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**:
  - $\mathrm{BReg}_n(\mathrm{TS}) = \widetilde{O}(\sqrt{n})$ in *1-D* convex bandits.
  - $\mathrm{BReg}_n(\mathrm{TS}) = \widetilde{O}\big(d^{2.5}\sqrt{n}\big)$ 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 $\widetilde{\Omega}(d^{1.5}\sqrt{n})$ for general stochastic & 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 $[0, 1]$.
- ξ is a known prior on $\mathcal{F}$.

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

Problem Setup

- Convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- $\mathcal{F}$ be a set of convex functions from $\mathcal{K}$ to $[0, 1]$.
- ξ 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]

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

Problem Setup

- Convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- $\mathcal{F}$ be a set of convex functions from $\mathcal{K}$ to $[0, 1]$.
- ξ 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]
- $\mathbb{E}\left[Y_t | X_1, Y_1, \cdots, X_t, f\right] = f(X_t)$, where $f \in \mathcal{F}$ is convex.

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

Problem Setup

- Convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- $\mathcal{F}$ be a set of convex functions from $\mathcal{K}$ to $[0, 1]$.
- ξ 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]
- $\mathbb{E}[Y_t | X_1, Y_1, \cdots, X_t, f] = f(X_t)$, where $f \in \mathcal{F}$ is convex.
- $\mathcal{A}$ (learner) is a mapping from histories & prior to $\mathcal{K}$.

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

Problem Setup

- Convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- $\mathcal{F}$ be a set of convex functions from $\mathcal{K}$ to $[0,1]$.
- ξ 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]
- $\mathbb{E}\left[Y_t | X_1, Y_1, \cdots, X_t, f\right] = f(X_t)$, where $f \in \mathcal{F}$ is convex.
- $\mathcal{A}$ (learner) is a 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 ξ over convex losses.
2: **for** $t = 1, \ldots, n$ **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

- Positive results for two sets of functions $f : \mathcal{K} \to [0,1]$:

$\mathcal{F}_{\texttt{bl}}$: $\mathcal{K} \subset \mathbb{R}$ and f is a 1-Lip and bounded convex function,

$\mathcal{F}_{\texttt{blrm}}$: $\mathcal{K} \subset \mathbb{R}^d$ and $f(x) = \ell(\langle x, \theta \rangle)$ for some $\theta \in \mathbb{R}^d$ and a convex 1-Lip bounded *monotone* $\ell : \mathbb{R} \to \mathbb{R}$.

Upper bounds – 1-D convex functions

TS achieves

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

And

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

Upper bounds – analysis

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

- Let $x_g = \operatorname{argmin}_{x \in \mathcal{K}} g(x)$. (analysis is really about x_f and $f(x_f)$.)
- Let $(F_1, F_2) \sim \xi \otimes \xi$ and $\bar{f} = \mathbb{E}[F_1]$.
- Define

$$\Delta(\xi) = \mathbb{E}[\bar{f}(x_{F_1}) - F_1(x_{F_1})] \quad \text{and} \quad \mathcal{I}(\xi) = \mathbb{E}[(F_1(x_{F_2}) - \bar{f}(X_{F_2}))^2]\,.$$

Generalized IR

Define generalized information ratio associated with π as

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

- $(0, \beta) \in \mathrm{IR}(\mathcal{F}) \Rightarrow \Delta(\xi)^2 / \mathcal{I}(\xi) \leq \beta$.

Upper bounds – regret & IR

Proposition 1 – regret bound through IR

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

$$\mathrm{BReg}_n(\mathrm{TS}, \xi_0) \leq n\alpha + O\left(\sqrt{\beta n d \log(n \operatorname{diam}(\mathcal{K}))}\right) .$$

Upper bounds – regret & IR

Proposition 1 – regret bound through IR

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

$$\mathrm{BReg}_n(\mathrm{TS}, \xi_0) \leq n\alpha + O\left(\sqrt{\beta n d \log(n \operatorname{diam}(\mathcal{K}))}\right) .$$

It remains to bound the IR of TS.

