# Thompson Sampling for Bandit Convex Optimisation

Alireza Bakhtiari[1] Tor Lattimore[2] Csaba Szepesvári[1,2]

University of Alberta[1] & Google DeepMind[2]

COLT 2025

# Paper in One Slide

- Goal: Understand how **Thompson Sampling (TS)** behaves in **bandit convex optimization (BCO)**.
- **Positive**:
  - $\mathrm{BReg}_n = \widetilde{O}(\sqrt{n})$ in *1-D* convex bandits.
  - $\mathrm{BReg}_n = \widetilde{O}(d^{2.5}\sqrt{n})$ for *d-D monotone convex ridge* losses.
- **Negative**:
  - TS can suffer $\Omega(\exp(d))$ regret for general d-D convex losses.
  - Classical info-ratio tricks $\Rightarrow$ no better than $\widetilde{O}(d^{1.5}\sqrt{n})$ for general adversarial BCO.

# Problem Setup

- Convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- $\mathcal{F}$ be a set of convex functions from $\mathcal{K}$ to $[0,1]$.
- $\xi$ is a known prior on $\mathcal{F}$.
- At round $t$:
  1. Learner plays $X_t \in \mathcal{K}$.
  2. Observes scalar loss $Y_t \in \{0,1\}$[1].
- $\mathbb{E}\left[Y_t | X_1, Y_1, \cdots, X_t, f\right] = f(X_t)$.
- $\mathcal{A}$ is a possibly random mapping from histories & prior to $\mathcal{K}$.
- **Bayesian Regret**

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

where $f \sim \xi$.

[1] The Bernoulli noise assumption is just for convenience.

# Thompson Sampling for BCO

**Algorithm 1** Thompson Sampling

1: Prior $\xi$ over convex losses.
2: **for** $t = 1, \dots, n$ **do**
3: Sample $f_t \sim \mathbb{P}(f = \cdot | X_1, Y_1, \cdots, X_{t-1}, Y_{t-1})$
4: Play $x_t \in \arg\min_{x \in \mathcal{K}} f_t(x)$ and observe $Y_t$
5: **end for**

- Our analysis studies how

$$\mathrm{BReg}^{\mathrm{TS}}(\mathcal{F}) = \sup_{\xi \in \mathcal{P}(\mathcal{F})} \mathrm{BReg}_n(\mathrm{TS}, \xi)\,,$$

behaves for various natural classes of convex functions $\mathcal{F}$.

# Upper bounds - main results

- Let $\mathcal{F}_{\mathtt{bl}}$ be the space of all 1-Lipschitz bounded convex functions $f : \mathcal{K} \to [0, 1]$.

**Theorem 1 – 1-D convex functions**

TS achieves

$$\mathrm{BReg}^{\mathrm{TS}}(\mathcal{F}_{\mathtt{bl}}) \;=\; \widetilde{O}\big(\sqrt{n}\big)\,.$$

- Let $\mathcal{F}_{\mathtt{blrm}}$ be the space of all 1-Lipschitz bounded convex functions $f : \mathcal{K} \to [0, 1]$, such that there exists a monotone convex function $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{R}^d$ such that $f(x) = \ell(\langle x, \theta \rangle)$.

**Theorem 2 – d-D convex ridge functions**

TS achieves

$$\mathrm{BReg}^{\mathrm{TS}}(\mathcal{F}_{\mathtt{blrm}}) \;=\; \widetilde{O}\big(d^{2.5}\sqrt{n}\big)\,.$$

# Upper bounds – Analysis

(Generalized) **Information Ratio (IR)**

- Let $(X, f)$ have law $\pi \otimes \xi$.
- Let $\bar{f} = \mathbb{E}[f]$, and $f_\star = \min_{x \in \mathcal{K}} f(x)$.
- Define

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

**Generalized IR**

Define generalized information ratio associated with $\mathrm{TS}$ as

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

- Note that $(0, \beta) \in \mathrm{IR}(\mathcal{F})$ is equivalent to $\Delta(\pi_{\mathrm{TS}}, \xi)^2 / \mathcal{I}(\pi_{\mathrm{TS}}, \xi) \leq \beta$, for all $\xi \in \mathcal{P}(\mathcal{F})$.

