Direct Contextual UCB under Exact Cauchy Noise

Anonymous Author(s)
Affiliation
Address
email

Abstract

Contextual bandits whose reward noise is so heavy-tailed that the mean does not exist break the usual pseudo-regret benchmark; we study exact symmetric Cauchy noise, targeting latent location regret. We analyze Direct-Cauchy-UCB, a one-pass arm-local projected gradient update driven by the Cauchy likelihood score, with optimism from the current iterate, and $\mathcal{O}(d)$ state per arm. Under bounded parameters, bounded contexts and directional excitation—every direction excited at every step—it attains high-probability location regret $\widetilde{\mathcal{O}}(K \log T + \sqrt{KT \log(KT/\delta)})$. Our lower bound of order $\sqrt{KT}$ matches this up to logarithmic factors once the horizon is large enough. The same update keeps that rate under a weaker windowed persistent-excitation assumption, covering deterministic adversarial context designs. Without any excitation assumption, an ellipsoidal variant is sublinear at $\widetilde{\mathcal{O}}(\sqrt{dK}T^{3/4})$ with $\mathcal{O}(d^2)$ per-arm state. Beyond exact Cauchy noise, a sign (median-score) variant handles symmetric heavy tails at the same rate, with no moment assumption, given a known lower bound on the noise law's slope at its median. Bounded per-arm deviations of size ρ from Cauchy cost the exact-score policy an additive $O(\rho T)$, against a floor $\Omega(\rho T/\sqrt{d})$ proved for that class. On synthetic exact-Cauchy environments our policy attains the lowest location regret among a strengthened suite of robust baselines at low and moderate noise scales.

1 Introduction

We study the stochastic linear contextual bandit problem with exact Cauchy noise. There are K arms and a horizon T. In every round $t \in [T]$, a context vector $y_{k,t} \in \mathbb{R}^d$ is revealed for every arm $k \in [K]$, the learner selects an arm k_t, and observes the reward

$$r_t = y_{k_t,t}^\top \theta_{k_t} + \xi_t. \tag{1}$$

where $\theta_k \in \mathbb{R}^d$ is the unknown parameter of arm k, and the noise is symmetric Cauchy with unknown scale $\gamma_\star > 0$:

$$\xi_t \mid \mathcal{F}_{t-1}, \{y_{j,t}\}_{j=1}^K \sim \mathrm{Cauchy}(0, \gamma_\star), \tag{2}$$

with $\mathcal{F}_{t-1}$ the history up to the conclusion of round $t-1$. The Cauchy law has no first moment, so the expected reward of an arm does not exist, and the standard benchmark, mean pseudo-regret, is undefined. Since no mean exists, we measure performance against the latent location parameter of the reward distribution, an approach taken in the super-heavy-tailed frameworks of Zhong et al. [2021], Dorn et al. [2025]. We evaluate policies by the cumulative *location regret*

$$R_{\mathrm{loc}}(T) = \sum_{t=1}^{T} \left(y_{k_t^\star,t}^\top \theta_{k_t^\star} - y_{k_t,t}^\top \theta_{k_t} \right), \qquad k_t^\star \in \arg\max_{k \in [K]} y_{k,t}^\top \theta_k. \tag{3}$$

Two properties of the Cauchy likelihood score drive the construction. The score is uniformly bounded, so a gradient step driven by it survives infinite-mean noise without hard truncation; and its expected

Submitted to 40th Conference on Neural Information Processing Systems (NeurIPS 2026). Do not distribute.

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.

drift points exactly at the true location (Lemma 1), so a one-pass arm-local projected update is analyzable directly, without refitting a ridge or a maximum-likelihood estimator. What is new here is the construction rather than the objective: exact symmetric Cauchy noise, a direct one-pass arm-local update, and a UCB rule built from the current iterate rather than a refitted ridge or maximum-likelihood estimator, together with an explicit confidence sequence, a matching minimax lower bound, and an excitation-free adversarial-context variant.

Our contributions are as follows:

- **Core guarantee.** A confidence sequence uniform over the local pull count (Proposition 1) and a high-probability location-regret bound $\widetilde{\mathcal{O}}(K \log T + \sqrt{KT \log(KT/\delta)})$ for the Direct-Cauchy update (Theorem 1), with a minimax lower bound of order $\gamma_\star \sqrt{KT}$ once the horizon is large enough (Theorem 2); gap-dependent refinements are in the appendices (Corollary 2, Theorem 7).
- **Beyond the idealized context model.** Windowed persistent excitation—covering deterministic adversarial designs—keeps the same rate with $\mathcal{O}(d)$ per-arm state (Theorem 3, Theorem 8); without any excitation assumption, an ellipsoidal variant guarantees $\widetilde{\mathcal{O}}(\sqrt{dK}\, T^{3/4})$ with $\mathcal{O}(d^2)$ per-arm state (Theorem 4).
- **Beyond Cauchy noise.** A sign-update variant attains the same rate for symmetric heavy-tailed noise with no moment assumption, given a known lower bound on the median-growth slope (Theorem 5); bounded per-arm deviations of size ρ cost an additive $O(\rho T)$ (Proposition 7), matching a floor $\Omega(\rho T/\sqrt{d})$ for that class (Proposition 10).
- **Scale robustness and adaptation.** The same rate under constant-factor score-scale misspecification (Corollary 1); adaptation to an unknown scale by a dyadic grid with a doubling schedule that removes the known upper bound (Theorem 6, Remark 1, Appendix B).
- **Experiments.** On synthetic exact-Cauchy environments the direct policy attains the lowest location regret across a strengthened robust baseline suite at noise scales up to 0.12 on both context families (and at 0.24 on Gaussian contexts); at the largest scale a ridge-style baseline leads—a difference in constants, not rates (Section 5).

Section 3 and Section 4 develop the exact-Cauchy results; within Section 4, Section 4.2 relaxes the excitation condition and Section 4.3 relaxes the noise model.

1.1 Related work

Classical linear and generalized-linear contextual bandits enforce mean pseudo-regret bounds using ridge-based or likelihood-based confidence sets [Li et al., 2010, Chu et al., 2011, Abbasi-Yadkori et al., 2011, Filippi et al., 2010, Li et al., 2017, Jun et al., 2017]. Heavy-tailed variants keep the mean-based objective and replace the estimator — truncation, Huber or Catoni robustification, one-pass updates, variance-adaptive intervals [Muñoz Medina and Yang, 2016, Shao et al., 2018, Kang and Kim, 2023, Xue et al., 2023, Ye et al., 2025, Li and Sun, 2024, Wang et al., 2025] — so they require at least a finite mean, and their objectives are ill posed under exact Cauchy noise. The nearest conceptual ancestors are bounded-influence robust location methods [Huber, 1964, Catoni, 2012], and Depersin and Lecué [2023] show that a location parameter is estimable with not even a first moment. The closest bandit counterparts are Zhong et al. [2021], who admit super heavy tails up to Cauchy but through a costly mean-of-medians index, and Dorn et al. [2025], who handle Cauchy noise with fast UCB-type updates but are not contextual. Our contribution fuses the two regimes: a direct one-pass Cauchy score with arm-local optimism under fresh contexts every round, yielding an explicit discrete-time confidence sequence and minimax-sharp location-regret guarantees. The idea of setting the objects of selection themselves as optimized functions is pursued in [Latypov et al., 2025, Dorn et al., 2026]; the present work takes a similar approach, defining one optimized function per arm.

The bandit literature on heavy-tailed rewards quantifies what a moment assumption buys: under a bounded $(1+\epsilon)$-th moment the minimax rate degrades from $\sqrt{T}$ to $T^{1/(1+\epsilon)}$ [Bubeck et al., 2013], and recent work sharpens its dimension and horizon dependence [Xue et al., 2020, Tajdini et al., 2025], makes it variance-aware and one-pass [Wang et al., 2025, Tani and Futami, 2026], extends it to adversarial loss vectors [Zhao et al., 2025] and batched feedback [Guo et al., 2025], and removes the need to know the noise magnitude [Xu et al., 2023, Jun and Kim, 2024] or even the moment index

itself [Genalti and Metelli, 2026]. A Cauchy law has no moment of any order: the rate $T^{1/(1+\epsilon)}$ degenerates to the trivial one as $\epsilon \downarrow 0$, and without knowledge of the moment parameters no strategy is uniformly sublinear over ϵ [Genalti and Metelli, 2026]. Adapting to misspecification is a different deviation notion [Krishnamurthy et al., 2021]: there the model class is wrong while the noise keeps its moments.

A separate line replaces heavy tails by an adversary: attack- and corruption-robust linear bandits [Bogunovic et al., 2021, Ding et al., 2022, He et al., 2022], fully adversarial linear contextual bandits [Neu and Olkhovskaya, 2020], and corruption-robust nonlinear and generalized-linear methods [Ye et al., 2023, Yu et al., 2025]. There the adversary has a bounded budget and the mean exists; the per-arm deviation model of Section 4.3 is a different primitive — the deviation is bounded at every round, but it is added to Cauchy noise, which has a location and no mean — and Proposition 10 shows that a linear-in-ρ term is unavoidable for it.

Two design axes of this paper have direct counterparts. The pathwise windowed excitation of Assumption 2 is the classical persistent-excitation condition of adaptive control, stated there pathwise as a sliding-window eigenvalue floor on the regressor Gram matrix [Karapetyan et al., 2024]; what is new here is the windowed cumulative requirement combined with the conditional-expectation form (23), which is strictly weaker and tolerates genuinely stochastic adaptive contexts. With no assumption on the context process, Theorem 4 attains the rate $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$; a similar rate is obtained by Muthirayan et al. [2020], who instead inject exploration noise into the contexts themselves to induce excitation. Sign-based updates are effective under heavy-tailed noise in stochastic optimization [Yu et al., 2026], though under a finite-moment condition ($p > 1$) that our median-growth condition removes; a one-pass adaptive-Huber line for nonlinear bandits exists as well [Zheng et al., 2026].

Paper organization. Section 2 introduces the problem setting and assumptions. Section 3 presents the direct Cauchy update, the proposed UCB scheme, and the associated computational overhead. Section 4 states the main confidence-sequence and regret guarantees, together with robustness results for score-scale misspecification, adversarial context processes (Section 4.2), and general symmetric heavy-tailed noise (Section 4.3); the dyadic-grid scale adaptation, the instance-dependent and gap-dependent refinements, and the derivation details of the robustness variants are deferred to the appendices (Theorem 6, Corollary 2, Theorem 7). Section 5 reports empirical results. Section 6 states the limitations of the study, with an extended discussion in Appendix H. Proofs and auxiliary lemmas are deferred to the appendices.

2 Problem Setting

The model is the exact Cauchy linear framework of Section 1; we collect here the standing assumptions and notation. We assume bounded parameters and bounded contexts: $\|\theta_k\|_2 \le S$ and $\|y_{k,t}\|_2 \le L$ for all arms and rounds. To analyze the localized updates, let τ denote the local pull count for a specific arm, and let $\mathcal{F}_{k,\tau}$ denote the localized filtration generated by the sequence of contexts and rewards associated with arm k up to its τ-th pull. We impose the following geometric condition on the contextual process.

Assumption 1 (Directional excitation)**.** *There exists a constant $\mu_{\mathrm{dir}} > 0$ such that for every arm k, every local time τ, and any $\mathcal{F}_{k,\tau}$-measurable vector $v \in \mathbb{R}^d$, the context vectors satisfy*

$$\mathbb{E}\big[(y_{k,\tau+1}^\top v)^2 \mid \mathcal{F}_{k,\tau}\big] \ge \mu_{\mathrm{dir}}\|v\|_2^2. \tag{4}$$

Assumption 1 is a conditional second-moment floor per direction given the arm-local history, not an i.i.d. or stationarity requirement, so it tolerates adaptive and history-dependent context processes. For discussion and examples we refer to Appendix A.3.

Section 4.2 relaxes this condition along two axes: windowed persistent excitation requires excitation only cumulatively over windows of consecutive pulls and already covers deterministic adversarial designs at the same $\widetilde{\mathcal{O}}(\sqrt{KT})$ rate, while an ellipsoidal variant of the same update drops excitation entirely, remaining valid under fully adversarial contexts at a degraded $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ rate.

3 The Direct Cauchy Policy

The standard approach to contextual linear bandits relies on regularized least squares or MLE refitting, both of which become unstable or computationally burdensome under exact Cauchy noise. We propose a direct, one-pass projected update driven by the Cauchy likelihood score.

Rather than maintaining a global empirical covariance matrix, the policy operates via arm-local updates: for each arm $k \in [K]$ the learner maintains a localized parameter iterate $z_{k,\tau} \in \mathbb{R}^d$, τ denoting the local pull count, and performs projected online gradient descent on the negative log-likelihood surface of the Cauchy distribution. Let $\mathcal{X} = \{x \in \mathbb{R}^d : \|x\|_2 \le S\}$ define the bounded parameter space, and let $\bar{\gamma} > 0$ denote the score scale hyperparameter. Upon observing a reward r_t for the selected arm k_t, the learner computes the gradient of the Cauchy log-likelihood and updates only the active arm via the mapping

$$z_{k_t,\tau+1} = \Pi_{\mathcal{X}}\left(z_{k_t,\tau} - \eta_\tau \frac{2(y_{k_t,t}^\top z_{k_t,\tau} - r_t)}{\bar{\gamma}^2 + (y_{k_t,t}^\top z_{k_t,\tau} - r_t)^2} y_{k_t,t} \right), \tag{5}$$

where $\Pi_{\mathcal{X}}$ denotes the Euclidean projection onto $\mathcal{X}$, and η_τ is the localized step size. The Cauchy score is bounded, $\sup_u |\varphi_{\bar{\gamma}}(u)| = 1/\bar{\gamma}$, so the stochastic gradient is bounded by $G := L/\bar{\gamma}$ in norm; this controls infinite-amplitude outliers without hard truncation.

The full procedure, termed Direct-Cauchy-UCB, is formalized in Algorithm 1, parameterized by a generic step size sequence $\{\eta_\tau\}_{\tau \ge 0}$ and a confidence radius schedule $\{\beta_\tau\}_{\tau \ge 0}$, which decouples the computational loop from the martingale analysis.

Algorithm 1 Direct-Cauchy-UCB

Require: Score scale $\bar{\gamma} > 0$, parameter radius $S > 0$. Step size sequence $\{\eta_\tau\}_{\tau \ge 0}$, confidence schedule $\{\beta_\tau\}_{\tau \ge 0}$.

Initialize local pull counts $T_k(0) = 0$ for all $k \in [K]$.
Initialize parameter iterates $z_{k,0} = \mathbf{0} \in \mathbb{R}^d$ for all $k \in [K]$.
for round $t = 1, \dots, T$ **do**
  Observe contextual features $\{y_{1,t}, \dots, y_{K,t}\} \subset \mathbb{R}^d$.
  for each arm $k \in [K]$ **do**
    Compute the optimistic index:

$$\mathrm{UCB}_{k,t} = y_{k,t}^\top z_{k,T_k(t-1)} + \beta_{T_k(t-1)} \|y_{k,t}\|_2$$

  end for
  Select action $k_t \in \arg\max_{k \in [K]} \mathrm{UCB}_{k,t}$.
  Observe heavy-tailed reward r_t.
  Execute the parameter update via Eq. (5) to obtain $z_{k_t,\tau+1}$ with $\tau = T_{k_t}(t-1)$.
  Update pull counts: $T_{k_t}(t) = \tau + 1$ and $T_k(t) = T_k(t-1)$ for all $k \ne k_t$.
end for

Computational footprint. The direct policy retains only the current arm iterate, a pull count, and the confidence radius: $\mathcal{O}(d)$ state and $\mathcal{O}(d)$ work per played-arm update, against at least one covariance or precision matrix per arm—$\mathcal{O}(d^2)$ state and typically $\mathcal{O}(d^2)$ work—for ridge-style robust baselines, a practical advantage in addition to the statistical one.

4 Main Theoretical Results

The template behind the guarantees. Every guarantee below instantiates a single template. A condition on the context process yields a contraction modulus κ, and the modulus fixes the update parameters, the step size η_τ and the burn-in τ_0. From these one obtains a one-arm confidence radius β_τ decaying as $1/\sqrt{\tau + \tau_0}$, and the deterministic reduction of Appendix A.1 turns the radius into a regret bound once and for all. Algorithm 1 is therefore stated with generic schedules $\{\eta_\tau\}$ and $\{\beta_\tau\}$, which each section instantiates in its own way. The results of this section and of Section 4.2 change only the first input, Section 4.3 changes the score, and the ellipsoidal variant is the single departure: it replaces the radius by an ellipsoidal confidence set.

Here and below, $\widetilde{\mathcal{O}}(\cdot)$ hides numerical constants and polylogarithmic factors in T, K and $1/\delta$, and $\lesssim$ hides a numerical constant. The regret guarantees in the body display the dependence on the horizon T and the number of arms K; the exact dependence on the problem parameters is deferred to the full statements in the appendices.

4.1 Core guarantees under exact Cauchy noise

Unless stated otherwise, the results below hold under the exact Cauchy linear framework of Section 1. We write $B := 2LS$ for the signal range throughout. The template begins with Lemma 1, which isolates the mechanism enabling one-pass stability.

Lemma 1 (Cauchy score compatibility)**.** *Let $\xi \sim \mathrm{Cauchy}(0, \gamma_\star)$ with $\gamma_\star > 0$, let $\bar{\gamma} > 0$, and define the likelihood score function $\varphi_{\bar{\gamma}}(u) = 2u/(\bar{\gamma}^2 + u^2)$. For every $\Delta \in \mathbb{R}$,*

$$\mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] = \frac{2\Delta}{\Delta^2 + (\gamma_\star + \bar{\gamma})^2}. \tag{6}$$

Consequently, the score is uniformly bounded and the induced map $m(\Delta) = \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)]$ is uniformly Lipschitz,

$$\sup_{u \in \mathbb{R}} |\varphi_{\bar{\gamma}}(u)| = \frac{1}{\bar{\gamma}}, \qquad \sup_{\Delta \in \mathbb{R}} \left| \frac{d}{d\Delta} \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] \right| \le \frac{2}{(\gamma_\star + \bar{\gamma})^2}, \tag{7}$$

and on $|\Delta| \le B$ the directional drift satisfies

$$\Delta\, \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] \ge \frac{2\Delta^2}{B^2 + (\gamma_\star + \bar{\gamma})^2}. \tag{8}$$

Proof in Appendix A.2.

Lemma 1 isolates the core mechanism that enables the direct update to survive infinite-mean noise: the Cauchy score guarantees a bounded gradient update while preserving an exact, positive-definite directional drift toward the true latent location, a stable stochastic approximation without moment truncation. Define the scale-dependent contraction modulus and prediction limits as

$$C_{\mathrm{Cau}} = \frac{2}{B^2 + (\gamma_\star + \bar{\gamma})^2}, \qquad \kappa_{\mathrm{dir}} = \mu_{\mathrm{dir}} C_{\mathrm{Cau}}, \qquad \ell = \frac{2L^2}{(\gamma_\star + \bar{\gamma})^2}. \tag{9}$$

where μ_{dir} is the constant of Assumption 1, which the results of this subsection impose on the contextual process.

The stability of the one-pass update necessitates a deterministic burn-in period, $\tau_0 = \left\lceil \max\left\{1, \frac{\ell}{\kappa_{\mathrm{dir}}}, \frac{\ell^2}{\kappa_{\mathrm{dir}}^2}\right\}\right\rceil$. With the deterministic reduction of the template stated once and for all, we concentrate here on the stochastic concentration of the arm-local iterates, which is where the Cauchy structure enters.