Upper bounds – regret & IR

Proposition 1 – regret bound through IR

Suppose that $\mathcal{F} \in \{\mathcal{F}_{\texttt{bl}}, \mathcal{F}_{\texttt{blrm}}\}$ and $(\alpha, \beta) \in \textsc{ir}(\mathcal{F})$. Then, the regret of $\textsc{ts}$ is at most

$$\mathrm{BReg}_n(\textsc{ts}, \xi_0) \leq n\alpha + O\left(\sqrt{\beta n d \log(n \,\mathrm{diam}(\mathcal{K}))}\right) .$$

Proposition 2 – $\textsc{ts}$-IR Partitioning

If $\exists k, m \in \mathbb{N}$ such that for all $\tilde{f} \in \mathrm{conv}(\mathcal{F})$ there is $\mathcal{F} = \cup_{i=1}^{m} \mathcal{F}_i$ for which

$$\sup_{f \in \mathcal{F}_i} (\tilde{f}(x_f) - f(x_f)) \leq \alpha + \sqrt{\beta \underbrace{\inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j \neq l \in [k]} (f_j(x_{f_l}) - \tilde{f}(x_{f_l}))^2}_{\textsc{dis}(\mathcal{F}_i, k)}} ,$$

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

Upper bounds – regret & IR

Proposition 1 – regret bound through IR

Suppose that $\mathcal{F} \in \{\mathcal{F}_{\texttt{bl}}, \mathcal{F}_{\texttt{blrm}}\}$ and $(\alpha, \beta) \in \textsc{ir}(\mathcal{F})$. Then, the regret of $\textsc{ts}$ is at most

$$\mathrm{BReg}_n(\textsc{ts}, \xi_0) \leq n\alpha + O\left(\sqrt{\beta n d \log(n \operatorname{diam}(\mathcal{K}))}\right).$$

Proposition 2 – $\textsc{ts}$-IR Partitioning

If $\exists k, m \in \mathbb{N}$ such that for all $\tilde{f} \in \operatorname{conv}(\mathcal{F})$ there is $\mathcal{F} = \cup_{i=1}^m \mathcal{F}_i$ for which

$$\boxed{\sup_{f \in \mathcal{F}_i} (\tilde{f}(x_f) - f(x_f))} \leq \alpha + \sqrt{\underbrace{\beta \inf_{f_1,\ldots,f_k \in \mathcal{F}_i} \sum_{j \neq l \in [k]} (f_j(x_{f_l}) - \tilde{f}(x_{f_l}))^2}_{\textsc{dis}(\mathcal{F}_i, k)}},$$

worst case *average* regret within $\mathcal{F}_i$

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

Upper bounds – regret & IR

Proposition 1 – regret bound through IR

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

$$\mathrm{BReg}_n(\mathrm{TS}, \xi_0) \leq n\alpha + O\left(\sqrt{\beta n d \log(n\,\mathrm{diam}(\mathcal{K}))}\right).$$

Proposition 2 – TS-IR Partitioning

If $\exists k, m \in \mathbb{N}$ such that for all $\tilde{f} \in \mathrm{conv}(\mathcal{F})$ there is $\mathcal{F} = \cup_{i=1}^m \mathcal{F}_i$ for which

$$\boxed{\sup_{f \in \mathcal{F}_i} (\tilde{f}(x_f) - f(x_f))} \leq \alpha + \sqrt{\beta \boxed{\underbrace{\inf_{f_1, \dots, f_k \in \mathcal{F}_i} \sum_{j \neq l \in [k]} (f_j(x_{f_l}) - \tilde{f}(x_{f_l}))^2}_{\mathrm{DIS}(\mathcal{F}_i, k)}}},$$

worst case *average* regret within $\mathcal{F}_i$

TS minimum inf. gain (discrepancy)

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

Upper bounds – IR bound in 1-D

Theorem 3 – IR bound for 1-D

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

Proof sketch:

1. Use the partitioning lemma $\Rightarrow$ partition based on x_f and $f(x_f)$.
2. Group functions with similar gaps (average regret) in $\mathcal{F}_i$:

   Similar gaps: $\tilde{f}(x_f) - f(x_f) \in [\epsilon, 2\epsilon]$,
   $\tilde{f}$ monotone: $x_{\tilde{f}} \leq x_f \Rightarrow \tilde{f}$ (begin 1-D is crucial here).
3. Prove that $\mathrm{DIS}(\mathcal{F}_i, k) \geq c^2 \epsilon_i^2$:

$$\mathrm{DIS}(\mathcal{F}_i, k) = \inf_{f_1, \ldots, f_k \in \mathcal{F}_i} \sum_{j \neq l \in [k]} (f_j(x_{f_l}) - \tilde{f}(x_{f_l}))^2 \geq c^2 \epsilon_i^2 \, .$$

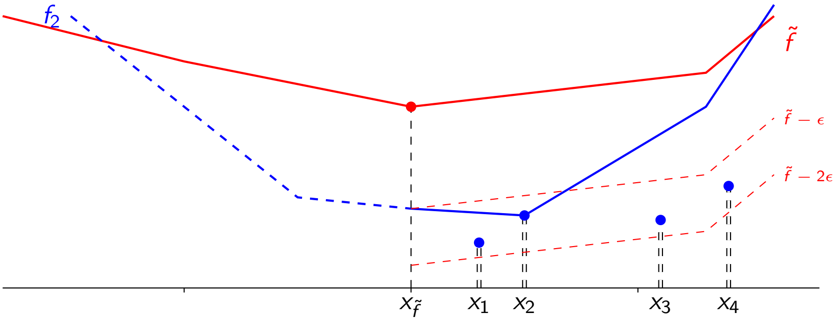
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\inf_{f_1, f_2, f_3, f_4 \in \mathcal{F}_i} \sum_{j \neq l \in [4]} (f_j(x_{f_l}) - \tilde{f}(x_{f_l}))^2 \leq c^2 \epsilon_i^2 \Rightarrow |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i, \forall j \neq l \in [4].$$

- WLOG, suppose that $x_{\tilde{f}} \leq x_1 \leq x_2 \leq x_3 \leq x_4$, for $x_i \triangleq x_{f_i}$.



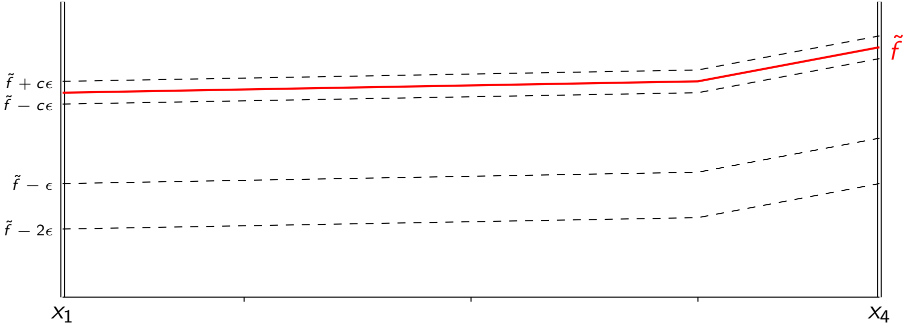
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \;\; |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \,.$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.



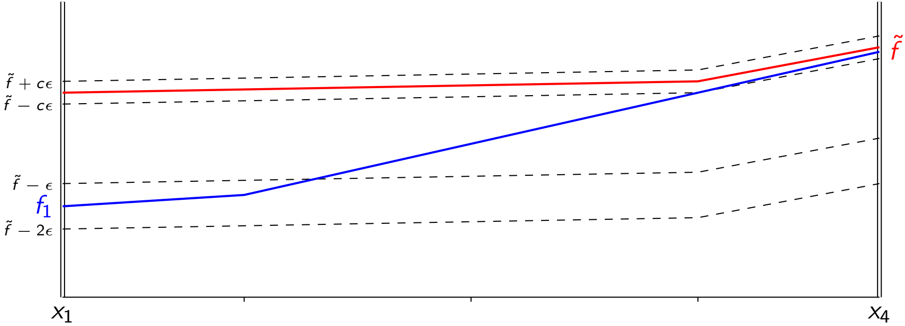
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \;\; |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \, .$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.