# Upper bounds – Regret & IR

**Proposition 3 – IR regret bound**

Suppose that $\mathcal{F} \in \{\mathcal{F}_{\mathtt{bl}}, \mathcal{F}_{\mathtt{blrm}}\}$ and $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F})$. Then, the regret of TS is at most

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

- Theorems 1,2 follow from Proposition 3 and bounding IR.
- Proposition 3 is somewhat subtle to prove though:
  1. The space of all convex function is not parametric; therefore, some form of a cover on $\mathcal{F}$ is needed to bound the entropy CITE.
  2. Notably, usual techniques rely on $\mathcal{F}$ being closed under convex combination. The space of convex ridge functions are not closed under convex combination as noticed by CITE.

# Upper bounds – Convex Cover

**Convex cover (informal)**

Let $\tilde{x}_f \approx \operatorname{argmin}_{x \in \mathcal{K}}$. Define $N(\mathcal{F}, \epsilon)$ to be the smallest number $N$ such that there exists $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ such that for all $k \in [N]$:

- *Closure:* $\mathcal{F}_k$ is a subset of $\mathcal{F}$ and $\operatorname{conv}(\mathcal{F}_k) \subset \mathcal{F}$.
- *Common near-minimiser:* There exists an $x_k \in \mathcal{K}$ such that $\|\tilde{x}_f - x_k\| \leq \epsilon$ for all $f \in \mathcal{F}_k$.
- *Approximation:* For all $f \in \mathcal{F}$ there exists a $k \in [N]$ and $g \in \mathcal{F}_k$ such that $\|f - g\|_\infty \leq \epsilon$ and $\|\tilde{x}_f - x_k\| \leq \epsilon$.

# Upper bounds – Convex Cover

## Convex cover (informal)

Let $\tilde{x}_f \approx \text{argmin}_{x \in \mathcal{K}}$. Define $N(\mathcal{F}, \epsilon)$ to be the smallest number $N$ such that there exists $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ such that for all $k \in [N]$:

- *Closure:* $\mathcal{F}_k$ is a subset of $\mathcal{F}$ and $\text{conv}(\mathcal{F}_k) \subset \mathcal{F}$.
- *Common near-minimiser:* There exists an $x_k \in \mathcal{K}$ such that $\|\tilde{x}_f - x_k\| \leq \epsilon$ for all $f \in \mathcal{F}_k$.
- *Approximation:* For all $f \in \mathcal{F}$ there exists a $k \in [N]$ and $g \in \mathcal{F}_k$ such that $\|f - g\|_\infty \leq \epsilon$ and $\|\tilde{x}_f - x_k\| \leq \epsilon$.

## Proposition 4 – Convex covering number

We have

$$\{N(\mathcal{F}_{\texttt{blrm}}, \epsilon)\ ,\ N(\mathcal{F}_{\texttt{bl}}, \epsilon)\} \subset O\left(d \log\left(\frac{\text{diam}(\mathcal{K})}{\epsilon}\right)\right) .$$

# Upper bounds – Decomposition Lemma

We introduce the following lemma to bound the IR of TS. ($x_f = \mathrm{argmin}_{x \in \mathcal{K}} f(x)$)

**Proposition 5 – TS-IR Decomposition**

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

$$\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f(x_f)) \leq \alpha + \sqrt{\beta \inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j,l \in \mathrm{PAIR}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2},$$

for all $i$, then $(\alpha, k(k-1)m\beta) \in \mathrm{IR}(\mathcal{F})$.

- The supremum term is the worst case *average* regret within $\mathcal{F}_i$.

# Upper bounds – Decomposition Lemma

We introduce the following lemma to bound the IR of $\textsc{ts}$. ($x_f = \operatorname{argmin}_{x \in \mathcal{K}} f(x)$)

**Proposition 5 – $\textsc{ts}$-IR Decomposition**

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

$$\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f(x_f)) \leq \alpha + \sqrt{\beta \inf_{f_1, \ldots, f_k \in \mathcal{F}_i} \sum_{j,l \in \textsc{pair}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$, then $(\alpha, k(k-1)m\beta) \in \textsc{ir}(\mathcal{F})$.

