spherical caps on sphere [Tkocz, 2018, Theorem B.1].

We now turn to the detailed argument. We start with some definitions.

**Definition 23** (Spherical caps)**.** For $\theta \in \mathbb{S}_1$ and $\varepsilon \in [0, 1/2]$, define

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

and $\bar{\mathcal{C}}_{\theta,\varepsilon} = \mathbb{S}_1 \setminus \mathcal{C}_{\theta,\varepsilon}$. Note that $\mathcal{C}_{\theta,\varepsilon} = \{x \in \mathbb{S}_1 : \langle \theta, x\rangle > \frac{1}{2} - \varepsilon\}$.

**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 [Tkocz, 2018] to bound the surface area of spherical caps.

**Theorem 25.** *[Tkocz, 2018, Theorem B.1] For all $\varepsilon \in [0, 1]$ and $\theta \sim \sigma$ we have*

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

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

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

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

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

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

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

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

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

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

**Step 1:** First we show that if the posterior distribution $\sigma_t$ at round $t \in [T]$ is uniform over $\Theta_t$, and the algorithm observes the loss $Y_t = \frac{3}{4}$ as a result of playing $X_t$, then the posterior distribution $\sigma_{t+1}$ at round $t+1$ is uniform over $\Theta_{t+1}$, i.e., $\sigma_{t+1} = \sigma_{\Theta_{t+1}}$. To this end, recall that $\mathbb{P}_{t-1}(\cdot) = \mathbb{P}(\cdot|X_1, Y_1, \ldots, X_{t-1}, Y_{t-1})$, and observe that if $Y_t = \frac{3}{4}$ then for any set $B \subseteq \Theta_t$,

$$\begin{aligned}
\sigma_{t+1}(B) = \mathbb{P}_{t-1}\left(\theta_\star \in B | Y_t = \frac{3}{4}, X_t\right) &= \frac{\mathbb{P}_{t-1}(\theta_\star \in B, Y_t = \frac{3}{4}|X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t)} \\
&= \frac{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t, \theta_\star \in B)\mathbb{P}_{t-1}(\theta_\star \in B|X_t)}{\mathbb{P}_{t-1}(Y_t = \frac{3}{4}|X_t)} \\
&= \frac{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4}|X_t, \theta_\star \in B)\mathbb{P}_{t-1}(\theta_\star \in B)}{\mathbb{P}_{t-1}(f_{\theta_\star}(X_t) = \frac{3}{4}|X_t)} . \qquad (8.2)
\end{aligned}$$

Note that TS samples $f_{\theta_t}$ from $\xi_t$, and then plays the minimizer of $f_{\theta_t}$, which from Lemma 21 is $\theta_t$, 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

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

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

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

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

**Step 2:** Let $\delta = \sigma(\mathcal{C}_{\theta_\star})$, and note that $\sigma(\mathcal{C}_\theta) = \delta$ for all $\theta \in \mathbb{S}_1$ due to the shape of the $\mathcal{C}_\theta$ which is a spherical cap with a fixed radius. Consider the event $\mathcal{E}_t$ where $X_1, \ldots, X_t \in \bar{\mathcal{C}}_{\theta_\star}$, which implies both that $Y_1, \ldots, Y_t = \frac{3}{4}$, and that $\xi_{t+1}$ is uniform over $\Theta_{t+1}$. On the event $\mathcal{E}_t$, 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 $t\delta \le n_0\delta \le 1/4$ for all $t < n_0$, we have

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

Therefore, the expected regret of TS is lower bounded by

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

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

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

which implies that

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

□

# Chapter 9

# IR Lower-Bound for General Convex Functions

Theorem 22 shows that Thompson sampling has large regret for general bandit convex optimization. The next theorem shows there exist priors for which the (regret-to-)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 optimization that is better than $\tilde{O}(d^{1.5}\sqrt{n})$

**Theorem 26.** *Suppose that $\mathcal{K} = \mathbb{B}_1$ and $d > 256$. Then there exists a prior $\xi$ on $\mathscr{F}_{b1}$ such that for all probability measures $\pi$ on $\mathcal{K}$, $\Delta(\pi, \xi) \geq 2^{-9} \frac{d}{\log(d)} \sqrt{\mathcal{I}(\pi, \xi)}$.*

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

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

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

Further, let $\xi$ be the law of $f_\theta$ where $\theta \sim \mathrm{Unif}(\mathbb{S}_1)$, 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 is the expected loss and the expected information gain at $x$. Therefore, for any policy $\pi$ 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 $\mathcal{I}_x=O(\log(d)^4d^{-4})$ for all $x\in\mathbb{B}_1$, which implies that $\Delta(\pi,\xi)=\Omega(\log(d)/d)$ and $\mathcal{I}(\pi,\xi)=\tilde{O}(\log(d)^4d^{-4})$ for any policy $\pi$, and hence the claimed lower bound.

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

**Lemma 27.** *For any $x\in\mathbb{B}_1$ such that $\tau\le\|x\|$, $\mathbb{P}\left(f_\theta(x)=f(x)\right)\ge 1-\dfrac{1}{d^4}$, where $\theta\sim$ Unif$(\mathbb{S}_1)$.*

*Proof.* From Lemma 21, $f_\theta(x)=f(x)$ if $\|\theta-x\|^2\ge 1+2\varepsilon$. Moreover, for $x\in\mathbb{B}_1$ with $r\triangleq\|x\|$,

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

which implies that $f(x)=f_\theta(x)$ if $\langle\theta,x\rangle\le\frac{r^2}{2}-\varepsilon$. We thus have

$$\begin{aligned}
\mathbb{P}\left(f_\theta(x)=f(x)\right)&=\mathbb{P}\left(\langle\theta,x\rangle\le\frac{r^2}{2}-\varepsilon\right)\\
&=\mathbb{P}\left(\langle\theta,e_1\rangle\le\left(\frac{r^2}{2}-\varepsilon\right)\|x\|^{-1}\right) && \text{(since } \theta\sim\mathrm{Unif}(\mathbb{S}_1))\\
&=1-\mathbb{P}\left(\langle\theta,e_1\rangle>\left(\frac{r}{2}-\frac{\varepsilon}{r}\right)\right)\\
&\ge 1-\exp\left(-\left(\frac{r}{2}-\frac{\varepsilon}{r}\right)^2\frac{d}{2}\right) && \text{(by Theorem 25)}\\
&\overset{\text{(a)}}{\ge} 1-\exp\left(-\frac{9\log(d)}{2}\right)\\
&\ge 1-\exp\left(-\log(d^4)\right)\\
&\ge 1-\frac{1}{d^4},
\end{aligned}$$

where (a) follows because $r\ge\tau\ge\sqrt{2\varepsilon}$, hence $\frac{r}{2}-\frac{\varepsilon}{r}\ge\frac{\tau}{2}-\frac{\varepsilon}{\tau}=4\sqrt{\frac{\log(d)}{d}}-\sqrt{\frac{\varepsilon}{8}}=3\sqrt{\frac{\log(d)}{d}}$. ☐

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

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

*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 $r = \|x\|$. Recall that $\Delta_x = \mathbb{E}\left[f_\theta(x)\right]$ where $\theta \sim \mathrm{Unif}(\mathbb{S}_1)$. We prove this result by considering two cases.

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

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

where (a) follows from Lemma 27, and the definition of $f$ and $r \geq \tau$.