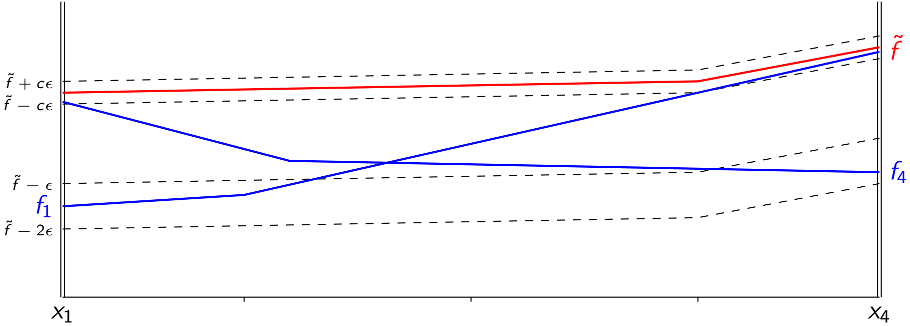
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \;\; |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \,.$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.



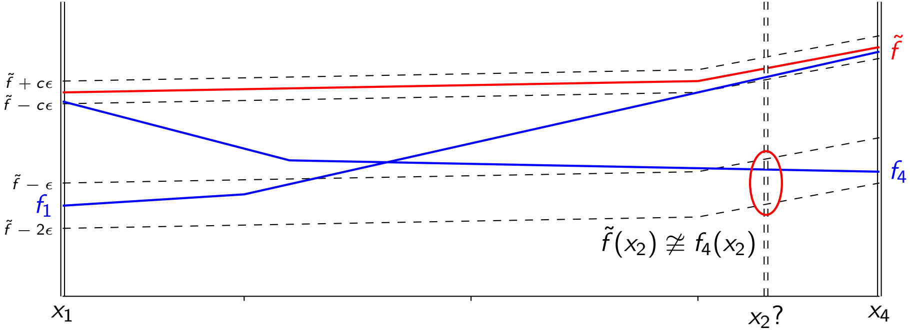
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \ \ |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \, .$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.



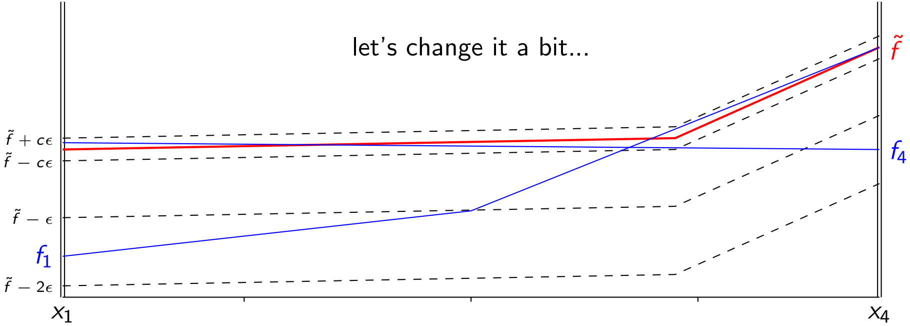
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \ \ |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \, .$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.

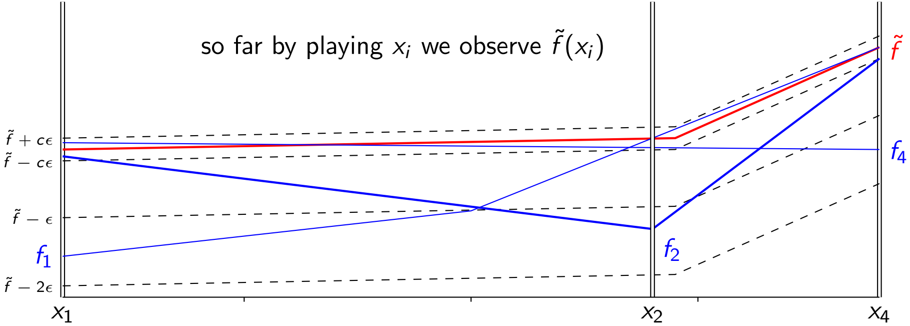


Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \ \ |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \, .$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.



