BEYOND MONOTONICITY: THOMPSON SAMPLING FOR CONVEX RIDGE BANDITS WITH POLYNOMIAL BAYESIAN REGRET

Anonymous authors
Paper under double-blind review

ABSTRACT

Thompson sampling (TS) is a canonical algorithm for Bayesian bandit optimization, yet its theoretical guarantees in nonlinear bandits remain confined to strongly structured settings. In particular, recent work established polynomial Bayesian regret for convex ridge bandits under a monotonicity assumption on the unknown convex link function and posed the open question of whether monotonicity is necessary. We answer this question negatively. We study Bayesian bandit convex optimization with convex ridge losses of the form $f(x) = \ell(\langle x, \theta \rangle)$, where the link ℓ is convex but not necessarily monotone. For arbitrary priors over bounded, Lipschitz convex ridge losses, any fixed measurable selection of minimizers, and exact-posterior Thompson sampling, we prove a polynomial-in-dimension Bayesian regret bound of $\widetilde{\mathcal{O}}(d^{9/2}\sqrt{n})$. The single-removal John-ellipsoid argument underlying the monotone analysis does not extend to the non-monotone setting. To overcome this obstacle, we develop a cardinality-based analysis of uninformative configurations that combines two-band geometry with a Boolean matrix-rounding argument, yielding an $\mathcal{O}(d^2)$ cardinality bound that is tight up to constants in the large-diameter-to-gap regime. Beyond resolving the open question, our results show that convex ridge structure alone is sufficient for exact-posterior Thompson sampling to achieve polynomial Bayesian regret without link monotonicity. This extends the class of nonlinear bandit problems for which Thompson sampling admits a provable Bayesian regret guarantee, while leaving the optimal dependence on dimension open.

1 INTRODUCTION

Single-index bandits model the expected loss of action x through a scalar link of an unknown linear projection $\langle x, \theta \rangle$. Standard generalised linear bandits typically assume known links; we study an unknown convex link in a Bayesian model. Thompson sampling (TS) plays a minimiser of a posterior sample each round. The polynomial-in-dimension guarantee of Bakhtiari et al. (2025) for exact TS assumes monotonicity; we remove this assumption for convex ridge losses in general dimension.

Using the notation of Section 2, Bakhtiari et al. (2025) show by information-ratio analysis that a prior ξ supported on *monotone* convex ridge functions, i.e. $f(x) = \ell(\langle x, \theta \rangle)$ with $\ell : \mathbb{R} \to \mathbb{R}$ convex and non-decreasing, gives $\mathrm{BReg}_n(\mathrm{TS}, \xi) = O(d^{2.5}\sqrt{n}\log^2(nd\,\mathrm{diam}K))$. They also show catastrophic TS failures on general convex losses and ask whether monotonicity is necessary in the ridge setting, suggesting it is not. The monotone ridge class is a Bayesian version of the generalised linear bandit with an unknown convex, increasing link. Without monotonicity, $\ell(s) = |s - s_0|$ has a whole hyperplane section as its minimising set on K when $d \geq 2$ and s_0 lies in the interior of the projection interval $\{\langle x, \theta \rangle : x \in K\}$. Thus minimisers can be non-unique and TS tie-breaking matters, while the ordering of minimisers along the ridge direction in the monotone analysis is lost.

Our question is algorithm-specific. Lattimore (2021) obtains $O(d\sqrt{n}\log(nD))$ minimax regret, under his domain normalisation, for adversarial ridge losses using another algorithm. Such learnability does not certify exact-posterior TS: Bakhtiari et al. (2025) show catastrophic failures on general high-dimensional convex losses. We identify convex ridge structure as sufficient without monotonicity, providing statistical foundations for posterior sampling with an unknown single-index loss.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Upper bounds only, not lower bounds or optimality claims: the certificate column is the β of the pair (α, β) each analysis constructs. Here $m_\alpha = 1 + \lceil \log_2(1/\alpha) \rceil$ for $0 < \alpha < 1$, and $\tilde{O}$ hides logarithms in n, d, $\mathrm{diam}K$ and, for the certificate of Bakhtiari et al. (2025), $1/\alpha$.

Convex ridge losses	Monotone link	Certificate β	Bayesian regret	Shown in
this paper	not required	$O((d+1)^8 m_\alpha)$	$\tilde{O}(d^{9/2}\sqrt{n})$	Thms. 3.4, 3.1
Bakhtiari et al. (2025)	required	$\tilde{O}(d^4)$	$\tilde{O}(d^{5/2}\sqrt{n})$	their Thms. 7 and 6

Our result concerns exact sampling and Bayesian regret, without claims of computational efficiency or minimax optimality.

Contributions. Let $\mathcal{F}_{\mathrm{blr}}$ denote the convex ridge functions $K \to [0,1]$ that are 1-Lipschitz, with arbitrary convex links (Section 2). The structural proof uses the masked matrix $M = U \circ H$ (Section 5), whose far-band entries vanish exactly.

- **Regret without monotonicity.** For every prior on $\mathcal{F}_{\mathrm{blr}}$ and every fixed measurable selection of minimisers, exact-posterior TS has Bayesian regret $\tilde{O}(d^{9/2}\sqrt{n})$ with an explicit constant (Theorem 3.1). This answers the qualitative question of Bakhtiari et al. (2025); matching the monotone $d^{5/2}$ dependence remains open (Section 9).
- **Structural core: cardinality.** The information-ratio analysis requires a bound on configurations whose cross-evaluations deviate from a common comparison function by less than δ, under Definition 3.2's gap and comparison-value conditions. For $c = 96(d+1)$ their size is at most $3(d+1)^2 - (d+1) - 1$ (Theorem 3.3), tight up to a constant in the large-diameter-to-gap regime (Proposition 3.5).
- **Boolean rounding and masking.** An elementary tool: a 0-1 matrix within $1/(4r)$ in max-norm of a rank-r matrix has rank at most $2r - 1$ (Lemma 5.3). This replaces the monotone proof's single-removal John-ellipsoid dichotomy. Proposition 3.6 refutes its literal non-monotone extension for every fixed shrink factor below one, without ruling out other ellipsoid arguments.
- **Transfer under arbitrary fixed selections.** To accommodate non-unique minimisers, we give a self-contained transfer from a selection-uniform information-ratio bound (Theorem 3.4) to Bayesian regret for exact-posterior TS under every fixed measurable selection (Theorem 7.2). History-dependent tie-breaking is not covered.

Related work. Rajaraman & Han (2026a) exhibit a non-monotone reward link on which their specified SGD dynamics and exploration provably stall in alignment (Proposition 1; Proposition 5.1 in the arXiv version). Their known-link reward-maximisation model and alignment guarantee differ from our convex-loss model and Bayesian regret criterion: this is a qualitative contrast, not a shared-class comparison. Luo & Bayati (2025) study frequentist regret of linear TS; Shufaro et al. (2025) give information-theoretic Bayesian regret bounds using accumulated information. Information-ratio analyses of TS go back to Russo & Van Roy (2016). See Appendix F for details. Our result resolves the main open question posed by Bakhtiari et al. (2025).

2 Setting and notation

Sequential decision problem. Bayesian bandit convex optimisation is the following game. A convex body $K \subset \mathbb{R}^d$ and a class $\mathcal{F}$ of convex functions $K \to [0,1]$ are known, together with a prior ξ on $\mathcal{F}$. The environment samples $f \sim \xi$ once. In each round $t = 1, \ldots, n$ the learner plays $X_t \in K$ and observes $Y_t \in \{0,1\}$ with $\mathbb{E}[Y_t \mid X_1, Y_1, \ldots, X_t, f] = f(X_t)$. The Bayesian regret of a learner $\mathcal{A}$ is

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

Thompson sampling (TS) samples f_t from the posterior in every round and plays a minimiser $X_t = x_{f_t}$ of the sampled function.

Function class. Throughout, $K \subset \mathbb{R}^d$ is a convex body (compact, convex, non-empty interior) with $0 \in K$, and $D = \operatorname{diam}(K)$. A function $f : K \to \mathbb{R}$ is a *convex ridge function* if $f(x) = \ell(\langle x, \theta \rangle)$ for some convex $\ell : \mathbb{R} \to \mathbb{R}$ and $\theta \in \mathbb{R}^d$ ($\theta = 0$ is allowed). We write

$$\mathcal{F}_{\mathrm{bl}} = \{f : K \to [0,1] \text{ convex}, \ \operatorname{Lip}(f) \le 1\}, \quad \mathcal{F}_{\mathrm{blr}} = \{f \in \mathcal{F}_{\mathrm{bl}} : \ f \text{ is a convex ridge function}\},$$

and $\mathcal{F}_{\mathrm{blrm}} \subset \mathcal{F}_{\mathrm{blr}}$ for the subclass with non-decreasing link. The classes $\mathcal{F}_{\mathrm{bl}}$ and $\mathcal{F}_{\mathrm{blrm}}$ are those of Bakhtiari et al. (2025); $\mathcal{F}_{\mathrm{blr}}$ is not treated there.

Measurability and tie-breaking (standing convention). We equip $\mathcal{F}_{\mathrm{blr}}$ with a σ-algebra $\mathcal{A}$ for which $f \mapsto f(x)$ is measurable for every $x \in K$ and a measurable selection $f \mapsto x_f \in \arg\min_K f$ has been fixed once and for all (such selections exist; see Bakhtiari et al. (2025, Appendix B)). Pointwise measurability makes the identity $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A}) \to (C(K), \mathcal{B})$ measurable, where $\mathcal{B}$ is the Borel σ-algebra of the uniform norm; $\mathcal{A}$ may be strictly finer than $\mathcal{B}$. All approximating functions constructed in Section 7 are measurable as $(C(K), \mathcal{B})$-valued maps, and there the information-ratio hypothesis is only applied to finitely-valued random elements of $\mathcal{F}_{\mathrm{blr}}$, which are $\mathcal{A}$-measurable automatically; this is what allows arbitrary priors and arbitrary fixed selections. All statements hold for *every* such selection rule, but the rule is fixed in advance: it may not depend on the history of the game. We write $f^\star = \min_K f = f(x_f)$. For a probability measure ξ on $\mathcal{F}_{\mathrm{blr}}$ we write $\bar{f} = \mathbb{E}_\xi[f]$ (pointwise), a convex function $K \to [0,1]$ which in general is *not* a finite convex combination of elements of $\mathcal{F}_{\mathrm{blr}}$ (for instance the average of $x \mapsto \max\{\langle x, \theta \rangle, 0\}$ over uniformly random directions in the plane is a multiple of the Euclidean norm); below we only use pointwise values of $\bar{f}$.

Thompson sampling. TS with prior ξ is Algorithm 1 of Bakhtiari et al. (2025): in round t sample f_t from the posterior $\mathbb{P}(f \in \cdot \mid X_1, Y_1, \ldots, X_{t-1}, Y_{t-1})$ and play $X_t = x_{f_t}$. The observation model is Bernoulli: $Y_t \in \{0,1\}$ with $\mathbb{E}[Y_t \mid X_1, Y_1, \ldots, X_t, f] = f(X_t)$.

Regret and information terms. Unless otherwise stated, a random function $F : K \to [0,1]$ is a Borel random element of $C(K)$ (so $(\omega, x) \mapsto F_\omega(x)$ is jointly measurable and $F^\star = \min_K F$ is measurable); laws on $\mathcal{F}_{\mathrm{bl}}$, and laws on $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$ through the identity map, are included. For the law ν of such a random function and a policy $\pi \in \mathcal{P}(K)$, let $(X, F) \sim \pi \otimes \nu$, $\bar{F} = \mathbb{E}_\nu[F]$, and

$$\Delta(\pi, \nu) = \mathbb{E}[\bar{F}(X) - F^\star], \qquad I(\pi, \nu) = \mathbb{E}\big[(F(X) - \bar{F}(X))^2\big].$$

For $\xi \in \mathcal{P}(\mathcal{F}_{\mathrm{blr}})$ let π^ξ_{TS} be the law of x_f under $f \sim \xi$. Following Bakhtiari et al. (2025),

$$\mathrm{IR}(\mathcal{F}) = \Big\{(\alpha, \beta) \in \mathbb{R}^2_+ : \sup_{\xi \in \mathcal{P}(\mathcal{F})} \big[\Delta(\pi^\xi_{\mathrm{TS}}, \xi) - \alpha - \sqrt{\beta I(\pi^\xi_{\mathrm{TS}}, \xi)}\big] \le 0\Big\}.$$

Since π^ξ_{TS} depends on the selection, so does $\mathrm{IR}(\mathcal{F})$; we say $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F})$ *uniformly* if it holds for every measurable selection.

The decomposition lemma. The following is a restatement, for the prior mean $\bar{f} = \mathbb{E}_\xi f$, of the decomposition device of Bakhtiari et al. (2025, Lemma 3), which is stated there for $\bar{f} \in \operatorname{conv}(\mathcal{F})$. Since $\mathbb{E}_\xi f$ need not be a finite convex combination, we include a proof in Appendix A.

Lemma 2.1 (Decomposition; after Bakhtiari et al. (2025, Lemma 3))**.** *Let $\mathcal{F} \subseteq \mathcal{F}_{\mathrm{blr}}$ and $\alpha, \beta_0 \ge 0$. Suppose there are integers $k \ge 2$ and $m \ge 1$ such that for every prior $\xi \in \mathcal{P}(\mathcal{F})$, with $\bar{f} = \mathbb{E}_\xi f$, there is a measurable partition $\mathcal{F} = \bigcup_{i=1}^m \mathcal{F}_i$ such that, with the maximum taken over the pieces of positive ξ-measure,*

$$\max_{i \in [m]} \Big[\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f^\star) - \alpha - \Big(\beta_0 \inf_{f_1, \ldots, f_k \in \mathcal{F}_i} \sum_{(j,l) \in \mathrm{PAIR}(k)} \big(f_j(x_{f_l}) - \bar{f}(x_{f_l})\big)^2\Big)^{1/2}\Big] \le 0,$$

where $\mathrm{PAIR}(k)$ is the set of ordered pairs of distinct indices. Then $(\alpha, \beta_0 k(k-1)m) \in \mathrm{IR}(\mathcal{F})$ for the selection used to define x_f.

The proof averages the hypothesis over k independent draws from each piece and applies Cauchy–Schwarz over the m pieces, using $I(\pi^\xi_{\mathrm{TS}}, \xi) \ge \sum_i \xi(\mathcal{F}_i)^2 J_i$ for the within-piece informations J_i; only pointwise values of $\bar{f}$ are used.

Notation. $h := d + 1$. For a finite set C of functions with selected minimisers and chosen unit ridge directions $(\theta_f)_{f \in C}$, $t_{fg} = \langle x_g, \theta_f \rangle$. $\mathrm{PAIR}(C)$ is the set of ordered pairs of distinct elements of C. Logarithms are natural.