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

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

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, with $r \leq \tau$, we get

$$1 + \frac{3r^2}{8} - \frac{r^4}{2} \geq \min(1, 1 + \frac{3\tau^2}{8} - \frac{\tau^4}{2}) = 1 + \min(0, 3\varepsilon - 32\varepsilon^2) \geq 1\,,$$

where the last inequality holds because $3\varepsilon - 32\varepsilon^2 \geq 0$ for our choice of $\varepsilon$ and $d$. This let us further lower bound $\mathbb{E}[f_\theta(x)]$ as

$$\mathbb{E}\left[f_\theta(x)\right] \geq \left(1 + \varepsilon - 2\varepsilon^2\right) - 1 = \varepsilon - 2\varepsilon^2\,.$$

The proof is finished by noting that $\varepsilon \leq \frac{1}{3}$ by the choice of $\varepsilon$ and $d$, hence

$$\varepsilon - 2\varepsilon^2 \geq \varepsilon - \frac{2}{3}\varepsilon \geq \frac{8\log(d)}{3d} \geq \frac{2\log(d)}{d} \,.$$

□

**Lemma 30.** *For all $x \in \mathbb{B}_1$, $d > 256$, and $\theta \in$ Uniform$(\mathbb{S}^{d-1})$, $\mathcal{I}_x \leq 2^{10} \frac{\log(d)^4}{d^4}$.*

*Proof.* Let $x \in \mathbb{B}_1$ and $r = \|x\|$. Note that

$$\mathcal{I}_x = \mathbb{E}\left[(f_\theta(x) - \mathbb{E}\left[f_\theta(x)\right])^2\right] \leq \mathbb{E}\left[(f_\theta(x) - f(x))^2\right] \,. \tag{9.1}$$

We prove this result by considering two cases.

**CASE 1** If $r \geq \tau$, then starting from Eq. (9.1), we get

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

where the first inequality holds since the mean minimizes the squared deviation, the second inequality holds since $0 \leq f_\theta \leq f$ by Lemma 21, $0 \leq f(x) - f_\theta(x) \leq \frac{1}{2}r^2 + \varepsilon$, where the last inequality is by the definition of $f$ and because $\|x\| = r \geq \tau$. The next inequality holds because by Lemma 27, $\mathbb{P}(f_\theta(x) \neq f(x)) \leq \frac{1}{d^4}$.

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

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

where (a) follows from $f(x) \geq f_\theta(x)$ which holds by Lemma 21; (b) follows from Lemma 28, (c) follows from the definition of $f$, and $\sqrt{1+x} \geq 1 + \frac{x}{2} - \frac{x^2}{2}$ for all $x \geq -1$, and (d) follows from the fact that $r^2, \varepsilon, |\langle \theta, x \rangle| \leq 1$, and the expression in the previous line is a homogeneous polynomial of these terms with degree at most 2 and coefficients at most 4.

Next, starting from the right-hand side of Eq. (9.1), we have

$$\begin{aligned}
\mathcal{I}_x \leq \mathbb{E}\left[(f(x) - f_\theta(x))^2\right] &\leq \mathbb{E}\left[4^2\left(r^2 + |\langle \theta, x \rangle| + \varepsilon\right)^4\right] \\
&\overset{\text{(a)}}{\leq} 4^2 \mathbb{E}\left[27\left(r^8 + \langle \theta, x \rangle^4 + \varepsilon^4\right)\right] \\
&\leq 4^3\left(r^8 + \mathbb{E}\left[\langle \theta, x \rangle^4\right] + \varepsilon^4\right) \\
&\overset{\text{(b)}}{\leq} 4^3\left(r^8 + \frac{3r^4}{d^2} + \varepsilon^4\right) \\
&\overset{\text{(c)}}{<} 4^3\left(8^4\varepsilon^4 + \frac{3 \cdot 8^2\varepsilon^2}{d^2} + \varepsilon^4\right) \\
&\leq 4^3\left(8^4\varepsilon^4 + 3\varepsilon^4 + \varepsilon^4\right) \\
&\leq 2^{10}\varepsilon^4 = 2^{10}\frac{\log(d)^4}{d^4},
\end{aligned}$$

where (a) follows from $(a+b+c)^4 \leq 16(a^4+b^4+c^4)$, (b) follows from the fact that $\mathbb{E}[\langle \theta, x \rangle^4] = \frac{3r^4}{d(d+2)} < \frac{3r^4}{d^2}$, and (c) follows from $r < \tau = \sqrt{8\varepsilon}$. □

*Proof of Theorem 26.* Let $\xi$ be the law of $f_\theta$ when $\theta$ has law Unif($\mathbb{S}_1$). Then for any policy $\pi$, and

$X \sim \pi$, from Lemma 29 we have

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

and from Lemma 30 we have

$$\mathcal{I}(\pi, \xi) = \mathbb{E}[\mathcal{I}_X] \leq \frac{2^{10}\log(d)^4}{d^4}\,,$$

which together imply

$$\frac{\Delta(\pi, \xi)}{\sqrt{\mathcal{I}(\pi, \xi)}} \geq \frac{d}{2^9 \log(d)}\,.$$

□

# Chapter 10

# TS for Known Link Function

In this chapter we provide a tighter upper bound for the information ratio of TS in $d$ dimensions, when the link function $\ell$ is convex, non-increasing, and known to the learner. Being known to the learner can be formalized as $\ell$ is fixed for all the functions in the function class. This function class is a subset of $\mathscr{F}_{\texttt{blrm}}$ that we studied in Chapter 7, but the significant difference is that the information ratio upper bound matches the one of linear bandits [Russo and Van Roy, 2016] with the same constant. This effectively shows that adding a known convex monotone ridge function is free in terms of the information ratio, and hence the Bayesian regret upper bound.

**Definition 31** (Known Ridge Function Class)**.** Let $\Theta \subseteq \mathbb{R}^d$ be a set of parameters and $\ell : \mathbb{R} \to \mathbb{R}$ be a convex, non-increasing function, for some $b > 0$. Define $\mathscr{F}_{\Theta,\ell}$ be the space of all functions $f : \mathcal{K} \to \mathbb{R}$ such that there exists $\theta \in \Theta$ such that

$$f(x) = \ell(\langle \theta, x \rangle) \quad \forall x \in \mathcal{K}\,.$$

Further, instead of defining the distributions over the space of convex function $f$, we can work with distributions over the parameter space $\Theta \in \mathbb{R}^d$. Precisely, let $f_\theta(\cdot) = \ell(\langle \theta, \cdot \rangle)$, and, for every subset $\Theta' \subseteq \Theta$, let $\mathscr{F}_{\Theta'} = \{f_\theta(\cdot) : \theta \in \Theta'\}$. Define $\sigma \in \mathcal{P}(\Theta)$ such that for every measurable subset $\Theta' \subseteq \Theta$,

$$\sigma\left(\Theta'\right) = \xi\left(\mathscr{F}_{\Theta'}\right)\,. \tag{10.1}$$

**Lemma 32.** *For every $\xi \in \mathscr{P}(\mathscr{F}_{\Theta,\ell})$, it holds that*

$$\Psi(\xi) < d,$$

*whenever for every $\theta \in \Theta$, it holds that*

$$\theta = \operatorname*{arg\,min}_{\theta' \in \Theta} \left\langle \theta', \operatorname*{arg\,min}_{x \in \mathcal{K}} \langle x, \theta \rangle \right\rangle .$$

First recall that

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

Therefore, to prove that $(0, \beta) \in \mathrm{IR}(\mathscr{F})$, it suffices to show that

$$\Psi(\xi) \triangleq \frac{\Delta(\pi_{\mathrm{TS}}^{\xi}, \xi)^2}{\mathcal{I}(\pi_{\mathrm{TS}}^{\xi}, \xi)} \leq \beta \,, \qquad \forall \xi \in \mathscr{P}(\mathscr{F}) \tag{10.2}$$

where the function $\Psi(\xi)$ was the original definition (up to the measure of information used) of information ratio in Russo and Van Roy [2016]. While the proof of Theorem 15 used the decomposition lemma (Lemma 4) the proof in this chapter avoids the need of this, thanks to the simpler setting. In particular, the proof uses a reduction to the linear bandits, i.e., the case when the link function $\ell$ is the identity function. For completeness, we start by presenting the analysis of the linear setting, borrowed from Russo and Van Roy [2016].