Proposition 1 (One-arm Cauchy confidence sequence)**.** *Let $\delta \in (0, 1)$. Fix an arm k and a local horizon n, and run the one-pass projected update defined in Eq.* (5) *with the localized step size $\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau + \tau_0 + 1)}$. Under Assumption 1, with probability at least $1 - \delta$, simultaneously for all local times $\tau \in \{0, \dots, n\}$, the estimation error satisfies $\|z_{k,\tau} - \theta_k\|_2 \le \beta_\tau$ and the prediction error satisfies $|y^\top z_{k,\tau} - y^\top \theta_k| \le \beta_\tau \|y\|_2$ for every $y \in \mathbb{R}^d$, where the radius satisfies*

$$\beta_\tau \;\lesssim\; \frac{S + L/(\kappa_{\mathrm{dir}}\bar{\gamma})}{\tau + \tau_0} + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}} \sqrt{\frac{\log(2n/\delta)}{\tau + \tau_0}}.$$

The exact constants behind the $\lesssim$ form of Proposition 1 are given, with the proof, in Appendix A.4.

Proof sketch. The drift inequality of Lemma 1 contracts the update error at the rate $1/(\tau + \tau_0)$, and a variance-smoothed potential with an Azuma–Hoeffding maximal inequality and a union bound over $\tau \le T$ promotes this to a time-uniform confidence sequence.

The following theorem gives the global minimax performance of the policy by combining the stochastic concentration of Proposition 1 with the deterministic reduction of Appendix A.1.

Theorem 1 (Direct Cauchy location regret)**.** *Let $\delta \in (0,1)$. Execute Algorithm 1 initialized with learning rate schedule η_τ and the confidence schedule β_τ from Proposition 1 instantiated with $n := T$ and $\delta' = \delta/K$. Under Assumption 1, with probability at least $1-\delta$, the cumulative location regret satisfies*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(KT/\delta)}\Big),$$

where the leading constants depend only on L, S, $\gamma_\star$, $\bar{\gamma}$ and μ_{dir}.

The exact constants behind the $\widetilde{\mathcal{O}}$ in Theorem 1 are given, with the proof, in Appendix A.5.

The policy is also rate-robust to a constant-factor misspecification of the score scale: if $\underline{c}\gamma_\star \le \bar{\gamma} \le \bar{c}\gamma_\star$ for absolute constants $0 < \underline{c} \le \bar{c} < \infty$, the rate of Theorem 1 is preserved with leading constants depending only on the structural parameters and the misspecification interval (Corollary 1, Appendix A.6).

An instance-dependent refinement of Theorem 1 is deferred to Appendix A.7: it bounds the same policy through the uniform contextual margins Δ_k of the arms, the infimum of an arm's realized disadvantage over the rounds in which it is strictly suboptimal.

If the true scale $\gamma_\star$ is not known exactly, we offer two ways to adapt. In the first, the learner knows a bounding interval $\gamma_{\min} \le \gamma_\star \le \Gamma_{\max}$ and runs a dyadic grid of scale candidates over it — one Direct-Cauchy branch per grid scale with computable surrogate schedules, and per-arm model selection by the smallest current confidence radius — preserving the rate of Theorem 1 up to a factor $\log(M+1)$ inside the logarithmic term, with $M+1$ the number of branches (Theorem 6, Appendix B). In the second, a doubling schedule restarts the grid on growing brackets, removing the need for a known upper bound at the price of a logarithmic overhead (Remark 1, Appendix B). Together, the results show that the direct Cauchy update is statistically robust, insensitive to moderate scale uncertainty, and adaptive to the difficulty of the instance it faces.

The derived $\sqrt{KT}$ worst-case rate is information-theoretically sharp: the following minimax lower bound matches it; here $\mathcal{N}$ is a subclass of constant-context instances, constructed in Appendix C.2. A gap-dependent lower bound for the constant-context subclass, based on a statistical-distance comparison (Appendix C.1), is given in Appendix C.3.

Theorem 2 (Minimax lower bound)**.** *For any $K \ge 2$ and $T \ge 1$, there exists a constant-context linear Cauchy subclass $\mathcal{N}$ such that every bandit policy satisfies*

$$\sup_{\nu \in \mathcal{N}} \mathbb{E}_\nu[R_{\mathrm{loc}}(T)] \ge \frac{T\Delta_T}{4}, \qquad \Delta_T = \min\Big\{LS,\, \gamma_\star\sqrt{\frac{K}{2T}}\Big\}.$$

Proof in Appendix C.2. In the square-root regime $T \ge K\gamma_\star^2/(2L^2S^2)$, this yields $\sup_{\nu\in\mathcal{N}} \mathbb{E}_\nu[R_{\mathrm{loc}}(T)] \ge \frac{\gamma_\star}{4\sqrt{2}}\sqrt{KT}$.

Theorem 2 matches the horizon dependence of Theorem 1 up to logarithmic factors whenever the $K \log T$ initialization term is not dominant; the hard subclass is one-dimensional and embeds into the class covered by Assumption 1 in any dimension d. A gap-dependent lower bound of a different kind—instance-dependent and applying to uniformly good policies—is Theorem 7 in Appendix C.3.

4.2 Adversarial contexts: persistent excitation and an ellipsoidal variant

The results above are derived under Assumption 1. This subsection derives guarantees under weaker context conditions: a windowed persistent-excitation condition, which requires excitation only cumulatively over windows, retains the $\widetilde{\mathcal{O}}(\sqrt{KT})$ rate and the $\mathcal{O}(d)$ per-arm state of Algorithm 1, and an ellipsoidal variant with no excitation condition at all remains valid at $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ with $\mathcal{O}(d^2)$ per-arm state.

Windowed persistent excitation. Assumption 1 requires every single context to excite every direction in conditional expectation. We now show that the direct Cauchy policy retains its $\widetilde{\mathcal{O}}(\sqrt{KT})$ guarantee when excitation is only required *cumulatively over windows* of w consecutive pulls of an arm. A similar windowed condition is used in [Karapetyan et al., 2024]. The pathwise form (Assumption 2) is the persistent-excitation condition of adaptive control for arm-local windows and covers arbitrary deterministic adversarial designs that enforce it (Remark 2, Appendix D).

Assumption 2 (Windowed persistent excitation, pathwise form)**.** *There exist an integer $w \geq 1$ and a constant $\mu > 0$ such that for every arm k, every local time τ, and every $v \in \mathbb{R}^d$,*

$$\sum_{j=1}^{w} (y_{k,\tau+j}^{\top} v)^2 \geq \mu w \|v\|_2^2 \qquad \textit{almost surely,} \tag{10}$$

The condition is equivalent to the eigenvalue floor $\lambda_{\min}\big(\sum_{j=1}^{w} y_{k,\tau+j} y_{k,\tau+j}^{\top}\big) \geq \mu w$ holding almost surely. As an eigenvalue condition, it defines a measurable event, so it also applies to random directions v. The window Gram matrix has rank at most $\min(w, d)$, so (10) is feasible only for $w \geq d$.

Proposition 2 (One-arm confidence sequence under pathwise windowed excitation)**.** *Let $\delta \in (0, 1)$. Fix an arm k and a local horizon n, and run the update* (5) *with the per-window step-size schedule of Appendix D. Under Assumption 2, with probability at least $1 - \delta$, simultaneously for all local times $\tau \in \{0, \dots, n\}$, the estimation error satisfies $\|z_{k,\tau} - \theta_k\|_2 \leq \beta_\tau^{\mathrm{pe}}$ and the prediction error satisfies $|y^{\top}(z_{k,\tau} - \theta_k)| \leq \beta_\tau^{\mathrm{pe}} \|y\|_2$ for every context y, where the radius satisfies*

$$\beta_\tau^{\mathrm{pe}} \;\lesssim\; \frac{A_w}{\tau + \tau_0^{\mathrm{pe}}} + \frac{G}{\kappa_{\mathrm{pe}}} \sqrt{\frac{\log(2n/\delta)}{\tau + \tau_0^{\mathrm{pe}}}},$$

where $A_w = \widetilde{\mathcal{O}}\big(Sw + (L^2/\mu + 1)\, G w^{3/2}/\kappa_{\mathrm{pe}}\big)$, and the schedule, the burn-in τ_0^{pe} and the modulus κ_{pe} are those of Appendix D.

The exact constants behind the $\lesssim$ form of Proposition 2 are given, with the proof, in Appendix D. The leading coefficient of $1/\sqrt{\tau + \tau_0^{\mathrm{pe}}}$ is independent of the window length: w enters only the lower-order harmonic constant A_w and the promotion factors, which are $\Theta(1)$ and absorbed into $\lesssim$.

Theorem 3 (Location regret under adversarial windowed-excited contexts)**.** *Let $\delta \in (0, 1)$. Execute Algorithm 1 with the step-size schedule and the confidence schedule β_τ^{pe} of Proposition 2, instantiated with local horizon $n := T$ and failure level δ/K. Under Assumption 2, with probability at least $1 - \delta$, the location regret satisfies*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(2KT/\delta)}\Big).$$

The exact constants behind the $\widetilde{\mathcal{O}}$ in Theorem 3 are given, with the proof, in Appendix D. The bound is uniform in the window length: w enters only through the burn-in and the lower-order terms (Proposition 2). Assumption 2 can be weakened to an expectation form (Assumption 3, Appendix D), which lower-bounds only the conditional expectation of the same windowed quantity: the same guarantees survive with an additional term in the radius and in the regret that scales with the square root of the window length (Theorem 8, Appendix D).

Fully adversarial contexts without excitation: an ellipsoidal direct policy. The guarantees of Proposition 1 and Theorem 1 still require excitation, albeit only cumulatively over windows. The same direct Cauchy OGD update, combined with an *ellipsoidal* confidence set instead of the ball-confidence schedule of Proposition 1, admits a high-probability location-regret guarantee with *no assumption on the contextual process whatsoever*: the contexts $\{y_{k,t}\}$ may be chosen by an adversary with full knowledge of the history, subject only to $\|y_{k,t}\|_2 \leq L$ and to being revealed before the noise (predictability). The price is a deterioration of the horizon rate from $\sqrt{T}$ to $T^{3/4}$, an explicit dependence on d, and an $\mathcal{O}(d^2)$ per-arm memory footprint. The construction is an online-to-confidence-set conversion [Abbasi-Yadkori et al., 2012], with the ridge predictor replaced by the direct Cauchy iterate; the full policy (Algorithm 3) and the conversion are given in Appendix E.

Theorem 4 (Location regret under fully adversarial contexts)**.** *Let $\delta \in (0, 1)$. In the exact Cauchy linear model of Section 1 with adversarial predictable contexts, $\|y_{k,t}\|_2 \leq L$, execute Algorithm 3. Then, with probability at least $1 - \delta$,*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\big(\sqrt{dK}\, T^{3/4}\big),$$

where the leading constants depend only on L, S, $\gamma_\star$, $\bar{\gamma}$ and the regularizer of Algorithm 3.

The exact constants behind the $\widetilde{\mathcal{O}}$ in Theorem 4 are given, with the proof, in Appendix E. The regret analysis there is carried out under the concentration event stated in Lemma 5. The discussion in that section also covers the cost of score-scale misspecification, which is confined to constants through an exact-KL identity (Remark 4), the behaviour of low-rank directions, which retain a wide confidence interval in the V-metric rather than spurious certainty (Remark 5), and the question of whether the $T^{3/4}$ rate is minimax, which remains open with a gap of $T^{1/4}$ to Theorem 2 (Remark 6).

Together, Theorems 3 and 4 describe how the policy behaves on low-rank or redundant context stretches. Under persistent excitation, such stretches inflate only the burn-in and the lower-order constants, while the regret rate is preserved. Without any excitation assumption, the ellipsoidal variant remains valid at the degraded $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ rate.

4.3 Beyond exact Cauchy noise

Theorems 1, 3 and 4 exploit the exact Cauchy likelihood through the drift identity of Lemma 1. This subsection asks how much of that structure the results actually need. We first replace the Cauchy score by the sign (median) score and show that the same $\widetilde{\mathcal{O}}(\sqrt{KT})$ rate survives under symmetry and a median-growth condition alone, with no moments of any order, given a known admissible lower bound p_0 on the median-growth slope. We then return to the exact-score update and quantify its sensitivity when the noise deviates from exact Cauchy: the regret degrades linearly in the deviation level, and a matching linear floor shows that this degradation is unavoidable.

A sign update for general symmetric heavy-tailed noise. Replacing the Cauchy score in Eq. (5) by the sign map $\psi(u) = 2\mathbf{1}\{u \geq 0\} - 1$ gives projected OGD on the absolute-deviation (median) loss: the stochastic gradient is bounded pathwise by L, so no moment of the noise needs to exist, and the only requirements are symmetry of the noise and a median-growth condition with constant p_0 (Assumption 4, Appendix F). The contraction modulus is $\kappa_s = 2p_0\mu_{\mathrm{dir}}$, with the gradient bound and the burn-in the exact analogues of Section 4 (Eq. (46)).

Theorem 5 (Sign-update location regret under general symmetric heavy-tailed noise)**.** *Let $\delta \in (0,1)$. Run Algorithm 1 with the update* (5) *replaced by the sign update* (45)*, step sizes $\eta_\tau = \frac{1}{\kappa_s(\tau+\tau_0^{(s)}+1)}$, and confidence schedule $\beta_\tau^{(s)}$ from Proposition 9 instantiated at failure level δ/K, under Assumptions 1 and 4 with a known admissible value of p_0. Then, with probability at least $1-\delta$,*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K\log T + \sqrt{KT\log(KT/\delta)}\Big)\,,$$

with leading constants depending on L, S and the median-growth modulus κ_s.

The exact constants behind the $\widetilde{\mathcal{O}}$ in Theorem 5 are given, with the proof, in Appendix F.

At the matched choice $\bar{\gamma} = \gamma_\star$ the exact score improves the leading regret constant over the sign update by up to a factor π, and it is never worse by more than $4/\pi$ (Remark 8, Appendix F.1). Underestimating p_0 is safe, and a dyadic grid over a known range of p_0 mirrors the scale adaptation of Theorem 6, with a branch-validity screening step left open (Remark 9, Appendix F.2).

Stability of the Direct-Cauchy update to deviations from Cauchy. We now quantify the sensitivity of the original Direct-Cauchy update (5) when the noise deviates from exact Cauchy. Consider the additive perturbation model

$$\xi_{k,t} = \xi_{C,k,t} + \zeta_{k,t}, \tag{11}$$

where, for each arm k, conditionally on $\mathcal{F}_{t-1}$ and the round-t contexts $\{y_{j,t}\}_{j=1}^K$, $\xi_{C,k,t} \sim \mathrm{Cauchy}(0,\gamma_\star)$, $\zeta_{k,t}$ is conditionally independent of $\xi_{C,k,t}$, and $|\zeta_{k,t}| \leq \rho$ almost surely for a deterministic $\rho \geq 0$. Only the played arm's reward is observed; the per-arm (potential-outcomes) perturbation $\zeta_{k,t}$ may otherwise be arbitrary and adaptive, with no symmetry or moment condition needed.

Under this model the exact-score update keeps its confidence sequence up to an error floor linear in ρ. The location regret of Theorem 1 acquires a single additive term, linear in the deviation level and the horizon (Proposition 7). This term is unavoidable: a matching floor linear in ρT holds for every policy on this class (Proposition 10). The exact statements and proofs, and the open arm-shared case (Remark 10), are deferred to Appendix F.

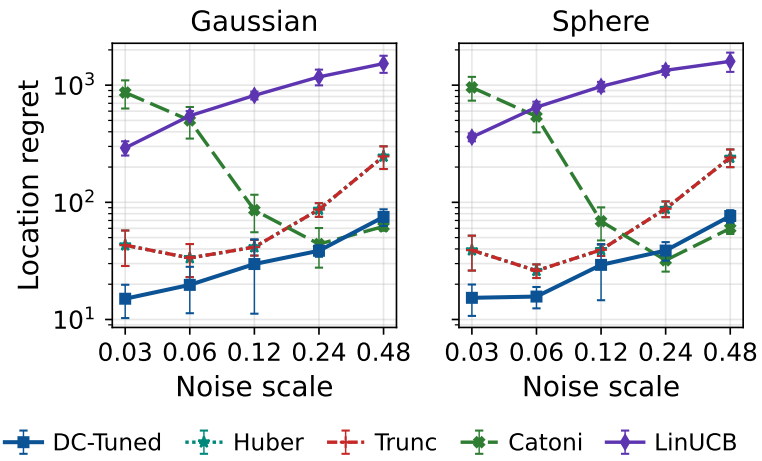


Figure 1: Exact Cauchy scale sweep, most competitive baselines, logarithmic regret axis. Error bars are environment-level standard deviations across the six environments. DC-Tuned is Direct-Cauchy-Tuned; Huber, Trunc and Catoni stand for the corresponding LinUCB variants.

5 Experiments

We evaluate the direct policy on synthetic exact-Cauchy environments with $K = 4$ arms, $d = 4$ features, horizon $T = 3000$, and six latent environments over two isotropic context families, Gaussian and sphere. The comparison suite covers robust baselines — Huber, truncated and Catoni variants of LinUCB, median-of-means and one-pass SGD variants, and classical LinUCB — of which the two clipping baselines are handed the true noise scale. The direct-policy row we report is Direct-Cauchy-Tuned, a practical configuration of Algorithm 1: it keeps the arm-local update (5) and the UCB index of the current iterate, but replaces the theorem-exact schedule with tuned constants (a constant step size, a proxy confidence bonus, an online-Newton preconditioner, and a box clip). The construction of this and the remaining configurations is described in Appendix G.1. The full protocol, the exact deviations, and the remaining studies — a nearly-Cauchy ablation that varies the noise law instead of its scale, and a walk-forward experiment on six decades of daily industry-portfolio returns — are presented in Appendix G.

At the reference noise scale the tuned configuration attains the lowest location regret of the suite on both context families, leading in five of the six environments; its closest competitor is Huber-LinUCB, and classical mean-oriented LinUCB degrades by roughly an order of magnitude. Figure 1 then varies the noise scale: the tuned configuration leads on both families from scale 0.03 through 0.12, at 0.24 it remains best on Gaussian contexts while Catoni-LinUCB overtakes it on spherical parameters, and at 0.48 Catoni-LinUCB leads on both. The advantage is therefore moderate-scale: the direct update's contraction modulus degrades quadratically in the scale-to-signal ratio, while ridge-style baselines reuse cross-direction information through full per-arm covariance matrices, the mechanism the ellipsoidal variant of Section 4.2 exploits. A companion sweep misspecifies the learner's score scale by constant factors: underestimating it degrades performance noticeably, while overestimating stays competitive, so the finite-sample sensitivity is asymmetric. Beyond the exact-Cauchy setting, the same suite on nearly-Cauchy noise laws keeps the advantage up to a measured tail index of about 1.5 and loses it beyond, and on six decades of daily industry-portfolio returns the tuned configuration attains the lowest regret of the suite on both evaluation metrics. The full protocol, all tables and figures, and the remaining configurations — the analysed schedule Direct-Cauchy-Scheduled and the grid and doubling adaptation rows — are presented in Appendix G.