3 MAIN RESULTS

Theorem 3.1 (Regret of TS on non-monotone convex ridge losses)**.** *Let $d \geq 1$, let $K \subset \mathbb{R}^d$ be a convex body with $0 \in K$ and diameter D, let ξ be any prior on $\mathcal{F}_{\mathrm{blr}}$ (with respect to any σ-algebra under which point evaluations and the selection are measurable, as in Section 2), let $f \mapsto x_f$ be any such fixed measurable selection of minimisers, and let observations be Bernoulli as above. Then for every $n \geq 1$,*

$$\mathrm{BReg}_n(\mathrm{TS}, \xi) \;\leq\; 7 + 73{,}728\sqrt{3}\,(d+1)^4\, d^{1/2}\,\sqrt{n}\,\log\big(e + nd\max\{1, D\}\big).$$

The information-ratio bound (Theorem 3.4) reduces, through the blocking argument of Bakhtiari et al. (2025), to a cardinality bound for the following objects.

Definition 3.2 (Uninformative configurations)**.** *Fix $c \geq 12$, $\varepsilon > 0$ and set $\delta = \varepsilon/(c(d+1))$. A* configuration *is a finite set $C \subset \mathcal{F}_{\mathrm{blr}}$ with pairwise distinct selected minimisers together with a convex function $\bar{f} : K \to [0, 1]$ (the* comparison function*) such that for all $f, g \in C$,*

$$r_f := \bar{f}(x_f) - f^\star \in [\varepsilon/2, \varepsilon], \qquad |\bar{f}(x_f) - \bar{f}(x_g)| \leq \delta.$$

A row $f \in C$ is uninformative *if $|f(x_g) - \bar{f}(x_g)| < \delta$ for all $g \in C \setminus \{f\}$; the configuration is* uninformative *if every row is. For fixed c and d we write $g_c(d, z)$ for the supremum of $|C|$ over all uninformative configurations, over all convex bodies $K \ni 0$ in $\mathbb{R}^d$, all $\varepsilon > 0$ with $\mathrm{diam}(K)/\varepsilon \leq z$, all comparison functions and all selections.*

Theorem 3.3 (Cardinality bound)**.** *Let $c = 96(d+1)$. Every uninformative configuration in $\mathcal{F}_{\mathrm{blr}}$ satisfies $|C| \leq 3(d+1)^2 - (d+1) - 1 < 4(d+1)^2$; that is, $g_c(d, z) \leq 3(d+1)^2 - (d+1) - 1$ for all $z > 0$.*

Theorem 3.4 (Information ratio)**.** *Let $h = d + 1$, $k = 3072\, h^4$ and, for $\alpha \in (0, 1)$, $m_\alpha = 1 + \lceil \log_2(1/\alpha) \rceil$. Then*

$$\big(\alpha,\; 12\, k(k-1)\, m_\alpha\big) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}}) \quad \textit{uniformly over measurable selections;}$$

in particular $\mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ contains a pair (α, β) with $\beta \leq 113{,}246{,}208\,(d+1)^8\, m_\alpha$.

Proposition 3.5 (Tightness of the cardinality bound)**.** *For every $d \geq 2$ and every $c \geq 12$ there is an uninformative configuration in $\mathcal{F}_{\mathrm{blr}}$, on a truncated simplex of diameter at most $\sqrt{2}$ in normalised coordinates, with $|C| = d(d+1)$ and $D/\varepsilon \leq 64\sqrt{2}\, c^2 d(d+1)^2$; rescaling the functions about the common comparison value shows that the same size is attained for every larger value of D/ε. In particular, for $c = 96(d+1)$ and $z \geq 589{,}824\sqrt{2}\, d(d+1)^4$, $d(d+1) \leq g_c(d, z) \leq 3(d+1)^2 - (d+1) - 1$. For $d = 1$, $g_c(1, z) \leq 2$ for every $z > 0$ and $g_c(1, z) = 2$ for $z \geq 4$.*

This tightness concerns the cardinality only, not the certificate of Theorem 3.4 or the regret bound (Section 9).

Proposition 3.6 (The single-removal fixed-shrink John-ellipsoid dichotomy fails for non-monotone convex ridge losses)**.** *For $d = 3$ and every $c \geq 12$ there is an uninformative configuration C of twelve functions in $\mathcal{F}_{\mathrm{blr}}$ such that, writing $J_\zeta(C) = \mathrm{conv}\{x_f : f \in C\} + B_\zeta$ and $E_\zeta(\cdot)$ for the maximum-volume inscribed ellipsoid, $E_\zeta(C \setminus \{f\}) = E_\zeta(C)$ for every $f \in C$ and every $\zeta > 0$.*

Proposition 3.6 concerns the proof strategy of Bakhtiari et al. (2025), not their results: their Lemma 8 ("some removal shrinks the John ellipsoid by a factor 0.85, or some ordered pair is informative") is proved for the monotone class only, and its literal analogue for $\mathcal{F}_{\mathrm{blr}}$ is false for every fixed shrink factor in place of 0.85; other uses of John ellipsoids are not excluded.

Proof architecture. Theorem 3.1 combines geometric information-ratio control (Theorem 3.4) and selection-preserving transfer (Theorem 7.2). Lemma 4.1 supports both branches; lower bounds and counterexamples delimit the results, but are not premises of the regret upper bound.

Write $h = d + 1$, $D = \mathrm{diam} K$, $c = 96h$ and $G = 4h^2$; in the assembly step below $n \geq 2$, the case $n = 1$ being covered by the trivial bound. The argument has five steps; we record for each its input, its output and its cost in the dimension, so that the exponent $9/2$ can be traced.

1. *Two thin bands (Section 4).* Every $f \in \mathcal{F}_{\mathrm{blr}}$ has a representation $\ell(\langle \cdot, \theta \rangle)$ with ℓ convex, 1-Lipschitz and globally minimised at the projections of the minimisers (Lemma 4.1). If a row f of a configuration is uninformative, convexity of ℓ along θ_f forces the projections of all other minimisers into two thin bands at the extreme projections, on either side of $\langle x_f, \theta_f \rangle$, each of relative width $\eta \approx 6/(c(d+1))$ (Lemma 4.2). For monotone links only one band can be non-empty; the second band is the whole difficulty.
2. *Balanced and unbalanced rows (Section 5).* Rows whose two band distances L, R are comparable are handled by a quadratic evaluation matrix of rank at most $h(h+1)/2$ with small off-diagonal entries, giving at most $h^2 - 1$ such rows (Lemma 5.1). For the remaining rows this quadratic estimate deteriorates when the two band distances are highly imbalanced, so we eliminate the far-band entries with an exactly low-rank Boolean mask instead: the near-band membership matrix H, a 0-1 matrix, is within $1/(8h)$ of an affine evaluation matrix of rank at most h, and Lemma 5.3 gives $\mathrm{rank} H \le 2h - 1$. Masking an affine matrix by H (Lemma 5.2) then yields a near-identity matrix of rank at most $h(2h-1)$, hence at most $2h^2 - h$ unbalanced rows. Altogether $|C| \le 3h^2 - h - 1 = O(d^2)$ (Theorem 3.3), which is tight up to constants in the large-diameter-to-gap regime (Proposition 3.5). The choice $c = 96h$ makes the rounding and trace–rank requirements hold together (Section 9).
3. *Counting (Section 6).* If a configuration has $G + q$ elements, deleting first endpoints of informative pairs q times shows that its ordered pairs carry energy at least $q\delta^2$ (Lemma 6.1). Within one dyadic level of the gap $\bar{f}(x_f) - f^\star$, after handling k-tuples with coincident selected minimisers or comparison-value spread at least 2ε by direct bounds, the remaining tuple is ordered by comparison value and cut into $b = 4ch$ consecutive blocks of $s = 2G$ functions; at least half the blocks are configurations, each contributing $G\delta^2$, so every k-tuple has energy at least $\varepsilon^2/12$ with $k = bs = 3072h^4$. Lemma 2.1 turns this into $(\alpha, 12k(k-1)m_\alpha) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$, i.e. $\beta = O(d^8 m_\alpha)$ with $m_\alpha = 1 + \lceil \log_2(1/\alpha) \rceil$ (Theorem 3.4). The exponent 8 is 2×4: k is the number of blocks $\Theta(cd) = \Theta(d^2)$ times the block size $\Theta(G) = \Theta(d^2)$.
4. *Transfer (Section 7).* A family of at most $N_\rho \le (1 + 2D/\rho)^d (1 + 8D/\rho)^d$ value-based classes approximating $\mathcal{F}_{\mathrm{blr}}$, built from infimal-convolution approximations that keep the selected minimiser exact (Lemma 7.1), together with the information-ratio bound applied to finitely-valued proxies, gives $\mathrm{BReg}_n \le n\alpha + n\rho(6 + 4\sqrt{\beta}) + \sqrt{\beta n \log N_\rho / 2}$ for any fixed measurable selection (Theorem 7.2); $\log N_{1/n} = O(d \log(e + nD))$ costs $\sqrt{d}$.
5. *Assembly.* With $\alpha = \rho = 1/n$, $\sqrt{\beta} = O(d^4 \sqrt{\log n})$ and $\sqrt{\log N_{1/n}} = O(\sqrt{d \log(e + nD)})$ give $\mathrm{BReg}_n = O(d^4 \sqrt{dn} \log(e + nd \max\{1, D\})) = \tilde{O}(d^{9/2}\sqrt{n})$ (Theorem 3.1).

4 Structure of uninformative rows

Lemma 4.1 (Ridge representation with a global minimiser)**.** *For every $f \in \mathcal{F}_{\mathrm{blr}}$ there exist a unit vector θ and a convex, 1-Lipschitz $\ell : \mathbb{R} \to \mathbb{R}$ such that $f = \ell(\langle \cdot, \theta \rangle)$ on K, $\ell \ge f^\star \ge 0$ on all of $\mathbb{R}$, and $\ell(\langle x, \theta \rangle) = f^\star = \min_{\mathbb{R}} \ell$ for every minimiser x of f on K. In particular $|f(x) - f(y)| \le |\langle \theta, x - y \rangle|$ for all $x, y \in K$.*

Appendix A proves the global-minimiser representation used in Section 7.

Lemma 4.2 (Two thin extreme bands)**.** *Let $c \ge 12$ and let C be a configuration (Definition 3.2). Fix $f \in C$ with representation (θ_f, ℓ_f) from Lemma 4.1, and write $s = \langle x_f, \theta_f \rangle$, $t_g = \langle x_g, \theta_f \rangle$, $a = \min_{g \in C} t_g$, $b = \max_{g \in C} t_g$, $L = s - a$, $R = b - s$, and*

$$\eta := \frac{3\delta}{\varepsilon/2 + \delta} = \frac{3}{c(d+1)/2 + 1}.$$

If the row f is uninformative then for every $g \ne f$:

(i) $|t_g - s| > \varepsilon/2 - 2\delta \ge 5\varepsilon/12$*;*

(ii) $t_g \in (b - \eta R,\, b]$ *if* $t_g > s$*, and* $t_g \in [a,\, a + \eta L)$ *if* $t_g < s$*.*

The proof (Appendix A) is a two-line convexity argument: uninformativeness pins $f(x_g)$ to within δ of $\bar{f}(x_g)$, which is at least $f^\star + \varepsilon/2 - 2\delta$, giving (i); and interpolating ℓ_f between its minimum at

s and its value at the extreme projection b shows that any t_g strictly inside (s, b) would make $f(x_g)$ smaller than $\bar{f}(x_g) - \delta$ unless t_g is within relative distance η of b, giving (ii).

5 THE CARDINALITY BOUND

5.1 THREE LINEAR-ALGEBRA TOOLS

Lemma 5.1 (Trace–rank bound: the classical symmetric bound of Alon (2009, Lemma 2.2) and its non-symmetric extension)**.** *For any real $N \times N$ matrix M, $|\mathrm{tr}M|^2 \le \mathrm{rank}(M)\,\|M\|_F^2$. Consequently, if $M_{ii} = 1$ for all i, $|M_{ij}| \le t$ for $i \ne j$, $\mathrm{rank}M \le Q$ and $Qt^2 < 1$, then*

$$N \le \frac{Q(1-t^2)}{1-Qt^2} \le \frac{Q}{1-Qt^2};$$

and if moreover $Q \ge 1$ is an integer and $t < 1/Q$, then $N \le Q$.

For symmetric M the second claim is Alon (2009, Lemma 2.2); the trace–rank inequality for arbitrary real matrices also appears as Russo & Van Roy (2016, Fact 10). The short proof (via the orthogonal projection onto the column space of M) is in Appendix B.

Lemma 5.2 (Hadamard products)**.** *For real matrices U, V of the same shape, $\mathrm{rank}(U \circ V) \le \mathrm{rank}(U)\mathrm{rank}(V)$, where $\circ$ is the entrywise product.*

Lemma 5.3 (Integer rounding preserves rank up to a factor two)**.** *Let $r \ge 1$ be an integer, let A be a real matrix with $\mathrm{rank}A \le r$, and let B be a 0-1 matrix of the same shape with $\|A - B\|_{\max} \le \gamma$. If $\gamma \le 1/(4r)$ then $\mathrm{rank}B \le 2r - 1$.*

The proof is in Appendix B.

We have not found this elementary statement in the literature in this form, but do not claim priority for it; the approximate-rank literature (e.g. Alon, 2009; Alon et al., 2013) studies both upper and lower bounds on the ε-approximate rank of real matrices.

5.2 PROOF SKETCH OF THEOREM 3.3

Let $h = d+1$, $c = 96h$, $\eta = 3/(48h^2+1) < 1/(16h^2)$ and $\delta = \varepsilon/(96h^2)$. Let C be uninformative and assume $|C| \ge 3$; smaller cases are trivial. For each row i use Lemmas 4.1 and 4.2, orienting the direction so that $0 \le L_i \le R_i$ and $R_i > 0$. Put $w_i = L_i + R_i$ and $\lambda_i = L_i/w_i \in [0, 1/2]$; all band quantities are computed on C, never on subsets. Split $I_B = \{i : \lambda_i > 1/(8h)\}$ and $I_U = C \setminus I_B$.

Balanced rows: quadratic evaluation. For $i, j \in I_B$ the matrix $P_{ij} = p_i(x_j)$, where $p_i(x) = (\langle x, \theta_i\rangle - a_i)(b_i - \langle x, \theta_i\rangle)/(L_iR_i)$, has unit diagonal and $0 \le P_{ij} < t_B := 8h\eta < 1/(2h)$ off the diagonal; its rank is at most $Q_B = h(h+1)/2$, so Lemma 5.1 gives $|I_B| \le h^2 - 1$.

Unbalanced rows: exact Boolean masking. All remaining matrices are indexed by $I_U \times I_U$. Let $H_{ii} = 1$ and, for $j \ne i$, let $H_{ij} = 1$ exactly on the near band and 0 otherwise. The affine matrix $A_{ij} = (b_i - t_{ij})/w_i$ has rank at most h and $\|A - H\|_{\max} \le 1/(8h)$, so Lemma 5.3 with $r = h$ gives $\mathrm{rank}H \le 2h - 1$. Define $U_{ij} = (t_{ij} - a_i)/L_i$ when $L_i > 0$ and $U_{ij} = 1$ otherwise; its rows are affine, so $\mathrm{rank}U \le h$. The mask $M = U \circ H$ has unit diagonal, zero far-band entries and near-band entries in $[0, \eta)$; rows with $L_i = 0$ have no near band. By Lemma 5.2, $\mathrm{rank}M \le Q_U := h(2h-1)$, and since $\eta < 1/Q_U$, Lemma 5.1 gives $|I_U| \le 2h^2 - h$. Thus $|C| \le 3h^2 - h - 1$. The full proof is in Appendix B.