## 10.1 Linear TS

**Theorem 33.** *If $\ell(x) = x$ for all $x \in \mathbb{R}$, then for all $\xi \in \mathscr{P}(\mathscr{F}_{\Theta,\ell})$, we have*

$$\Psi(\xi) \leq d \, .$$

*Proof.* Let $\sigma$ be the corresponding distribution over the parameter space $\Theta$, as defined in Eq. (10.1). Let $(\theta, X)$ be jointly distributed such that $\theta \sim \sigma$, $X = \operatorname*{arg\,min}_{x \in \mathcal{K}} \langle x, \theta \rangle$. Then $X \sim \pi_{\mathrm{TS}}^{\xi}$. First, we have

$$\bar{f}(\cdot) = \mathbb{E}[f(\cdot)] = \mathbb{E}[\langle \cdot, \theta \rangle] = \langle \cdot, \mathbb{E}[\theta] \rangle \,,$$

from which we define $\bar{\theta} = \mathbb{E}[\theta]$. Moreover, let

$$V = \mathbb{E}[XX^\top].$$

For the remainder of the proof we suppose that $V$ is invertible, which holds as long $X$ is not supported on any subspace of $\mathbb{R}^d$. If $X$ was supported on such a subspace, the proof can be repeated by replacing $\mathbb{R}^d$ with that subspace.

Let $(\theta', X')$ be an independent copy of $(\theta, X)$. We have

$$\begin{aligned}
\Delta(\pi_{\mathrm{TS}}^{\xi}, \xi) &= \mathbb{E}[\ell(X'^\top \theta) - \ell(X^\top \theta)] \\
&\overset{\text{(a)}}{=} \mathbb{E}[X'^\top \theta - X^\top \theta] \\
&\overset{\text{(b)}}{=} \mathbb{E}[X^\top \theta' - X^\top \theta] \\
&= \mathbb{E}[X^\top (\bar{\theta} - \theta)] \\
&\overset{\text{(c)}}{\leq} \mathbb{E}[\|X\|_{V^{-1}} \|\bar{\theta} - \theta\|_V] \\
&\overset{\text{(d)}}{\leq} \sqrt{\mathbb{E}[\|X\|_{V^{-1}}^2] \mathbb{E}[\|\bar{\theta} - \theta\|_V^2]} \,,
\end{aligned}$$

where `(a)` is by $\ell$ being identity, `(b)` is by $X'$ and $\theta'$ being independent copies of $X$ and $\theta$, and `(c)` and `(d)` are by Cauchy-Schwarz. Further, we have

$$\begin{aligned}
\mathbb{E}\left[\|X\|_{V^{-1}}^2\right] &= \mathbb{E}\left[X^\top \mathbb{E}[XX^\top]^{-1} X\right] \\
&= \mathbb{E}\left[\mathrm{trace}(X^\top \mathbb{E}[XX^\top]^{-1} X)\right] \\
&= \mathbb{E}\left[\mathrm{trace}(XX^\top \mathbb{E}[XX^\top]^{-1})\right] \\
&= \mathrm{trace}(\mathbb{E}[XX^\top] \mathbb{E}[XX^\top]^{-1}) && \text{(linearity of trace and expectation)} \\
&= d.
\end{aligned}$$

Moreover, we have

$$
\begin{aligned}
\mathcal{I}(\pi_{\mathrm{TS}}^{\xi}, \xi) &= \mathbb{E}[(f_\theta(X') - \bar{f}(X'))^2] \\
&= \mathbb{E}[(X^\top \theta' - \mathbb{E}[X^\top \theta' | X])^2] \qquad \text{(since } (X', \theta') \text{ is an independent copy of } (X, \theta)) \\
&= \mathbb{E}[(X^\top (\theta' - \bar{\theta}))^2] \\
&= \mathbb{E}[(\theta' - \bar{\theta})^\top X X^\top (\theta' - \bar{\theta})] \\
&\overset{\text{(a)}}{=} \mathbb{E}[(\theta' - \bar{\theta})^\top V (\theta' - \bar{\theta})] \\
&= \mathbb{E}[\|\theta' - \bar{\theta}\|_V^2]\,,
\end{aligned}
$$

where (a) follows because $\theta'$ and $X$ are independent, the tower rule and the linearity of expectation. Putting these together, we have

$$
\begin{aligned}
\Psi(\xi) &= \frac{\Delta(\pi_{\mathrm{TS}}^{\xi}, \xi)^2}{\mathcal{I}(\pi_{\mathrm{TS}}^{\xi}, \xi)} \\
&\leq \frac{\mathbb{E}[\|X\|_{V^{-1}}^2] \mathbb{E}[\|\bar{\theta} - \theta\|_V^2]}{\mathbb{E}[\|\bar{\theta} - \theta\|_V^2]} \\
&= \mathbb{E}[\|X\|_{V^{-1}}^2] = d\,.
\end{aligned}
$$

□

## 10.2 Inequalities for Convex Functions

We first provide three lemmas about convex functions, the first two of which are only used to prove the third one.

**Lemma 34.** *Let $h : [0, \infty) \to [0, \infty)$ be convex, nondecreasing with $h(0) = 0$. Let $X \geq 0$ be a random variable with $\mathbb{E}[h(X)] = \mathbb{E}[X]$ and $\mathbb{E}[X^2] < \infty$. Then*

$$\mathbb{E}[h(X)^2] \;\geq\; \mathbb{E}[X^2].$$

*Proof.* Define $\phi(x) = h(x) - x$. Then $\phi$ is convex and continuous, with $\phi(0) = 0$. Convexity of $\phi$ implies that $I = \{x \geq 0 : \phi(x) \leq 0\}$ is a closed interval. Since $\phi(0) = 0$, we know that the left endpoint of $I$ is zero. Let $x_0 = \max I$ be its right end-point, possibly $x_0 = \infty$. Then,

$$h(x) > x \;\Rightarrow\; x > x_0, \qquad h(x) \leq x \;\Rightarrow\; x \leq x_0\,.$$

We consider three cases based on whether $x_0 = 0$, $0 < x_0 < \infty$ or $x_0 = \infty$. In the first case, $\phi$ is positive on $(0, \infty)$ and since $\mathbb{P}(X > 0) > 0$, $\mathbb{E}[\phi(X)] > 0$, a contradiction to our assumption that $\mathbb{E}[h(X)] = \mathbb{E}[X]$. In the last case when $x_0 = \infty$, $\phi$ is is zero on $[0, \infty)$, which implies that $h$ is the identity function, in which case the statement trivially holds.