- The supremum term is the worst case *average* regret within $\mathcal{F}_i$.
- The infimum term is a kind of bound on the minimum amount of information obtained by $\textsc{ts}$.

# Upper bounds – Decomposition Lemma

We introduce the following lemma to bound the IR of $\textsc{ts}$. ($x_f = \operatorname{argmin}_{x \in \mathcal{K}} f(x)$)

**Proposition 5 – $\textsc{ts}$-IR Decomposition**

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

$$\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f(x_f)) \leq \alpha + \sqrt{\beta \inf_{f_1,\ldots,f_k \in \mathcal{F}_i} \sum_{j,l \in \textsc{pair}(k)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2},$$

for all $i$, then $(\alpha, k(k-1)m\beta) \in \textsc{ir}(\mathcal{F})$.

- The supremum term is the worst case *average* regret within $\mathcal{F}_i$.
- The infimum term is a kind of bound on the minimum amount of information obtained by $\textsc{ts}$.
- The price of $m$ comes from the same Cauchy-Schwarz(CS) that is somehow the 'same' CS in the multi-armed setting.

# Upper bounds – IR bound in 1-D

**Theorem 6 – IR bound for 1-D**

Suppose that $d = 1$ and $\alpha \in (0, 1)$. Then $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \mathrm{IR}(\mathcal{F}_{\mathtt{bl}})$.

**Proof.** We use our decomposition lemma with

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

and the proof is done through Prop 5 if we show

$$\sup_{f \in \mathcal{F}_i} \left(\bar{f}(x_f) - f_\star\right) \leq \epsilon_i \leq \alpha + \sqrt{230 \inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j, l \in \mathrm{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

# Upper bounds – IR bound in 1-D

**Theorem 6 – IR bound for 1-D**

Suppose that $d = 1$ and $\alpha \in (0, 1)$. Then $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \mathrm{IR}(\mathcal{F}_{\mathtt{bl}})$.

**Proof.** We use our decomposition lemma with

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

and the proof is done through Prop 5 if we show

$$\sup_{f \in \mathcal{F}_i} \left(\bar{f}(x_f) - f_\star\right) \underbrace{\leq}_{immediate} \epsilon_i \leq \alpha + \sqrt{230 \inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j, l \in \mathrm{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

# Upper bounds – IR bound in 1-D

**Theorem 6 – IR bound for 1-D**

Suppose that $d = 1$ and $\alpha \in (0,1)$. Then $(\alpha, 10^4 \lceil \log(1/\alpha) \rceil) \in \mathrm{IR}(\mathcal{F}_{\mathtt{bl}})$.

**Proof.** We use our decomposition lemma with

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

and the proof is done through Prop 5 if we show

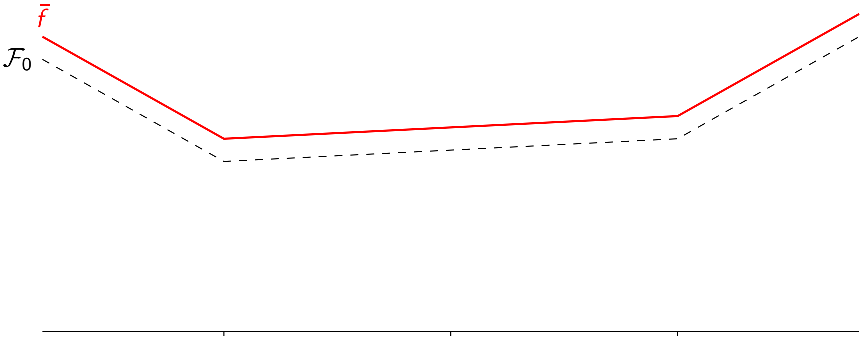
$$\epsilon_i \underbrace{\le}_{\text{remaining...}} \alpha + \sqrt{230 \inf_{f_1,f_2,f_3,f_4 \in \mathcal{F}_i} \sum_{j,l \in \mathrm{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

# Upper bounds – IR bound in 1-D

How does it look like?