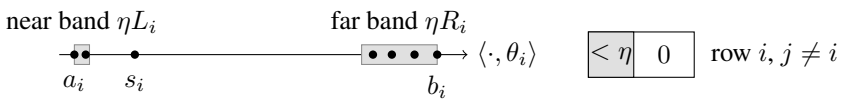

Figure 1: An unbalanced row i, $L_i \ll R_i$ (Lemma 4.2; schematic, not to scale). For $j \ne i$ the entries of $M = U \circ H$ are $< \eta$ in the near band and exactly 0 in the far band; $M_{ii} = 1$.

6 FROM THE CARDINALITY BOUND TO THE INFORMATION RATIO

Lemma 6.1 (Counting informative pairs). *Fix $c \geq 12$ and suppose every uninformative configuration (Definition 3.2, with this c) has at most G elements. Let $q \geq 1$ be an integer. If C satisfies the hypotheses of Definition 3.2 and $|C| \geq G + q$, then*

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

The proof (Appendix B) deletes, q times, the first endpoint of an ordered pair with $|f(x_g) - \bar{f}(x_g)| \geq \delta$; such a pair exists while more than G elements remain, the hypotheses of Definition 3.2 survive deletion, and the q recorded pairs have distinct first endpoints.

Proof sketch of Theorem 3.4. Fix a measurable selection and a prior ξ with mean $\bar{f}$, and let $h = d + 1$, $c = 96h$ and $G = 4h^2$, so that by Theorem 3.3 every uninformative configuration for this c has at most G elements. Partition $\mathcal{F}_{\mathrm{blr}}$ by the dyadic level of the gap $\bar{f}(x_f) - f^\star$: the piece $\mathcal{F}_0$ where the gap is at most α, and the pieces $\mathcal{F}_i$ where it lies in $(\alpha 2^{i-1}, \alpha 2^i]$; there are m_α pieces and the condition of Lemma 2.1 is trivial on $\mathcal{F}_0$.

From a tuple to configurations. On $\mathcal{F}_i$ put $\varepsilon = \alpha 2^i$, $\delta = \varepsilon/(ch)$, and consider any k-tuple with $k = bs$, $b = 4ch$ and $s = 2G$; write S for its sum of squared errors over ordered pairs. If two functions share the selected minimiser then $S \geq \varepsilon^2/4$; if the comparison values $\bar{f}(x_{f_j})$ spread by at least 2ε then the extreme pair gives $S \geq \varepsilon^2$; otherwise order the tuple by comparison value and cut it into b consecutive blocks of s functions. Fewer than $2\varepsilon/\delta = 2ch$ blocks have spread larger than δ, so at least $2ch$ blocks have spread at most δ; each of these, with $\bar{f}$, is a configuration in the sense of Definition 3.2 with $G + G$ elements, and Lemma 6.1 with $q = G$ gives it energy at least $G\delta^2$.

From pairwise energy to the certificate. Hence $S \geq 2ch \cdot G\delta^2 = \varepsilon^2/12$ for every tuple, and Lemma 2.1 with $\beta_0 = 12$ and $k = 8chG = 3072h^4$ gives $(\alpha, 12k(k-1)m_\alpha) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$, where $12k^2 = 113{,}246{,}208\, h^8$. Nothing depends on which selection was fixed. The full proof is in Appendix B.

7 FROM INFORMATION RATIO TO REGRET UNDER A FIXED SELECTION

The transfer theorem of Bakhtiari et al. (2025, Theorem 28) bounds the regret of (approximate) TS from an information-ratio bound and a cover of the function class, and its covering argument for ridge classes uses, as the authors note, that ties among minimisers are broken consistently. For $\mathcal{F}_{\mathrm{blr}}$ minimisers can be non-unique (Section 1), so we give a self-contained transfer theorem for exact TS that is valid for any fixed measurable selection rule. Two ingredients differ from Bakhtiari et al. (2025): the cover is built from infimal-convolution approximations, so that the selected minimiser of f is an exact minimiser of its approximation (Lemma 7.1); and the information-ratio hypothesis is required uniformly over selection rules and extended to finitely-supported randomised minimisers (Lemma C.1), which is what the proof actually uses. Three lemmas supply measurable representations, randomised minimisers, and perturbation control. They are summarised next and proved in Appendix C.

For $\mathcal{F} \subseteq \mathcal{F}_{\mathrm{blr}}$, Lemma C.1 gives $\Delta(\mathcal{L}(X), \xi) \leq \alpha + \sqrt{\beta I(\mathcal{L}(X), \xi)}$ whenever $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F})$ uniformly over measurable selections, $\xi \in \mathcal{P}(\mathcal{F})$, and (f, X) is given by a measurable kernel with $f \sim \xi$ and $X \in \arg\min_K f$ almost surely; in law this is a mixture of selections, and linearity together with concavity of the square root gives the bound. Lemma C.2 has three parts: (a) and (b) say that $\sqrt{I}$ changes by at most ρ under a uniform ρ-perturbation of the random function, or of the action when the function is almost surely 1-Lipschitz, provided the perturbed pair is independent of the other argument; (c) says that if $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ uniformly over selections and (f, X) is a finitely-valued pair with $f \in \mathcal{F}_{\mathrm{blr}}$ and $f(X) \leq f^\star + \rho$ almost surely, then $\Delta(\mathcal{L}(X), \mathcal{L}(f)) \leq \alpha + \sqrt{\beta I(\mathcal{L}(X), \mathcal{L}(f))} + \rho(1 + \sqrt{\beta})$. Lemma C.3 provides a Borel choice of ridge direction $f \mapsto \theta_f \in S^{d-1}$ on $\mathcal{F}_{\mathrm{blr}}$ and the jointly continuous map $L(f, \theta, t) = \min_{y\in K}[f(y) + |t - \langle \theta, y\rangle|]$ such that $(\theta_f, L(f, \theta_f, \cdot))$ is a representation of f as in Lemma 4.1.

Lemma 7.1 (Cover by infimal-convolution approximations). *Let $0 < \rho < 1$, and choose a ρ-net C_K of K and a $\rho/(2D)$-net C_S of S^{d-1} with $|C_K| \leq (1 + 2D/\rho)^d$ and $|C_S| \leq (1 + 8D/\rho)^d$ (such*

nets exist by the standard volume argument), so $N_\rho := |C_K||C_S| \le (1+2D/\rho)^d(1+8D/\rho)^d$. *For* $(x,\theta) \in C_K \times C_S$ *let*

$$\mathcal{F}^{\mathrm{val}}_{x,\theta} = \{g \in \mathcal{F}_{\mathrm{blr}} :\ g \text{ is a ridge function in direction } \theta,\ g(x) \le g^\star + 2\rho\}.$$

Then (i) every probability average of functions in $\mathcal{F}^{\mathrm{val}}_{x,\theta}$ *lies in* $\mathcal{F}_{\mathrm{blr}}$*, is a ridge function in direction* θ*, and is* 2ρ*-near-minimal at* x*; (ii) for every measurable selection there are measurable maps* $f \mapsto \kappa(f) = (x_\kappa, \theta_\kappa) \in C_K \times C_S$ *and* $f \mapsto g_f \in \mathcal{F}^{\mathrm{val}}_{\kappa(f)}$ *with* $\|f - g_f\|_\infty \le \rho/2$*,* $\|x_f - x_\kappa\| \le \rho$*, and such that* x_f *is an exact minimiser of* g_f *with* $g_f^\star = f^\star$.

Preserving the selected minimiser. (i) is closed under averaging because all functions in a piece share the direction θ, and near-minimality at x is preserved since the minimum of an average is at least the average of the minima. For (ii), let θ_f and L be as in Lemma C.3, let $\kappa(f)$ be the first pair in a fixed enumeration of $C_K \times C_S$ with $\|x_\kappa - x_f\| \le \rho$ and $\|\theta_\kappa - \theta_f\| \le \rho/(2D)$, and set $g_f(x) = L(f, \theta_\kappa, \langle x, \theta_\kappa\rangle) = \min_{y\in K}[f(y) + |\langle \theta_\kappa, x - y\rangle|]$, the infimal convolution of f with the ridge kernel $|\langle \theta_\kappa, \cdot\rangle|$. Its profile is convex and 1-Lipschitz, so g_f is a ridge function in direction θ_κ; $y = x$ gives $g_f \le f$ and $f \ge f^\star$ gives $g_f \ge f^\star$, so no rescaling is needed; the inequality $f(x) - f(y) \le |\langle \theta_f, x - y\rangle|$ of Lemma 4.1 and $\|\theta_\kappa - \theta_f\| \le \rho/(2D)$ give $f - \rho/2 \le g_f$; and $f^\star \le g_f(x_f) \le f(x_f) = f^\star$ shows that x_f is an exact minimiser of g_f. The details, including measurability, are in Appendix C.

Theorem 7.2 (Transfer under any fixed measurable selection rule)**.** *Suppose* $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ *uniformly over measurable selections. Then for every measurable selection, every prior* ξ *on* $\mathcal{F}_{\mathrm{blr}}$*, every* $n \ge 1$ *and every* $\rho \in (0,1)$*, exact Thompson sampling satisfies*

$$\mathrm{BReg}_n(\mathrm{TS}, \xi) \le n\alpha + n\rho(6 + 4\sqrt{\beta}) + \sqrt{\tfrac{1}{2}\beta n \log N_\rho}.$$

Construct the finite proxy. Fix a round, let $f \sim \xi$ be the current posterior, $X^\star = x_f$, and let X be an independent copy of $X^\star$, so $\mathcal{L}(X) = \pi_{\mathrm{TS}}$. With $\kappa = \kappa(f)$, $g = g_f$ and $z = x_\kappa$ from Lemma 7.1, set $H_\kappa = \mathbb{E}[f \mid \kappa]$ and $G_\kappa = \mathbb{E}[g \mid \kappa]$. Both take finitely many values (one per label). G_κ is a finitely-valued $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$-valued random element, a ridge function in direction θ_κ that is 2ρ-near-minimal at z (Lemma 7.1(i)); H_κ is only a convex 1-Lipschitz function in $\mathcal{F}_{\mathrm{bl}}$ (an average of ridge functions with nearby but different directions is in general not a ridge function), viewed as a Borel $C(K)$-valued element, and $\|G_\kappa - H_\kappa\|_\infty \le \rho/2$. The information-ratio hypothesis is applied, via Lemmas C.1 and C.2(c), only to the finitely-valued flattened proxy $\max\{G_\kappa, G_\kappa(z)\}$, never to H_κ.

Regret and information comparisons. Since $g \le f$, $\|f - g\|_\infty \le \rho/2$ and $\|X^\star - z\| \le \rho$, the one-step regret satisfies $\Delta(\pi_{\mathrm{TS}}, \xi) \le \Delta(\pi_z, \nu_G) + \frac{5}{2}\rho$ with $\nu_G = \mathcal{L}(G_\kappa)$ and $\pi_z = \mathcal{L}(z)$; Lemma C.2(c) gives $\Delta(\pi_z, \nu_G) \le \alpha + \sqrt{\beta I(\pi_z, \nu_G)} + 2\rho(1 + \sqrt{\beta})$; and Lemma C.2(a),(b), applied on a product space built from two independent posterior draws, give $\sqrt{I(\pi_z, \nu_G)} \le \sqrt{I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \frac{3}{2}\rho$. Altogether $\Delta(\pi_{\mathrm{TS}}, \xi) \le \alpha + \sqrt{\beta I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \rho(6 + 4\sqrt{\beta})$.

Accumulate information over rounds. For Bernoulli observations $I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa)) \le \frac{1}{2}\mathsf{I}(\kappa; X, Y)$ by Pinsker's inequality; the label $\kappa(f)$ is a fixed function of the environment, so the chain rule for mutual information gives $\mathbb{E}\sum_t \mathsf{I}_{t-1}(\kappa; X_t, Y_t) \le \log N_\rho$, and Cauchy–Schwarz over rounds finishes. The full proof is in Appendix C.

Proof of Theorem 3.1. For $n = 1$ the regret is at most $1 \le 7$. For $n \ge 2$ take $\alpha = \rho = 1/n$ and $L = \log(e + nd\max\{1, D\})$: Theorem 3.4 gives $\sqrt{\beta} \le 12{,}288\sqrt{3}\,(d+1)^4\sqrt{L}$ (using $m_{1/n} \le 4L$), Lemma 7.1 gives $\log N_{1/n} \le 8dL$, and Theorem 7.2 yields $\mathrm{BReg}_n \le 7 + 6\sqrt{\beta}\sqrt{ndL} = 7 + 73{,}728\sqrt{3}\,(d+1)^4 d^{1/2}\sqrt{n}\,L$; the arithmetic is in Appendix C.

8 The lower-bound family and the failure of the John dichotomy

The proofs of Propositions 3.5 and 3.6 are in Appendix D; we describe the constructions. For Proposition 3.5 take the standard d-simplex with vertices $v_0, \dots, v_d$, a small $\tau = 1/(4c(d+1))$, and the $d(d+1)$ points $x_{ij} = (1-\tau)v_i + \tau v_j$ $(i \ne j)$, which are the vertices of a truncated simplex K. The loss f_{ij} is a V-shaped ridge function of the affine coordinate $u_{ij} = 1 - \lambda_i + \tau\lambda_j$ (barycentric

coordinates λ), with a steep left slope and a shallow right slope, minimised at x_{ij} and equal to the constant comparison function $\bar{f} \equiv 1/4$ at every other point up to a cross error below δ; the gap is $\varepsilon = \tau^2/(4d)$. Along its own direction, row (i, j) sees the $d-1$ points x_{il} with the same base vertex in a near band at u-distance τ^2 and the other d^2 points in a far band at u-distance about 1; every row is two-sided, and the ratio of the band distances is of order $1/\tau^2$, unbounded in the family.

The ellipsoid obstruction. For Proposition 3.6, use the centred regular tetrahedron with inradius r. Every single-vertex deletion of the truncated tetrahedron preserves B_r. For every $\zeta > 0$, both smoothed bodies contain $B_{r+\zeta}$ and lie in the enlarged tetrahedron, whose maximum-volume inscribed ellipsoid is $B_{r+\zeta}$; hence the maximum-volume inscribed ellipsoids of both bodies coincide.

A prior-level certificate obstruction. The truncated-simplex construction also gives a statistical obstruction under an actual prior: for every $d \geq 2$ there are a convex body K, a uniform prior on $d(d+1)$ functions in $\mathcal{F}_{\mathrm{blr}}$ and a fixed measurable selection such that every $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ for that selection with $\alpha \leq 1/(36{,}864\, d(d+1)^2)$ has $\beta \geq d(d+1)/8$ (Proposition D.1 with $c = 12$); this is a one-step certificate bound, not a regret lower bound or a separation from the monotone class.

In the monotone case an uninformative row has all other minimisers on one side of its own along the ridge direction, and Bakhtiari et al. (2025) show that they then lie beyond the centre of the John ellipsoid, so removing the row cuts the ellipsoid through its centre; for non-monotone links both bands of Lemma 4.2 can be non-empty and arbitrarily imbalanced, and Proposition 3.6 shows that the single-removal, fixed-shrink-factor dichotomy cannot hold.