Now, let us consider the case when $0 < x_0 < \infty$. Define the complementary events $A = \{h(X) > X\}$, $B = \{h(X) \leq X\}$. On $A$ we have $X > x_0$ and $h(X) > X > x_0$, hence $h(X) + X > 2x_0$. On $B$, we have $X \leq x_0$ and $h(X) \leq X \leq x_0$, hence $h(X) + X \leq 2x_0$. Thus,

$$(h(X)^2 - X^2)\mathbf{1}_A = (h(X) - X)(h(X) + X)\mathbf{1}_A \;\geq\; 2x_0\,(h(X) - X)\mathbf{1}_A\,,$$
$$(h(X)^2 - X^2)\mathbf{1}_B = (h(X) - X)(h(X) + X)\mathbf{1}_B \;\geq\; 2x_0\,(h(X) - X)\mathbf{1}_B\,.$$

Adding the two inequalities and taking expectations,

$$\mathbb{E}[h(X)^2 - X^2] \;\geq\; 2x_0\,\mathbb{E}[h(X) - X] \;=\; 0,$$

since $\mathbb{E}[h(X)] = \mathbb{E}[X]$. Therefore $\mathbb{E}[h(X)^2] \geq \mathbb{E}[X^2]$. ☐

**Lemma 35.** *Assume that $g : \mathbb{R} \to \mathbb{R}$ is a convex and non-decreasing function such that $g(0) = 0$. Let $X \geq 0$ be a random variable with $\mathbb{E}[X^2] \leq \infty$. Let $\frac{0}{0} = 1$, then it holds that*

$$\frac{\mathbb{E}[g(X)^2]}{\mathbb{E}[g(X)]^2} \geq \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2}\,.$$

*Proof.* First, suppose that $\mathbb{E}[X] = 0$. Then since $X$ is supported on $\mathbb{R}_+$, $X = 0$ almost surely, $\mathbb{E}[X^2] = 0$, and $\frac{\mathbb{E}[X^2]}{E[X]^2} = \frac{0}{0} = 1$. Further, since $g(0) = 0$, we have $g(X) = 0$ almost surely. Therefore, $\mathbb{E}[g(X)]^2 = \mathbb{E}[g(X)^2] = 0$, and the statement of the lemma holds with equality.

Define the function $h(x) = \frac{\mathbb{E}[x]}{\mathbb{E}[g(x)]} g(x)$, and observe that

$$\frac{\mathbb{E}[g(X)^2]}{\mathbb{E}[g(X)]^2} = \frac{\mathbb{E}[h(X)^2]}{\mathbb{E}[h(X)]^2},$$

therefore it suffices to show the inequality for the function $h$. Note that $\mathbb{E}[h(X)] = \mathbb{E}[X]$ and the desired inequality is equivalent to

$$\mathbb{E}[h(X)^2] \geq \mathbb{E}[X^2],$$

which holds by Lemma 34 since $h$ is convex, non-decreasing, and $h(0) = 0$. ☐

**Lemma 36.** *Assume that $f : \mathbb{R} \to \mathbb{R}$ is a convex and non-decreasing function. Further, let $X$ be* *a random variable supported on* $[0, b]$ *for some* $b > 0$ *and* $x_1 = \operatorname{ess\,inf} X$*. Then, the following inequality holds*

$$\frac{\mathbb{E}[f(X)] - f(x_1)}{\sqrt{\mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])^2\right]}} \le \frac{\mathbb{E}[X] - x_1}{\sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}},$$

*where we let* $\frac{0}{0} = 1$*.*

*Proof.* First, note that both quotients are well-defined. To see that suppose that $\mathbb{E}[(x - \mathbb{E}[X])^2] = 0$, which means that $X$ is almost surely constant, which makes $\mathbb{E}[X] = x_1$, and hence the fraction would be $\frac{0}{0} = 1$. Same argument applies for the other quotient.

Let $X' = X - x_1$, and observe that

$$\frac{\mathbb{E}[X] - x_1}{\sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}} = \frac{\mathbb{E}[X']}{\sqrt{\mathbb{E}\left[(X' - \mathbb{E}[X'])^2\right]}}.$$

Next define $g(x') = f(x' + x_1) - f(x_1)$ so that $g(X') = f(X) - f(x_1)$, and observe that

$$\frac{\mathbb{E}[f(X)] - f(x_1)}{\sqrt{\mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])^2\right]}} = \frac{\mathbb{E}[g(X')]}{\sqrt{\mathbb{E}\left[(g(X') - \mathbb{E}[g(X')])^2\right]}}.$$

Therefore, the desired inequality is proved once we show

$$\frac{\mathbb{E}[g(X')]}{\sqrt{\mathbb{E}\left[(g(X') - \mathbb{E}[g(X')])^2\right]}} \ge \frac{\mathbb{E}[X']}{\sqrt{\mathbb{E}\left[(X' - \mathbb{E}[X'])^2\right]}}. \tag{10.3}$$

Note that $g$ is convex and non-decreasing. Also, we know that $\operatorname{ess\,inf} X' = \operatorname{ess\,inf} X - x_1 = 0$, and $g(0) = f(x_1) - f(x_1) = 0$, which implies that both $\mathbb{E}[X']$, $\mathbb{E}[g(X')]$ are non-negative, so Eq. (10.3) is equivalent to

$$\frac{\mathbb{E}[g(X')]^2}{\mathbb{E}\left[(g(X') - \mathbb{E}[g(X')])^2\right]} \ge \frac{\mathbb{E}[X']^2}{\mathbb{E}\left[(X' - \mathbb{E}[X'])^2\right]},$$

which is equivalent to

$$\frac{\mathbb{E}[g(X')^2] - \mathbb{E}[g(X')]^2}{\mathbb{E}[g(X')]^2} \le \frac{\mathbb{E}[X'^2] - \mathbb{E}[X']^2}{\mathbb{E}[X']^2}$$

or equivalently

$$\frac{\mathbb{E}[g(X')^2]}{\mathbb{E}[g(X')]^2} \geq \frac{\mathbb{E}[X'^2]}{\mathbb{E}[X']^2} \tag{10.4}$$

which holds by Lemma 35 and the fact that $X'$ is supported on $\mathbb{R}_+$. $\square$

## 10.3 Information Ratio of TS for Known Ridge

We are now in position to prove Lemma 32.

**Lemma 32.** *For every $\xi \in \mathscr{P}(\mathscr{F}_{\Theta,\ell})$, it holds that*

$$\Psi(\xi) < d,$$

*whenever for every $\theta \in \Theta$, it holds that*

$$\theta = \operatorname*{arg\,min}_{\theta' \in \Theta} \left\langle \theta', \operatorname*{arg\,min}_{x \in \mathcal{K}} \langle x, \theta \rangle \right\rangle .$$

*Proof.* Similar to the proof of Theorem 33, let $\sigma$ be the corresponding distribution over the parameter space $\Theta$, as defined in Eq. (10.1). Let $(\theta, X)$ be jointly distributed such that $\theta \sim \sigma$, $X = \arg\min_{x \in \mathcal{K}} \langle x, \theta \rangle$. Then $X \sim \pi^{\xi}_{\text{TS}}$ as $\ell$ is assumed to be monotone.