6 Limitations

Several scope boundaries of this study are worth stating; an extended discussion is given in Appendix H. The theoretical scope is symmetric noise with a median-growth condition, sharpest for exact symmetric Cauchy noise. The main guarantees rely on excitation conditions, which persistently degenerate context processes can violate; dropping excitation entirely is possible only through the ellipsoidal variant, at $\mathcal{O}(d^2)$ per-arm memory and a degraded $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ rate. The sign update requires a known lower bound on the median-growth constant p_0, and the empirical benchmarks do not yet cover the adversarial-context and perturbation regimes of Sections 4.2 and 4.3.

7 AI use statement

Generative AI tools were used in the preparation of this work: to aid and polish writing, through iterative language editing and clarity improvements of the manuscript; for retrieval and discovery, assisting the search and review of related work; and for proving mathematical claims, as a brainstorming and cross-checking aid while developing the proofs. All suggested content was reviewed by the authors, who take full responsibility for the final manuscript.

8 Reproducibility statement

The supplementary material provides code that reproduces all experiments of this paper. See Appendix G for the full protocol, the entry points, and the reproduction commands.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems 24*, pages 2312–2320, 2011. URL `https://papers.nips.cc/paper/4417-improved-algorithms-for-linear-stochastic-bandits`.

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 22 of *Proceedings of Machine Learning Research*, pages 1–9, 2012.

Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67(3):1329–1349, 2021.

Ilija Bogunovic, Arpan Losalka, Andreas Krause, and Jonathan Scarlett. Stochastic linear bandits robust to adversarial attacks. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 991–999. PMLR, 2021. URL `https://proceedings.mlr.press/v130/bogunovic21a.html`.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013. doi: 10.1109/TIT.2013.2273101.

Olivier Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4):1148–1185, 2012. doi: 10.1214/11-AIHP454. URL `https://www.numdam.org/articles/10.1214/11-AIHP454/`.

Wei Chu, Lihong Li, Lev Reyzin, and Robert E. Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 208–214. JMLR, 2011. URL `https://proceedings.mlr.press/v15/chu11a.html`.

Jules Depersin and Guillaume Lecué. On the robustness to adversarial corruption and to heavy-tailed data of the Stahel–Donoho median of means. *Information and Inference: A Journal of the IMA*, 12 (2):814–850, 2023. doi: 10.1093/imaiai/iaac026. arXiv:2101.09117.

Qin Ding, Cho-Jui Hsieh, and James Sharpnack. Robust stochastic linear contextual bandits under adversarial attacks. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7111–7123. PMLR, 2022. URL `https://proceedings.mlr.press/v151/ding22c.html`.

Yuriy Dorn, Aleksandr Katrutsa, Ilgam Latypov, and Andrey Pudovikov. Fast ucb-type algorithms for stochastic bandits with heavy and super heavy symmetric noise. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '25, pages 649–657, Richland, SC, 2025. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9798400714269.

Yuriy Dorn, Aleksandr Katrutsa, Ilgam Latypov, and Anastasiia Soboleva. Functional multi-armed bandit and the best function identification problems. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '26, pages 2913–2921, Richland, SC, 2026. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9798400723179. doi: 10.65109/SFMN9947. URL https://doi.org/10.65109/SFMN9947.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems 23*, pages 586–594, 2010. URL https://papers.nips.cc/paper/4166-parametric-bandits-the-generalized-linear-case.

Gianmarco Genalti and Alberto Maria Metelli. Parameter-free heavy-tailed bandits. *arXiv preprint*, 2026. arXiv:2607.29460.

Alexander Goldenshluger and Assaf Zeevi. A linear response bandit problem. *Stochastic Systems*, 3 (1):230–261, 2013.

Yunwen Guo, Yunlu Shu, Gongyi Zhuo, and Tianyu Wang. Batched bandits with heavy-tailed rewards. *arXiv preprint*, 2025. doi: 10.1093/imaiai/iaag031. arXiv:2510.03798; journal version in Information and Inference.

Jiafan He, Dongruo Zhou, Tong Zhang, and Quanquan Gu. Nearly optimal algorithms for linear contextual bandits with adversarial corruptions. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Kwang-Sung Jun and Jungtaek Kim. Noise-adaptive confidence sets for linear bandits and application to bayesian optimization. *OpenReview*, 2024. URL https://openreview.net/forum?id=W8hBNk1FhQ.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. In *Advances in Neural Information Processing Systems 30*, pages 99–109, 2017. URL https://papers.nips.cc/paper/6615-scalable-generalized-linear-bandits-online-computation-and-hashing.

Minhyun Kang and Gi-Soo Kim. Heavy-tailed linear bandit with Huber regression. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 1027–1036. PMLR, 2023. URL https://proceedings.mlr.press/v216/kang23a.html.

Sampath Kannan, Jamie Morgenstern, Aaron Roth, and Zhiwei Steven Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Aren Karapetyan, Efe C. Balta, Anastasios Tsiamis, Andrea Iannelli, and John Lygeros. On the regret of recursive methods for discrete-time adaptive control with matched uncertainty. *arXiv preprint*, 2024. arXiv:2404.02023.

Sanath Kumar Krishnamurthy, Vitor Hadad, and Susan Athey. Adapting to misspecification in contextual bandits with offline regression oracles. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5805–5814. PMLR, 2021. URL https://proceedings.mlr.press/v139/krishnamurthy21a.html.

Ilgam Latypov, Alexandra Suvorikova, Alexey Kroshnin, Alexander Gasnikov, and Yuriy Dorn. Ucb-type algorithm for budget-constrained expert learning. *arXiv preprint arXiv:2510.22654*, 2025.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, pages 661–670. ACM, 2010. doi: 10.1145/1772690.1772758. URL https://dl.acm.org/doi/10.1145/1772690.1772758.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2071–2080. PMLR, 2017. URL https://proceedings.mlr.press/v70/li17c.html.

Xiang Li and Qiang Sun. Variance-aware decision making with linear function approximation under heavy-tailed rewards. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=8bnsoL2IyJ.

Andrés Muñoz Medina and Scott Yang. No-regret algorithms for heavy-tailed linear bandits. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1642–1650. PMLR, 2016. URL https://proceedings.mlr.press/v48/medina16.html.

Deepan Muthirayan, Jianjun Yuan, and Pramod P. Khargonekar. Balancing exploration for online receding horizon learning control with provable regret guarantees. *arXiv preprint*, 2020. arXiv:2010.07269.

Gergely Neu and Julia Olkhovskaya. Efficient and robust algorithms for adversarial linear contextual bandits. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3049–3068. PMLR, 2020. URL https://proceedings.mlr.press/v125/neu20b.html.

Han Shao, Xiaotian Yu, Irwin King, and Michael R. Lyu. Almost optimal algorithms for linear stochastic bandits with heavy-tailed payoffs. In *Advances in Neural Information Processing Systems 31*, pages 8430–8439, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/173f0f6bb0ee97cf5098f73ee94029d4-Abstract.html.

Vidyashankar Sivakumar, Zhiwei Steven Wu, and Arindam Banerjee. Structured linear contextual bandits: A sharp and geometric smoothed analysis. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9026–9035, 2020.

Artin Tajdini, Jonathan Scarlett, and Kevin Jamieson. Improved regret bounds for linear bandits with heavy-tailed rewards. *arXiv preprint*, 2025. arXiv:2506.04775.

Naoto Tani and Futoshi Futami. Robust and computationally efficient linear contextual bandits under adversarial corruption and heavy-tailed noise. *arXiv preprint*, 2026. arXiv:2603.15596.

Jing Wang, Yu-Jie Zhang, Peng Zhao, and Zhi-Hua Zhou. Heavy-tailed linear bandits: Huber regression with one-pass update. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 65595–65617. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25et.html.

Ruitu Xu, Yifei Min, and Tianhao Wang. Noise-adaptive thompson sampling for linear contextual bandits. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

Bo Xue, Guanghui Wang, Yimu Wang, and Lijun Zhang. Nearly optimal regret for stochastic linear bandits with heavy-tailed payoffs. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI)*, 2020. doi: 10.24963/ijcai.2020/406.

Bo Xue, Yimu Wang, Yuanyu Wan, Jinfeng Yi, and Lijun Zhang. Efficient algorithms for generalized linear bandits with heavy-tailed rewards. In *Advances in Neural Information Processing Systems 36*, pages 60584–60612, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/e0982cbc81401df3430ee1ff780dc7a2-Abstract-Conference.html.

Chenlu Ye, Wei Xiong, Quanquan Gu, and Tong Zhang. Corruption-robust algorithms with uncertainty weighting for nonlinear contextual bandits and markov decision processes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 39834–39863. PMLR, 2023. URL https://proceedings.mlr.press/v202/ye23d.html.

Chenlu Ye, Yujia Jin, Alekh Agarwal, and Tong Zhang. Catoni contextual bandits are robust to heavy-tailed rewards. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 71998–72040. PMLR, 2025. URL https://proceedings.mlr.press/v267/.

Dingzhi Yu, Hongyi Tao, Yuanyu Wan, Luo Luo, and Lijun Zhang. Sign-based optimizers are effective under heavy-tailed noise. *arXiv preprint*, 2026. arXiv:2602.07425.

Qingyuan Yu, Euijin Baek, Xiang Li, and Qiang Sun. Corruption-robust variance-aware algorithms for generalized linear bandits under heavy-tailed rewards. *OpenReview*, 2025. URL https://openreview.net/forum?id=319ORvbXYS.

Canzhe Zhao, Shinji Ito, and Shuai Li. Heavy-tailed linear bandits: Adversarial robustness, best-of-both-worlds, and beyond. *arXiv preprint*, 2025. arXiv:2508.13679.

Tianshuo Zheng, Ting Wu, Zhi-Hua Zhou, and Keqin Liu. Nonlinear bandit. *arXiv preprint*, 2026. arXiv:2607.07304.

Han Zhong, Jiayi Huang, Lin Yang, and Liwei Wang. Breaking the moments condition barrier: No-regret algorithm for bandits with super heavy-tailed payoffs. In *Advances in Neural Information Processing Systems 34*, pages 15710–15720, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/843a4d7fb5b1641b0bb8e3c2b2e75231-Abstract.html.

A Upper-Bound Proofs

This appendix states and proves the upper-bound results of Section 4: the deterministic reduction, the one-arm ingredients, the main regret bound, the scale-misspecification corollary, and the instance-dependent refinement. The dyadic-grid adaptation is Appendix B, and the lower bounds are Appendix C.

Standing notation. Throughout the appendix, $B = 2LS$ and $\Gamma := \gamma_\star + \bar{\gamma}$; the scale-dependent constants are

$$C_{\mathrm{Cau}} = \frac{2}{B^2 + (\gamma_\star + \bar{\gamma})^2}, \qquad \kappa_{\mathrm{dir}} = \mu_{\mathrm{dir}} C_{\mathrm{Cau}}, \qquad \ell = \frac{2L^2}{(\gamma_\star + \bar{\gamma})^2}.$$

as in Section 4, the burn-in is

$$\tau_0 = \left\lceil \max\left\{1, \frac{\ell}{\kappa_{\mathrm{dir}}}, \frac{\ell^2}{\kappa_{\mathrm{dir}}^2}\right\} \right\rceil,$$

and the leading constants of the confidence radius are $A_{\mathrm{Cau}} = 2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\tau_0}$ and $B_{\mathrm{Cau}} = \frac{2L}{\kappa_{\mathrm{dir}}\bar{\gamma}} + \frac{4\sqrt{2}L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\log(2KT/\delta)}$.

A.1 The deterministic reduction, and its proof

Proposition 3 (Deterministic Direct-UCB reduction)**.** *Assume a deterministic, non-increasing schedule $\{\beta_\tau\}_{\tau\geq 0}$ provides a valid upper confidence bound, such that $|y_{k,t}^\top(z_{k,T_k(t-1)} - \theta_k)| \leq \beta_{T_k(t-1)}\|y_{k,t}\|_2$ holds simultaneously for all arms and rounds. Then any policy executing the optimistic selection rule of Algorithm 1 satisfies*

$$R_{\mathrm{loc}}(T) \leq 2L \sum_{k=1}^{K} \sum_{\tau=0}^{T_k-1} \beta_\tau.$$

Proof. Fix round t. Since k_t maximizes the direct index,

$$\begin{aligned} y_{k_t^\star,t}^\top \theta_{k_t^\star} &\leq y_{k_t^\star,t}^\top z_{k_t^\star,T_{k_t^\star}(t-1)} + \beta_{T_{k_t^\star}(t-1)}\|y_{k_t^\star,t}\|_2 \\ &\leq y_{k_t,t}^\top z_{k_t,T_{k_t}(t-1)} + \beta_{T_{k_t}(t-1)}\|y_{k_t,t}\|_2 \\ &\leq y_{k_t,t}^\top \theta_{k_t} + 2\beta_{T_{k_t}(t-1)}\|y_{k_t,t}\|_2. \end{aligned}$$

Thus the instantaneous regret is at most $2L\beta_{T_{k_t}(t-1)}$. Summing over rounds and regrouping by local pulls gives the claim. □

A.2 Proof of Lemma 1

Lemma 1 (Restatement). *Let $\xi \sim \mathrm{Cauchy}(0, \gamma_\star)$ with $\gamma_\star > 0$, let $\bar{\gamma} > 0$, and define*

$$\varphi_{\bar{\gamma}}(u) = \frac{2u}{\bar{\gamma}^2 + u^2}.$$

Then for every $\Delta \in \mathbb{R}$,

$$\mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] = \frac{2\Delta}{\Delta^2 + (\gamma_\star + \bar{\gamma})^2}.$$

Consequently, $\sup_{u \in \mathbb{R}} |\varphi_{\bar{\gamma}}(u)| = 1/\bar{\gamma}$,

$$\sup_{\Delta \in \mathbb{R}} \left| \frac{d}{d\Delta} \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] \right| \leq \frac{2}{(\gamma_\star + \bar{\gamma})^2},$$

and on any bounded interval $|\Delta| \leq B$,

$$\Delta\, \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] \geq \frac{2\Delta^2}{B^2 + (\gamma_\star + \bar{\gamma})^2}.$$

Proof. Let

$$p_c(x) = \frac{c}{\pi(c^2 + x^2)}, \qquad q_c(x) = \frac{x}{\pi(c^2 + x^2)}.$$

Then ξ has density $p_{\gamma_\star}$ and $\varphi_{\bar{\gamma}}(u) = 2\pi q_{\bar{\gamma}}(u)$. Hence

$$\mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] = 2\pi (q_{\bar{\gamma}} * p_{\gamma_\star})(\Delta).$$

The Fourier transforms satisfy

$$\widehat{p_c}(\omega) = e^{-c|\omega|}, \qquad \widehat{q_c}(\omega) = -i \operatorname{sign}(\omega) e^{-c|\omega|},$$

so

$$\widehat{q_{\bar{\gamma}} * p_{\gamma_\star}}(\omega) = -i \operatorname{sign}(\omega) e^{-(\gamma_\star + \bar{\gamma})|\omega|} = \widehat{q_{\gamma_\star + \bar{\gamma}}}(\omega).$$

Here the transform of q_c is understood in the principal-value sense, since $q_c \notin L^1(\mathbb{R})$; the product $-i \operatorname{sign}(\omega) e^{-(\gamma_\star + \bar{\gamma})|\omega|}$ is integrable, so the distributional route is legitimate. Therefore

$$q_{\bar{\gamma}} * p_{\gamma_\star} = q_{\gamma_\star + \bar{\gamma}}$$

almost everywhere, and since both sides are continuous in Δ, the equality holds pointwise, which proves the identity. The drift bound follows on $|\Delta| \leq B$ because $\Delta\, \mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] = 2\Delta^2/(\Delta^2 + (\gamma_\star + \bar{\gamma})^2)$ and the denominator is at most $B^2 + (\gamma_\star + \bar{\gamma})^2$. The score bound is the elementary fact that $2|u|/(\bar{\gamma}^2 + u^2)$ is maximized at $u = \pm\bar{\gamma}$ with value $1/\bar{\gamma}$. Finally, differentiating

$$m(\Delta) = \frac{2\Delta}{\Delta^2 + (\gamma_\star + \bar{\gamma})^2}$$

gives

$$m'(\Delta) = \frac{2((\gamma_\star + \bar{\gamma})^2 - \Delta^2)}{(\Delta^2 + (\gamma_\star + \bar{\gamma})^2)^2},$$

whose absolute value is maximized at $\Delta = 0$ with value $2/(\gamma_\star + \bar{\gamma})^2$, proving the stated Lipschitz bound. □

A.3 Discussion of Assumption 1: sufficient condition, examples, and failure modes

This section collects the discussion of Assumption 1 announced in the body: a verifiable sufficient condition (Proposition 4, with proof), examples of designs that satisfy the assumption, and the way it can fail.

What the assumption asks, and what it does not. Assumption 1 is a per-direction conditional second-moment floor given the arm-local history $\mathcal{F}_{k,\tau}$. It is not an i.i.d. or stationarity requirement: the bound is conditional, so the context process may be adaptive and history-dependent. It is also weaker than a full covariance floor: it constrains only the second moment along each measurable direction, not the entire conditional covariance matrix. A simple verifiable sufficient condition is a uniform lower bound on the conditional covariance, which we now state and prove; it is a sufficient condition, not a characterization — the assumption is strictly broader and also covers designs whose conditional covariance is degenerate along some realizations but not along the directions that matter, such as the smoothed-adversarial and forced-exploration designs discussed below.

Proposition 4 (Bounded exploratory designs)**.** *Suppose the local arm contexts satisfy* $\|y_{k,\tau+1}\|_2 \le L$ *and*

$$\mathbb{E}[y_{k,\tau+1} y_{k,\tau+1}^\top \mid \mathcal{F}_{k,\tau}] \succeq \mu I_d \tag{12}$$

for every arm k *and local time* τ*, for some* $\mu > 0$*. Then Assumption 1 holds with* $\mu_{\mathrm{dir}} = \mu$*.*

Proof of Proposition 4. Fix an arm k, local time τ, and any $\mathcal{F}_{k,\tau}$-measurable vector $v \in \mathbb{R}^d$. Expanding the quadratic term,

$$\mathbb{E}\big[(y_{k,\tau+1}^\top v)^2 \mid \mathcal{F}_{k,\tau}\big] = v^\top \mathbb{E}[y_{k,\tau+1} y_{k,\tau+1}^\top \mid \mathcal{F}_{k,\tau}] v \ge \mu \|v\|_2^2,$$

which satisfies Assumption 1 given $\mu_{\mathrm{dir}} = \mu$.

For unconditioned contexts drawn from the fixed-norm sphere model $y_{k,t} = \sigma_x \sqrt{d}\, U_{k,t}$, rotational symmetry gives $\mathbb{E}[y_{k,t} y_{k,t}^\top] = \sigma_x^2 I_d$, and because $\|U_{k,t}\|_2 = 1$, the model is bounded with $\|y_{k,t}\|_2 = \sigma_x \sqrt{d}$. Under adaptive play, conditioning on the chosen arm can break exact isotropy, so Assumption 1 should be verified through a uniform conditional covariance lower bound as required by the proposition. □

Examples. The fixed-norm sphere model used in the experiments is one bounded exploratory design: by the computation in the proof above it satisfies the covariance floor of Proposition 4 with $\mu = \sigma_x^2$ under unconditioned sampling. One caveat is in order: under the adaptive arm selection of an actual run, the conditional law of the next context given the arm-local history is no longer exactly isotropic, so certified excitation constants are not available for the measured runs; Appendix G accordingly records those runs as numerical instantiations of the setting rather than instances with verified constants. Beyond isotropic designs, two non-i.i.d. context models retain a per-direction conditional floor. In the *smoothed-adversarial* model [Kannan et al., 2018, Sivakumar et al., 2020] an adversary chooses the contexts, but each is then perturbed by independent noise with a non-degenerate covariance; the perturbation alone supplies the conditional second-moment floor along every direction, regardless of the adversary's choices, and the unperturbed part may be fully history-dependent. In *forced-exploration* designs a positive fraction of rounds is drawn from a fixed exploratory law, as in the forced-sampling schemes of Goldenshluger and Zeevi [2013] and Bastani et al. [2021]: conditionally on the history, each round is exploratory with nonvanishing probability, so the per-direction floor of the exploratory law survives at a proportional level.