9 Consequences, quantitative limits, and open questions

Theorem 3.1 answers Bakhtiari et al. (2025)'s qualitative question: exact-posterior TS achieves Bayesian regret polynomial in d and $\tilde{O}(\sqrt{n})$ in n without link monotonicity. The key conceptual implication is that convex ridge structure, rather than monotonicity, is the ingredient that makes exact Thompson sampling learnable. Optimal dimension dependence remains open (Table 1). Theorem 3.1 improves on the trivial bound $\mathrm{BReg}_n \leq n$ only once n exceeds d^9 up to constants and logarithmic factors (Appendix C); it certifies polynomial learnability, not a sharp finite-horizon bound.

Two steps determine dimension dependence. Tight cardinality $\Theta(d^2)$ in the large-diameter-to-gap regime (Proposition 3.5) entails $\Theta(d^2)$ functions per block in Lemma 2.1 and $\Theta(cd)$ blocks; $c = 96(d+1)$ meets the rounding-tolerance and trace–rank requirements of Section 5. A uniform improvement from order $1/r$ to $1/\sqrt{r}$ fails for arbitrary 0-1 matrices: Proposition E.1 (Appendix E) exhibits, for every $r = 2^j - 1$ with $j \geq 2$, a rank-r matrix within $2/(r+1)$ in max-norm of a 0-1 matrix of rank $2r$. This obstruction concerns arbitrary 0-1 matrices; an exponent improvement through larger rounding tolerance must exploit the additional structure of Section 5's band-membership matrices.

Scope of lower bounds. Attaining the monotone exponent $5/2$ here, and optimal dimension dependence, remain open. Our construction lower-bounds configuration cardinality, not regret; the conversion to an information-ratio certificate ($\Theta(d^2)$ to $O(d^8 m_\alpha)$) is not known to be tight. Its comparison function is constant, not the prior mean: for $c = 96(d+1)$, the uniform prior on that family has cross errors relative to its mean of order ε/N, with $N = d(d+1)$, above δ (Appendix D); its fixed-selection certificates with $\alpha \leq \varepsilon/4$ satisfy $\beta \geq d(d+1)/8$ (Proposition D.1). This does not exclude a posterior-aware analysis.

Limitations. (1) We bound Bayesian regret for arbitrary priors, not frequentist regret for every fixed environment. The guarantee concerns exact-posterior TS; computation remains open. (2) Selection is arbitrary but fixed in advance: deterministic, or randomised with fresh independent randomness (Remark C.4). History-dependent rules are not covered (a limitation of the proof, not a known failure). (3) Observations are Bernoulli; Appendix C gives the precise conditions for the bounded-observation extension. (4) The loss is single, stationary and in $\mathcal{F}_{\mathrm{blr}}$, on a compact convex body $K \ni 0$ with non-empty interior, with measurable point evaluations and a measurable selection. General convex losses and approximate-posterior implementations are excluded. The information-ratio question for a *known* convex link in the same paragraph of Bakhtiari et al. (2025) is not addressed.

AI USE STATEMENT

In this work, we used generative AI tools to help develop theoretical models or conceptual frameworks, assist in writing proofs, design or provide feedback on the research methodology, implement methods, assist with translation, and interpret the theoretical results and their scope. We have not used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, or propose or refine hypotheses. Generating synthetic datasets, cleaning or reformatting datasets, and supporting qualitative or thematic data analysis are not applicable to this work, which is theoretical and does not involve datasets.

Additionally, we used generative AI tools to create or modify scientific figures; create or edit software code; draft parts of the paper; summarize or analyse existing literature; discover research topics or identify gaps; brainstorm; source or search for information; improve the paper's readability; identify relevant literature; format references; suggest the paper's structure; and propose titles or keywords.

We have reviewed all AI-assisted work. All mathematical derivations and any AI-assisted code were manually checked for correctness by three human reviewers, with no errors identified in the final versions. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All results are theoretical. The regret, cardinality and information-ratio theorems, the transfer theorem, and the tightness and John-ellipsoid propositions are stated explicitly in the main text with explicit constants; Section 9 summarises the rounding obstruction whose proposition and proof are in Appendix E. Complete proofs are given in the appendices: Lemmas 2.1, 4.1 and 4.2 in Appendix A; Lemmas 5.1, 5.3 and 6.1 together with the full proofs of Theorems 3.3 and 3.4 in Appendix B; Lemmas C.1, C.2, C.3 and 7.1, Theorem 7.2 and the arithmetic of Theorem 3.1 in Appendix C; Propositions 3.5, D.1 and 3.6 in Appendix D; and Proposition E.1 in Appendix E. The constants in Theorems 3.1 and 3.4 follow by exact rational arithmetic from the parameters $c = 96(d+1)$, $G = 4(d+1)^2$, $b = 4c(d+1)$ and $s = 2G$ fixed in the proofs; the exact-arithmetic scripts are provided as supplementary material.

ETHICS STATEMENT

This is a theoretical study of a learning algorithm on an abstract model. It involves no data, human subjects or deployed system, and we are not aware of direct ethical concerns arising from it.

REFERENCES

Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability and Computing*, 18(1–2):3–15, 2009.

Noga Alon, Troy Lee, Adi Shraibman, and Santosh Vempala. The approximate rank of a matrix and its algorithmic applications. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 675–684. Association for Computing Machinery, 2013.

Alireza Bakhtiari, Tor Lattimore, and Csaba Szepesvári. Thompson sampling for bandit convex optimisation. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 231–263. PMLR, 2025.

SeyedAlireza Bakhtiari. Thompson sampling for bandit convex optimization. *Master Thesis of University of Alberta*, 2025. doi: https://doi.org/10.7939/83050.

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

Sébastien Bubeck, Ofer Dekel, Tomer Koren, and Yuval Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pp. 266–278. PMLR, 2015.

Devdan Dey, Sujoy Bhore, and Avishek Ghosh. Optimal regret for single index bandits. *arXiv preprint arXiv:2605.09454*, 2026. Version 3.

Olav Kallenberg. *Foundations of Modern Probability*. Springer, New York, NY, 1997.

Yue Kang, Mingshuo Liu, Bongsoo Yi, Jing Lyu, Zhi Zhang, Doudou Zhou, and Yao Li. Single index bandits: Generalized linear contextual bandits with unknown reward functions. In *International Conference on Learning Representations*, 2026. arXiv:2506.12751.

Kazimierz Kuratowski and Czesław Ryll-Nardzewski. A general theorem on selectors. *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques*, 13:397–403, 1965. URL `https://mathscinet.ams.org/mathscinet-getitem?mr=188994`. MR0188994, Zbl 0152.21403.

Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2(3/4):311–334, 2019.

Tor Lattimore. Minimax regret for bandit convex optimisation of ridge functions. *arXiv preprint arXiv:2106.00444*, 2021. Version 2.

Yuwei Luo and Mohsen Bayati. Geometry-aware approaches for balancing performance and theoretical guarantees in linear bandits. In *International Conference on Learning Representations*, 2025.

Nived Rajaraman and Yanjun Han. Interactive learning of single-index models via stochastic gradient descent. In *International Conference on Learning Representations*, 2026a. arXiv:2602.17876.

Nived Rajaraman and Yanjun Han. The price of hidden curvature: Improved lower bounds for bandit convex optimization. *arXiv preprint arXiv:2607.18652*, 2026b. Version 3.

Nived Rajaraman, Yanjun Han, Jiantao Jiao, and Kannan Ramchandran. Statistical complexity and optimal algorithms for nonlinear ridge bandits. *The Annals of Statistics*, 52(6):2557–2582, 2024.

Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17(68):1–30, 2016.

Itai Shufaro, Nadav Merlis, Nir Weinberger, and Shie Mannor. On bits and bandits: Quantifying the regret-information trade-off. In *International Conference on Learning Representations*, 2025. arXiv:2405.16581v4.

A Proofs for Sections 2 and 4

This appendix contains the common foundations of the two proof branches. The decomposition lemma supports the information-ratio argument; the representation lemma supports both the two-band geometry and the approximation used in transfer.

Proof of Lemma 2.1. Let $w_i = \xi(\mathcal{F}_i)$, discard pieces with $w_i = 0$, let ξ_i be the conditional law on $\mathcal{F}_i$, and $J_i = \mathbb{E}_{\xi_i \otimes \xi_i}[(f(x_g) - \bar{f}(x_g))^2]$. Taking expectations of the hypothesis over k independent draws from ξ_i gives $\sup_{f \in \mathcal{F}_i}(\bar{f}(x_f) - f^\star) \le \alpha + \sqrt{\beta_0 k(k-1) J_i}$. Hence

$$\begin{aligned}\Delta(\pi^{\xi}_{\mathrm{TS}}, \xi) &= \sum_i w_i \, \mathbb{E}_{\xi_i}[\bar{f}(x_f) - f^\star] \le \alpha + \sqrt{\beta_0 k(k-1)} \sum_i w_i \sqrt{J_i} \\ &\le \alpha + \sqrt{\beta_0 k(k-1) m \textstyle\sum_i w_i^2 J_i} \le \alpha + \sqrt{\beta_0 k(k-1) m \, I(\pi^{\xi}_{\mathrm{TS}}, \xi)},\end{aligned}$$

by Cauchy–Schwarz and because $I(\pi^{\xi}_{\mathrm{TS}}, \xi) = \sum_{i,j} w_i w_j J_{ij} \ge \sum_i w_i^2 J_i$, with

$$J_{ij} = \mathbb{E}_{\xi_i \otimes \xi_j}[(f(x_g) - \bar{f}(x_g))^2] \ge 0.$$

□

Proof of Lemma 4.1. If f is constant, any unit θ and the constant link work. Otherwise normalise the ridge direction and let u be the link on $I = \{\langle x, \theta\rangle : x \in K\} = [a, b]$. For $t \in (a, b)$ there is $z \in \operatorname{int} K$ with $\langle z, \theta\rangle = t$ and a short segment through z in direction θ inside K, along which f is 1-Lipschitz, so u is locally 1-Lipschitz on (a, b), hence 1-Lipschitz on (a, b); and since f is continuous on K and every point of K is a limit of interior points of K, u is continuous at a and b as well, so u is 1-Lipschitz on all of I. Define $\ell(t) = u(a) + a - t$ for $t < a$, $\ell = u$ on I, and $\ell(t) = u(b) + t - b$ for $t > b$. The slopes -1 on the left and $+1$ on the right bracket every secant slope of u, so ℓ is convex and 1-Lipschitz; outside I it exceeds the endpoint values, so $\min_{\mathbb{R}} \ell = \min_I u = f^\star$, attained at the projections of the minimisers of f. The last claim is the 1-Lipschitz property of ℓ. □

Proof of Lemma 4.2. Uninformativeness and the comparison condition give $f(x_g) > \bar{f}(x_g) - \delta \geq \bar{f}(x_f) - 2\delta = f^\star + r_f - 2\delta \geq f^\star + \varepsilon/2 - 2\delta$; with 1-Lipschitzness of ℓ_f this is (i), and $\varepsilon/2 - 2\delta \geq 5\varepsilon/12$ because $c \geq 12$. For (ii) take g with $t_g > s$ and let $g^+ \in C$ attain $t_{g^+} = b$ (so $g^+ \neq f$ and $b > s$). With $\lambda = (b - t_g)/(b - s) \in [0, 1)$, convexity of ℓ_f gives $f(x_g) \leq \lambda f^\star + (1 - \lambda) f(x_{g^+})$. Using $f^\star \leq \bar{f}(x_g) + \delta - \varepsilon/2$ and $f(x_{g^+}) < \bar{f}(x_{g^+}) + \delta \leq \bar{f}(x_g) + 2\delta$,

$$f(x_g) < \bar{f}(x_g) + 2\delta - \lambda(\varepsilon/2 + \delta),$$

and combining with $f(x_g) > \bar{f}(x_g) - \delta$ yields $\lambda < \eta$. The left side is symmetric. □

B Proofs for Sections 5 and 6

The first three proofs establish the algebraic tools used by the cardinality theorem. The counting lemma then connects that theorem to the full information-ratio proof. The two full theorem proofs below follow this dependency order.

Proof of Lemma 5.1. For symmetric M the second claim is Alon (2009, Lemma 2.2), which Alon attributes to earlier sources; the trace–rank inequality for arbitrary real matrices also appears as Russo & Van Roy (2016, Fact 10), and we include the short argument. Let Π be the orthogonal projection onto the column space of M, of rank $q = \operatorname{rank} M$. Then $|\operatorname{tr} M| = |\operatorname{tr}(\Pi M)| = |\langle \Pi, M\rangle_F| \leq \|\Pi\|_F \|M\|_F = \sqrt{q}\|M\|_F$. Inserting $\operatorname{tr} M = N$ and $\|M\|_F^2 \leq N + N(N-1)t^2$ gives the second claim. For the third, $(Q+1) - Q(1-t^2)/(1-Qt^2) = (1 - Q^2t^2)/(1 - Qt^2) > 0$ when $t < 1/Q$, so $N < Q + 1$. □

Proof of Lemma 5.2. Write $U = \sum_{a \leq p} u_a v_a^\top$ and $V = \sum_{b \leq q} y_b z_b^\top$ with $p = \operatorname{rank} U$, $q = \operatorname{rank} V$; then $U \circ V = \sum_{a,b} (u_a \circ y_b)(v_a \circ z_b)^\top$. □

Proof of Lemma 5.3. Suppose $\operatorname{rank} B \geq 2r$ and take a non-singular $2r \times 2r$ submatrix B_0 of B; let A_0 be the corresponding submatrix of A and $E = B_0 - A_0$, so $\operatorname{rank} A_0 \leq r$ and $\|E\|_F^2 \leq 4r^2\gamma^2$. Let $\sigma_1 \geq \cdots \geq \sigma_{2r} > 0$ be the singular values of B_0. Since $\dim \ker A_0 \geq r$, let Π be the orthogonal projection onto an r-dimensional subspace of $\ker A_0$; then $B_0\Pi = E\Pi$, so $\|B_0\Pi\|_F^2 \leq \|E\|_F^2$. In an orthonormal basis of right singular vectors of B_0, $\|B_0\Pi\|_F^2 = \sum_k \sigma_k^2 w_k$ with $0 \leq w_k \leq 1$ and $\sum_k w_k = r$, which is at least the sum of the r smallest σ_k^2. Hence

$$\sum_{k=r+1}^{2r} \sigma_k^2 \leq 4r^2\gamma^2, \qquad \sum_{k=1}^{r} \sigma_k^2 \leq \|B_0\|_F^2 \leq 4r^2 - 1,$$

the latter because a non-singular $2r \times 2r$ 0-1 matrix has at least one zero entry. By the arithmetic–geometric mean inequality applied to each group of r squared singular values,

$$|\det B_0| = \prod_{k=1}^{2r} \sigma_k \leq \Big(\frac{4r^2 - 1}{r} \cdot \frac{4r^2\gamma^2}{r}\Big)^{r/2} = (4r\gamma)^r \Big(1 - \frac{1}{4r^2}\Big)^{r/2} < 1 \qquad \text{for } \gamma \leq \frac{1}{4r},$$

contradicting $|\det B_0| \geq 1$ for a non-singular integer matrix. □

Proof of Lemma 6.1. As long as the remaining set has more than G elements it is not uninformative, so it contains an ordered pair (f, g) with $|f(x_g) - \bar{f}(x_g)| \geq \delta$; delete f (only the first endpoint). The hypotheses of Definition 3.2 are per-element and pairwise, so they survive deletion, and before the $(j+1)$-st deletion ($0 \leq j \leq q-1$) at least $G+q-j \geq G+1$ elements remain. The q recorded ordered pairs have distinct first endpoints, hence are distinct, and each contributes at least δ^2. □