Also note that because of the assumption of the theorem, we have both

$$X = \operatorname*{arg\,min}_{x \in \mathcal{K}} \ell(\langle \theta, x \rangle), \quad \text{and} \quad \theta = \operatorname*{arg\,min}_{y \in \Theta} \ell(\langle y, X \rangle) . \tag{10.5}$$

Next, we let $\frac{0}{0} = 1$, and we define the function

$$\phi(X) = \frac{\ell(X^\top \theta) - \mathbb{E}[\ell(X^\top \theta')|X]}{X^\top \theta - \mathbb{E}[X^\top \theta'|X]}$$

Note that in this quotient, if $X^\top \theta - \mathbb{E}[X^\top \theta'|X] = 0$, then $X^\top \theta = X^\top \nu$ for all $\nu \in \text{supp}(\sigma)$, which makes the numerator $\ell(X^\top \theta) - \mathbb{E}[\ell(X^\top \theta')|X] = 0$ as well. Therefore, the function $\phi(X)$

is well-defined. Further, we have

$$\frac{\ell(X^\top\theta)-\mathbb{E}[\ell(X^\top\theta')|X]}{\sqrt{\mathbb{E}\left[\left(\ell(X^\top\theta')-\mathbb{E}[\ell(X^\top\theta')|X]\right)^2|X\right]}} \overset{\text{(a)}}{\leq} \frac{X^\top\theta-\mathbb{E}[X^\top\theta'|X]}{\sqrt{\mathbb{E}\left[\left(X^\top\theta'-\mathbb{E}[X^\top\theta'|X]\right)^2|X\right]}}$$
$$\overset{\text{(b)}}{=} \frac{\phi(X)X^\top\theta-\mathbb{E}[\phi(X)X^\top\theta'|X]}{\sqrt{\mathbb{E}\left[\left(\phi(X)X^\top\theta'-\mathbb{E}[\phi(X)X^\top\theta'|X]\right)^2|X\right]}}$$
$$\overset{\text{(c)}}{=} \frac{\ell(X^\top\theta)-\mathbb{E}[\ell(X^\top\theta')|X]}{\sqrt{\mathbb{E}\left[\left(\phi(X)X^\top\theta'-\mathbb{E}[\phi(X)X^\top\theta'|X]\right)^2|X\right]}},$$

where `(a)` follows from Eq. (10.5) and Lemma 36, `(b)` follows from multiplying both numerator and denominator by $\phi(X)$, and `(c)` follows from the definition of $\phi(X)$. Thus, by rearranging and simplifying, we obtain

$$\mathbb{E}\left[\left(\ell(X^\top\theta')-\mathbb{E}[\ell(X^\top\theta')|X]\right)^2|X\right] \geq \mathbb{E}\left[\left(\phi(X)X^\top\theta'-\mathbb{E}[\phi(X)X^\top\theta'|X]\right)^2|X\right]. \tag{10.6}$$

Further, we can write

$$\Psi(\xi)^{\frac{1}{2}} = \frac{\mathbb{E}\left[\ell(X^\top\theta)-\mathbb{E}[\ell(X^\top\theta')|X]\right]}{\sqrt{\mathbb{E}\left[\left(\ell(X^\top\theta')-\mathbb{E}[\ell(X^\top\theta')|X]\right)^2\right]}} = \frac{\mathbb{E}\left[\phi(X)X^\top\theta-\mathbb{E}[\phi(X)X^\top\theta'|X]\right]}{\sqrt{\mathbb{E}\left[\mathbb{E}\left[\left(\ell(X^\top\theta')-\mathbb{E}[\ell(X^\top\theta')|X]\right)^2|X\right]\right]}}$$
$$\leq \frac{\mathbb{E}\left[\phi(X)X^\top\theta-\mathbb{E}[\phi(X)X^\top\theta'|X]\right]}{\sqrt{\mathbb{E}\left[\mathbb{E}\left[\left(\phi(X)X^\top\theta'-\mathbb{E}[\phi(X)X^\top\theta'|X]\right)^2|X\right]\right]}}$$

where the first equality follows from the definition of $\phi(X)$ and the second inequality follows from 10.6. Further, define $\tilde{X}=\phi(X)X$ and rewrite the inequality as

$$\Psi(\xi)^{\frac{1}{2}} = \frac{\mathbb{E}\left[\ell(X^\top\theta)-\mathbb{E}[\ell(X^\top\theta')|X]\right]}{\sqrt{\mathbb{E}\left[\left(\ell(X^\top\theta')-\mathbb{E}[\ell(X^\top\theta')|X]\right)^2\right]}} \leq \frac{\mathbb{E}\left[\tilde{X}^\top\theta-\mathbb{E}[\tilde{X}^\top\theta'|X]\right]}{\sqrt{\mathbb{E}\left[\left(\tilde{X}^\top\theta'-\mathbb{E}[\tilde{X}^\top\theta'|X]\right)^2\right]}} \leq \sqrt{d},$$

where the $\sqrt{d}$ upper bound follows from Theorem 33. $\square$

# Chapter 11

# Empirical Performance of TS

In this chapter we study the empirical performance of TS for 1-dimensional BCO problems. The main challenge of implementing TS for BCO is sampling from the posterior distribution. This empirical analysis has multiple goals:

- it forces us to think about efficiency,
- it gives us a glimpse into the performance of TS in relation to some competing designs.

In particular, sampling from the posterior defined over some function space in general is intractable. Here, we design an approximate sampling method for a special case when $\mathcal{K} = [-1, 1]$. We don't provide any theoretical guarantees for our sampling scheme, but instead we show that it works well in practice.

## 11.1 A Probability Distribution on $\mathscr{F}_{\mathtt{bl}}$

We consider the case where $\mathcal{K} = [-1, 1]$ and $d = 1$. We define a prior $\xi$ on $\mathscr{F}_{\mathtt{bl}}^{+} = \mathscr{F}_{\mathtt{bl}} \cap \{f : f \geq 0\}$ as follows. Let

$$\phi_y(x) = |x - y| \,,$$

which is a convex function. We use $\phi_y$s as building blocks for making convex functions. Let $\varepsilon > 0$ be a small constant, which specify the level of discretization of convex functions. Now let

$$\Phi = \{\phi_{\varepsilon y} : \varepsilon y \in \mathcal{K}, y \in \mathbb{Z}\} . \tag{11.1}$$

Let $d = |\Phi|$. Then for any $\theta \in \mathbb{R}_+^d$,

$$f_\theta(x) = \sum_{i=1}^{d} \theta_i \phi_{y_i}(x) ,$$

where $y_i$ is the $i$-th element of $\Phi$. Since $\theta \in \mathbb{R}_+^d$, the function $f_\theta$ is convex and non-negative. As it is well-known, the set $\left\{\sum_{i=1}^d \theta_i \phi_{y_i} : d \geq 1, \theta_i \geq 0, y_i \in Y\right\}$ is dense in $\mathscr{F}_{\mathtt{b1}}^+$ in the uniform norm, provided that $Y \subseteq \mathbb{R}$ is dense [Billingsley, 2013], justifying our choice. This construction allows us to define a probability distribution over the space of convex functions, by defining a probability measure on $\mathbb{R}_+^d$.