How the assumption can fail. The failure mode is specific and identifiable: persistently low-rank context sequences, which never excite some direction. Along such a direction the arm's rewards carry no information about the corresponding parameter component, and no method can identify it; the regret benchmarks of this paper are therefore unaffected only because those directions never influence any reward either (see the discussion after Theorem 4). Section 4.2 relaxes the per-step requirement to cumulative excitation over windows.

A.4 Proof of the one-arm confidence ingredient

Proposition 1 (full form). *Let* $\delta \in (0,1)$*. Fix an arm* k *and run the one-pass projected update defined in Eq.* (5) *with the localized step size* $\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau+\tau_0+1)}$*. Under Assumption 1, with probability at least* $1-\delta$*, simultaneously for all local times* $\tau \in \{0,\dots,T\}$*, the estimation error is bounded by*

$$\beta_\tau := \frac{2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\tau_0}}{\tau+\tau_0} + \frac{2L}{\kappa_{\mathrm{dir}}\bar{\gamma}\sqrt{\tau+\tau_0}} + \frac{4\sqrt{2}L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\frac{\log(2T/\delta)}{\tau+\tau_0}}. \tag{13}$$

Thus $\|z_{k,\tau} - \theta_k\|_2 \le \beta_\tau$ *and* $|y^\top z_{k,\tau} - y^\top \theta_k| \le \beta_\tau \|y\|_2$. With only numerical constants hidden and the τ-dependence kept explicit, the radius at the local horizon $n = T$ satisfies $\beta_\tau \lesssim \frac{S + L/(\kappa_{\mathrm{dir}}\bar{\gamma})}{\tau+\tau_0} + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\log(2T/\delta)/(\tau+\tau_0)}$ uniformly over $\tau \le T$, which is the form stated in the body.

The informal argument behind this radius, sketched in the body, is recorded here for completeness.

Proof idea. Write the update error as $e_{k,\tau} = z_{k,\tau} - \theta_k$ and decompose the stochastic gradient into its mean drift and martingale noise,

$$g_{k,\tau+1} = h_{k,\tau}(z_{k,\tau}) + M_{k,\tau+1}, \qquad \mathbb{E}[M_{k,\tau+1} \mid \mathcal{F}_{k,\tau}] = 0,$$

with $\|M_{k,\tau+1}\|_2 \le 2L/\bar{\gamma}$. By nonexpansiveness of Euclidean projection,

$$\|e_{k,\tau+1}\|_2^2 \le \|e_{k,\tau} - \eta_\tau g_{k,\tau+1}\|_2^2.$$

Expanding the square and using the directional drift condition from Assumption 1 gives a contraction term of order

$$-2\eta_\tau \langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \le -2\kappa_{\mathrm{dir}}\eta_\tau \|e_{k,\tau}\|_2^2,$$

plus a bounded-noise term of order $\eta_\tau^2 L^2/\bar{\gamma}^2$. With the choice $\eta_\tau = 1/(\kappa_{\mathrm{dir}}(\tau+\tau_0+1))$, this yields a Robbins–Monro type recursion with $1/(\tau+\tau_0)$ decay in expectation, matching Proposition 5 of Appendix A.4.

To obtain a *uniform-in-time* confidence sequence, one then tracks the variance-smoothed potential

$$q_\tau = (\tau+\tau_0)\sqrt{\|e_{k,\tau}\|_2^2 + c/(\tau+\tau_0)},$$

for a suitable constant $c = \Theta\big(L^2/(\kappa_{\mathrm{dir}}^2\bar{\gamma}^2)\big)$, made explicit in the proof of Proposition 1 (Appendix A.4). The recursion for q_τ separates into a deterministic drift part and a martingale increment with bounded magnitude. Applying Azuma–Hoeffding (or an equivalent maximal martingale inequality) and a union bound over all $\tau \le T$ controls the noise term simultaneously for all times. Finally, unwrapping the smoothing and using the initial bound $\|z_{k,0} - \theta_k\|_2 \le 2S$ yields the stated radius

$$\beta_\tau = \frac{2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\tau_0}}{\tau+\tau_0} + \frac{2L}{\kappa_{\mathrm{dir}}\bar{\gamma}\sqrt{\tau+\tau_0}} + \frac{4\sqrt{2}L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\frac{\log(2T/\delta)}{\tau+\tau_0}}.$$

Since $|y^\top(e_{k,\tau})| \le \|y\|_2\|e_{k,\tau}\|_2$, the same bound immediately gives the prediction guarantee. □

Proposition 5 (One-arm contraction)**.** *Fix an arm* k *and let* $e_{k,\tau} = z_{k,\tau} - \theta_k$*. Under Assumption 1, the direct projected update with step size*

$$\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau+\tau_0+1)} \tag{14}$$

satisfies

$$\mathbb{E}[\|e_{k,\tau+1}\|_2^2 \mid \mathcal{F}_{k,\tau}] \le \left(1 - \frac{2}{\tau+\tau_0+1}\right)\|e_{k,\tau}\|_2^2 + \frac{L^2}{\kappa_{\mathrm{dir}}^2\bar{\gamma}^2(\tau+\tau_0+1)^2}. \tag{15}$$

Consequently,

$$\mathbb{E}\|e_{k,\tau}\|_2^2 \le \frac{D_{\mathrm{Cau}}}{\tau+\tau_0}, \qquad D_{\mathrm{Cau}} = \max\left\{4S^2\tau_0, \frac{L^2}{\kappa_{\mathrm{dir}}^2\bar{\gamma}^2}\right\}. \tag{16}$$

Proof. Define

$$g_{k,\tau+1} = \varphi_{\bar{\gamma}}(y_{k,\tau+1}^\top z_{k,\tau} - r_{k,\tau+1})y_{k,\tau+1}$$

and

$$h_{k,\tau}(x) = \mathbb{E}\big[\varphi_{\bar{\gamma}}(y_{k,\tau+1}^\top x - r_{k,\tau+1})y_{k,\tau+1} \mid \mathcal{F}_{k,\tau}\big].$$

By Lemma 1, applied conditionally on $(\mathcal{F}_{k,\tau}, y_{k,\tau+1})$ — legitimate because the noise model of Section 2 makes $\xi_{k,\tau+1}$ independent of $(\mathcal{F}_{k,\tau}, y_{k,\tau+1})$ —

$$h_{k,\tau}(x) = \mathbb{E}\left[\frac{2y_{k,\tau+1}^\top(x-\theta_k)}{(y_{k,\tau+1}^\top(x-\theta_k))^2 + (\gamma_\star+\bar{\gamma})^2}y_{k,\tau+1} \mid \mathcal{F}_{k,\tau}\right].$$

Since $\|x\|_2, \|\theta_k\|_2 \le S$ and $\|y_{k,\tau+1}\|_2 \le L$, we have $|y_{k,\tau+1}^\top (x - \theta_k)| \le B$. Thus, the directional drift condition gives

$$\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \ge \mu_{\mathrm{dir}} c_{\gamma_\star,\bar\gamma}(B)\|e_{k,\tau}\|_2^2 = \kappa_{\mathrm{dir}}\|e_{k,\tau}\|_2^2.$$

Projection nonexpansiveness yields

$$\|e_{k,\tau+1}\|_2^2 \le \|e_{k,\tau} - \eta_\tau g_{k,\tau+1}\|_2^2.$$

Conditioning on $\mathcal{F}_{k,\tau}$,

$$\begin{aligned}\mathbb{E}[\|e_{k,\tau+1}\|_2^2 \mid \mathcal{F}_{k,\tau}] &\le \|e_{k,\tau}\|_2^2 - 2\eta_\tau\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle + \eta_\tau^2\mathbb{E}[\|g_{k,\tau+1}\|_2^2 \mid \mathcal{F}_{k,\tau}] \\ &\le (1 - 2\kappa_{\mathrm{dir}}\eta_\tau)\|e_{k,\tau}\|_2^2 + \eta_\tau^2\frac{L^2}{\bar\gamma^2},\end{aligned}$$

since $|\varphi_{\bar\gamma}(u)| \le 1/\bar\gamma$. Substituting η_τ gives the local recursion and the resulting mean-square $1/(\tau+\tau_0)$ contraction.

Taking the unconditional expectation of both sides yields

$$\mathbb{E}[\|e_{k,\tau+1}\|_2^2] \le (1 - 2\kappa_{\mathrm{dir}}\eta_\tau)\mathbb{E}[\|e_{k,\tau}\|_2^2] + \eta_\tau^2\frac{L^2}{\bar\gamma^2}.$$

Substituting the step size $\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau+\tau_0+1)}$ gives the explicit recursive bound:

$$\mathbb{E}[\|e_{k,\tau+1}\|_2^2] \le \left(1 - \frac{2}{\tau+\tau_0+1}\right)\mathbb{E}[\|e_{k,\tau}\|_2^2] + \frac{L^2}{\kappa_{\mathrm{dir}}^2\bar\gamma^2(\tau+\tau_0+1)^2}.$$

We now prove $\mathbb{E}[\|e_{k,\tau}\|_2^2] \le \frac{D_{\mathrm{Cau}}}{\tau+\tau_0}$ by induction. For the base case $\tau = 0$, since $z_{k,0}, \theta_k \in \mathcal{X}$ and the radius of $\mathcal{X}$ is S, the maximum squared distance is $\|z_{k,0} - \theta_k\|_2^2 \le (2S)^2 = 4S^2$. Because $D_{\mathrm{Cau}} \ge 4S^2\tau_0$ by definition, it follows that $\mathbb{E}[\|e_{k,0}\|_2^2] \le \frac{D_{\mathrm{Cau}}}{\tau_0}$, which proves the base case.

For the inductive step, assume the bound holds for local time $\tau \ge 0$. Let $m = \tau + \tau_0$. Using the inductive hypothesis and the fact that the constant $D_{\mathrm{Cau}} \ge \frac{L^2}{\kappa_{\mathrm{dir}}^2\bar\gamma^2}$, we substitute into the recursion:

$$\begin{aligned}\mathbb{E}[\|e_{k,\tau+1}\|_2^2] &\le \left(1 - \frac{2}{m+1}\right)\frac{D_{\mathrm{Cau}}}{m} + \frac{D_{\mathrm{Cau}}}{(m+1)^2} \\ &= D_{\mathrm{Cau}}\left(\frac{m-1}{m(m+1)} + \frac{1}{(m+1)^2}\right) \\ &= D_{\mathrm{Cau}}\left(\frac{(m-1)(m+1)+m}{m(m+1)^2}\right) \\ &= D_{\mathrm{Cau}}\left(\frac{m^2+m-1}{m(m+1)^2}\right) \\ &= D_{\mathrm{Cau}}\left(\frac{m(m+1)-1}{m(m+1)^2}\right) \\ &= D_{\mathrm{Cau}}\left(\frac{1}{m+1} - \frac{1}{m(m+1)^2}\right) \\ &\le \frac{D_{\mathrm{Cau}}}{m+1} = \frac{D_{\mathrm{Cau}}}{\tau+\tau_0+1}.\end{aligned}$$

This completes the induction and establishes the stated mean-square contraction.

□

Proposition 1 (Restatement). *Fix an arm k. Under the model in Section 2, Assumption 1, and the step size $\eta_\tau = 1/(\kappa_{\mathrm{dir}}(\tau+\tau_0+1))$, with probability at least $1-\delta$, simultaneously for all local times $\tau \in \{0, \dots, n\}$,*

$$\|z_{k,\tau} - \theta_k\|_2 \le \frac{2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}}\bar\gamma}\sqrt{\tau_0}}{\tau+\tau_0} + \frac{2L}{\kappa_{\mathrm{dir}}\bar\gamma\sqrt{\tau+\tau_0}} + \frac{4\sqrt{2}L}{\kappa_{\mathrm{dir}}\bar\gamma}\sqrt{\frac{\log(2n/\delta)}{\tau+\tau_0}}.$$

Proof. Let

$$M_{k,\tau+1} = g_{k,\tau+1} - h_{k,\tau}(z_{k,\tau}).$$

Then $M_{k,\tau+1}$ is a zero-mean martingale difference with respect to the local filtration, satisfying

$$\|M_{k,\tau+1}\|_2 \le \|g_{k,\tau+1}\|_2 + \|h_{k,\tau}(z_{k,\tau})\|_2 \le \frac{2L}{\bar{\gamma}}.$$

The structural one-step inequality is derived from the pathwise projection recursion before evaluating the conditional expectation mapped in Proposition 5. Set

$$c_\star = \frac{L^2}{\kappa_{\mathrm{dir}}^2 \bar{\gamma}^2}, \qquad m_\tau = \tau + \tau_0, \qquad a_{k,\tau} = \|e_{k,\tau}\|_2^2, \qquad \xi_{k,\tau+1} = \langle e_{k,\tau}, M_{k,\tau+1}\rangle.$$

The projected update gives $z_{k,\tau+1} = \Pi_{\mathcal{X}}(z_{k,\tau} - \eta_\tau g_{k,\tau+1})$ given that $\theta_k \in \mathcal{X}$. Hence, by nonexpansiveness of Euclidean projections,

$$\begin{aligned} a_{k,\tau+1} &\le \|z_{k,\tau} - \eta_\tau g_{k,\tau+1} - \theta_k\|_2^2 \\ &= \|e_{k,\tau} - \eta_\tau g_{k,\tau+1}\|_2^2 \\ &= a_{k,\tau} - 2\eta_\tau \langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle - 2\eta_\tau \xi_{k,\tau+1} + \eta_\tau^2 \|g_{k,\tau+1}\|_2^2. \end{aligned}$$

The drift condition gives $\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \ge \kappa_{\mathrm{dir}} a_{k,\tau}$. Also, $\|g_{k,\tau+1}\|_2 \le L/\bar{\gamma}$ holds because $2|u|/(\bar{\gamma}^2 + u^2) \le 1/\bar{\gamma}$ and $\|y_{k,\tau+1}\|_2 \le L$. Thus,

$$a_{k,\tau+1} \le (1 - 2\kappa_{\mathrm{dir}}\eta_\tau) a_{k,\tau} - 2\eta_\tau \xi_{k,\tau+1} + \eta_\tau^2 \frac{L^2}{\bar{\gamma}^2}.$$

Substituting $\eta_\tau = 1/(\kappa_{\mathrm{dir}}(m_\tau + 1))$ and $c_\star = L^2/(\kappa_{\mathrm{dir}}^2 \bar{\gamma}^2)$ gives the pathwise bound:

$$a_{k,\tau+1} \le \left(1 - \frac{2}{m_\tau + 1}\right) a_{k,\tau} - \frac{2}{\kappa_{\mathrm{dir}}(m_\tau + 1)} \xi_{k,\tau+1} + \frac{c_\star}{(m_\tau + 1)^2}.$$

We now translate this squared-norm recursive form into a linear dynamic. Initialize the variance-smoothed state

$$s_{k,\tau} = \sqrt{\|e_{k,\tau}\|_2^2 + \frac{c_\star}{m_\tau}}.$$

Adding the smoothing term $c_\star/(m_\tau + 1)$ to $a_{k,\tau+1}$ and using the identity $1 - \frac{2}{m_\tau+1} = \frac{m_\tau - 1}{m_\tau + 1} = \frac{m_\tau^2 - 1}{(m_\tau+1)^2}$, we expand:

$$s_{k,\tau+1}^2 \le \frac{m_\tau^2 - 1}{(m_\tau + 1)^2} \left(s_{k,\tau}^2 - \frac{c_\star}{m_\tau}\right) - \frac{2\xi_{k,\tau+1}}{\kappa_{\mathrm{dir}}(m_\tau + 1)} + \frac{c_\star}{(m_\tau + 1)^2} + \frac{c_\star}{m_\tau + 1}.$$

Combining the $c_\star$ coefficients via the identity $-\frac{m_\tau^2 - 1}{m_\tau(m_\tau+1)^2} + \frac{1}{(m_\tau+1)^2} + \frac{1}{m_\tau+1} = \frac{2m_\tau + 1}{m_\tau(m_\tau+1)^2}$ yields $c_\star \frac{2m_\tau+1}{m_\tau(m_\tau+1)^2}$. Furthermore, because $s_{k,\tau}^2 \ge c_\star/m_\tau$, we can bound the leading fractional term by extracting the implicit negative quadratic component:

$$\frac{m_\tau^2 - 1}{(m_\tau + 1)^2} s_{k,\tau}^2 = \frac{m_\tau^2}{(m_\tau + 1)^2} s_{k,\tau}^2 - \frac{s_{k,\tau}^2}{(m_\tau + 1)^2} \le \frac{m_\tau^2}{(m_\tau + 1)^2} s_{k,\tau}^2 - \frac{c_\star}{m_\tau(m_\tau + 1)^2}.$$

Combining this $-\frac{c_\star}{m_\tau(m_\tau+1)^2}$ residual with the previously aggregated $c_\star$ multiplier gives $c_\star \frac{2m_\tau+1-1}{m_\tau(m_\tau+1)^2} = \frac{2c_\star}{(m_\tau+1)^2}$. Thus,

$$s_{k,\tau+1}^2 \le \frac{m_\tau^2}{(m_\tau + 1)^2} s_{k,\tau}^2 - \frac{2}{\kappa_{\mathrm{dir}}(m_\tau + 1)} \xi_{k,\tau+1} + \frac{2c_\star}{(m_\tau + 1)^2}.$$

We now use the concave square root tangent upper bound, $\sqrt{x + r} \le \sqrt{x} + r/(2\sqrt{x})$, at $x = m_\tau^2 s_{k,\tau}^2/(m_\tau + 1)^2$. Given that $s_{k,\tau} \ge \sqrt{c_\star/m_\tau}$, this gives

$$s_{k,\tau+1} \le \frac{m_\tau}{m_\tau + 1} s_{k,\tau} - \frac{\xi_{k,\tau+1}}{\kappa_{\mathrm{dir}} m_\tau s_{k,\tau}} + \frac{c_\star}{m_\tau(m_\tau + 1) s_{k,\tau}}.$$

Let $u_{k,\tau} = e_{k,\tau}/s_{k,\tau}$ and $q_{k,\tau} = m_\tau s_{k,\tau}$. Multiplying through by the term $m_\tau + 1$ yields the single-step inequality

$$q_{k,\tau+1} \le q_{k,\tau} - \frac{m_\tau + 1}{\kappa_{\mathrm{dir}} m_\tau} \langle u_{k,\tau}, M_{k,\tau+1} \rangle + \frac{\sqrt{c_\star}}{\sqrt{m_\tau}},$$

using $c_\star/(m_\tau s_{k,\tau}) \le \sqrt{c_\star/m_\tau}$. The martingale increment is bounded as

$$\left| \frac{m_\tau + 1}{\kappa_{\mathrm{dir}} m_\tau} \langle u_{k,\tau}, M_{k,\tau+1} \rangle \right| \le \frac{m_\tau + 1}{\kappa_{\mathrm{dir}} m_\tau} \frac{2L}{\bar{\gamma}} \le 4\sqrt{c_\star},$$

where $m_\tau \ge 1$. Azuma-Hoeffding and a union bound yield

$$q_{k,\tau} \le q_{k,0} + 4\sqrt{2 c_\star \tau \log(2n/\delta)} + \sqrt{c_\star} \sum_{s=0}^{\tau-1} \frac{1}{\sqrt{m_s}}$$

for all $\tau \le n$. Using the inequalities

$$q_{k,0} \le 2S\tau_0 + \sqrt{c_\star \tau_0}, \qquad \sum_{s=0}^{\tau-1} \frac{1}{\sqrt{m_s}} \le 2\sqrt{\tau + \tau_0},$$

and normalizing by $\tau + \tau_0$ completes the proof of the displayed confidence parameter sequence. □

A.5 Proof of Theorem 1

Theorem 1 (full form). *Let $\delta \in (0,1)$. Execute Algorithm 1 initialized with learning rate schedule η_τ and the confidence schedule β_τ from Proposition 1 with $\delta' = \delta/K$. Under Assumption 1, with probability at least $1-\delta$, the cumulative location regret satisfies*

$$\begin{aligned} R_{\mathrm{loc}}(T) \le 2L \left(2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}} \bar{\gamma}} \sqrt{\tau_0} \right) K(1 + \log(T + \tau_0)) \\ + \frac{8L^2}{\kappa_{\mathrm{dir}} \bar{\gamma}} \sqrt{KT} + \frac{16\sqrt{2} L^2}{\kappa_{\mathrm{dir}} \bar{\gamma}} \sqrt{KT \log(2KT/\delta)}. \end{aligned} \tag{17}$$

Proof of Theorem 1. Apply Proposition 1 to each arm with local horizon $n = T$ and failure level δ/K, then take a union bound. This constructs the displayed simultaneous confidence sequence. Consequently, Proposition 3 guarantees

$$R_{\mathrm{loc}}(T) \le 2L \sum_{k=1}^{K} \sum_{\tau=0}^{T_k - 1} \beta_\tau^{\mathrm{Cau}}.$$

The harmonic term is bounded by

$$2L \left(2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}} \bar{\gamma}} \sqrt{\tau_0} \right) K(1 + \log(T + \tau_0)).$$

The square-root terms are bounded using

$$\sum_{\tau=0}^{T_k - 1} \frac{1}{\sqrt{\tau + \tau_0}} \le 2\sqrt{T_k} \qquad \text{and} \qquad \sum_{k=1}^{K} \sqrt{T_k} \le \sqrt{KT}.$$

Combining these bounds gives Eq. (17). □

A.6 Proof of Corollary 1

The body summarizes this corollary; its statement is repeated here.