FULL PROOF OF THEOREM 3.3

Let $h = d+1$, $c = 96h$, so that

$$\eta = \frac{3}{48h^2+1} < \frac{1}{16h^2}, \qquad \delta = \frac{\varepsilon}{96h^2}. \tag{1}$$

Let C be an uninformative configuration with $N = |C| \geq 3$ (for $N \leq 2$ there is nothing to prove). For each row $i \in C$ take the representation of Lemma 4.1 and the quantities $t_{ij} = \langle x_j, \theta_i \rangle$, s_i, a_i, b_i, L_i, R_i of Lemma 4.2. By Lemma 4.2(i), $L_i + R_i > 0$. Replacing (θ_i, ℓ_i) by $(-\theta_i, \ell_i(-\cdot))$ if necessary (this changes neither the function nor its selected minimiser), we may assume

$$0 \leq L_i \leq R_i, \qquad R_i > 0, \qquad w_i := L_i + R_i, \qquad \lambda_i := L_i/w_i \in [0, 1/2].$$

By Lemma 4.2(ii) every $j \neq i$ is either in the *near band* $t_{ij} \in [a_i, a_i + \eta L_i)$ (empty if $L_i = 0$) or in the *far band* $t_{ij} \in (b_i - \eta R_i, b_i]$. Throughout, a_i, b_i, s_i, L_i, R_i refer to the whole configuration C and are never recomputed for subsets. Split

$$I_B = \{i : \lambda_i > 1/(8h)\}, \qquad I_U = \{i : \lambda_i \leq 1/(8h)\}.$$

Balanced rows. For $i \in I_B$ we have $L_i R_i > 0$ and $R_i/L_i < 8h - 1$. Consider the quadratic polynomial $p_i(x) = (\langle x, \theta_i \rangle - a_i)(b_i - \langle x, \theta_i \rangle)/(L_i R_i)$ and the matrix $P = (p_i(x_j))_{i,j \in I_B}$. Then $P_{ii} = 1$, and for $j \neq i$ in the near band $0 \leq P_{ij} < \eta(L_i + R_i)/R_i \leq 2\eta$, in the far band $0 \leq P_{ij} < \eta(L_i + R_i)/L_i < 8h\eta$; so all off-diagonal entries are below $t_B := 8h\eta < 1/(2h)$. Each row of P is a polynomial of total degree at most two in $x_j \in \mathbb{R}^d$, so $\operatorname{rank} P \leq Q_B := 1 + d + d(d+1)/2 = h(h+1)/2$. Since $Q_B t_B^2 < (h+1)/(8h) \leq 3/16$, Lemma 5.1 gives

$$|I_B| \leq \frac{Q_B}{1 - Q_B t_B^2} < \frac{4h^2(h+1)}{7h-1} < h^2, \qquad \text{i.e. } |I_B| \leq h^2 - 1, \tag{2}$$

where the last strict inequality uses $7h - 1 - 4(h+1) = 3h - 5 > 0$ for $h \geq 2$.

Unbalanced rows. All matrices below are indexed by $I_U \times I_U$. Define the 0-1 matrix H by $H_{ii} = 1$ and, for $j \neq i$, $H_{ij} = 1$ if t_{ij} lies in the near band of row i and $H_{ij} = 0$ if it lies in the far band. Define the affine evaluation matrix $A_{ij} = (b_i - t_{ij})/w_i$; each row is an affine function of x_j, so $\operatorname{rank} A \leq h$. Entrywise, $|A_{ii} - 1| = \lambda_i \leq 1/(8h)$; for j in the near band $|A_{ij} - 1| = (t_{ij} - a_i)/w_i < \eta\lambda_i$; for j in the far band $|A_{ij}| = (b_i - t_{ij})/w_i < \eta(1 - \lambda_i)$. By (1), $\|A - H\|_{\max} \leq 1/(8h)$, so Lemma 5.3 with $r = h$ gives $\operatorname{rank} H \leq 2h - 1$.

Next define $U_{ij} = (t_{ij} - a_i)/L_i$ if $L_i > 0$ and $U_{ij} = 1$ if $L_i = 0$; again each row is affine in x_j, so $\operatorname{rank} U \leq h$. Let $M = U \circ H$. Then $M_{ii} = 1$; for j in the far band $M_{ij} = 0$ exactly (the mask removes the possibly huge values U_{ij} there); for j in the near band $0 \leq M_{ij} < \eta$; and rows with $L_i = 0$ have no near band, so their off-diagonal entries vanish. By Lemma 5.2, $\operatorname{rank} M \leq h(2h-1) =: Q_U$, and $Q_U \eta < h(2h-1)/(16h^2) < 1/8$, so $\eta < 1/Q_U$ and the integer form of Lemma 5.1 gives

$$|I_U| \leq Q_U = 2h^2 - h. \tag{3}$$

Conclusion. By (2) and (3), $|C| = |I_B| + |I_U| \leq 3h^2 - h - 1 < 4h^2$. □

Only Lemmas 4.1 and 4.2 and finite-dimensional linear algebra are used: no monotonicity, no John ellipsoid, no bound on D/ε, and only the fact that x_f is some minimiser, so the bound holds for every measurable selection. On the family of Proposition 3.5 every row is unbalanced, H is the "same base vertex" mask, and M is the identity.

We do not obtain the required uniform cardinality estimate from polynomial approximation (a univariate polynomial that is 1 at the row's own projection and small on both bands); the mask argument of Lemma 5.3 avoids the need to control the imbalance ratio of the two band distances.

Table 2: Algebraic roles in the unbalanced-row argument of Theorem 3.3, complementing Figure 1. All matrices are indexed by $I_U \times I_U$; band endpoints are computed on the full configuration. Near- and far-band entries below are off-diagonal.

Object	**Entrywise property used**	**Rank bound**
Affine approximation A	$\|A - H\|_{\max} \le 1/(8h)$	$\mathrm{rank} A \le h$
Boolean mask H	$H_{ii} = 1$; near band: 1; far band: 0	$\mathrm{rank} H \le 2h - 1$ (Lemma 5.3)
Affine normalisation U	Each row is affine in x_j. If $L_i = 0$, set $U_{ij} = 1$; the near band is empty.	$\mathrm{rank} U \le h$
Masked matrix $M = U \circ H$	$M_{ii} = 1$; near band: $[0, \eta)$; far band: exactly 0	$\mathrm{rank} M \le h(2h-1)$ (Lemma 5.2)

Lemma 5.1 converts the entrywise and rank bounds for M into $|I_U| \le 2h^2 - h$.

Full proof of Theorem 3.4

Proof of Theorem 3.4. Fix a measurable selection, a prior ξ on $\mathcal{F}_{\mathrm{blr}}$, and $\bar{f} = \mathbb{E}_\xi f$. Let $h = d + 1$, $c = 96h$ and $G = 4h^2$; by Theorem 3.3 every uninformative configuration for this c has at most G elements. With $\varepsilon_i = \alpha 2^i$ partition $\mathcal{F}_{\mathrm{blr}}$ into $\mathcal{F}_0 = \{f : \bar{f}(x_f) - f^\star \le \alpha\}$ and $\mathcal{F}_i = \{f : \bar{f}(x_f) - f^\star \in (\alpha 2^{i-1}, \alpha 2^i]\}$ for $1 \le i \le \lceil \log_2(1/\alpha) \rceil$; these are measurable (they only involve $f \mapsto \bar{f}(x_f)$ and $f \mapsto f^\star$) and there are m_α of them. The condition of Lemma 2.1 holds trivially on $\mathcal{F}_0$. Fix $i \ge 1$, $\varepsilon = \varepsilon_i$, $\delta = \varepsilon/(ch)$, and set

$$b = 4ch, \qquad s = 2G, \qquad k = bs = 8chG = 3072\, h^4$$

(over $b = 2ch + u$, $s = G + v$ the argument below yields energy $uv\delta^2$, and $k^2/(uv)$ is minimised at $u = 2ch$, $v = G$). Let $f_1, \dots, f_k \in \mathcal{F}_i$ be a k-tuple. We show

$$S := \sum_{(j,l) \in \mathrm{PAIR}(k)} \big(f_j(x_{f_l}) - \bar{f}(x_{f_l})\big)^2 \ge \frac{\varepsilon^2}{12}, \tag{4}$$

which with $\sup_{f \in \mathcal{F}_i} (\bar{f}(x_f) - f^\star) \le \varepsilon \le \sqrt{12S}$ and Lemma 2.1 (applied with $\beta_0 = 12$) proves the theorem. If two functions of the tuple share the selected minimiser (in particular if they coincide), the pair contributes $(\bar{f}(x_f) - f^\star)^2 \ge \varepsilon^2/4$. Otherwise order the tuple so that $v_j = \bar{f}(x_{f_j})$ is non-increasing. If $v_1 - v_k \ge 2\varepsilon$ then $f_1(x_{f_k}) \ge f_1^\star \ge v_1 - \varepsilon \ge v_k + \varepsilon$ and the pair $(1, k)$ contributes ε^2. Otherwise cut the ordered tuple into b consecutive blocks of s functions; the sum of the value spreads of the blocks is less than 2ε, so fewer than $2\varepsilon/\delta = 2ch$ blocks have spread larger than δ, and at least $b - 2ch = 2ch$ blocks have spread at most δ. Each such block, together with $\bar{f}$, satisfies Definition 3.2 (its elements have $r_f \in (\varepsilon/2, \varepsilon]$ and distinct minimisers) and has $s = 2G = G + G$ elements, so Lemma 6.1 with $q = G$ shows that it contributes at least $G\delta^2$; the blocks are disjoint, so

$$S \ge 2ch \cdot G\delta^2 = \frac{2G}{ch}\,\varepsilon^2 = \frac{8h^2}{96h^2}\,\varepsilon^2 = \frac{\varepsilon^2}{12}.$$

Since $\varepsilon^2/4$ and ε^2 are not smaller, (4) holds for every tuple. Lemma 2.1 gives $(\alpha, 12k(k-1)m_\alpha) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$, and $12k^2 = 12 \cdot 3072^2 h^8 = 113{,}246{,}208\, h^8$. Nothing in the argument depends on which measurable selection was fixed. □

In Bakhtiari et al. (2025) the role of Lemma 6.1 is played by their Lemma 9, which iterates their John-ellipsoid dichotomy (Lemma 8); Proposition 3.6 shows that no such dichotomy holds for $\mathcal{F}_{\mathrm{blr}}$, which is why we count through a cardinality bound instead.

C Proofs for Section 7

The transfer branch has two preparatory components: measurable approximation through Lemmas C.3 and 7.1, and stability of the information-ratio certificate through Lemmas C.1 and C.2. The posterior construction then justifies conditioning before the proof of Theorem 7.2.

The three preparatory lemmas used in Section 7 are stated here in full.

Lemma C.1 (Uniform bounds extend to randomised minimisers)**.** *Suppose $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F})$ for every measurable selection, where $\mathcal{F} \subseteq \mathcal{F}_{\mathrm{blr}}$. Let $\xi \in \mathcal{P}(\mathcal{F})$ and let (f, X) be a random pair with $f \sim \xi$ and $X \in \arg\min_K f$ almost surely, given by a measurable kernel. Then $\Delta(\mathcal{L}(X), \xi) \le \alpha + \sqrt{\beta I(\mathcal{L}(X), \xi)}$.*

Lemma C.2 (Continuity and near-minimisers)**.** *(a) Let f, g be Borel random elements of $C(K)$ with values in $[0,1]$ (no convexity or ridge assumption), coupled with $\|f - g\|_\infty \le \rho$ a.s., and let $X \sim \pi$ be independent of (f, g). Then $\sqrt{I(\pi, \mathcal{L}(g))} \le \sqrt{I(\pi, \mathcal{L}(f))} + \rho$.*

(b) Let ν be the law of a Borel random element of $C(K)$ that is a.s. 1-Lipschitz with values in $[0,1]$, and let actions X, Z be coupled with $\|X - Z\| \le \rho$ a.s. and independent of the function. Then $\sqrt{I(\mathcal{L}(Z), \nu)} \le \sqrt{I(\mathcal{L}(X), \nu)} + \rho$.

(c) Suppose $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ uniformly over selections. Let (f, X) be a random pair taking finitely many values, with $f \in \mathcal{F}_{\mathrm{blr}}$ and $f(X) \le f^\star + \rho$ a.s. Then $\Delta(\mathcal{L}(X), \mathcal{L}(f)) \le \alpha + \sqrt{\beta I(\mathcal{L}(X), \mathcal{L}(f))} + \rho(1 + \sqrt{\beta})$.

Lemma C.3 (Measurable representations)**.** *There are a Borel map $f \mapsto \theta_f \in S^{d-1}$ on $(\mathcal{F}_{\mathrm{blr}}, \mathcal{B})$ and a jointly continuous map $(f, \theta, t) \mapsto L(f, \theta, t)$ on $\mathcal{F}_{\mathrm{blr}} \times S^{d-1} \times \mathbb{R}$,*

$$L(f, \theta, t) = \min_{y \in K} \big[f(y) + |t - \langle \theta, y \rangle| \big],$$

such that, for every f, $(\theta_f, L(f, \theta_f, \cdot))$ is a representation of f as in Lemma 4.1.