We suppose that $f_\star$, the true loss function, has the form

$$f_\star(x) = f_{\theta_\star}(x) = \sum_{i=1}^{d} \theta_{\star,i} \phi_{y_i}(x) , \tag{11.2}$$

and we assume that $\theta_\star$ is sampled from a $\mathbb{R}_+-$truncated multivariate Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$. We denote the law of $\theta_\star$ by $\sigma$, and the law of $f_\star$ by $\xi$. Further, we assume that

$$Y_t = f_{\theta_\star}(X_t) + \sigma_0 \varepsilon_t , \tag{11.3}$$

where $\varepsilon_t$ is a standard Gaussian noise, and $\sigma$ is a positive constant. The posterior density of $\theta \in \mathbb{R}_+^d$ is proportional to

$$\pi_t(\theta) := \exp\left(-\frac{1}{2c}\|\theta - \mu\|^2 \sum_{s=1}^{t} (Y_s - f_\theta(X_s))^2\right)$$

### 11.1.1 Hit-and-Run Samplers

We use Hit-and-run (HAR) methods to sample from the posterior $\xi_t$. HAR methods generate proposals by first choosing a random direction and then moving along that line according to the target density restricted to the line. Originally developed for uniform sampling from convex bodies [e.g. Smith, 1984, Lovász, 1999], they extend naturally to general log-concave (and even non log-concave) densities by sampling from the one-dimensional conditional along each line.

Let $\pi(\theta) \propto \tilde{\pi}(\theta)$ be the target on a convex subset $\mathcal{X} \subseteq \mathbb{R}^d$ (here $\mathcal{X} = \mathbb{R}^d_+$). Given the current state $\theta^{(m)}$, HAR proceeds by

(a) Draw a direction $v$ from a spherically symmetric distribution (commonly $v \sim \mathcal{N}(0, I_d)$ then normalise $v \leftarrow v/\|v\|$).

(b) Determine the feasible interval

$$I(\theta^{(m)}, v) = \{t \in \mathbb{R} : \theta^{(m)} + tv \in \mathcal{X}\}.$$

In the positive orthant this is simply

$$t_{\min} = \max_{i:v_i<0} \frac{-\theta_i^{(m)}}{v_i}, \qquad t_{\max} = \min_{i:v_i>0} \frac{-\theta_i^{(m)}}{v_i}, \qquad I = [t_{\min}, t_{\max}] .$$

(c) Sample $t$ from the one-dimensional density proportional to $\pi(\theta^{(m)} + tv)$ on $I$.

(d) Set $\theta^{(m+1)} = \theta^{(m)} + tv$.

The only non-trivial step is the one-dimensional draw in Step 3; different variants of HAR use different inner samplers. In our setting, we can sample from the truncated Gaussian posterior restricted to $I$.

## 11.2 Bayesian Regret

First we study Bayesian regret. We first sample $\theta_\star$ from a multivariate Gaussian distribution that is known to the learner. We set the observation noise $\mathcal{N}(0, \sigma_0^2)$ with $\sigma_0 = 0.1$. The number of basis functions is set to $d = 100$. We run the experiment for $n = 100$ rounds, and report the average performance over 10 runs in Figure 11.1. The performance is compared to the Bisection Method [Agarwal et al., 2010], the Exponential Weights with kernel estimation [Bubeck et al., 2017], and Online Newton Step [Fokkema et al., 2024b]. The superiority of TS can be explained by the fact that it is able to use the knowledge of prior, and also the fact that most of the other algorithms are designed for the adversarial setting primarily.

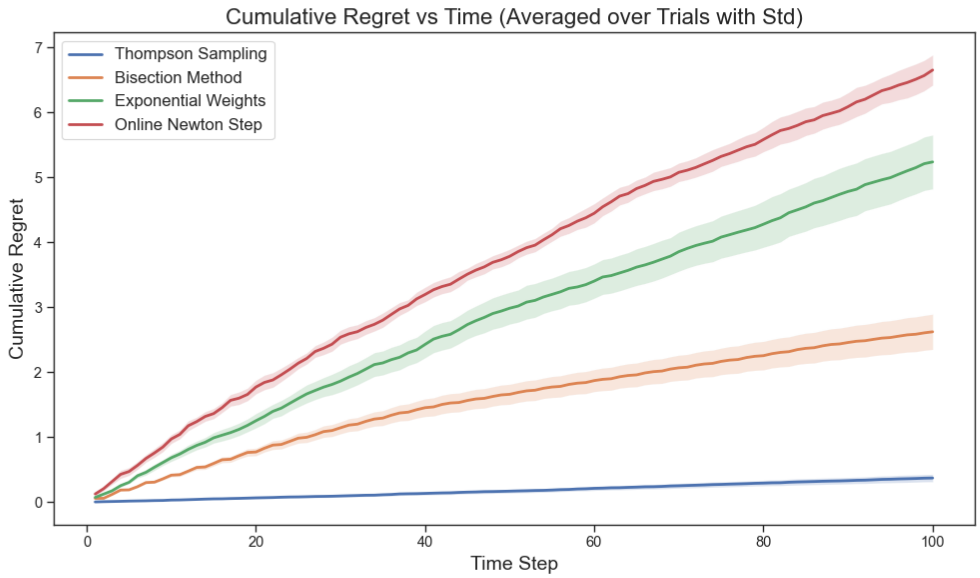



**Figure 11.1:** Bayesian regret for 1-dimensional bandit convex optimization. The plot shows the cumulative regret over $n = 100$ rounds, averaged over 10 independent runs. The loss functions are sampled from a multivariate Gaussian prior with $d = 100$ basis functions, and observations are corrupted with Gaussian noise $\mathcal{N}(0, 0.1^2)$, i.e. $\sigma_0 = 0.1$.

## 11.3 Regret Against Specific Losses

One might wonder how well TS will perform if the loss function is not sampled from the prior $\xi$, and is instead fixed to a specific function $f_\star$. We study three loss functions:

(a) Absolute loss: $f_\star(x) = m\,|x - c|$, for some $c \in [-1, 1]$ and $m > 0$.

(b) Quadratic loss: $f_\star(x) = m(x - c)^2$, for some $c \in [-1, 1]$ and $m > 0$.

(c) Linear loss: $f_\star(x) = mx + b$, such that $\inf_{x \in [-1,1]} f_\star(x) \geq 0$.

We show the behavior of TS alongside other algorithms in Fig 11.2. The dashed lines are functions $f_t$ that are sampled from $\xi_t$ for $t \in [100]$, with newer samples been drawn with higher color density. The circles are the actions taken by different algorithms (the color specify the algorithm) with bigger circles showing more recent actions.

The regret of different algorithms against these losses is pictured in Fig 11.3. It can be seen that TS remains highly competitive even when the losses are chosen in an arbitrary way.