Corollary 1 (Constant-factor misspecification)**.** *Under the conditions of Theorem 1, suppose the learner's chosen score scale $\bar{\gamma}$ exhibits a constant-factor mismatch from the true scale, such that $\underline{c}\gamma_\star \le \bar{\gamma} \le \overline{c}\gamma_\star$ for absolute constants $0 < \underline{c} \le \overline{c} < \infty$. The policy maintains the optimal asymptotic rate $R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}(K \log T + \sqrt{KT \log(KT/\delta)})$, where the leading constants depend only on the structural parameters and the misspecification interval.*

Proof of Corollary 1. The proof is that of Theorem 1, which never schedules updates, so the step size for arm k is its local-time factor

$$\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau + \tau_0 + 1)}, \qquad \tau = T_k(t-1).$$

Equivalently, if the step size is indexed by the global round after the selected arm is known, then $\eta_t = \eta_{T_{k_t}(t-1)}$. The explicit burn-in offset τ_0 remains identical to Section 4:

$$\tau_0 = \left\lceil \max\left\{1, \frac{\ell}{\kappa_{\mathrm{dir}}}, \frac{\ell^2}{\kappa_{\mathrm{dir}}^2}\right\}\right\rceil,$$

restricting the penalty of score scale misspecification to the boundaries dictated by κ_{dir} and ℓ.

Define the worst-case constants induced by the interval $\underline{c}\gamma_\star \le \bar{\gamma} \le \overline{c}\gamma_\star$:

$$\kappa_{\min} := \frac{2\mu_{\mathrm{dir}}}{B^2 + (1+\overline{c})^2\gamma_\star^2}, \qquad \ell_{\max} := \frac{2L^2}{(1+\underline{c})^2\gamma_\star^2}.$$

Then

$$\kappa_{\mathrm{dir}} = \frac{2\mu_{\mathrm{dir}}}{B^2 + (\gamma_\star + \bar{\gamma})^2} \ge \kappa_{\min}, \qquad \ell = \frac{2L^2}{(\gamma_\star + \bar{\gamma})^2} \le \ell_{\max}, \qquad \frac{1}{\bar{\gamma}} \le \frac{1}{\underline{c}\,\gamma_\star}.$$

Hence the confidence sequence from Proposition 1 admits the explicit envelope

$$\beta_\tau \le \beta_\tau^{\mathrm{mis}} := \frac{2S\tau_0 + \frac{L}{\kappa_{\min}\underline{c}\,\gamma_\star}\sqrt{\tau_0}}{\tau + \tau_0} + \frac{2L}{\kappa_{\min}\underline{c}\,\gamma_\star\sqrt{\tau + \tau_0}} + \frac{4\sqrt{2}L}{\kappa_{\min}\underline{c}\,\gamma_\star}\sqrt{\frac{\log(2T/\delta)}{\tau + \tau_0}}.$$

Since $\tau_0 = \lceil \max\{1, \ell/\kappa_{\mathrm{dir}}, \ell^2/\kappa_{\mathrm{dir}}^2\}\rceil$ and $\kappa_{\mathrm{dir}}, \ell$ are bounded by $\kappa_{\min}, \ell_{\max}$, the burn-in is also a fixed constant depending only on $L, S, \mu_{\mathrm{dir}}, \gamma_\star, \underline{c}, \overline{c}$.

Applying Proposition 3 with $\beta_\tau^{\mathrm{mis}}$ yields

$$R_{\mathrm{loc}}(T) \le 2L \sum_{k=1}^{K} \sum_{\tau=0}^{T_k - 1} \beta_\tau^{\mathrm{mis}}.$$

Using the standard harmonic and square-root sums from Theorem 1,

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(KT/\delta)}\Big),$$

with constants depending only on $L, S, \mu_{\mathrm{dir}}, \gamma_\star, \underline{c}, \overline{c}$. □

A.7 The instance-dependent bound, and its proof

Corollary 2 (Uniform contextual margins)**.** *Define the uniform contextual gap for arm k as*

$$\Delta_k = \inf\left\{ y_{k_t^\star,t}^\top \theta_{k_t^\star} - y_{k,t}^\top \theta_k \;:\; t \in [T],\; y_{k,t}^\top \theta_k < y_{k_t^\star,t}^\top \theta_{k_t^\star} \right\},$$

with $\Delta_k = +\infty$ if arm k is never strictly suboptimal. Under Theorem 1, with probability at least $1 - \delta$ the location regret admits the explicit instance-dependent bound

$$R_{\mathrm{loc}}(T) \le \sum_{k:\Delta_k < \infty} \left(2LA_{\mathrm{Cau}}\left(1 + \log(u_k^\Delta + \tau_0)\right) + 4LB_{\mathrm{Cau}}\sqrt{u_k^\Delta} \right),$$

where

$$u_k^\Delta = \left\lceil \max\left\{ \frac{8LA_{\mathrm{Cau}}}{\Delta_k}, \frac{64L^2B_{\mathrm{Cau}}^2}{\Delta_k^2} \right\} \right\rceil.$$

In particular, the location regret admits the instance-dependent asymptotic bound

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\left(\sum_{k:\Delta_k < \infty} \left(LA_{\mathrm{Cau}} + \frac{L^2 B_{\mathrm{Cau}}^2}{\Delta_k} \right) \right).$$

Proof of Corollary 2. Work on the confidence event of probability at least $1-\delta$ constructed in the proof of Theorem 1. Fix an arm k with $\Delta_k < \infty$; then $\Delta_k > 0$ by construction, since the infimum ranges over finitely many strictly positive realized gaps and is attained. If the policy chooses arm k during round t, Proposition 3 demonstrates

$$y_{k_t^\star,t}^\top \theta_{k_t^\star} - y_{k,t}^\top \theta_k \le 2L\beta_{T_k(t-1)}^{\mathrm{Cau}}.$$

If $T_k(t-1) \ge u_k^\Delta$, then $\tau + \tau_0 \ge u_k^\Delta$ and the definition of u_k^Δ, coupled with the limit

$$\beta_\tau^{\mathrm{Cau}} \le \frac{A_{\mathrm{Cau}}}{\tau+\tau_0} + \frac{B_{\mathrm{Cau}}}{\sqrt{\tau+\tau_0}},$$

ensures

$$2L\beta_{T_k(t-1)}^{\mathrm{Cau}} \le \frac{2LA_{\mathrm{Cau}}}{u_k^\Delta} + \frac{2LB_{\mathrm{Cau}}}{\sqrt{u_k^\Delta}} \le \frac{\Delta_k}{4} + \frac{\Delta_k}{4} = \frac{\Delta_k}{2} < \Delta_k.$$

Since every round in which arm k is suboptimal contributes a gap of at least Δ_k, no such round can have $T_k(t-1) \ge u_k^\Delta$: every suboptimal pull of arm k occurs at a distinct local count $\tau < u_k^\Delta$, so arm k is pulled suboptimally at most u_k^Δ times.

The regret contributed by arm k is therefore at most the sum of $2L\beta_\tau^{\mathrm{Cau}}$ over the local counts $\{0, \dots, u_k^\Delta - 1\}$ (an over-count, since not all of these pulls need be suboptimal, but all summands are nonnegative):

$$\sum_{\tau=0}^{u_k^\Delta-1} 2L\beta_\tau^{\mathrm{Cau}} \le 2LA_{\mathrm{Cau}} \sum_{\tau=0}^{u_k^\Delta-1} \frac{1}{\tau+\tau_0} + 2LB_{\mathrm{Cau}} \sum_{\tau=0}^{u_k^\Delta-1} \frac{1}{\sqrt{\tau+\tau_0}}.$$

Applying the bounds

$$\sum_{\tau=0}^{u-1} \frac{1}{\tau+\tau_0} \le 1 + \log(u+\tau_0), \qquad \sum_{\tau=0}^{u-1} \frac{1}{\sqrt{\tau+\tau_0}} \le 2\sqrt{u},$$

and summing over all arms with $\Delta_k < \infty$ proves the explicit gap bound. The asymptotic form follows from

$$\sqrt{u_k^\Delta} \le 1 + \sqrt{\frac{8LA_{\mathrm{Cau}}}{\Delta_k}} + \frac{8LB_{\mathrm{Cau}}}{\Delta_k}$$

($\sqrt{\lceil \max\{x,y\} \rceil} \le 1 + \sqrt{x} + \sqrt{y}$), the absorption of the mixed term $\sqrt{LA_{\mathrm{Cau}} \cdot L^2 B_{\mathrm{Cau}}^2/\Delta_k}$ into $LA_{\mathrm{Cau}} + L^2 B_{\mathrm{Cau}}^2/\Delta_k$ by AM–GM, and the logarithmic growth of $\log(u_k^\Delta + \tau_0)$. □

B Dyadic-Grid Scale Adaptation

This section treats the adaptation problem left open by the body: the confidence schedule of Proposition 1 requires a score scale $\bar{\gamma}$, while Corollary 1 covers only a constant-factor mismatch around the true scale $\gamma_\star$. The arm-wise dyadic-grid policy below removes the need to tune the scale accurately: each branch of a geometric grid runs the Direct-Cauchy update at its own scale with computable surrogate schedules, and a per-arm model-selection rule picks the branch with the smallest current confidence radius. Theorem 6 shows that this costs only a factor $\log(M+1)$ inside the logarithmic term, and Remark 1 then removes the known upper bound $\Gamma_{\max}$ by a restarting doubling schedule. The section gives the algorithm, the full statement with the surrogate schedules, the proof, and the doubling analysis.

Theorem 6 (Dyadic scale adaptation)**.** *Let $\delta \in (0,1)$. Assume only a known bounding interval $\gamma_{\min} \le \gamma_\star \le \Gamma_{\max}$. Let*

$$\mathcal{G} = \left\{\gamma_{\min} 2^j \,:\, j = 0, \dots, M\right\}, \qquad M = \max\left\{1, \left\lceil \log_2 \frac{\Gamma_{\max}}{\gamma_{\min}} \right\rceil\right\},$$

so that the top grid point satisfies $\gamma_{\min} 2^M \ge \Gamma_{\max} \ge \gamma_\star$ and $|\mathcal{G}| = M+1 \ge 2$. Execute Algorithm 2 with total failure budget δ, where each branch m runs the computable surrogate schedules of Eq. (18) *— the contraction modulus, prediction limit and burn-in obtained by evaluating the constants of*

Algorithm 2 Arm-wise dyadic-grid Direct-Cauchy-UCB

Require: Dyadic grid $\mathcal{G} = \{\bar{\gamma}_0, \dots, \bar{\gamma}_M\}$ ($M+1$ branches), total failure budget δ.

For each scale $m \in \{0, \dots, M\}$ and each arm $k \in [K]$, initialize $T_k^{(m)}(0) = 0$ and $z_{k,0}^{(m)} = \mathbf{0}$.
for round $t = 1, \dots, T$ **do**
  Observe contexts $\{y_{1,t}, \dots, y_{K,t}\}$.
  for each arm $k \in [K]$ **do**
    for each scale $m \in \{0, \dots, M\}$ **do**
      Compute $\mathrm{UCB}_{k,t}^{(m)} = y_{k,t}^\top z_{k,T_k^{(m)}(t-1)}^{(m)} + \beta_{T_k^{(m)}(t-1)}^{(m)} \|y_{k,t}\|_2$.
    end for
    Select $m_{k,t} \in \arg\min_{m \in \{0,\dots,M\}} \beta_{T_k^{(m)}(t-1)}^{(m)}$.
    Let $\mathrm{UCB}_{k,t} = \mathrm{UCB}_{k,t}^{(m_{k,t})}$.
  end for
  Play $k_t \in \arg\max_{k \in [K]} \mathrm{UCB}_{k,t}$.
  Observe reward r_t.
  for each scale $m \in \{0, \dots, M\}$ **do**
    Update the played arm only, using scale $\bar{\gamma}_m$:

$$z_{k_t,\tau+1}^{(m)} = \Pi_{\mathcal{X}}\left(z_{k_t,\tau}^{(m)} - \eta_\tau^{(m)} \frac{2(y_{k_t,t}^\top z_{k_t,\tau}^{(m)} - r_t)}{\bar{\gamma}_m^2 + (y_{k_t,t}^\top z_{k_t,\tau}^{(m)} - r_t)^2} y_{k_t,t} \right),$$

    where $\tau = T_{k_t}^{(m)}(t-1)$.
    Set $T_{k_t}^{(m)}(t) = \tau + 1$ and keep all other $T_k^{(m)}(t)$ unchanged.
  end for
end for

Section 4 at the interval endpoints — together with the confidence radius $\beta_\tau^{(m)}$ of Eq. (13) *evaluated with that branch's surrogate constants at failure level $\delta/(K(M+1))$. Then, with probability at least $1-\delta$,*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(KT(M+1)/\delta)} \Big).$$

Theorem 6 (full form). *Let $\delta \in (0,1)$. Assume only a known bounding interval $\gamma_{\min} \le \gamma_\star \le \Gamma_{\max}$. Let*

$$\mathcal{G} = \{\gamma_{\min} 2^j \,:\, j = 0, \dots, M\}, \qquad M = \max\Big\{1, \Big\lceil \log_2 \frac{\Gamma_{\max}}{\gamma_{\min}} \Big\rceil \Big\},$$

so that the top grid point satisfies $\gamma_{\min} 2^M \ge \Gamma_{\max} \ge \gamma_\star$ and $|\mathcal{G}| = M+1 \ge 2$. Execute Algorithm 2 with total failure budget δ, where each branch m runs the computable surrogate schedules

$$\begin{aligned} \widehat{\kappa}_m &= \frac{2\mu_{\mathrm{dir}}}{B^2 + (\Gamma_{\max} + \bar{\gamma}_m)^2}, \qquad \widehat{\ell}_m = \frac{2L^2}{(\gamma_{\min} + \bar{\gamma}_m)^2}, \\ \widehat{\tau}_0^{(m)} &= \Big\lceil \max\Big\{1, \frac{\widehat{\ell}_m}{\widehat{\kappa}_m}, \frac{\widehat{\ell}_m^2}{\widehat{\kappa}_m^2}\Big\} \Big\rceil, \qquad \eta_\tau^{(m)} = \frac{1}{\widehat{\kappa}_m\big(\tau + \widehat{\tau}_0^{(m)} + 1\big)}, \end{aligned} \tag{18}$$

and the confidence radius $\beta_\tau^{(m)}$ of Eq. (13) *evaluated with $(\bar{\gamma}_m, \widehat{\kappa}_m, \widehat{\tau}_0^{(m)})$ at failure level $\delta/(K(M+1))$. Then, with probability at least $1-\delta$,*

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(KT(M+1)/\delta)} \Big).$$

Proof of Theorem 6. We analyze the arm-wise aggregation in Algorithm 2. For each grid point $\bar{\gamma}_m$, define $\beta_\tau^{(m)}$ by Eq. (13) with score scale $\bar{\gamma} = \bar{\gamma}_m$ and the computable surrogate constants $\widehat{\kappa}_m, \widehat{\ell}_m, \widehat{\tau}_0^{(m)}, \eta_\tau^{(m)}$ of the theorem statement, and the failure level $\delta/(K(M+1))$ for each arm-scale copy. Equivalently, each scale receives budget $\delta/(M+1)$, distributed over the K arms. The

surrogates sandwich the true (oracle) constants: since $\gamma_{\min} \le \gamma_\star \le \Gamma_{\max}$,

$$\widehat{\kappa}_m = \frac{2\mu_{\mathrm{dir}}}{B^2 + (\Gamma_{\max} + \bar{\gamma}_m)^2} \;\le\; \kappa_{\mathrm{dir}}^{(m)} := \frac{2\mu_{\mathrm{dir}}}{B^2 + (\gamma_\star + \bar{\gamma}_m)^2},$$
$$\widehat{\ell}_m = \frac{2L^2}{(\gamma_{\min} + \bar{\gamma}_m)^2} \;\ge\; \ell^{(m)} := \frac{2L^2}{(\gamma_\star + \bar{\gamma}_m)^2},$$

hence $\widehat{\ell}_m/\widehat{\kappa}_m \ge \ell^{(m)}/\kappa_{\mathrm{dir}}^{(m)}$, so $\widehat{\tau}_0^{(m)} \ge \tau_0^{(m)}$, and the contraction step satisfies $\kappa_{\mathrm{dir}}^{(m)}\eta_\tau^{(m)} \ge \widehat{\kappa}_m\,\eta_\tau^{(m)} = 1/\big(\tau + \widehat{\tau}_0^{(m)} + 1\big)$. Therefore the analysis of Propositions 5 and 1 replays verbatim with $(\widehat{\kappa}_m, \widehat{\tau}_0^{(m)})$ in place of $(\kappa_{\mathrm{dir}}, \tau_0)$ — those proofs use the schedule only through $\kappa_{\mathrm{dir}}\eta_\tau \ge 1/(\tau+\tau_0+1)$ and $\tau_0 \ge 1$, with $c_\star$ and D_{Cau} redefined in terms of $(\widehat{\kappa}_m, \widehat{\tau}_0^{(m)})$ and the drift inequality still taken with the true $\kappa_{\mathrm{dir}}^{(m)}$ — and the radius $\beta_\tau^{(m)}$ is valid for branch m on the shared ensemble stream. (Note that Assumption 1 is imposed on this ensemble-generated stream; since the ensemble's selection rule inspects the current contexts, this is a genuine restriction, as discussed after Proposition 4.) At round t, each scale branch observes the same contexts and forms

$$\mathrm{UCB}_{k,t}^{(m)} = y_{k,t}^\top z_{k,T_k^{(m)}(t-1)}^{(m)} + \beta_{T_k^{(m)}(t-1)}^{(m)} \|y_{k,t}\|_2.$$

The algorithm then performs the scale choice separately for each arm,

$$m_{k,t} \in \arg\min_{m\in\{0,\dots,M\}} \beta_{T_k^{(m)}(t-1)}^{(m)}, \qquad \mathrm{UCB}_{k,t} = \mathrm{UCB}_{k,t}^{(m_{k,t})},$$

and plays the arm with largest aggregated upper confidence value,

$$k_t \in \arg\max_{k\in[K]} \mathrm{UCB}_{k,t}.$$

After observing r_t, every scale branch receives the same played-arm sample and updates only the component corresponding to k_t. Hence, for every arm k and scale m, the local clock is identical:

$$T_k^{(m)}(t) = T_k(t) := \sum_{s\le t} \mathbf{1}\{k_s = k\}.$$

The branches differ only through their score scale and confidence radius.