Proof of Lemma C.1. By the randomisation lemma for kernels into Borel spaces (Kallenberg, 1997, Lemma 2.22) there is a jointly measurable $a(f, u)$ with $X = a(f, U)$ for U uniform on $[0,1]$ independent of f. Let a_0 be a fixed measurable selection, $B = \{(f, u) : f(a(f, u)) > f^\star\}$, and $\tilde{a} = a$ off B, $\tilde{a} = a_0$ on B; then $\tilde{a}$ is jointly measurable, $f \mapsto \tilde{a}(f, u)$ is a measurable selection for every u, and $(f, \tilde{a}(f, U))$ has the same law as (f, X). Let π_u be the law of $\tilde{a}(f, u)$. Both $\Delta(\cdot, \xi)$ and $I(\cdot, \xi)$ are linear in the policy, so $\Delta(\mathcal{L}(X), \xi) = \int \Delta(\pi_u, \xi)\, du \le \alpha + \int \sqrt{\beta I(\pi_u, \xi)}\, du \le \alpha + \sqrt{\beta \int I(\pi_u, \xi)\, du}$ by concavity of the square root. □

Proof of Lemma C.2. For $(X, F) \sim \pi \otimes \nu$ one has $\bar{F}(X) = \mathbb{E}[F(X) \mid X]$, so $I(\pi, \nu)^{1/2} = \|F(X) - \mathbb{E}[F(X) \mid X]\|_{L^2}$. (a): realise $X \sim \pi$ independently of the coupled pair (f, g) and let $QW = W - \mathbb{E}[W \mid X]$, an L^2 contraction; then $\sqrt{I(\pi, \mathcal{L}(f))} = \|Qf(X)\|_{L^2}$, $\sqrt{I(\pi, \mathcal{L}(g))} = \|Qg(X)\|_{L^2}$, and the two differ by at most $\|Q(g(X) - f(X))\|_{L^2} \le \|g(X) - f(X)\|_{L^2} \le \rho$. (b): realise $F \sim \nu$ independently of the coupled pair (X, Z) and use the single centring operator $QW = W - \mathbb{E}[W \mid X, Z]$; independence gives $QF(X) = F(X) - \bar{F}(X)$ and $QF(Z) = F(Z) - \bar{F}(Z)$, so $\sqrt{I(\mathcal{L}(Z), \nu)}$ and $\sqrt{I(\mathcal{L}(X), \nu)}$ differ by at most $\|Q(F(Z) - F(X))\|_{L^2} \le \|F(Z) - F(X)\|_{L^2} \le \rho$ by the Lipschitz bound. (c): let $g = \max(f, f(X))$; truncation at a level in $[0,1]$ preserves convexity, range, Lipschitzness and the ridge direction, so $g \in \mathcal{F}_{\mathrm{blr}}$; X is an exact minimiser of g, $\|g - f\|_\infty \le \rho$, $\bar{f} \le \bar{g}$ and $f^\star \ge g^\star - \rho$. Since (f, X) takes finitely many values, (g, X) is a finitely-valued measurable kernel of exact minimisers (distinct values of (f, X) may share g but not X, which is why a kernel rather than a selection is needed), and Lemma C.1 gives $\Delta(\mathcal{L}(X), \mathcal{L}(f)) \le \Delta(\mathcal{L}(X), \mathcal{L}(g)) + \rho \le \alpha + \sqrt{\beta I(\mathcal{L}(X), \mathcal{L}(g))} + \rho$. Finally, since $I(\pi, \nu)$ depends only on the marginal law of the policy and the marginal law of the random function, part (a) may be applied on a product coupling in which an independent $X' \sim \mathcal{L}(X)$ is used together with the coupled pair (f, g), $\|g - f\|_\infty \le \rho$; this gives $\sqrt{I(\mathcal{L}(X), \mathcal{L}(g))} \le \sqrt{I(\mathcal{L}(X), \mathcal{L}(f))} + \rho$ and finishes the proof. □

Proof of Lemma C.3. A unit vector θ is a ridge direction of f if and only if $|f(x) - f(y)| \le |\langle \theta, x - y \rangle|$ for all $x, y \in K$ (the forward direction is Lemma 4.1, the backward direction yields a profile that is constant on fibres, convex and 1-Lipschitz). Let $\Phi(f)$ be the set of such θ. Profiles extended as in Lemma 4.1 satisfy $\ell(0) = f(0) \in [0,1]$ and $\mathrm{Lip}(\ell) \le 1$, so they are uniformly bounded and

equicontinuous on $[-D, D]$; if $f_j \to f$ uniformly and $\theta_j \to \theta$, a uniformly convergent subsequence of profiles shows $f = \ell(\langle \cdot, \theta \rangle)$. Hence $\mathcal{F}_{\mathrm{blr}}$ is compact in $C(K)$ (Arzelà–Ascoli) and the graph of Φ is closed in $\mathcal{F}_{\mathrm{blr}} \times S^{d-1}$, so Φ has non-empty closed values. For an open $O \subseteq S^{d-1}$ write $O = \bigcup_j C_j$ with C_j compact; then $\{f : \Phi(f) \cap O \neq \emptyset\} = \bigcup_j \mathrm{proj}(\mathrm{Graph}(\Phi) \cap (\mathcal{F}_{\mathrm{blr}} \times C_j))$ is a countable union of compact sets, hence Borel, so Φ is weakly measurable and the Kuratowski–Ryll-Nardzewski selection theorem (Kuratowski & Ryll-Nardzewski, 1965) gives a Borel selection $\theta_f \in \Phi(f)$. The map L satisfies $|L(f, \theta, t) - L(g, \psi, s)| \leq \|f - g\|_\infty + D\|\theta - \psi\| + |t - s|$, hence is jointly continuous, and for $\theta \in \Phi(f)$ it coincides with the profile of f on the projection interval and with the ∓ 1-slope extension of Lemma 4.1 outside it. □

Proof of Lemma 7.1. (i) Averages of convex ridge functions with a common direction are convex ridge functions with that direction; range and Lipschitz bounds are preserved; and if $g_\nu = \int g \, d\nu$ then $g_\nu(x) \leq \int g^\star d\nu + 2\rho \leq g_\nu^\star + 2\rho$ since the minimum of an average is at least the average of the minima. (ii) Let θ_f and L be as in Lemma C.3 and let $\kappa(f)$ be the first pair in a fixed enumeration of $C_K \times C_S$ with $\|x_\kappa - x_f\| \leq \rho$ and $\|\theta_\kappa - \theta_f\| \leq \rho/(2D)$; since x_f is $\mathcal{A}$-measurable and θ_f is Borel, κ is $\mathcal{A}$-measurable. Define

$$g_f(x) = L(f, \theta_\kappa, \langle x, \theta_\kappa \rangle) = \min_{y \in K} \big[f(y) + |\langle \theta_\kappa, x - y \rangle| \big].$$

The profile $t \mapsto L(f, \theta_\kappa, t)$ is convex (the minimand is jointly convex in (t, y) and partial minimisation over the convex set K preserves convexity) and 1-Lipschitz (each y contributes a 1-Lipschitz function of t), so g_f is a convex ridge function in direction θ_κ with $\mathrm{Lip}(g_f) \leq 1$. Taking $y = x$ gives $g_f \leq f \leq 1$, and $f \geq f^\star$ gives $g_f \geq f^\star \geq 0$; hence $g_f \in \mathcal{F}_{\mathrm{blr}}$ with no rescaling. By Lemma 4.1, for all $x, y \in K$, $f(x) - f(y) \leq |\langle \theta_f, x - y \rangle| \leq |\langle \theta_\kappa, x - y \rangle| + D\|\theta_\kappa - \theta_f\| \leq |\langle \theta_\kappa, x - y \rangle| + \rho/2$ (using $\|x - y\| \leq D$); minimising over y gives $f(x) - \rho/2 \leq g_f(x) \leq f(x)$, i.e. $\|g_f - f\|_\infty \leq \rho/2$. At the selected minimiser, $f^\star \leq g_f(x_f) \leq f(x_f) = f^\star$, so x_f minimises g_f and $g_f^\star = f^\star$; by 1-Lipschitzness $g_f(x_\kappa) - g_f^\star \leq \|x_\kappa - x_f\| \leq \rho \leq 2\rho$, so $g_f \in \mathcal{F}^{\mathrm{val}}_{\kappa(f)}$. Finally $f \mapsto g_f$ is measurable as a $(C(K), \mathcal{B})$-valued map and $(f, x) \mapsto g_f(x)$ is jointly measurable, by the joint continuity of L and the measurability of κ. □

Existence of the posterior. The proof below conditions on the history. Since $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$ is not assumed to be a standard Borel space, we construct the posterior directly for Bernoulli observations rather than invoke a standard-Borel disintegration theorem. For a history $H = ((x_s, y_s))_{s<t}$ with $x_s \in K$ and $y_s \in \{0, 1\}$ put

$$L_H(f) = \prod_{s<t} \big[y_s f(x_s) + (1 - y_s)(1 - f(x_s)) \big], \qquad Z_H = \int L_H \, d\xi,$$

and define $\xi_H(A_0) = Z_H^{-1} \int_{A_0} L_H \, d\xi$ when $Z_H > 0$ and $\xi_H = \xi$ when $Z_H = 0$. The map $(f, H) \mapsto L_H(f)$ is jointly measurable because $(f, x) \mapsto f(x)$ is (Section 2), so $(H, A_0) \mapsto \xi_H(A_0)$ is a probability kernel by Fubini's theorem; histories with $Z_H = 0$ have probability zero; and the action kernels Q_s satisfy $\mathcal{L}(X_s \mid f, H_{s-1}) = Q_s(\cdot \mid H_{s-1})$, where $H_{s-1} = ((X_r, Y_r))_{r<s}$. Let R_t be the history law obtained from these action kernels and independent $\mathrm{Bernoulli}(1/2)$ labels. Then

$$\mathbb{P}(df, dH) = 2^{t-1} L_H(f) \, \xi(df) R_t(dH), \qquad \mathbb{P}_H(dH) = 2^{t-1} Z_H \, R_t(dH),$$

so Bayes' formula verifies that ξ_H is a version of the conditional law of f given H. Writing $S(f) = x_f$, the Thompson-sampling action has conditional law $S_\# \xi_H$ and is independent of f given H. The proof uses ξ_H as the conditional law of f, the resulting conditional product law $\xi_H \otimes (S_\# \xi_H)$, independent posterior copies, and conditioning on the finite label $\kappa(f)$; no additional disintegration theorem is needed.

Proof of Theorem 7.2. The proof has six steps: (i) discretise the selected minimiser and the ridge direction by the nets C_K, C_S of Lemma 7.1; (ii) replace f by the ridge proxy g_f, which keeps the selected minimiser exact; (iii) average within a cell of the label κ, giving the finitely-valued G_κ; (iv) apply the information-ratio hypothesis, through Lemmas C.1 and C.2(c), only to a finitely-valued flattened proxy; (v) compare its information term with the observation mutual information by Pinsker's inequality; (vi) sum over rounds by the chain rule. We bound the one-step regret under an arbitrary posterior and then sum. Let $f \sim \xi$ (a posterior at some round), $X^\star = x_f$, and let X be

an independent copy of $X^\star$, so $\mathcal{L}(X) = \pi_{\mathrm{TS}}$. Let $\kappa = \kappa(f)$, $g = g_f$, $z = x_\kappa$ be as in Lemma 7.1, and

$$H_\kappa = \mathbb{E}[f \mid \kappa], \qquad G_\kappa = \mathbb{E}[g \mid \kappa], \qquad \nu_G = \mathcal{L}(G_\kappa), \qquad \pi_z = \mathcal{L}(z).$$

Both H_κ and G_κ take finitely many values (one per label). G_κ is a finitely-valued $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$-valued random element; H_κ is viewed only as a Borel $C(K)$-valued random element with values in $\mathcal{F}_{\mathrm{bl}}$. By Lemma 7.1(i), $G_\kappa \in \mathcal{F}_{\mathrm{blr}}$ is a ridge function in direction θ_κ and is 2ρ-near-minimal at z; H_κ is merely a convex 1-Lipschitz function $K \to [0,1]$ (an average of ridge functions with nearby but different directions is in general not a ridge function), and $\|G_\kappa - H_\kappa\|_\infty \le \rho/2$. In this proof the information-ratio hypothesis is applied, via Lemmas C.1 and C.2(c), to the finitely-valued flattened proxy $\max\{G_\kappa, G_\kappa(z)\}$, never to H_κ; H_κ enters only through the continuity statements of Lemma C.2(a),(b), which do not require the ridge structure.

First, $\mathbb{E}[f^\star \mid \kappa] = \mathbb{E}[f(X^\star) \mid \kappa] \ge \mathbb{E}[g(X^\star) \mid \kappa] \ge G_\kappa(z) - \rho \ge G^\star_\kappa - \rho$ (using $g \le f$, $\|X^\star - z\| \le \rho$ and Lipschitzness), and $\|\mathbb{E}f - \mathbb{E}g\|_\infty \le \rho/2$; hence

$$\Delta(\pi_{\mathrm{TS}}, \xi) = \mathbb{E}[\bar{f}(X) - f^\star] \le \mathbb{E}[\bar{g}(X) - G^\star_\kappa] + \tfrac{3}{2}\rho \le \mathbb{E}[\bar{g}(z) - G^\star_\kappa] + \tfrac{5}{2}\rho = \Delta(\pi_z, \nu_G) + \tfrac{5}{2}\rho.$$

Here the second inequality is $\mathbb{E}\bar{g}(X) = \mathbb{E}\bar{g}(X^\star) \le \mathbb{E}\bar{g}(z) + \rho$: X and $X^\star$ have the same law, $\bar{g}$ is 1-Lipschitz and $\|X^\star - z\| \le \rho$; no coupling of X with z is used. By Lemma C.2(c) applied to the finitely-valued pair (G_κ, z), $\Delta(\pi_z, \nu_G) \le \alpha + \sqrt{\beta I(\pi_z, \nu_G)} + 2\rho(1 + \sqrt{\beta})$. The information terms depend only on the product of the marginal laws, so the continuity statements may be applied on a product space: let F_1, F_2 be independent draws from the current posterior, build $(G_1, H_1) = (G_{\kappa(F_1)}, H_{\kappa(F_1)})$ from F_1 and $(Z_2, A_2) = (x_{\kappa(F_2)}, x_{F_2})$ from F_2. Lemma C.2(a) with the coupling (G_1, H_1), $\|G_1 - H_1\|_\infty \le \rho/2$, independent of $Z_2 \sim \pi_z$, gives $\sqrt{I(\pi_z, \nu_G)} \le \sqrt{I(\pi_z, \mathcal{L}(H_\kappa))} + \rho/2$; Lemma C.2(b) with the coupling (Z_2, A_2), $\|Z_2 - A_2\| \le \rho$, independent of $H_1 \sim \mathcal{L}(H_\kappa)$, gives $\sqrt{I(\pi_z, \mathcal{L}(H_\kappa))} \le \sqrt{I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \rho$. Together $\sqrt{I(\pi_z, \nu_G)} \le \sqrt{I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \frac{3}{2}\rho$. Altogether

$$\Delta(\pi_{\mathrm{TS}}, \xi) \le \alpha + \sqrt{\beta I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \rho\big(\tfrac{9}{2} + \tfrac{7}{2}\sqrt{\beta}\big) \le \alpha + \sqrt{\beta I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa))} + \rho(6 + 4\sqrt{\beta}).$$

In the one-step analysis X is an independent copy of $X^\star = x_f$, so X is independent of f and hence of the pair (κ, H_κ), which is a function of f; in the process this is the statement, recorded in the paragraph on the existence of the posterior, that conditionally on the history the action is drawn from an independent posterior sample and is independent of f. Consequently, with all conditional expectations taken under the current posterior and $\mathbb{E}[Y \mid X, f] = f(X)$, one has $H_\kappa(X) = \mathbb{E}[f(X) \mid X, \kappa] = \mathbb{E}[Y \mid X, \kappa]$ and $\mathbb{E}[H_\kappa(X) \mid X] = \mathbb{E}[f(X) \mid X] = \bar{f}(X) = \mathbb{E}[Y \mid X]$; this identifies $I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa)) = \mathbb{E}[(H_\kappa(X) - \mathbb{E}[H_\kappa(X) \mid X])^2]$ with $\mathbb{E}[(\mathbb{E}[Y \mid X, \kappa] - \mathbb{E}[Y \mid X])^2]$, the quantity to which Pinsker's inequality is applied next. For Bernoulli observations, $I(\pi_{\mathrm{TS}}, \mathcal{L}(H_\kappa)) = \mathbb{E}[(\mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid X, \kappa])^2] \le \frac{1}{2}\mathbb{E}[\mathrm{KL}(P_{Y|X,\kappa} \,\|\, P_{Y|X})] = \frac{1}{2}\mathsf{I}(\kappa; X, Y)$ by Pinsker's inequality. The label $\kappa = \kappa(f)$ is a fixed finite-valued function of the environment (it depends on the fixed nets, the fixed direction map of Lemma C.3 and the fixed selection, not on the round), so summing over rounds and using the chain rule for mutual information and Cauchy–Schwarz,

$$\begin{aligned}\mathrm{BReg}_n(\mathrm{TS}, \xi) &\le n\alpha + n\rho(6 + 4\sqrt{\beta}) + \sqrt{\tfrac{1}{2}\beta n\, \mathbb{E}\textstyle\sum_t \mathsf{I}_{t-1}(\kappa; X_t, Y_t)} \\ &\le n\alpha + n\rho(6 + 4\sqrt{\beta}) + \sqrt{\tfrac{1}{2}\beta n \log N_\rho}.\end{aligned}$$