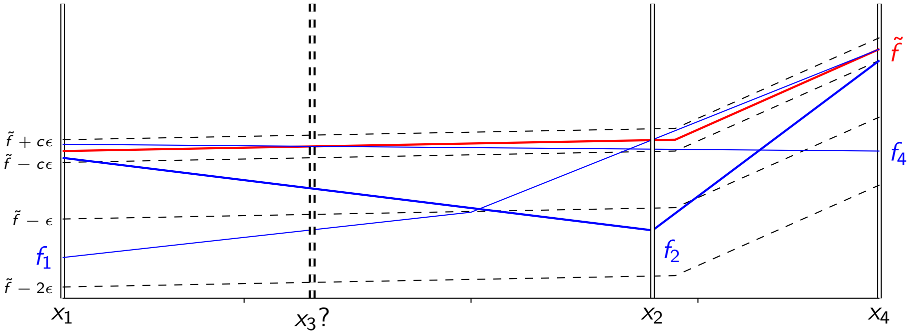
Upper bounds – IR bound in 1-D

Proof sketch:(proof by contradiction)

- Fix $k = 4, c \approx 1/16$, suppose that

$$\forall j \neq l \in [4], \;\; |f_j(x_{f_l}) - \tilde{f}(x_{f_l})| \leq c\epsilon_i \, .$$

- f_j needs to be close to $\tilde{f}$, at minimizers x_{f_l} (double lines $||$) for $l \neq j$.

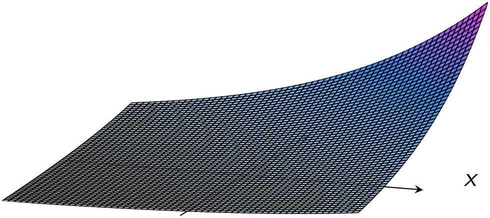


Upper bounds – IR bound in d-D

Theorem 7 – IR bound in d-D

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

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



$f(x,y) = e^{x+y}$ — a convex ridge function with $\theta = (1,1)$ and $\ell(t) = e^t$.

Lower bounds – main results

The following lower bounds share the same construction.

Lower Bound 4

When $\mathcal{K} = \mathbb{B}_1 \subset \mathbb{R}^d$, there exists a prior ξ on $\mathcal{F}_{\mathtt{bl}}$ such that

$$\mathrm{BReg}_n(\mathrm{TS}) \;\geq\; \tfrac{1}{2} \min\Big\{n,\; \frac{1}{4} e^{\Omega(d/32)}\Big\}.$$

Lower Bound 2

Suppose that $\mathcal{K} = \mathbb{B}_1 \subset \mathbb{R}^d$ and $d > 256$. Then there exists a prior ξ on $\mathcal{F}_{\mathtt{bl}}$ such that

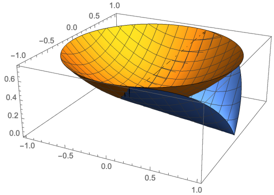
$$\Delta(\pi, \xi) \geq 2^{-19} \frac{d}{\log(d)} \sqrt{\mathcal{I}(\pi, \xi)}\,,$$

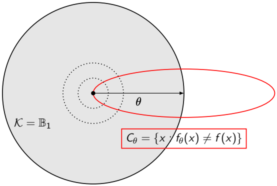
for all policies π on K.

Lower bound 2 shows that for general convex losses in $\mathbb{R}^d$ the **classical info-ratio machinery** cannot give regret better than the known $\widetilde{O}(d^{1.5}\sqrt{n})$.

Lower bounds – construction

- Idea:
  1. All $f_\theta \sim \xi$ are mostly equal to a fixed f on $\mathbb{B}_1$.
  2. Uniform prior and no noise!
- TS lower bound:
  1. TS plays on the "perimeter" $\mathbb{S}_1$ where the minimizers are.
  2. Suffers constant regret for $\exp(d)$ rounds.
- IR lower bound:
  1. Radial symmetry.
  2. Optimize for r in $\frac{\Delta(r)^2}{\mathcal{I}(r)}$:
     1. Closer to 0, very low inf. with good prob.
     2. Closer to $\mathbb{S}_1$, high inf. with very low prob.





Paper in One Slide

- Goal: Understand how **Thompson Sampling (TS)** behaves in **bandit convex optimization (BCO)**.
- **Positive**:
  - $\mathrm{BReg}_n(\text{TS}) = \widetilde{O}(\sqrt{n})$ in *1-D* convex bandits.
  - $\mathrm{BReg}_n(\text{TS}) = \widetilde{O}\big(d^{2.5}\sqrt{n}\big)$ 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 stochastic & adversarial BCO.

Thanks!