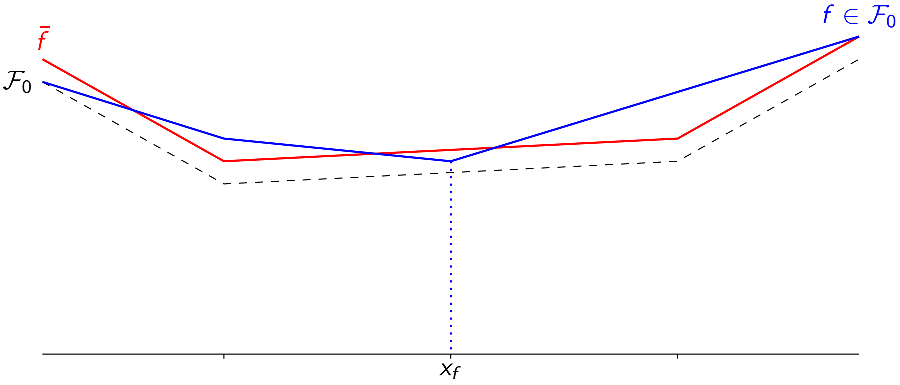
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \le \alpha\} & \text{if } i = 0 \\ & \end{cases}$$



# Upper bounds – IR bound in 1-D

How does it look like?

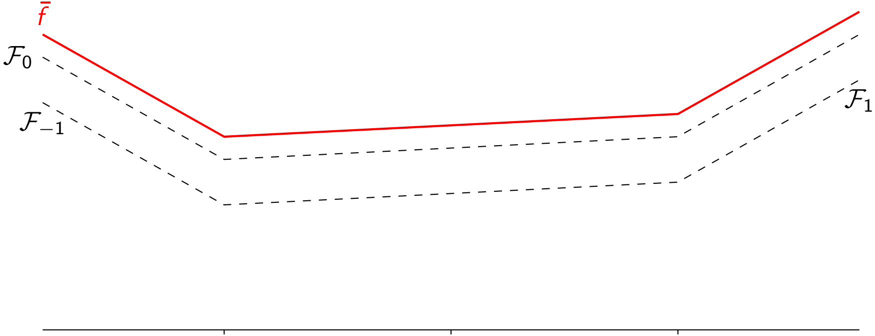
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \leq \alpha\} & \text{if } i = 0 \\ & \end{cases}$$



# Upper bounds – IR bound in 1-D

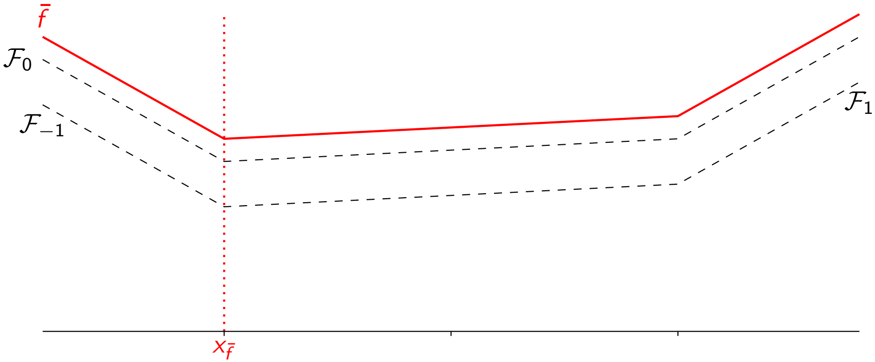
How does it look like?

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



How does it look like?

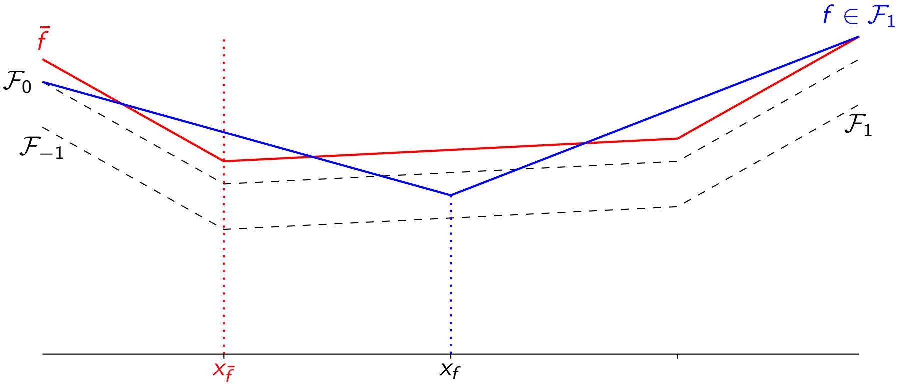
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \leq \alpha\} & \text{if } i = 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f \geq x_{\bar{f}}\} & \text{if } i > 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f < x_{\bar{f}}\} &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



# Upper bounds – IR bound in 1-D

How does it look like?