□

Remark C.4 (Randomised fixed tie-breaking). Theorem 7.2, and hence Theorem 3.1, also covers exact-posterior TS with a selection rule that is randomised but fixed in advance: a probability kernel from $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$ to K supported on minimisers, represented as in the proof of Lemma C.1 by a jointly measurable $a(f, u)$ with $a(f, u) \in \arg\min_K f$, where TS plays $a(f_t, U_t)$ for a posterior sample f_t and fresh U_t uniform on $[0,1]$. Apply the proof above to the augmented environment (f, U), U uniform and independent of f, with the measurable selection $(f, u) \mapsto a(f, u)$. The loss and the observation law depend on f alone, so the posterior of (f, U) given the history is $\xi_H \otimes \mathrm{Unif}[0,1]$ and the TS action $a(f', U')$ of an independent posterior copy is independent of (f, U) given the history. Lemma 7.1(ii) applies with (f, U) in place of f (its proof uses only the direction map of Lemma C.3, which depends on f, and the selected point, now $a(f, U)$), giving a label $\kappa = \kappa(f, U)$ with at most N_ρ values and a proxy g of which $a(f, U)$ is an exact minimiser; $f^\star = \min_K f$ is unchanged. The information-ratio hypothesis enters only through Lemma C.2(c), applied with tolerance 2ρ to the finitely-valued pair (G_κ, z). This pair may depend on the selection through κ,

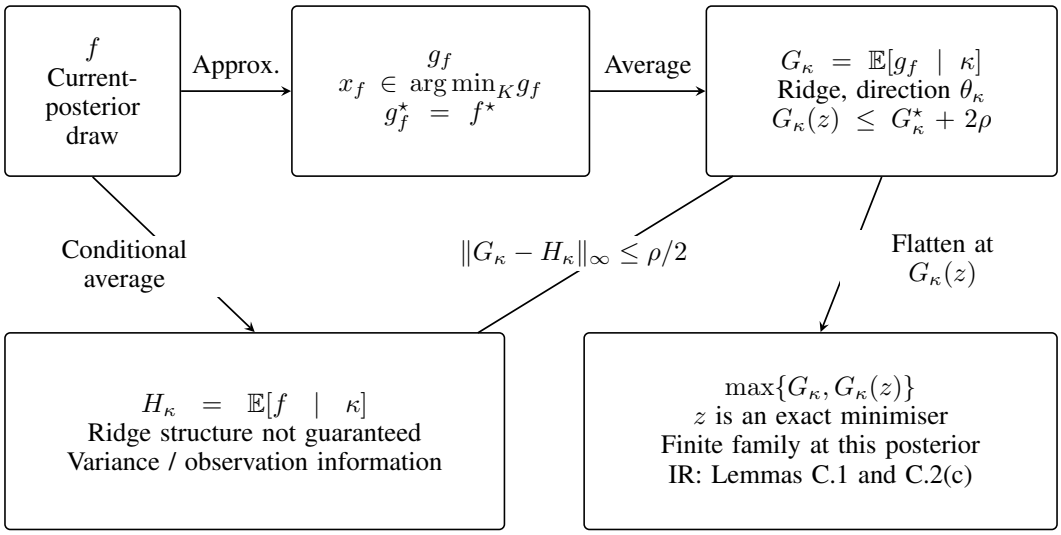

Figure 2: Proof objects in Theorem 7.2 at a fixed posterior, with $z = x_\kappa$. Approximation and averaging within a fixed-direction cell use Lemma 7.1(ii) and (i), respectively. Arrows denote object constructions; the undirected line gives the comparison bound. The information-ratio hypothesis enters through the flattened ridge proxy, not through H_κ. The finite family is indexed by κ; cell averages may change with the posterior, whereas the label map is fixed across rounds. These are analytical objects, not modifications of TS; information comparisons retain the independent posterior copies used in the proof.

but, under the hypothesis uniform over measurable selections, the requirements on the pair are finite range, $G_\kappa \in \mathcal{F}_{\mathrm{blr}}$, and $G_\kappa(z) \le G_\kappa^\star + 2\rho$, not agreement with any prescribed selection rule. The chain-rule step uses only that κ is a fixed finite-valued function of the augmented environment. The constants are unchanged. General history-dependent selection rules remain outside the scope of this proof: conditional on a fixed history, the one-step construction can be repeated for a jointly measurable selector under the stated uniform hypothesis, but the resulting cover label may depend on the history. The conditional cell averages can still be defined at each round; what the argument does not provide in general is one common finite-valued environment label to which the cross-round chain-rule bound by $\log N_\rho$ applies. This is a limitation of the proof; we do not know whether such rules can make exact-posterior TS fail on $\mathcal{F}_{\mathrm{blr}}$.

Proof of Theorem 3.1. For $n = 1$ the regret is at most $1 \le 7$. Let $n \ge 2$, $\alpha = \rho = 1/n$, $L = \log(e + nd\max\{1, D\})$, and $h = d+1$. From Theorem 3.4, $k = 3072h^4$ and $m_{1/n} \le 4L$, so $\sqrt{\beta} \le \sqrt{12 \cdot 4L}\, k = 4\sqrt{3} \cdot 3072\, h^4\sqrt{L} = 12{,}288\sqrt{3}\, h^4\sqrt{L}$. From Lemma 7.1, $\log N_{1/n} \le d\log(1 + 2nD) + d\log(1 + 8nD) \le 8dL$. Theorem 7.2 gives $\mathrm{BReg}_n \le 1 + 6 + 4\sqrt{\beta} + \sqrt{\beta}\sqrt{4ndL} \le 7 + \sqrt{\beta}\,(4 + 2\sqrt{ndL}) \le 7 + 6\sqrt{\beta}\sqrt{ndL}$ because $\sqrt{ndL} \ge 1$, i.e. $\mathrm{BReg}_n \le 7 + 73{,}728\sqrt{3}\, h^4 d^{1/2}\sqrt{n}\, L$. □

Range of the bound. Since the losses take values in $[0, 1]$, $\mathrm{BReg}_n \le n$ trivially, so Theorem 3.1 is informative only once n exceeds d^9 times a large constant and logarithmic factors; it certifies polynomial-in-dimension learnability rather than sharp numbers at moderate horizons. The constants are not optimised, and the choice $\alpha = \rho = 1/n$ is made for a closed form, not claimed to be the best choice at any finite horizon; the bound may always be combined with the trivial one as $\min\{\cdot, n\}$.

Observation-model scope. Bernoulli observations provide the posterior probability kernel constructed in Appendix C on $(\mathcal{F}_{\mathrm{blr}}, \mathcal{A})$. Information comparison uses only the history-conditional mean and Pinsker's inequality for $[0, 1]$-valued observations. Thus the argument extends with unchanged constants to any complete observation model with $Y_t \in [0, 1]$ and $\mathbb{E}[Y_t \mid \mathcal{H}_{t-1}, X_t, f] = f(X_t)$ almost surely, where $\mathcal{H}_{t-1} = \sigma(X_1, Y_1, \ldots, X_{t-1}, Y_{t-1})$, provided such a posterior kernel

exists and TS uses fresh internal randomness conditionally on the history, making X_t conditionally independent of f given $\mathcal{H}_{t-1}$. $\mathbb{E}[Y_t \mid X_t, f] = f(X_t)$ alone is insufficient.

D PROOFS FOR SECTION 8

The truncated-simplex construction is the common source of three side results: cardinality tightness, a prior-level information-ratio lower bound, and the tetrahedral counterexample to the John-ellipsoid dichotomy. These results delimit the upper-bound argument; they are not used to prove the regret theorem.

Proof of Proposition 3.5. Let $d \geq 2$, $c \geq 12$, $\tau = 1/(4c(d+1))$. Take the standard d-simplex with vertices $v_0, \ldots, v_d$ and barycentric coordinates $\lambda_0, \ldots, \lambda_d$ (translated so that the average of the points below is 0). For each ordered pair $i \neq j$ let $x_{ij} = (1-\tau)v_i + \tau v_j$, and let $K = \mathrm{conv}\{x_{ij}\}$ (full-dimensional, containing 0, $D \leq \sqrt{2}$). Let $u_{ij}(x) = 1 - \lambda_i(x) + \tau\lambda_j(x)$, an affine function, and set $v = 1/4$, $\varepsilon = \tau^2/(4d)$, $\bar{f} \equiv v$, $s^u = \tau + \tau^2$, and

$$f_{ij}(x) = v - \varepsilon + \varepsilon \max\left\{\frac{s^u - u_{ij}(x)}{\tau^2},\ \frac{u_{ij}(x) - s^u}{1 - 2\tau^2}\right\}, \qquad x_{f_{ij}} = x_{ij}.$$

Each f_{ij} is the maximum of two affine functions, hence a convex ridge function, minimised exactly where $u_{ij} = s^u$, in particular at x_{ij}, with minimum $v - \varepsilon$. Evaluating u_{ij} at the $d(d+1)$ points gives exactly the values $\tau + \tau^2$ (at x_{ij}), τ (the $d-1$ points x_{il}), $1 - \tau^2$ (x_{ji}), $1 + \tau - \tau^2$ (x_{jl}), $1 - \tau$ (x_{ki}), $1 + \tau^2$ (x_{kj}) and 1 (the remaining $(d-1)(d-2)$ points), so on K the values of f_{ij} lie in $[v - \varepsilon, v]$, the points x_{il} have $f_{ij} = v$ exactly, and all other cross values satisfy $|f_{ij}(x_{kl}) - v| \leq \varepsilon E_\tau$ with $E_\tau = (2\tau - \tau^2)/(1 - 2\tau^2) < 3\tau$. Since $\delta = \varepsilon/(c(d+1)) = 4\tau\varepsilon$, every ordered pair is uninformative; the gaps are $r_f = \varepsilon$ and the comparison values coincide. Lipschitzness: $\|\nabla u_{ij}\| \leq (1+\tau)\sqrt{d}$ in the standard simplex coordinates, so $\mathrm{Lip}(f_{ij}) \leq \varepsilon(1+\tau)\sqrt{d}/\tau^2 = (1+\tau)/(4\sqrt{d}) < 1$. The functions are pairwise distinct because only f_{ij} attains $v - \varepsilon$ at x_{ij}. Along its own direction each row sees the $d-1$ points x_{il} at distance τ^2 in u_{ij}-coordinates (the near band; the unit-direction distances of Lemma 4.2 are these values divided by $\|\nabla u_{ij}\|$) and the other d^2 points at u_{ij}-distance about 1 (the far band). Finally $D/\varepsilon \leq \sqrt{2} \cdot 4d/\tau^2 = 64\sqrt{2}\, c^2 d(d+1)^2$; for every $0 < \varepsilon' \leq \varepsilon$ keep K, the selected points and the ridge directions fixed and replace each function by $f'_{ij} = v + (\varepsilon'/\varepsilon)(f_{ij} - v)$, with comparison function $\bar{f}' \equiv v$: the diagonal gaps and the cross errors are multiplied by ε'/ε while the Lipschitz constants do not increase, so the new configuration is uninformative at scale ε' (with $\delta' = \varepsilon'/(c(d+1))$) and every larger ratio D/ε' is attained. The stated parameter cost for $c = 96(d+1)$ follows by substituting $\tau = 1/(384(d+1)^2)$ and $\varepsilon = \tau^2/(4d)$; the upper bound is Theorem 3.3. For $d = 1$ take $K = [0,1]$, $\bar{f} \equiv 1/4$, $f_1 = 1/4 - \varepsilon + \varepsilon x$, $f_2 = 1/4 - \varepsilon + \varepsilon(1-x)$ with minimisers $0, 1$ and $0 < \varepsilon \leq 1/4$, which requires $D/\varepsilon \geq 4$. The bound $|C| \leq 2$ in dimension one follows from Lemma 4.2(ii): if $x_1 < x_2 < x_3$ are three minimisers of an uninformative configuration, the row of x_1 sees x_2 in its far band, $x_2 \geq x_3 - \eta(x_3 - x_1)$, and the row of x_3 sees x_2 in its far band, $x_2 \leq x_1 + \eta(x_3 - x_1)$; together $(1 - 2\eta)(x_3 - x_1) \leq 0$, contradicting $\eta < 1/2$. (For $D/\varepsilon \leq 5/12$ not even two points fit, since Lemma 4.2(i) forces a distance larger than $5\varepsilon/12$.) □

Proposition D.1 (A prior-level information-ratio lower bound)**.** *Let $d \geq 2$ and $c \geq 12$, and let $C = \{f_{ij}\}$ be the uninformative configuration constructed in the proof of Proposition 3.5, with $N = |C| = d(d+1)$, $\tau = 1/(4c(d+1))$, $\varepsilon = \tau^2/(4d)$ and the points x_{ij}. Let ξ_0 be the uniform distribution on C, and let the fixed measurable selection satisfy $x_{f_{ij}} = x_{ij}$ for all $i \neq j$ (any measurable selection can be modified on the finite set C to achieve this: singletons of $\mathcal{F}_{\mathrm{blr}}$ are measurable, since point evaluations at a countable dense subset of K separate continuous functions). Then*

$$\Delta(\pi^{\xi_0}_{\mathrm{TS}}, \xi_0) \geq \tfrac{3}{4}\varepsilon, \qquad I(\pi^{\xi_0}_{\mathrm{TS}}, \xi_0) \leq \frac{2\varepsilon^2}{N}.$$

Consequently, if $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ for this selection (in particular, if $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ uniformly over measurable selections) and $\alpha \leq \varepsilon/4$, then $\beta \geq N/8 = d(d+1)/8$.