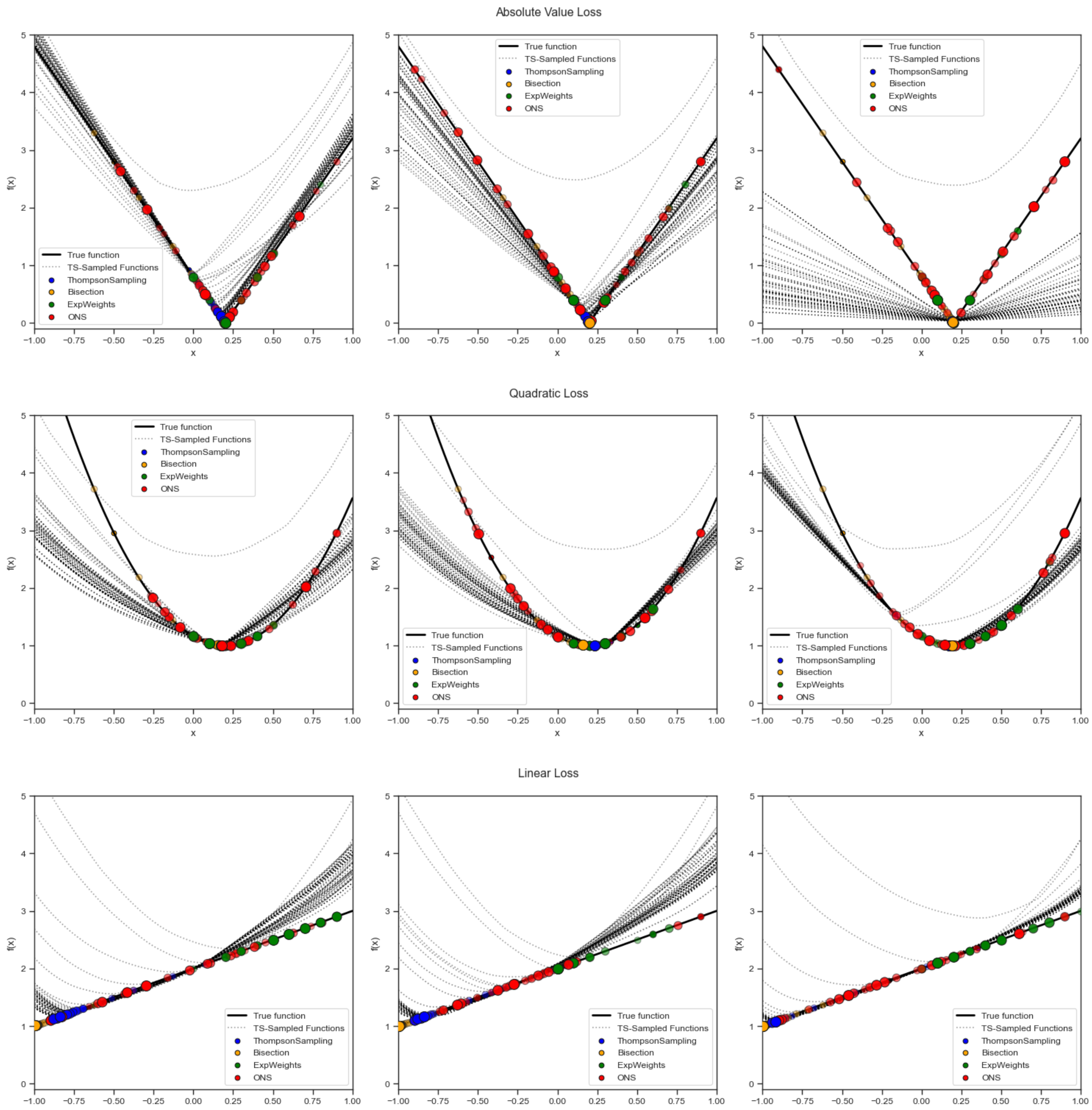
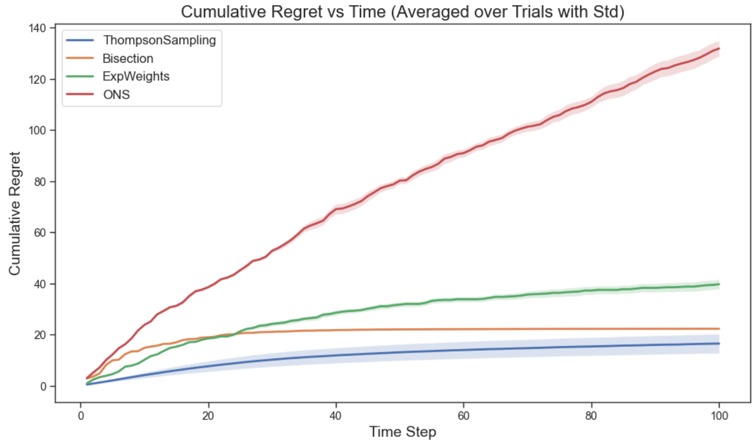
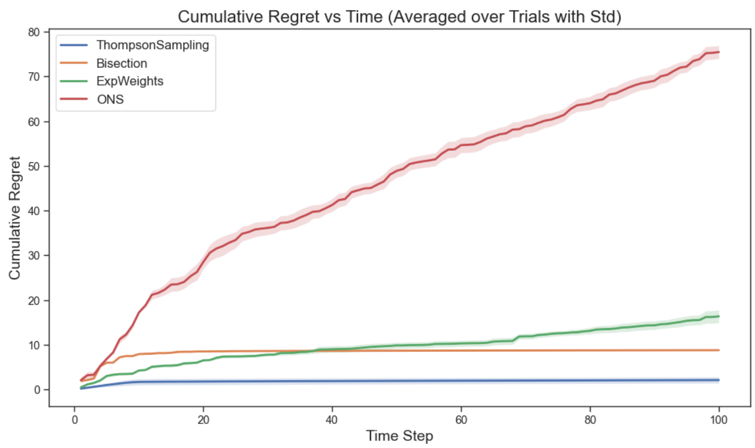


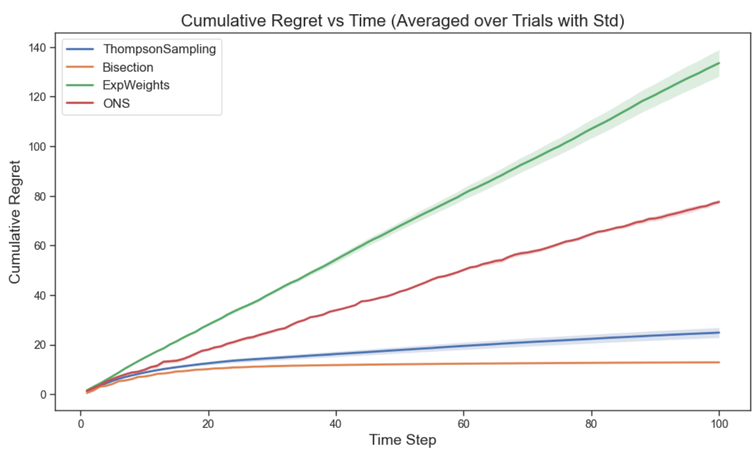

**Figure 11.2:** Performance of `BCO` algorithms against specific loss functions. Top: Absolute loss $f_\star(x) = m|x-c|$. Middle: Quadratic loss $f_\star(x) = m(x-c)^2$. Bottom: Linear loss $f_\star(x) = mx+b$. Solid line shows $f_\star$, dashed lines show TS sampled functions, and circles show actions taken by different algorithms.




**(a)** Average regret against absolute loss: $f_\star(x) = m|x - c|$




**(b)** Average regret against quadratic loss: $f_\star(x) = m(x - c)^2$




**(c)** Average regret against linear loss: $f_\star(x) = mx + b$

**Figure 11.3:** Average regret of TS and other algorithms against specific loss functions.

# Chapter 12

# 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 generalizations of TS in this setting are the following:

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

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

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

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 minimize $f_t$. This must be reflected in the definition of the information ratio. Let $\xi$ be a probability measure on $\mathscr{F} \times \mathcal{K}$ and $\pi$ be a probability measure on $\mathcal{K}$ and let

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

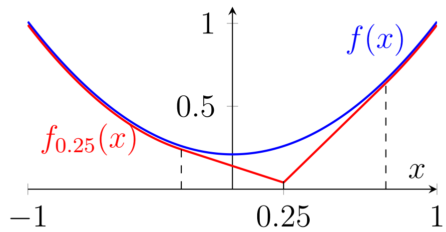



**Figure 12.1:** The function $f(x) = 0.2 + 0.8x^2$ and the function $f_{0.25}$ with $\varepsilon = 0.2$. The function $f_{0.25}$ is the largest convex function that is smaller than $f$ and has $f_{0.25}(0.25) = f(0.25) - 0.2$.