Apply Proposition 1 to every pair (k, m) with local horizon T and failure level $\delta/(K(M+1))$. A union bound over the $K(M+1)$ pairs gives an event $\mathcal{E}$ with probability at least $1-\delta$ on which, simultaneously for all $k \in [K]$, $m \in \{0,\dots,M\}$, local times $\tau \le T$, and contexts y,

$$|y^\top(z_{k,\tau}^{(m)} - \theta_k)| \le \beta_\tau^{(m)}\|y\|_2.$$

Therefore every branch UCB is valid on $\mathcal{E}$. Since the arm-wise selected branch $m_{k,t}$ is one of these valid branches,

$$y_{k,t}^\top\theta_k \le y_{k,t}^\top z_{k,T_k(t-1)}^{(m_{k,t})} + \beta_{T_k(t-1)}^{(m_{k,t})}\|y_{k,t}\|_2 = \mathrm{UCB}_{k,t}.$$

Moreover, by the definition of $m_{k,t}$,

$$\beta_{T_k(t-1)}^{(m_{k,t})} = \min_{m\in\{0,\dots,M\}} \beta_{T_k(t-1)}^{(m)}.$$

The deterministic reduction in Proposition 3 therefore applies with the non-increasing effective radius $\widetilde{\beta}_\tau := \min_m \beta_\tau^{(m)}$ and yields

$$R_{\mathrm{loc}}(T) \le 2L \sum_{k=1}^{K}\sum_{\tau=0}^{T_k-1} \widetilde{\beta}_\tau.$$

It remains to bound the effective radius by that of an anchor branch. Let $\bar{\gamma}^\sharp$ be the smallest grid point satisfying $\bar{\gamma}^\sharp \ge \gamma_\star$; it exists because the top grid point obeys $\gamma_{\min}2^M \ge \Gamma_{\max} \ge \gamma_\star$, and minimality gives $\gamma_\star \le \bar{\gamma}^\sharp < 2\gamma_\star$. The anchor's radius is built from the surrogate constants, and

$$\frac{1}{\widehat{\kappa}_\sharp\,\bar{\gamma}^\sharp} = \frac{B^2 + (\Gamma_{\max} + \bar{\gamma}^\sharp)^2}{2\mu_{\mathrm{dir}}\,\bar{\gamma}^\sharp} \;\le\; \frac{B^2 + (\Gamma_{\max} + 2\gamma_\star)^2}{2\mu_{\mathrm{dir}}\,\gamma_\star} \;\le\; \frac{B^2 + 9\Gamma_{\max}^2}{2\mu_{\mathrm{dir}}\,\gamma_{\min}},$$

(the last step uses $\gamma_\star \le \Gamma_{\max}$ in the numerator and $\gamma_\star \ge \gamma_{\min}$ in the denominator), while $1/(\widehat{\kappa}_\sharp \bar{\gamma}^\sharp)$ and $\widehat{\tau}_0^\sharp$ are bounded above by functions of $L, S, \mu_{\mathrm{dir}}, \gamma_{\min}, \Gamma_{\max}$ only — the former by the displayed chain, the latter since $\widehat{\ell}_\sharp/\widehat{\kappa}_\sharp \le L^2\big(B^2 + 9\Gamma_{\max}^2\big)/\big(\mu_{\mathrm{dir}}\gamma_{\min}^2\big)$ (using $\gamma_{\min} \le \bar{\gamma}^\sharp < 2\gamma_\star \le 2\Gamma_{\max}$). Thus the burn-in $\widehat{\tau}_0^\sharp$ and the constants in $\beta_\tau^\sharp$ depend only on the structural parameters and the bounding interval, and not on M or T, while the logarithmic term becomes $\log(2KT(M+1)/\delta)$. Since $\widetilde{\beta}_\tau \le \beta_\tau^\sharp$, the same summation used in the proof of Theorem 1 gives

$$R_{\mathrm{loc}}(T) \le 2LA_\sharp K\big(1 + \log\big(T + \widehat{\tau}_0^\sharp\big)\big) + C_\sharp\sqrt{KT} + C'_\sharp\sqrt{KT\log(2KT(M+1)/\delta)}$$

for constants $A_\sharp, C_\sharp, C'_\sharp$ depending only on $L, S, \mu_{\mathrm{dir}}, \gamma_{\min}, \Gamma_{\max}$. Consequently,

$$R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\Big(K\log T + \sqrt{KT\log(KT(M+1)/\delta)}\Big),$$

where the tilde hides constants depending on $L, S, \mu_{\mathrm{dir}}, \gamma_{\min}, \Gamma_{\max}$ and polylogarithmic factors. □

Remark 1 (Doubling removes the known upper bound)**.** *A restarting doubling schedule removes the need for a known upper bound $\Gamma_{\max}$; a lower end $\gamma_{\min}$ is still used throughout. Phase i runs Algorithm 2 on the interval $[\gamma_{\min}, \Gamma_i]$ with $\Gamma_i = \Gamma_0 2^i$, with phase length coupled to the guess, $n_i = n_0 2^i$, and with a phase failure level δ_i from any summable allocation with $\sum_i \delta_i \le \delta$. A phase with $\Gamma_i < \gamma_\star$ is not covered by Theorem 6: its surrogates are anti-conservative, $\widehat{\kappa}_m > \kappa_{\mathrm{dir}}^{(m)}$ for every branch, so the surrogate radii carry no guarantee. The rounds of such phases are charged the trivial per-round cost $2LS$. Since the phase lengths grow geometrically, the uncovered rounds total at most $\sum_{i<i^\star} n_i \le n_{i^\star} \approx n_0\,\gamma_\star/\Gamma_0$, where $i^\star = \lceil \log_2(\gamma_\star/\Gamma_0)\rceil$ is the first covering phase; this is an additive, instance-dependent constant. For each covering phase, Theorem 6 applies to its own n_i rounds, and the dominant square-root terms sum geometrically, $\sum_{i\le I}\sqrt{n_i} \le \sqrt{n_I}/(1-2^{-1/2})$, a constant overhead over the last phase. The per-phase burn-in terms $K\log n_i$ accumulate over all $I - i^\star + 1 \approx \log_2(T/n_0) - \log_2(\gamma_\star/\Gamma_0)$ covering phases, with the last-phase index set by the horizon through $\sum_{i\le I} n_i \approx T$; the $K\log T$ term is therefore inflated by an additional factor $\Theta(\log T)$, absorbed by the $\widetilde{\mathcal{O}}$ notation. Altogether, removing $\Gamma_{\max}$ costs a constant factor on the $\sqrt{KT}$ term, an extra logarithmic factor on the $K\log T$ term, and the additive instance-dependent charge $2LS\,n_{i^\star}$.*

Phase accounting for Remark 1. Allocate the failure budget across phases by any summable schedule, e.g. $\delta_i = 6\delta/(\pi^2(i+1)^2)$, so that $\sum_i \delta_i \le \delta$; on the resulting intersection event (probability at least $1-\delta$), every covering phase's instance of Theorem 6 holds simultaneously for its own n_i rounds. Note that the per-phase logarithm $\log(2Kn_i(M_i+1)/\delta_i)$ exceeds $\log(2KT(M+1)/\delta)$ only by an additive $O(\log(i+1)) \le O(\log\log T)$, which is absorbed by the $\widetilde{\mathcal{O}}$ notation. The dominant square-root contributions then sum geometrically:

$$\begin{aligned}\sum_{i=i^\star}^{I}\sqrt{Kn_i\log\big(2Kn_i(M_i+1)/\delta_i\big)} \;&\le\; \sqrt{K\log\big(2KT(M+1)/\delta\big) + \mathrm{polylog}(T)}\;\sum_{i=i^\star}^{I}\sqrt{n_i}\\ &\le\; \frac{\sqrt{Kn_I\,\big(\log(2KT(M+1)/\delta) + \mathrm{polylog}(T)\big)}}{1-2^{-1/2}},\end{aligned}$$

a constant factor over the last phase (and $n_I = \Theta(T)$). The burn-in parts contribute $K\sum_{i=i^\star}^{I}\log\big(n_i + \widehat{\tau}_0\big) \le K\,(I - i^\star + 1)\,\log\big(T + \widehat{\tau}_0\big) = \mathcal{O}\big(K\log^2 T\big)$, i.e. an additional factor $\Theta(\log T)$ on the $K\log T$ term, since the covering phases number $I - i^\star + 1 \approx \log_2(T/n_0) - \log_2(\gamma_\star/\Gamma_0)$ (the last-phase index is set by the horizon through $\sum_{i\le I} n_i \approx T$, and $i^\star = \lceil\log_2(\gamma_\star/\Gamma_0)\rceil$). Finally, the uncovered phases ($\Gamma_i < \gamma_\star$) contribute at most $2LS\sum_{i<i^\star} n_i \le 2LS\,n_{i^\star}$, an additive constant independent of T.

C Lower-Bound Proof

C.1 The Cauchy statistical distance, and its proof

Lemma 2 (Cauchy statistical distance)**.** *The exact Kullback-Leibler divergence between two Cauchy distributions sharing a scale parameter $\gamma_\star$ but differing in location by Δ is analytically given by*

$$\mathrm{KL}(\mathrm{Cauchy}(\mu,\gamma_\star)\,\|\,\mathrm{Cauchy}(\lambda,\gamma_\star)) = \log(1 + \frac{(\mu-\lambda)^2}{4\gamma_\star^2}). \tag{19}$$

Proof in Appendix C.1.

Proof. By scaling invariance it suffices to evaluate

$$I(\delta) = \mathrm{KL}(\mathrm{Cauchy}(0,1) \,\|\, \mathrm{Cauchy}(\delta,1)) = \int p_1(u) \log \frac{1+(u-\delta)^2}{1+u^2}\, du.$$

Differentiating under the integral gives

$$I'(\delta) = \int p_1(u) \frac{2(\delta-u)}{1+(u-\delta)^2}\, du = 2\pi(p_1 * q_1)(\delta) = 2\pi q_2(\delta) = \frac{2\delta}{\delta^2+4},$$

using the same Poisson-kernel convolution identity as in Lemma 1. Integrating back from $I(0) = 0$ yields $\log(1+\delta^2/4)$. □

C.2 Proof of Theorem 2

Proof of Theorem 2. Consider the one-dimensional constant-context subclass with $y_{k,t} \equiv L$. Let $\mathbb{P}_0$ denote the instance in which every arm has reward law $\mathrm{Cauchy}(0,\gamma_\star)$. For each $j \in [K]$, let $\mathbb{P}_j$ denote the instance in which arm j has reward law $\mathrm{Cauchy}(\Delta_T,\gamma_\star)$ and all other arms have reward law $\mathrm{Cauchy}(0,\gamma_\star)$. Since $\Delta_T \le LS$, this subclass is realized by $\theta_j = \Delta_T/L$ and $\theta_k = 0$ for $keqj$.

Under $\mathbb{P}_j$, arm j is optimal and

$$R_{\mathrm{loc}}(T) = \Delta_T(T - N_j(T)),$$

where $N_j(T) = \sum_{t=1}^T \mathbf{1}\{k_t = j\}$. Since $\sum_{j=1}^K N_j(T) = T$ under $\mathbb{P}_0$, the pigeonhole principle gives an arm $j_\star$ such that

$$\mathbb{E}_0[N_{j_\star}(T)] \le \frac{T}{K}.$$

Because divergence between $\mathbb{P}_0$ and $\mathbb{P}_{j_\star}$ is accumulated only when the policy pulls $j_\star$, the chain rule gives

$$\mathrm{KL}(\mathbb{P}_0 \,\|\, \mathbb{P}_{j_\star}) = \mathbb{E}_0[N_{j_\star}(T)]\, \mathrm{KL}(\mathrm{Cauchy}(0,\gamma_\star) \,\|\, \mathrm{Cauchy}(\Delta_T,\gamma_\star)).$$

Using Lemma 2, this is at most

$$\mathrm{KL}(\mathbb{P}_0 \,\|\, \mathbb{P}_{j_\star}) \le \frac{T}{K} \cdot \frac{\Delta_T^2}{4\gamma_\star^2} \le \frac{1}{8},$$

using $\log(1+x) \le x$ and $\Delta_T \le \gamma_\star\sqrt{K/(2T)}$. Pinsker's inequality gives

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_{j_\star}) \le \sqrt{\frac{1}{2}\,\mathrm{KL}(\mathbb{P}_0 \,\|\, \mathbb{P}_{j_\star})} \le \frac{1}{4}.$$

Because $0 \le N_{j_\star}(T) \le T$, we obtain

$$\mathbb{E}_{j_\star}[N_{j_\star}(T)] \le \mathbb{E}_0[N_{j_\star}(T)] + T\ \mathrm{TV}(\mathbb{P}_0, \mathbb{P}_{j_\star}) \le \frac{T}{K} + \frac{T}{4} \le \frac{3T}{4},$$

where the last step uses $K \ge 2$. Hence,

$$\mathbb{E}_{j_\star}[R_{\mathrm{loc}}(T)] = \Delta_T\,(T - \mathbb{E}_{j_\star}[N_{j_\star}(T)]) \ge \frac{T\Delta_T}{4}.$$

Since $\mathbb{P}_{j_\star}$ is one of the instances $\{\mathbb{P}_0, \dots, \mathbb{P}_K\}$ of the subclass constructed above, the supremum over the subclass satisfies $\sup_\nu \mathbb{E}_\nu[R_{\mathrm{loc}}(T)] \ge T\Delta_T/4$ for every policy. The square-root regime follows by substituting $\Delta_T = \gamma_\star\sqrt{K/(2T)}$. □

C.3 The gap-dependent lower bound, and its proof

Theorem 7 (Gap-dependent lower bound)**.** *Let $\mathcal{P}$ denote the family of all constant-context exact-Cauchy translation instances of the model of Section 2: $r_t = \mu_{k_t} + \xi_t$, where $\xi_t \sim \mathrm{Cauchy}(0,\gamma_\star)$ conditionally on $\mathcal{F}_{t-1}$, $y_{k,t} \equiv y$ with $\|y\|_2 \le L$, and $\mu_k = y^\top\theta_k$ with $\|\theta_k\|_2 \le S$ (so $R_{\mathrm{loc}}(T) = \sum_{t=1}^T(\mu_{k_t^\star} - \mu_{k_t})$). A policy is* uniformly good on $\mathcal{P}$ *if $\mathbb{E}_\nu[R_{\mathrm{loc}}(T)] = o(T^\alpha)$ for every $\alpha > 0$ and*

every instance $\nu \in \mathcal{P}$. Fix an instance $\nu \in \mathcal{P}$ with unique optimal arm 1 and gaps $\Delta_k = \mu_1 - \mu_k > 0$ for $k \neq 1$, fixed independently of T. Then any policy uniformly good on $\mathcal{P}$ must obey the fundamental asymptotic limit

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[R_{\mathrm{loc}}(T)]}{\log T} \geq \sum_{k\neq 1} \frac{\Delta_k}{\log(1 + \frac{\Delta_k^2}{4\gamma_\star^2})}. \tag{20}$$

Proof in Appendix C.3.

Proof of Theorem 7. Fix the instance $\nu = (\mathrm{Cauchy}(\mu_1, \gamma_\star), \ldots, \mathrm{Cauchy}(\mu_K, \gamma_\star))$ with unique optimal arm 1; as is standard for lower bounds, the policy's internal randomization is taken independent of the instance. For an alternate suboptimal arm $k \neq 1$ and a hypothetical margin $\lambda > \mu_1$, let $\nu^{(k,\lambda)} \in \mathcal{P}$ denote the alternative instance in which arm k is replaced by $\mathrm{Cauchy}(\lambda, \gamma_\star)$, and let $R_{\mathrm{loc}}^{(k,\lambda)}(T)$ denote the location regret under $\nu^{(k,\lambda)}$.

Define the event

$$A_{k,T} = \{N_k(T) < T/2\},$$

where $N_k(T) = \sum_{t=1}^T \mathbf{1}\{k_t = k\}$. Under ν, the regret identity of the theorem statement reads

$$R_{\mathrm{loc}}(T) = \sum_{j\neq 1} \Delta_j N_j(T) \geq \Delta_k N_k(T),$$

so on $A_{k,T}^c$ we have $R_{\mathrm{loc}}(T) \geq \frac{\Delta_k T}{2}$. Markov's inequality therefore gives $\mathbb{E}_\nu[R_{\mathrm{loc}}(T)] \geq (\Delta_k T/2)\,\mathbb{P}_\nu(A_{k,T}^c)$, and since the policy is uniformly good and $\Delta_k > 0$ is fixed, for every $\alpha > 0$,

$$\mathbb{P}_\nu(A_{k,T}^c) \leq \frac{2\,\mathbb{E}_\nu[R_{\mathrm{loc}}(T)]}{\Delta_k T} = o(T^{\alpha-1}).$$

Under the alternative instance $\nu^{(k,\lambda)}$, arm k is uniquely optimal and every pull of a different arm incurs instantaneous location regret at least $\lambda - \mu_1$; therefore, on the event $A_{k,T}$,

$$R_{\mathrm{loc}}^{(k,\lambda)}(T) \geq \frac{(\lambda - \mu_1)T}{2}.$$

The alternative $\nu^{(k,\lambda)}$ itself belongs to $\mathcal{P}$, so uniform goodness applies to it as well, and the same Markov step yields, for every $\alpha > 0$,

$$\mathbb{P}_{\nu^{(k,\lambda)}}(A_{k,T}) \leq \frac{2\,\mathbb{E}_{\nu^{(k,\lambda)}}[R_{\mathrm{loc}}^{(k,\lambda)}(T)]}{(\lambda - \mu_1)T} = o(T^{\alpha-1}).$$

Let $\mathbb{P}_\nu^T$ and $\mathbb{P}_{\nu^{(k,\lambda)}}^T$ denote the distributions of the length-T interaction history under the two instances. The two instances differ only through observations from arm k, hence the chain rule yields

$$\mathrm{KL}(\mathbb{P}_\nu^T \,\|\, \mathbb{P}_{\nu^{(k,\lambda)}}^T) = \mathbb{E}_\nu[N_k(T)]\,\mathrm{KL}(\mathrm{Cauchy}(\mu_k, \gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda, \gamma_\star))\,.$$

By data processing for the indicator of $A_{k,T}$,

$$\mathrm{KL}(\mathbb{P}_\nu^T \,\|\, \mathbb{P}_{\nu^{(k,\lambda)}}^T) \geq \mathrm{kl}(\mathbb{P}_\nu(A_{k,T}), \mathbb{P}_{\nu^{(k,\lambda)}}(A_{k,T}))\,,$$

where kl denotes the binary Kullback–Leibler divergence.