Proof. Number the functions $f_1, \ldots, f_N$ and their selected points $x_1, \ldots, x_N$, and write $f_i(x_j) = v - \varepsilon D_{ij}$. By the proof of Proposition 3.5, the values of each f_i on K lie in $[v - \varepsilon, v]$, $f_i(x_i) = v - \varepsilon$,

and $|f_i(x_j) - v| \le \varepsilon E_\tau$ with $E_\tau < 3\tau$ for $j \neq i$; hence $D_{ii} = 1$ and $0 \le D_{ij} < 3\tau$ for $i \neq j$, where $\tau \le 1/144$ because $c \ge 12$ and $d \ge 2$. Under ξ_0 the posterior sample is uniform on C, so $\pi_{\mathrm{TS}}^{\xi_0}$ is the uniform distribution on $\{x_1, \dots, x_N\}$; every $f_i^\star = v - \varepsilon$; and $\bar{f} = \mathbb{E}_{\xi_0} f$ satisfies $\bar{f}(x_j) = v - \varepsilon\mu_j$ with $\mu_j = N^{-1}\sum_i D_{ij} \in [N^{-1}, N^{-1} + 3\tau]$. Therefore

$$\Delta(\pi_{\mathrm{TS}}^{\xi_0}, \xi_0) = \frac{1}{N}\sum_j \bar{f}(x_j) - (v - \varepsilon) = \varepsilon\Big(1 - \frac{1}{N}\sum_j \mu_j\Big) \ge \varepsilon\big(1 - N^{-1} - 3\tau\big) \ge \tfrac{3}{4}\varepsilon,$$

using $N \ge 6$ and $\tau \le 1/144$. Next,

$$\begin{aligned} I(\pi_{\mathrm{TS}}^{\xi_0}, \xi_0) &= \frac{1}{N^2}\sum_{i,j}\big(f_i(x_j) - \bar{f}(x_j)\big)^2 = \frac{\varepsilon^2}{N^2}\sum_{i,j}(D_{ij} - \mu_j)^2 \le \frac{\varepsilon^2}{N^2}\sum_{i,j} D_{ij}^2 \\ &\le \frac{\varepsilon^2}{N^2}\big(N + N(N-1)\cdot 9\tau^2\big) \le \varepsilon^2\big(N^{-1} + 9\tau^2\big), \end{aligned}$$

where the first inequality is $\sum_i (D_{ij} - \mu_j)^2 = \sum_i D_{ij}^2 - N\mu_j^2$. Since $9\tau^2 \le 9/(2304(d+1)^2) = 1/(256(d+1)^2) \le 1/(d(d+1)) = N^{-1}$, we get $I \le 2\varepsilon^2/N$. Finally, $(\alpha, \beta) \in \mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ for this selection gives $\Delta(\pi_{\mathrm{TS}}^{\xi_0}, \xi_0) - \alpha \le \sqrt{\beta I(\pi_{\mathrm{TS}}^{\xi_0}, \xi_0)}$ at the prior ξ_0, so for $\alpha \le \varepsilon/4$,

$$\beta \;\ge\; \frac{(\Delta - \alpha)^2}{I} \;\ge\; \Big(\frac{\varepsilon}{2}\Big)^2 \cdot \frac{N}{2\varepsilon^2} = \frac{N}{8}. \qquad \square$$

Remark. Proposition D.1 is a one-step statement about certificates (α, β) that are valid for every prior, as membership in $\mathrm{IR}(\mathcal{F}_{\mathrm{blr}})$ requires; it is witnessed by one prior and one admissible fixed selection. It depends on the gap scale: it constrains β only for $\alpha \le \varepsilon/4$ with $\varepsilon = \tau^2/(4d)$ (for $c = 12$, $\varepsilon = 1/(9{,}216\, d(d+1)^2)$). It is not a regret lower bound and not a separation between monotone and non-monotone links. Theorem 3.4's $(d+1)^8 m_\alpha$-order certificate exceeds this lower bound by a factor $\Theta(d^6 m_\alpha)$; a $d^{5/2}$ regret bound remains open. The family of Proposition 3.5 is uninformative for $\bar{f} \equiv v$; for the prior mean $f_{\xi_0} = \mathbb{E}_{\xi_0} f$, the cross errors satisfy $|f_i(x_j) - f_{\xi_0}(x_j)| = \varepsilon|D_{ij} - \mu_j| \ge \varepsilon(N^{-1} - 3\tau)$ for $i \neq j$, exceeding $\delta = 4\tau\varepsilon$ whenever $4c > 7d$, in particular for $c = 96(d+1)$.

Proof of Proposition 3.6. Take the regular tetrahedron $v_0 = \frac{1}{2}(1,1,1)$, $v_1 = \frac{1}{2}(1,-1,-1)$, $v_2 = \frac{1}{2}(-1,1,-1)$, $v_3 = \frac{1}{2}(-1,-1,1)$, so $\lambda_i(x) = \frac{1}{4} + \langle v_i, x\rangle$, and use the configuration of Proposition 3.5 with $K = P := \mathrm{conv}\{x_{ij}\}$ (a truncated tetrahedron; the twelve points are its vertices). The inscribed ball of the tetrahedron $S = \mathrm{conv}\{v_i\}$ is B_r with $r = 1/(2\sqrt{3})$, and on B_r one has $0 \le \lambda_i \le 1/2$, so $B_r \subset P$. Fix a vertex x_{ij} and let $b_0 = (1 - 2\tau)/(1 - \tau)$. The affine inequality $\lambda_i + b_0\lambda_j \le 1 - \tau$ is violated by x_{ij}, holds with equality at its three neighbours x_{ik} ($k \neq i, j$) and x_{ji}, and holds at all other vertices; since the cutting plane meets every edge from x_{ij} exactly at a neighbour, $P_{-ij} := \mathrm{conv}\{x_{kl} : (k,l) \neq (i,j)\} = P \cap \{\lambda_i + b_0\lambda_j \le 1 - \tau\}$. On B_r, $\max(\lambda_i + b_0\lambda_j) = \big(1 + b_0 + \sqrt{1 + b_0^2 - 2b_0/3}\big)/4 \le (2 + \sqrt{2})/4 < 1 - \tau$ for $\tau \le 1/192$, which holds for all $c \ge 12$; hence $B_r \subset P_{-ij}$ for all $i \neq j$. Now let $\zeta > 0$. Both $J_\zeta(C) = P + B_\zeta$ and $J_\zeta(C \setminus \{f_{ij}\}) = P_{-ij} + B_\zeta$ contain $B_{r+\zeta}$ and are contained in the enlarged tetrahedron $S_\zeta = \{x : \langle u_k, x\rangle \le r + \zeta,\ k = 0, \dots, 3\}$, where u_k are the unit outer normals of S, which satisfy $\sum_k u_k = 0$ and $\sum_k u_k u_k^\top = \frac{4}{3} I$. If an ellipsoid $z + AB_1$ ($A \succ 0$ symmetric) lies in S_ζ then $\langle u_k, z\rangle + \|Au_k\| \le r + \zeta$ for each k; summing and using $\|Au_k\| \ge u_k^\top A u_k$ gives $\frac{4}{3}\mathrm{tr}A \le 4(r + \zeta)$, so $\det A \le (\mathrm{tr}A/3)^3 \le (r + \zeta)^3$. Hence the maximum-volume inscribed ellipsoid of both bodies is $B_{r+\zeta}$ (equality in the arithmetic–geometric mean inequality forces $A = (r + \zeta)I$ and then $z = 0$), and every single-vertex removal leaves it unchanged. $\square$

E A LIMIT OF MAX-NORM ROUNDING

The tolerance $1/(4r)$ in Lemma 5.3 has the order $1/r$. The following proposition shows that, for arbitrary 0-1 matrices, no tolerance of order $1/\sqrt{r}$ can work; it is used only in the discussion of Section 9.

Proposition E.1 (Max-norm rounding cannot have a uniform $1/\sqrt{r}$ tolerance)**.** *For every $r = 2^j - 1$ with $j \geq 2$ there are $A, B \in \mathbb{R}^{2r \times 2r}$ such that $\mathrm{rank}A = r$, B is a 0-1 matrix with $\mathrm{rank}B = 2r$, and $\|A - B\|_{\max} = 2/(r+1)$. Consequently, for every constant $a > 0$ there are infinitely many r for which the implication "$\mathrm{rank}A \leq r$ and $\|A - B\|_{\max} \leq a/\sqrt{r}$ for a 0-1 matrix B imply $\mathrm{rank}B \leq 2r - 1$" fails.*

Proof. Let $s = r + 1 = 2^j$ and let S_s be the Sylvester–Hadamard matrix of order s, defined by $S_1 = (1)$ and $S_{2m} = \left(\begin{smallmatrix} S_m & S_m \\ S_m & -S_m \end{smallmatrix}\right)$; it is symmetric with entries ± 1, its first row and column consist of ones, and $S_s S_s^\top = sI_s$. Write $S_s = \left(\begin{smallmatrix} 1 & \mathbf{1}^\top \\ \mathbf{1} & C \end{smallmatrix}\right)$ with C a symmetric $r \times r$ matrix with entries ± 1. Orthogonality of each of the rows $2, \ldots, s$ to the first row gives $C\mathbf{1} = -\mathbf{1}$, and orthogonality among the rows $2, \ldots, s$ gives $J + CC^\top = sI_r$, where $J = \mathbf{1}\mathbf{1}^\top$; that is, $CC^\top = sI_r - J$. Let $W = (J - C)/2$, a 0-1 matrix. Using $JC^\top = \mathbf{1}(C\mathbf{1})^\top = -J$,

$$W\Big(-\frac{2}{s}C^\top\Big) = -\frac{1}{s}\big(JC^\top - CC^\top\big) = -\frac{1}{s}\big(-J - sI_r + J\big) = I_r,$$

so W is non-singular with $W^{-1} = -\frac{2}{s}C^\top$, whose entries are $\pm 2/s$; hence $\|W^{-1}\|_{\max} = 2/(r+1)$. Now set

$$A = \begin{pmatrix} I_r & W \\ W^{-1} & I_r \end{pmatrix} = \begin{pmatrix} I_r \\ W^{-1} \end{pmatrix} (I_r \quad W), \qquad B = \begin{pmatrix} I_r & W \\ 0 & I_r \end{pmatrix}.$$

The factorisation shows $\mathrm{rank}A \leq r$, and the top-left block I_r shows $\mathrm{rank}A \geq r$. B is a 0-1 matrix, block upper triangular with identity diagonal blocks, so $\det B = 1$ and $\mathrm{rank}B = 2r$. Finally $A - B$ has the single non-zero block W^{-1}, so $\|A - B\|_{\max} = \|W^{-1}\|_{\max} = 2/(r+1)$. For the consequence, $2/(r+1) \leq a/\sqrt{r}$ as soon as $2\sqrt{r}/(r+1) \leq a$, which holds for all large r since $2\sqrt{r}/(r+1) \to 0$. □

The smallest instance is $r = 3$, where the construction gives $W = J_3 - I_3$ up to a row permutation, with $\|A - B\|_{\max} = 1/2 < 1/\sqrt{3}$. The proposition concerns arbitrary 0-1 matrices only; the band-membership matrices of Section 5 have additional structure (two bands, one of them extremely thin), and whether that structure admits a rounding statement with a larger tolerance is open.

F Extended related work

Ridge (single-index) bandits have been studied under different learning criteria and for different algorithms. Lattimore (2021) proves $O(d\sqrt{n}\log(nD))$ minimax regret, under his domain normalisation, for an adversarial ridge model with a hidden fixed direction and convex links that may vary over time, by information-theoretic arguments and minimax duality; as noted above, this establishes learnability of the ridge structure without bearing on standard TS under an arbitrary prior. There D denotes the diameter of that normalised domain. Bakhtiari et al. (2025) analyse exact TS directly and prove the monotone ridge bound; the present paper removes monotonicity within the ridge structure. Rajaraman et al. (2024), Rajaraman & Han (2026a) and Kang et al. (2026) study frequentist nonlinear ridge, single-index and contextual single-index bandits with algorithms other than TS, and Rajaraman & Han (2026b) give minimax lower bounds for general stochastic bandit convex optimisation, which quantify the difficulty of general convex losses but are not specific to non-monotone ridge losses or to TS. For contextual single-index bandits with a possibly non-monotone link and i.i.d. contexts, under their regularity and non-degeneracy assumptions, Dey et al. (2026) obtain frequentist $\tilde{O}(T^{2/3})$ regret for fixed dimension with the UCB-based `ZoomSIB-UCB` algorithm and a matching minimax lower bound of order $T^{2/3}$ up to logarithmic factors, with the lower-bound construction allowing the number of available arms to grow with T; their model, criterion and algorithm differ from ours, and these results establish neither a Bayesian regret guarantee nor a failure result for exact-posterior TS in our non-contextual convex ridge model. None of these results provides a Bayesian regret guarantee for exact-posterior TS over arbitrary convex non-monotone ridge links. We do not address the *second* question of the same paragraph of Bakhtiari et al. (2025) (the information ratio with a *known* link); see also the MSc thesis of Bakhtiari (2025). Information-ratio analyses of TS go back to Russo & Van Roy (2016); the convex-bandit machinery we build on is from Bubeck et al. (2015); Bubeck & Eldan (2018); Lattimore (2019); Bakhtiari et al. (2025).

Interactive learning and single-index models. Rajaraman & Han (2026a) analyse stochastic gradient descent for the interactive learning of single-index models (generalised linear or ridge bandits): under their link regularity assumptions and specified initialisation, exploration and learning-rate schedules, their SGD procedure exhibits a burn-in phase followed by a learning phase and achieves near-optimal or best-known sample-complexity and regret guarantees, depending on the link class. For the known reward link $g(m) = 0$ for $m \leq 0$, $g(m) = -m$ for $0 < m \leq 1/3$ and $g(m) = m - 2/3$ for $1/3 < m \leq 1$, they show that their SGD dynamics satisfy $\mathbb{P}(\max_{t\in[T]} m_t \leq 0.2) \geq 1 - \delta$ whenever the initial alignment satisfies $m_1 \leq 0.1$, the exploration schedule satisfies $\sigma_t \leq 0.1$, and the learning rate satisfies $\eta_t \leq c/\log(T/\delta)$ for a sufficiently small absolute constant $c > 0$ (their Proposition 1; Proposition 5.1 in the arXiv version). This link is non-monotone and is not convex on the full domain $[-1, 1]$. Their result concerns alignment in a known-link reward model, not the Bayesian regret of TS on our convex-loss class; the comparison with Theorem 3.1 is a qualitative contrast between settings. Kang et al. (2026) study single-index contextual bandits with an unknown, continuously differentiable increasing reward link under their stochastic-design assumptions; they propose the Stein-estimator-based STOR and ESTOR algorithms, extend the analysis to sparse high-dimensional index parameters, and treat general non-monotone links with GSTOR under Gaussian design and their stated regularity conditions. Their regret criterion is frequentist, and arm sets are sampled i.i.d. rather than chosen adversarially.

Thompson sampling and information. Luo & Bayati (2025) develop data-driven frequentist regret bounds for POFUL, a framework containing OFUL, LinTS, TS-Freq and Greedy and allowing non-optimistic samples, and use these bounds to construct course-corrected LinTS and Greedy variants with minimax-optimal frequentist regret; the model is linear. Shufaro et al. (2025) derive Fano-type Bayesian regret lower bounds in their contextual-information framework and also give an accumulated-information upper bound for Thompson sampling in a multi-armed bandit environment (their Proposition 4.3); these results do not supply a uniform information-ratio certificate for our convex ridge class.

Learnability and the exact scope of the open question. Lattimore (2021) establishes the stated adversarial minimax bound for bounded, 1-Lipschitz ridge losses using a Bayesian information-theoretic argument and a policy other than exact TS. The information-ratio question for a known convex link raised in the discussion of Bakhtiari et al. (2025) is outside our scope; the MSc thesis of Bakhtiari (2025) is cited for completeness, and we have not been able to consult its full text.