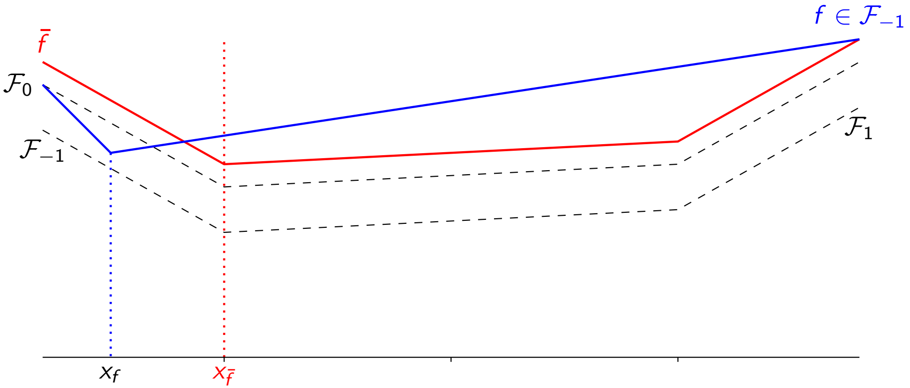
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \le \alpha\} & \text{if } i = 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\, x_f \ge x_{\bar{f}}\} & \text{if } i > 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\, x_f < x_{\bar{f}}\} &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



# Upper bounds – IR bound in 1-D

How does it look like?

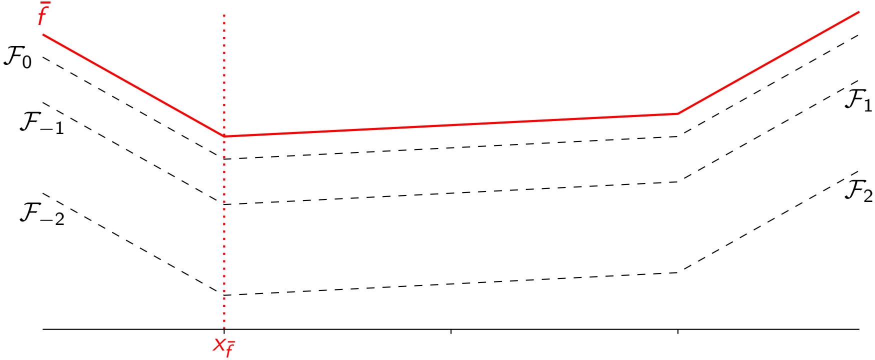
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \le \alpha\} & \text{if } i = 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f \ge x_{\bar{f}}\} & \text{if } i > 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f < x_{\bar{f}}\} & \text{if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



# Upper bounds – IR bound in 1-D

How does it look like?

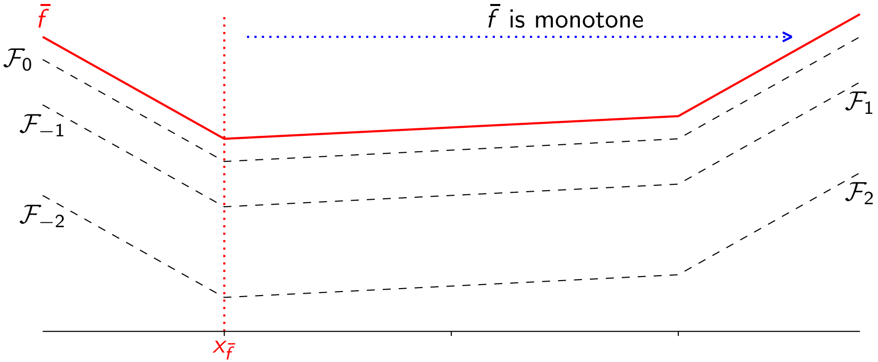
$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \leq \alpha\} & \text{if } i = 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f \geq x_{\bar{f}}\} & \text{if } i > 0 \\ \{f : \bar{f}(x_f) - f(x_f) \in (\alpha 2^{|i|-1}, \alpha 2^{|i|}],\ x_f < x_{\bar{f}}\} &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



# Upper bounds – IR bound in 1-D

How does it look like?