Taking any $\alpha < 1$ in the preceding displays gives $p_T = \mathbb{P}_\nu(A_{k,T}) \to 1$ while $q_T = \mathbb{P}_{\nu^{(k,\lambda)}}(A_{k,T}) = o(T^{\alpha-1})$; the elementary bound $\mathrm{kl}(p, q) \geq p\log(1/q) - \log 2$ then yields

$$\liminf_{T\to\infty} \frac{\mathrm{kl}(\mathbb{P}_\nu(A_{k,T}), \mathbb{P}_{\nu^{(k,\lambda)}}(A_{k,T}))}{\log T} \geq 1 - \alpha \qquad \text{for every } \alpha > 0,$$

and since $\alpha > 0$ is arbitrary, the liminf is at least 1. Dividing the chain-rule identity by $\mathrm{KL}(\mathrm{Cauchy}(\mu_k, \gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda, \gamma_\star))$ and passing to the limit therefore gives, for every $\lambda > \mu_1$,

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\log T} \geq \frac{1}{\mathrm{KL}(\mathrm{Cauchy}(\mu_k, \gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda, \gamma_\star))},$$

and optimizing over λ tightens it to $1/\inf_{\lambda>\mu_1} \mathrm{KL}(\mathrm{Cauchy}(\mu_k,\gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda,\gamma_\star))$:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\log T} \geq \frac{1}{\inf_{\lambda>\mu_1} \mathrm{KL}(\mathrm{Cauchy}(\mu_k,\gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda,\gamma_\star))}.$$

By translation invariance and Lemma 2,

$$\begin{aligned}\inf_{\lambda>\mu_1} \mathrm{KL}(\mathrm{Cauchy}(\mu_k,\gamma_\star) \,\|\, \mathrm{Cauchy}(\lambda,\gamma_\star)) &= \mathrm{KL}(\mathrm{Cauchy}(0,\gamma_\star) \,\|\, \mathrm{Cauchy}(\Delta_k,\gamma_\star)) \\ &= \log\left(1+\frac{\Delta_k^2}{4\gamma_\star^2}\right),\end{aligned}$$

where the infimum over the open half-line $\lambda > \mu_1$ is approached as $\lambda \downarrow \mu_1$, since $x \mapsto \log(1 + x^2/(4\gamma_\star^2))$ is increasing for $x > 0$. Finally, since $R_{\mathrm{loc}}(T) = \sum_{j\neq 1} \Delta_j N_j(T)$ with nonnegative fixed weights, and the liminf of a finite sum dominates the sum of the liminf's,

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[R_{\mathrm{loc}}(T)]}{\log T} \geq \sum_{j\neq 1} \frac{\Delta_j}{\log\left(1+\frac{\Delta_j^2}{4\gamma_\star^2}\right)},$$

which is the claimed bound. □

D Proofs for Adversarial Contexts

This section states and proves the two one-arm results accompanying Theorems 3 and 8: a mean-square contraction under the conditional windowed excitation (23) (a sanity benchmark for the window analysis; the theorems below do not consume it directly) and a uniform confidence sequence under the pathwise windowed excitation (10) (the ingredient the theorems do consume). The notation G, Γ, κ_{pe}, i_0^{pe}, τ_0^{pe}, $c^{(2)}$, $c^{(3)}$, Q, Q_w is as in Eqs. (21)–(22). The deterministic-design reading of the pathwise condition, cited by both theorems, is recorded first.

Remark 2 (Deterministic adversarial contexts)**.** *If the context sequence $\{y_{k,t}\}$ is chosen obliviously, or by an adaptive rule that enforces the design condition on every realized path, condition* (10) *is the purely deterministic design condition $\lambda_{\min}\big(\sum_{j=1}^{w} y_{k,\tau+j} y_{k,\tau+j}^\top\big) \geq \mu w$ for all k, τ. No stochasticity on the contexts is required, and the only randomness is the exact Cauchy noise. This is the persistent-excitation condition of adaptive control, stated here for arm-local windows. By the rank bound recorded after Assumption 2,* (10) *is feasible only for $w \geq d$, and then $\tau_0^{\mathrm{pe}} = i_0^{\mathrm{pe}} w \geq 2d$. The mean-square result (Proposition 6 below) needs only the weaker conditional form* (23)*, which carries no window-length restriction; its $w = 1$ case is exactly Assumption 1 with $\mu_{\mathrm{dir}} = \mu$.*

The windowed constants, and the two one-arm results proved here. We use the per-window step-size schedule

$$\kappa_{\mathrm{pe}} := \mu\, C_{\mathrm{Cau}}, \qquad i_0^{\mathrm{pe}} := 2, \qquad \tau_0^{\mathrm{pe}} := i_0^{\mathrm{pe}} w, \qquad \eta_\tau := \frac{2}{\kappa_{\mathrm{pe}} w\,(i(\tau)+i_0^{\mathrm{pe}})}, \;\; i(\tau) := \lfloor \tau/w \rfloor, \tag{21}$$

(the step size is held constant within each window of w consecutive local updates and decays as $1/i$ across windows, with $i_0^{\mathrm{pe}} = 2$). We also define the window forcing constants

$$\begin{aligned} c^{(2)} &:= \frac{4G^2}{\kappa_{\mathrm{pe}}^2\, w}, \qquad c^{(3)} := \frac{16L^2G^2}{3\kappa_{\mathrm{pe}}^2\,\mu}, \\ Q &:= \max\left\{4S^2\big(i_0^{\mathrm{pe}}\big)^2,\; 2c^{(3)} i_0^{\mathrm{pe}},\; \frac{\big(c^{(2)}+c^{(3)}\big)^2}{6c^{(2)}}\right\}, \qquad Q_w := w^2 Q. \end{aligned} \tag{22}$$

The full form of Theorem 3 follows. Theorem 8 uses the expectation form of the windowed condition, stated next:

Assumption 3 (Windowed persistent excitation, expectation form)**.** *There exist an integer $w \geq 1$ and a constant $\mu > 0$ such that for every arm k, every local time τ, and every $\mathcal{F}_{k,\tau}$-measurable vector $v \in \mathbb{R}^d$,*

$$\sum_{j=1}^{w} \mathbb{E}\big[(y_{k,\tau+j}^\top v)^2 \,\big|\, \mathcal{F}_{k,\tau}\big] \geq \mu w \|v\|_2^2. \tag{23}$$

The full form of Theorem 8, which the body summarizes in one sentence, then reads as follows.

Theorem 3 (full form). *Let $\delta \in (0,1)$. Execute Algorithm 1 with the step-size schedule* (21) *and the confidence schedule β_τ^{pe} from Proposition 2, instantiated with local horizon $n = T$ and failure level δ/K. Under the pathwise windowed excitation* (10)*, with probability at least $1-\delta$,*

$$R_{\mathrm{loc}}(T) \le 3L\,K\,A'_w\big(1+\log(T+\tau_0^{\mathrm{pe}})\big) + \sqrt{\tfrac{3}{2}}\,\frac{\big(64+128\sqrt{\log(2KT/\delta)}\big)L^2}{\kappa_{\mathrm{pe}}\,\bar\gamma}\sqrt{KT}, \tag{24}$$

where

$$A'_w := 2S\tau_0^{\mathrm{pe}} + \frac{2G\sqrt{\tau_0^{\mathrm{pe}}}}{\kappa_{\mathrm{pe}}} + \frac{16L^2G\,w^{3/2}}{\kappa_{\mathrm{pe}}\,\mu\,\sqrt{i_0^{\mathrm{pe}}}} + \frac{32Gw\sqrt{\big(1+\log\frac{T+\tau_0^{\mathrm{pe}}}{w}\big)\log(2KT/\delta)}}{\kappa_{\mathrm{pe}}} + \frac{2Gw}{\kappa_{\mathrm{pe}}}$$

(the constant A_w of Eq. (28) *evaluated at $n = T$ and $\delta \leftarrow \delta/K$; the factors $3/2$ and $\sqrt{3/2}$ are $1 + w/\tau_0^{\mathrm{pe}}$ and its square root at $i_0^{\mathrm{pe}} = 2$).* The bound implies the stated $\widetilde{\mathcal{O}}\big(K\log T + \sqrt{KT\log(2KT/\delta)}\big)$ form. In particular, the pathwise hypothesis covers any deterministic adversarial context design satisfying Remark 2.

Theorem 8 (Variant: stochastic contexts under conditional windowed excitation)**.** *Suppose only the conditional form* (23) *holds and contexts are genuinely stochastic. Then the confidence sequence of Proposition 2 remains valid with the single modification that the leading coefficient of $1/\sqrt{\tau+\tau_0^{\mathrm{pe}}}$ is increased by the additive term*

$$\sqrt{1+\frac{w}{\tau_0^{\mathrm{pe}}}}\;\frac{8\sqrt{2}\,L^2\big(2S+\sqrt{c^{(2)}}\big)\sqrt{w\,\log(2n/\delta)}}{\mu}, \tag{25}$$

and Theorem 3 remains valid with the regret increased by

$$\sqrt{1+\frac{w}{\tau_0^{\mathrm{pe}}}}\;\frac{32\sqrt{2}\,L^3\big(2S+\sqrt{c^{(2)}}\big)\sqrt{w\log(2KT/\delta)}}{\mu}\sqrt{KT}.$$

The square-root factor in (25) is the same endpoint-to-interior promotion as in Proposition 2.

Proposition 6 (One-arm contraction under windowed excitation)**.** *Fix an arm k, let $e_{k,\tau} = z_{k,\tau} - \theta_k$, and run the update* (5) *with the step-size schedule* (21)*. Under the conditional windowed excitation* (23)*, for every window index $i \ge 0$,*

$$\mathbb{E}\|e_{k,iw}\|_2^2 \le \frac{8G^2}{\kappa_{\mathrm{pe}}^2\big(iw+\tau_0^{\mathrm{pe}}\big)} + \frac{Q_w}{\big(iw+\tau_0^{\mathrm{pe}}\big)^2}, \tag{26}$$

with Q_w the burn-in constant of (22)*.*

The leading constant $8G^2/\kappa_{\mathrm{pe}}^2$ is independent of w: the window length enters the first-order term only through the burn-in $\tau_0^{\mathrm{pe}} = i_0^{\mathrm{pe}} w$, while the within-window movement forcing is correctly demoted to the $O\big((iw+\tau_0^{\mathrm{pe}})^{-2}\big)$ second-order term. The leading constant matches the second branch $L^2/(\kappa_{\mathrm{dir}}^2\bar\gamma^2)$ of D_{Cau} of Proposition 5 up to a factor 8 whenever $\mu = \mu_{\mathrm{dir}}$ and that branch is the dominant one (i.e. $4S^2\tau_0 \le G^2/\kappa_{\mathrm{pe}}^2$).

Proposition 2 (full form). *Let $\delta \in (0,1)$. Fix an arm k and run the update* (5) *with the step-size schedule* (21)*. Under the pathwise windowed excitation* (10)*, with probability at least $1-\delta$, simultaneously for all local times $\tau \in \{0,\dots,n\}$,*

$$\|z_{k,\tau} - \theta_k\|_2 \le \beta_\tau^{\mathrm{pe}} := \Big(1+\frac{w}{\tau_0^{\mathrm{pe}}}\Big)\frac{A_w}{\tau+\tau_0^{\mathrm{pe}}} + \sqrt{1+\frac{w}{\tau_0^{\mathrm{pe}}}}\;\frac{\big(16+32\sqrt{\log(2n/\delta)}\big)\,G}{\kappa_{\mathrm{pe}}\sqrt{\tau+\tau_0^{\mathrm{pe}}}}, \tag{27}$$

where, with $c^{(2)}, c^{(3)}$ from (22)*,*

$$A_w := 2S\tau_0^{\mathrm{pe}} + \frac{2G\sqrt{\tau_0^{\mathrm{pe}}}}{\kappa_{\mathrm{pe}}} + \frac{16L^2G\,w^{3/2}}{\kappa_{\mathrm{pe}}\,\mu\,\sqrt{i_0^{\mathrm{pe}}}} + \frac{32Gw\sqrt{\big(1+\log\frac{n+\tau_0^{\mathrm{pe}}}{w}\big)\log(2n/\delta)}}{\kappa_{\mathrm{pe}}} + \frac{2Gw}{\kappa_{\mathrm{pe}}}. \tag{28}$$

Consequently $|y^\top(z_{k,\tau}-\theta_k)| \le \beta_\tau^{\mathrm{pe}}\|y\|_2$ *for every context* y. With the choice $i_0^{\mathrm{pe}}=2$ of (21), the promotion factors are $1+w/\tau_0^{\mathrm{pe}}=3/2$ and $\sqrt{1+w/\tau_0^{\mathrm{pe}}}=\sqrt{3/2}$; in particular, the pathwise form covers arbitrary deterministic context designs (Remark 2).

The leading $1/\sqrt{\tau+\tau_0^{\mathrm{pe}}}$ coefficient is independent of w: the window length appears only in the burn-in τ_0^{pe}, in the $O(1/(\tau+\tau_0^{\mathrm{pe}}))$ constant A_w (polynomially, as $w^{3/2}$), and in the constant promotion factor $\sqrt{1+w/\tau_0^{\mathrm{pe}}}=\sqrt{3/2}$ paid in Step 5 for extending the bound from window endpoints to all local times. Relative to Proposition 1 the leading constants degrade by explicit factors (16 versus 2; 32 versus $4\sqrt{2}$; and the further $\sqrt{3/2}$), originating in the smoothing choice $\bar{c}=c^{(2)}$, the factor $(m_i+1)^2 \le 4m_i^2$, the window sum $\sum_i m_i^{-1/2} \le 2\sqrt{m_I}$, and the endpoint-to-interior promotion (the factor-$1/2$ discount paid for within-window iterate movement enters only through $c^{(3)}$).

Remark 3 (Where the window length does and does not enter)**.** *Under the conditional assumption* (23)*, the realized excitation* $\sum_{j=1}^{w}(y_{k,\tau+j}^\top v)^2$ *may deviate from its* $\mathcal{F}_{k,\tau}$*-conditional mean by up to* $2wL^2\|v\|_2^2$*, and the assumption provides no within-window martingale structure for this deviation: conditioning is only at the window start. The deviation therefore enters the confidence recursion as a window-indexed martingale increment. Its pathwise range is* w*-free after the step-size discount,* $|\Delta_i^\chi| \le 8L^2\big(2S+\sqrt{c^{(2)}}\big)/\mu$*, because* $\eta_i \propto 1/w$ *cancels the* $O(w)$ *range of the raw excitation deviation. The* $\sqrt{w}$ *factor in* (25) *comes instead from Azuma accumulation over the* $I=\lfloor\tau/w\rfloor$ *windows, followed by the normalization by* $m_I \propto \tau/w$*. For deterministic designs (Remark 2) the deviation vanishes identically, and the rate is exactly that of Theorem 1 with* w*-independent leading constants. Whether the* $\sqrt{w}$ *factor in the stochastic variant is removable is open: a multiplicative (exponential-martingale) analysis suggests the true effect of the excitation fluctuation on the error is an* $O(1)$ *factor, but the linearized potential argument used here cannot certify that.*

Proofs. The following lemma is the pathwise window recursion at the heart of both propositions. We suppress the arm index k throughout; all local times, filtrations, and iterates refer to the fixed arm.

Lemma 3 (Window recursion)**.** *Fix an arm and let* $a_\tau := \|e_\tau\|_2^2$*,* $m_i := i + i_0^{\mathrm{pe}}$*. With the step-size schedule* (21)*, for every window* $i \ge 0$ *and every realization,*

$$a_{(i+1)w} \le \Big(1-\frac{2}{m_i}\Big)a_{iw} + \frac{c^{(2)}}{m_i^2} + \frac{c^{(3)}}{m_i^3} - 2\eta_i\sum_{j=1}^{w}\psi_{iw+j}, \tag{29}$$

where $\eta_i := 2/(\kappa_{\mathrm{pe}} w m_i)$*, and*

$$\psi_{s+1} := \big\langle e_s,\, g_{s+1}\big\rangle - \widetilde{h}_{s+1}, \qquad \widetilde{h}_{s+1} := \frac{2\,(y_{s+1}^\top e_s)^2}{(y_{s+1}^\top e_s)^2+\Gamma^2}, \qquad g_{s+1} := \varphi_{\bar\gamma}(y_{s+1}^\top z_s - r_{s+1})\,y_{s+1},$$

is a martingale difference: $\mathbb{E}[\psi_{s+1} \mid \mathcal{F}_s]=0$*, bounded pathwise by* $|\psi_{s+1}| \le 2G\|e_s\|_2$*. Under the pathwise condition* (10) *the recursion holds as stated; under only the conditional condition* (23) *the same recursion holds after replacing the drift term by its conditional expectation, i.e. taking* $\mathbb{E}[\,\cdot \mid \mathcal{F}_{iw}]$ *of both sides.*

Proof. Step 1 (per-step pathwise recursion). Since $\theta_k \in \mathcal{X}$, nonexpansiveness of $\Pi_{\mathcal{X}}$ gives

$$a_{s+1} \le \|e_s - \eta_{i(s)} g_{s+1}\|_2^2 = a_s - 2\eta_{i(s)}\langle e_s, g_{s+1}\rangle + \eta_{i(s)}^2\|g_{s+1}\|_2^2.$$

Since $|\varphi_{\bar\gamma}(u)| = 2|u|/(\bar\gamma^2+u^2) \le 1/\bar\gamma$ and $\|y_{s+1}\|_2 \le L$, we have $\|g_{s+1}\|_2 \le L/\bar\gamma = G$ pathwise.