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

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

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

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

which is convex and smaller than $f$ for all $x \in \mathcal{K}$. Essentially, $f_\theta$ should be thought of as the largest convex function that is smaller than $f$ and has $f_\theta(\theta) = f(\theta) - \varepsilon$ (see Fig. 12.1). Moreover, an elementary calculation shows that $\max_{x \in \mathcal{K}} |f(x) - f_\theta(x)| = \varepsilon$ for all $\theta \in [-1, 1]$. Let $\xi$ be the law of $(f_\theta, \theta)$ when $\theta$ is sampled uniformly from $[-1, 1]$ and $\pi$ be uniform on $[-1, 1]$ which is the TS policy as defined in `(b)`. Then, by letting $\theta'$ be an i.i.d. copy of $\theta$ we have

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

where the second equality follows from the definition of $f_\theta$ and the third equality follows from

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

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

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

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

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

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

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

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

which can be further lower bounded by

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

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

# Chapter 13

# Discussion

In this thesis, we explored the performance of Thompson sampling (TS) for Bayesian bandit convex optimization (BCO), revealing a nuanced picture of its capabilities. We established its near-optimal performance in one-dimensional settings and for the class of monotone ridge functions, demonstrating its efficacy in structured scenarios. In stark contrast, we also showed that TS can fail dramatically in general high-dimensional problems and that standard analytical tools face fundamental limitations. This chapter synthesizes these findings, discusses their broader implications for algorithm design in online optimization, and explores the structural properties that determine the success or failure of TS.

## 13.1 Adversarial setup

In the Bayesian adverarial setting a sequence of loss functions $f_1, \ldots, f_n$ 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 $\text{BReg}(\mathscr{A}, \xi) = \mathbb{E}[\sup_{x \in \mathcal{K}} \sum_{t=1}^n (f_t(X_t) - f_t(x))]$. 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 $g_1, \ldots, g_n$ from the posterior and plays $X_t$ as the minimizer of $\sum_{t=1}^n g_t$. The former has linear regret, while Bubeck et al. [2015] notes that the latter has an unbounded information ratio. The situation was discussed in more details in Chapter 12.

## 13.2 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 generalized linear bandits are no harder than linear bandits.

## 13.3 TS vs IDS

Theorem 22 shows that TS can have more-or-less linear regret in high-dimensional problems. On the other hand, Bubeck and Eldan [2018] and Lattimore [2020] show that IDS has a well-controlled information ratio, but is much harder to compute. An obvious question is whether some simple adaptation of Thompson sampling has a well-controlled information ratio.

## 13.4 Applications

Many problems are reasonably modelled as 1-dimensional convex bandits, with the classical example being dynamic pricing where $\mathcal{K}$ 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 $\mathcal{K} = \{x \geq \mathbf{0} : \|x\|_1 \leq 1\}$.

## 13.5 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 $\mathcal{K}$. Concretely, suppose that $\mathcal{K}$ is a convex body and $f : \mathcal{K} \to [0, 1]$ is convex and $\mathbb{B}_r \subset \mathcal{K}$ and $\mathcal{K}_\varepsilon = (1 - \varepsilon)\mathcal{K}$. Then $\min_{x \in K_\varepsilon} f(x) \leq \inf_{x \in \mathcal{K}} f(x) + \varepsilon$ and $f$ is $1/(r\varepsilon)$-Lipschitz on $K_\varepsilon$ [Lattimore, 2024, Chapter

3]. Hence, you can run TS on $K_\varepsilon$ with $\varepsilon = 1/n$ and the Lipschitz constant is at most $n/r$. Moreover, if $\mathcal{K}$ is in (approximate) isotropic or John's position, then $\mathbb{B}_1 \subset \mathcal{K} \subset \mathbb{B}_{2d}$ by Kannan et al. [1995] and John's theorem, respectively.

## 13.6 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 [Kaufmann et al., 2012, Agrawal and Goyal, 2012] and linear bandits [Agrawal and Goyal, 2013] and is out of reach of the techniques developed here. On the other hand, the Bayesian algorithm has the advantage of being able to specify a prior that makes use of background knowledge and the theoretical guarantees for TS provide a degree of comfort.

## 13.7 Choice of prior

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

# References


M. Abeille and A. Lazaric. Linear Thompson sampling revisited. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 176–184, Fort Lauderdale, FL, USA, 2017. JMLR.org.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Colt*, pages 28–40. Citeseer, 2010.

S. Agrawal and N. Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Proceedings of the 25th Conference on Learning Theory*, 2012.

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *Proceedings of the 30th International Conference on Machine Learning*, pages 127–135, Atlanta, GA, USA, 2013. JMLR.org.

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

Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://web.stanford.edu/~boyd/cvxbook/.

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

S. Bubeck, O. Dekel, T. Koren, and Y. Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Proceedings of the 28th Conference on Learning Theory*, pages 266–278, Paris, France, 2015. JMLR.org.

S. Bubeck, Y.T. Lee, and R. Eldan. Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 72–85, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4528-6.

I-S. Chang, L-C. Chien, C. Hsiung, C-C. Wen, and Y-J. Wu. Shape restricted regression with random bernstein polynomials. *Lecture Notes-Monograph Series*, pages 187–202, 2007.

V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Conference on Learning Theory*, pages 355–366, 2008.

S. Dong, T. Ma, and B. Van Roy. On the performance of thompson sampling on logistic bandits. In *Conference on Learning Theory*, pages 1158–1160. PMLR, 2019.

S. Filippi, O. Cappé, A. Garivier, and Cs. Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, pages 586–594. Curran Associates, Inc., 2010.

A Flaxman, A Kalai, and HB McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *SODA'05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394, 2005.

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

H. Fokkema, D. van der Hoeven, T. Lattimore, and J. Mayo. Improved online Newton method for bandit convex optimisation. 2024b.

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.

Fritz John. Extremum problems with inequalities as subsidiary conditions. pages 187–204, 1948.

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

E. Kaufmann, N. Korda, and R. Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory*, volume 7568 of *Lecture Notes in Computer Science*, pages 199–213. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-34105-2.

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

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

B. Kveton, M. Zaheer, Cs. Szepesvári, L. Li, M. Ghavamzadeh, and C. Boutilier. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2066–2076. PMLR, 2020.

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

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

T. Lattimore. *Bandit Convex Optimisation*. 2024.

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

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

László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

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

P. Ramgopal, P. Laud, and A. Smith. Nonparametric bayesian bioassay with prior constraints on the shape of the potency curve. *Biometrika*, 80(3):489–498, 1993.

D. Russo and B. Van Roy. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems*, pages 1583–1591. Curran Associates, Inc., 2014.

D. Russo and B. Van Roy. An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17(1):2442–2471, 2016. ISSN 1532-4435.

D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen. A tutorial on Thompson sampling. *Foundations and Trends in Machine Learning*, 11(1):1–96, 2018.

A. Saha, N. Natarajan, P. Netrapalli, and P. Jain. Optimal regret algorithm for pseudo-1d bandit convex optimization. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9255–9264. PMLR, 2021.

T. Shively, S. Walker, and P. Damien. Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. *Journal of Econometrics*, 161(2):166–181, 2011.

Robert L. Smith. Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308, 1984.

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

W. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

T. Tkocz. Asymptotic convex geometry lecture notes. 2018.

Michael J Todd. *Minimum-volume ellipsoids: Theory and algorithms*. SIAM, 2016.

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.