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



**Back to the proof**

$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \leq \alpha\} & \text{if } i = 0 \\ & \\ & \end{cases}$$

... and the proof is done through Prop 5 if we show

$$\epsilon_i \leq \alpha + \sqrt{230 \inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j,l \in \mathrm{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2},$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

**Back to the proof**

$$\mathcal{F}_i = \begin{cases} \{f : \bar{f}(x_f) - f(x_f) \leq \alpha\} & \text{if } i = 0 \\ & \end{cases}$$

... and the proof is done through Prop 5 if we show

$$\epsilon_i \leq \alpha + \sqrt{230 \inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j, l \in \text{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

**Which trivially holds for $i = 0$ since $\epsilon_0 = \alpha$.**

**Back to the proof**

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

... and the proof is done through Prop 5 if we show

$$\epsilon_i \leq \alpha + \sqrt{230 \inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j, l \in \mathrm{PAIR}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2}\,,$$

for all $i$ show that with $\epsilon_i = \alpha 2^{|i|}$.

**For $i \neq 0$ we can use the monotonicity of $\bar{f}$.**

# Upper bounds – IR bound in 1-D

**Back to the proof for $i \neq 0$ ($\Rightarrow$ monotone $\bar{f}$)**

- Only 4 functions are enough to get non-negligible disparity.
- The proof follows by contradiction. Suppose that $x_1 \leq x_2 \leq x_3 \leq x_4$ and

$$\sum_{j,l \in \textsc{pair}(4)} (f_j(x_{f_l}) - \bar{f}(x_{f_l}))^2 < c^2 \epsilon^2 \quad \text{for some } c\,,$$

  which means that all $f_1, f_2, f_3, f_4$ are $\epsilon$ close to $\bar{f}$ at all the points $x_1, x_2, x_3, x_4$.
- This has to happen while $\bar{f}(x_j) - f_j(x_j) \leq \epsilon_i$ for all $j \in [4]$.
- The convexity of $f_1, f_2, f_3, f_4$ and the convexity+monotonicity of $\bar{f}$ in $[x_1, x_4]$ gives the contradiction.

# Upper bounds – IR bound in d-D

**Theorem 7 – IR bound in d-D**

$(\alpha, \beta \lceil \log(1/\alpha) \rceil) \in \mathrm{IR}(\mathcal{F}_{\texttt{blrm}})$ whenever $\alpha \in (0, 1)$ and

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

with the Big-O hiding only a universal constant.

# When TS Fails (High-D General Convex)

**Lower Bound 1**

There is a prior on bounded Lipschitz convex functions such that

$$\mathrm{BReg}_T(\mathrm{TS}) \;\geq\; \tfrac{1}{2}\min\Big\{T,\; e^{\Omega(d)}\Big\}.$$

- Construction: hide a sharp "valley" in a random direction.
- TS keeps sampling valleys it hasn't discovered → linear regret until exponential time.

# Information-Ratio Barrier

**Lower Bound 2**

For general convex losses the classical info-ratio machinery cannot give regret better than $\widetilde{O}(d^{1.5}\sqrt{T})$.

- Matches best known algorithmic upper bound.
- Suggests new ideas needed for $\widetilde{O}(d\sqrt{T})$ in adversarial BCO.

# Conclusion

- Thompson sampling Bayesian upper bound for 1-D convex bandits.
- Structured high-D (monotone ridge) $\Rightarrow \tilde{O}(d^{2.5}\sqrt{n})$ regret.
- TS suffers *exponential in d* for general convex losses in $\mathbb{R}^d$.
- Classic IR analysis can't be used to beat the $O(d^{1.5}\sqrt{n})$ upper bound for adversarial BCO.

**Thanks!**