Step 2 (score decomposition). Write $r_{s+1} = y_{s+1}^\top\theta_k + \xi_{s+1}$, so $y_{s+1}^\top z_s - r_{s+1} = y_{s+1}^\top e_s - \xi_{s+1}$. Conditionally on $\mathcal{F}_s$ and y_{s+1}, Lemma 1 with $\Delta = y_{s+1}^\top e_s$ gives

$$\mathbb{E}[\langle e_s, g_{s+1}\rangle \,|\, \mathcal{F}_s, y_{s+1}] = \frac{2(y_{s+1}^\top e_s)^2}{(y_{s+1}^\top e_s)^2+\Gamma^2} = \widetilde{h}_{s+1}.$$

(For deterministic contexts y_{s+1} is $\mathcal{F}_s$-measurable and this is Lemma 1 directly.) Hence $\mathbb{E}[\psi_{s+1} \mid \mathcal{F}_s]=0$ by the tower property. The bound $|\psi_{s+1}| \le 2G\|e_s\|_2$ follows from $|\langle e_s, g_{s+1}\rangle| \le G\|e_s\|_2$ and $|\widetilde{h}_{s+1}| \le |y_{s+1}^\top e_s|\cdot\max_u \frac{2|u|}{u^2+\Gamma^2} = |y_{s+1}^\top e_s|/\Gamma \le G\|e_s\|_2$, using $\Gamma \ge \bar\gamma$. Moreover $|y_{s+1}^\top e_s| \le L\cdot 2S = B$ pathwise (projection keeps $z_s \in \mathcal{X}$), so

$$\widetilde{h}_{s+1} \ge \frac{2(y_{s+1}^\top e_s)^2}{B^2+\Gamma^2} = C_{\mathrm{Cau}}\,(y_{s+1}^\top e_s)^2 \qquad \text{pathwise.} \tag{30}$$

Step 3 (movement within a window). Within window i the step size is constant, $\eta_{i(s)} = \eta_i$. Since $z_s \in \mathcal{X}$ and $\|\Pi_{\mathcal{X}}(x) - z_s\|_2 \le \|x - z_s\|_2$, each update moves the iterate by at most $\eta_i G$; hence for $s = iw + j - 1$, $1 \le j \le w$,

$$\|e_s - e_{iw}\|_2 \le (j-1)\,\eta_i G \qquad \text{pathwise.}$$

Using $(a+b)^2 \ge a^2/2 - b^2$ with $a = y_{s+1}^\top e_{iw}$ and $b = y_{s+1}^\top (e_s - e_{iw})$, and $(y_{s+1}^\top (e_s - e_{iw}))^2 \le L^2 (j-1)^2 \eta_i^2 G^2$ (Cauchy–Schwarz and $\|y_{s+1}\|_2 \le L$),

$$(y_{s+1}^\top e_s)^2 \ge \tfrac{1}{2}\,(y_{s+1}^\top e_{iw})^2 - L^2 (j-1)^2 \eta_i^2 G^2.$$

Summing over $j = 1, \dots, w$ and using $\sum_{j=1}^{w} (j-1)^2 = \frac{(w-1)w(2w-1)}{6} \le \frac{w^3}{3}$,

$$\sum_{j=1}^{w} (y_{iw+j}^\top e_{iw+j-1})^2 \ge \frac{1}{2} \sum_{j=1}^{w} (y_{iw+j}^\top e_{iw})^2 - \frac{L^2 \eta_i^2 G^2 w^3}{3} \qquad \text{pathwise.} \tag{31}$$

Step 4 (window recursion). Summing Step 1 over the window, inserting $\langle e_s, g_{s+1} \rangle = \widetilde{h}_{s+1} + \psi_{s+1}$ and (30),

$$a_{(i+1)w} \le a_{iw} - 2\eta_i C_{\mathrm{Cau}} \sum_{j=1}^{w} (y_{iw+j}^\top e_{iw+j-1})^2 - 2\eta_i \sum_{j=1}^{w} \psi_{iw+j} + \eta_i^2 w G^2.$$

Under (10), apply (31) and then (10) with the $\mathcal{F}_{iw}$-measurable vector $v = e_{iw}$:

$$\sum_{j=1}^{w} (y_{iw+j}^\top e_{iw+j-1})^2 \ge \frac{\mu w}{2}\, a_{iw} - \frac{L^2 \eta_i^2 G^2 w^3}{3} \qquad \text{pathwise.}$$

Therefore, using $\kappa_{\mathrm{pe}} = \mu C_{\mathrm{Cau}}$,

$$a_{(i+1)w} \le \big(1 - \eta_i \kappa_{\mathrm{pe}} w\big) a_{iw} + \frac{2\eta_i^3 C_{\mathrm{Cau}} L^2 G^2 w^3}{3} + \eta_i^2 w G^2 - 2\eta_i \sum_{j=1}^{w} \psi_{iw+j}.$$

Substituting $\eta_i = 2/(\kappa_{\mathrm{pe}} w m_i)$: the contraction coefficient is $1 - \frac{2}{m_i}$ (nonnegative since $m_i \ge i_0^{\mathrm{pe}} \ge 2$); the noise forcing is $\eta_i^2 w G^2 = \frac{4G^2}{\kappa_{\mathrm{pe}}^2 w\, m_i^2} = \frac{c^{(2)}}{m_i^2}$; the movement forcing is $\frac{2\eta_i^3 C_{\mathrm{Cau}} L^2 G^2 w^3}{3} = \frac{16 L^2 G^2}{3 \kappa_{\mathrm{pe}}^2 \mu\, m_i^3} = \frac{c^{(3)}}{m_i^3}$, where we used $\kappa_{\mathrm{pe}}^2 \mu = C_{\mathrm{Cau}}^2 \mu^3$. This is (29). Under only (23), take $\mathbb{E}[\cdot \mid \mathcal{F}_{iw}]$ instead: $\mathbb{E}[\psi_{s+1} \mid \mathcal{F}_{iw}] = 0$ by the tower property (since $\mathcal{F}_{iw} \subseteq \mathcal{F}_s$), and (23) with $v = e_{iw}$ applied to the conditional expectation of (31) yields the same deterministic terms. □

Proof of Proposition 6. Let $A_i := \mathbb{E}[a_{iw}]$. Taking expectations in Lemma 3 (conditional form),

$$A_{i+1} \le \Big(1 - \frac{2}{m_i}\Big) A_i + \frac{c^{(2)}}{m_i^2} + \frac{c^{(3)}}{m_i^3}.$$

A single-scale induction $A_i \le D/m_i$ would already work with a constant satisfying $D \ge \max_{m \ge i_0^{\mathrm{pe}}} \big(c^{(2)} + c^{(3)}/m\big) \frac{m+1}{m+2}$ (a sufficient choice being $D \ge 2c^{(2)} + 2c^{(3)}/i_0^{\mathrm{pe}}$); the reason to avoid it is that $c^{(3)} \propto 1/\mu$ is w-free while $c^{(2)} \propto 1/w$, so a single-scale constant $D \gtrsim c^{(3)}$ would push a w-free amount into the leading term, which after conversion to local time ($m_i = (iw + \tau_0^{\mathrm{pe}})/w$) would make the leading constant grow with w. The two-scale form instead demotes the $c^{(3)}$ part to the w^2-scaled second-order term. We therefore induct on the two-scale claim

$$A_i \le \frac{P}{m_i} + \frac{Q}{m_i^2}, \qquad P := 2c^{(2)}, \qquad Q = \max\Big\{4S^2 \big(i_0^{\mathrm{pe}}\big)^2,\ 2c^{(3)} i_0^{\mathrm{pe}},\ \frac{\big(c^{(2)} + c^{(3)}\big)^2}{6c^{(2)}}\Big\},$$

with Q from (22). Base case: $A_0 = a_0 \le (2S)^2 \le Q/(i_0^{\mathrm{pe}})^2$ since $z_0, \theta_k \in \mathcal{X}$ and $Q \ge 4S^2 (i_0^{\mathrm{pe}})^2$. Inductive step: assume the claim at window i and write $m := m_i \ge i_0^{\mathrm{pe}} \ge 2$. Since $1 - \frac{2}{m} \ge 0$, the recursion and the inductive hypothesis give

$$A_{i+1} \le \Big(1 - \frac{2}{m}\Big)\Big(\frac{P}{m} + \frac{Q}{m^2}\Big) + \frac{c^{(2)}}{m^2} + \frac{c^{(3)}}{m^3},$$

so $A_{i+1} \le \frac{P}{m+1} + \frac{Q}{(m+1)^2}$ follows once

$$\frac{P(m+2)}{m^2(m+1)} + \frac{Q(3m+2)}{m^3(m+1)^2} \ge \frac{c^{(2)}}{m^2} + \frac{c^{(3)}}{m^3}, \tag{32}$$

where the left-hand terms are exactly $\frac{P}{m+1} - (1-\frac{2}{m})\frac{P}{m}$ and $\frac{Q}{(m+1)^2} - (1-\frac{2}{m})\frac{Q}{m^2}$, using $m^2 - (m-2)(m+1) = m+2$ and $m^3 - (m-2)(m+1)^2 = 3m+2$. Multiply (32) by m^3 and insert $P = 2c^{(2)}$ together with $\frac{m(m+2)}{m+1} = m + \frac{m}{m+1} \ge m + \frac{1}{2}$ and $\frac{3m+2}{(m+1)^2} \ge \frac{3}{2(m+1)}$ (both for $m \ge 1$):

$$\frac{Pm(m+2)}{m+1} + \frac{Q(3m+2)}{(m+1)^2} \ge 2c^{(2)}m + c^{(2)} + \frac{3Q}{2(m+1)},$$

so, comparing with the right-hand side $c^{(2)}m + c^{(3)}$ of the multiplied (32), it suffices that

$$c^{(2)}(m+1) + \frac{3Q}{2(m+1)} \ge c^{(3)} \qquad \text{for all } m \ge i_0^{\mathrm{pe}}.$$

Set $u := m+1 \ge i_0^{\mathrm{pe}} + 1$ and $h(u) := c^{(2)}u + \frac{3Q}{2u}$. If the unconstrained minimizer $u_\star = \sqrt{3Q/(2c^{(2)})}$ satisfies $u_\star \ge i_0^{\mathrm{pe}} + 1$, then $h(u) \ge h(u_\star) = \sqrt{6Qc^{(2)}} \ge c^{(2)} + c^{(3)} \ge c^{(3)}$, using $Q \ge (c^{(2)} + c^{(3)})^2/(6c^{(2)})$. Otherwise h is increasing on $[i_0^{\mathrm{pe}} + 1, \infty)$ and $h(u) \ge h(i_0^{\mathrm{pe}} + 1) \ge \frac{3Q}{2(i_0^{\mathrm{pe}}+1)} \ge \frac{3c^{(3)}i_0^{\mathrm{pe}}}{i_0^{\mathrm{pe}}+1} \ge c^{(3)}$, using $Q \ge 2c^{(3)}i_0^{\mathrm{pe}}$ and $3i_0^{\mathrm{pe}} \ge i_0^{\mathrm{pe}} + 1$. This completes the induction.

Finally, substituting $m_i = (iw + \tau_0^{\mathrm{pe}})/w$,

$$\mathbb{E}\|e_{k,iw}\|_2^2 \le \frac{Pw}{iw + \tau_0^{\mathrm{pe}}} + \frac{Qw^2}{(iw + \tau_0^{\mathrm{pe}})^2} = \frac{8G^2}{\kappa_{\mathrm{pe}}^2\,(iw + \tau_0^{\mathrm{pe}})} + \frac{Q_w}{(iw + \tau_0^{\mathrm{pe}})^2},$$

since $Pw = 2c^{(2)}w = 8G^2/\kappa_{\mathrm{pe}}^2$ exactly and $Q_w = Qw^2$. (For τ inside a window, $\|e_\tau\|_2 \le \|e_{i(\tau)w}\|_2 + w\eta_{i(\tau)}G$ pathwise, so the same $O(1/\tau)$ rate holds at all local times up to constant factors.) □

Proof of Proposition 2. We linearize the squared-norm recursion of Lemma 3 around the smoothed potential used in the proof of Proposition 1, now indexed by windows. Set

$$\bar{c} := c^{(2)}, \qquad s_i := \sqrt{a_{iw} + \frac{\bar{c}}{m_i}}, \qquad q_i := m_i s_i, \qquad \text{so that} \qquad s_i \ge \sqrt{\bar{c}/m_i}.$$

Step 1 (smoothed window recursion). From Lemma 3 and $a_{iw} = s_i^2 - \bar{c}/m_i$,

$$s_{i+1}^2 = a_{(i+1)w} + \frac{\bar{c}}{m_i+1} \le \frac{m_i - 2}{m_i}\Big(s_i^2 - \frac{\bar{c}}{m_i}\Big) + \frac{c^{(2)}}{m_i^2} + \frac{c^{(3)}}{m_i^3} + \frac{\bar{c}}{m_i+1} + \mathcal{M}_i,$$
$$\mathcal{M}_i := -2\eta_i \sum_{j=1}^{w} \psi_{iw+j}.$$

Since $\frac{m_i-2}{m_i} \le \frac{m_i^2}{(m_i+1)^2}$ (equivalently $(m_i-2)(m_i+1)^2 = m_i^3 - 3m_i - 2 \le m_i^3$) and $s_i^2 \ge \bar{c}/m_i$,

$$\frac{m_i - 2}{m_i} s_i^2 = \frac{m_i^2}{(m_i+1)^2} s_i^2 - \frac{(3m_i+2)}{m_i(m_i+1)^2}\, s_i^2 \le \frac{m_i^2}{(m_i+1)^2} s_i^2 - \frac{(3m_i+2)\bar{c}}{m_i^2(m_i+1)^2}.$$

The total $\bar{c}$-coefficient is

$$-\frac{(m_i-2)\bar{c}}{m_i^2} + \frac{\bar{c}}{m_i+1} - \frac{(3m_i+2)\bar{c}}{m_i^2(m_i+1)^2} = \frac{\bar{c}}{(m_i+1)^2} \le \frac{\bar{c}}{m_i^2},$$

where the equality follows from $-(m_i-2)(m_i+1)^2 + m_i^2(m_i+1) - (3m_i+2) = m_i^2$ after collecting all summands over the common denominator $m_i^2(m_i+1)^2$ (expanding $-(m_i-2)(m_i+1)^2 = -m_i^3 + 3m_i + 2$). Hence, recalling $\bar{c} = c^{(2)}$,

$$s_{i+1}^2 \le \frac{m_i^2}{(m_i+1)^2} s_i^2 + \frac{2c^{(2)}}{m_i^2} + \frac{c^{(3)}}{m_i^3} + \mathcal{M}_i.$$

Apply the concave tangent bound $\sqrt{x+r} \le \sqrt{x} + \frac{r}{2\sqrt{x}}$ at $x = \frac{m_i^2 s_i^2}{(m_i+1)^2} > 0$ (valid since the right-hand side above is nonnegative), then multiply by $m_i + 1$:

$$q_{i+1} \le q_i + \underbrace{\frac{(m_i+1)^2\big(2c^{(2)}/m_i^2 + c^{(3)}/m_i^3\big)}{2m_i s_i}}_{\text{deterministic drift}} + \underbrace{\frac{(m_i+1)^2}{2m_i s_i}\,\mathcal{M}_i}_{\text{noise}}.$$

Step 2 (deterministic drift sum). Using $1/s_i \le \sqrt{m_i/\bar{c}}$ and $(m_i+1)^2 \le 4m_i^2$ for $m_i \ge 1$,

$$\frac{(m_i+1)^2\,2c^{(2)}}{2m_i^3 s_i} \le \frac{4c^{(2)}}{m_i s_i} \le \frac{4\sqrt{c^{(2)}}}{\sqrt{m_i}}, \qquad \frac{(m_i+1)^2 c^{(3)}}{2m_i^4 s_i} \le \frac{2c^{(3)}}{m_i^2 s_i} \le \frac{2c^{(3)}}{\sqrt{c^{(2)}}\,m_i^{3/2}}.$$

Summing over windows $i = 0, \dots, I-1$ with $\sum_i m_i^{-1/2} \le 2\sqrt{m_I}$ and $\sum_i m_i^{-3/2} \le 3\big(i_0^{\mathrm{pe}}\big)^{-1/2}$,

$$q_I \le q_0 + 8\sqrt{c^{(2)} m_I} + \frac{6c^{(3)}}{\sqrt{c^{(2)} i_0^{\mathrm{pe}}}} + \sum_{i=0}^{I-1} \frac{(m_i+1)^2}{2m_i s_i}\mathcal{M}_i.$$

Step 3 (noise: a step-level martingale). The accumulated noise expands as a sum over *individual local steps*,

$$W_t := \sum_{s=0}^{t-1} (-c_s\,\psi_{s+1}), \qquad c_s := \frac{(m_{i(s)}+1)^2\,\eta_{i(s)}}{m_{i(s)}\,s_{i(s)}},$$

and the noise term in q_I is exactly W_{Iw}. For each step s in window i, the coefficient c_s is $\mathcal{F}_{iw}$-measurable (hence $\mathcal{F}_s$-measurable), so by Lemma 3 the sequence $\{-c_s\psi_{s+1}\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_s\}$. It is bounded *pathwise*: since $|\psi_{s+1}| \le 2G\|e_s\|_2$ and, by the movement bound of Lemma 3 Step 3, $\|e_s\|_2 \le \|e_{iw}\|_2 + w\eta_i G \le s_i + w\eta_i G$,

$$|c_s\psi_{s+1}| \le \frac{(m_i+1)^2\eta_i}{m_i}\,2G\Big(1 + \frac{w\eta_i G}{s_i}\Big) \le \frac{16G}{\kappa_{\mathrm{pe}} w}\Big(1 + \sqrt{\frac{w}{m_i}}\Big) =: b_s,$$

where we used $\frac{(m_i+1)^2\eta_i}{m_i} \le \frac{4m_i\cdot 2}{\kappa_{\mathrm{pe}} w m_i} = \frac{8}{\kappa_{\mathrm{pe}} w}$ and $\frac{w\eta_i G}{s_i} \le w\eta_i G\sqrt{\frac{m_i}{c^{(2)}}} = \frac{2G\sqrt{m_i}}{\kappa_{\mathrm{pe}}\sqrt{m_i c^{(2)}}} = \sqrt{\frac{w}{m_i}}$ (the last equality is the definition $c^{(2)} = \frac{4G^2}{\kappa_{\mathrm{pe}}^2 w}$). Summing the variance proxies over the Iw steps, with $H_I := \sum_{i=0}^{I-1} m_i^{-1} \le 1 + \log m_I$,

$$\sum_{s=0}^{Iw-1} b_s^2 \le \frac{512\,G^2}{\kappa_{\mathrm{pe}}^2 w^2} \sum_{i=0}^{I-1} w\Big(1 + \frac{w}{m_i}\Big) = \frac{512\,G^2}{\kappa_{\mathrm{pe}}^2 w^2}\big(wI + w^2 H_I\big),$$

using $(1+x)^2 \le 2(1+x^2)$. Azuma–Hoeffding applied to W and a union bound over the at most n window endpoints $t = Iw \le n$ give, with probability at least $1-\delta$, simultaneously for all I with $Iw \le n$,

$$W_{Iw} \le \sqrt{2\Big(\textstyle\sum_{s<Iw} b_s^2\Big)\log(2n/\delta)} \le \frac{32G}{\kappa_{\mathrm{pe}} w}\sqrt{\big(wI + w^2 H_I\big)\log(2n/\delta)},$$

Step 4 (unwrapping). Since $q_0 = i_0^{\mathrm{pe}} s_0 \le i_0^{\mathrm{pe}}\big(\|e_0\|_2 + \sqrt{\bar{c}/i_0^{\mathrm{pe}}}\big) \le 2Si_0^{\mathrm{pe}} + \sqrt{c^{(2)} i_0^{\mathrm{pe}}}$ (using $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ and $\|e_0\|_2 \le 2S$), for every window endpoint $\tau = Iw$,

$$\|e_\tau\|_2 \le s_I = \frac{q_I}{m_I} \le \frac{2Si_0^{\mathrm{pe}} + \sqrt{c^{(2)} i_0^{\mathrm{pe}}}}{m_I} + \frac{8\sqrt{c^{(2)}}}{\sqrt{m_I}} + \frac{6c^{(3)}}{\sqrt{c^{(2)} i_0^{\mathrm{pe}}}\,m_I} + \frac{32G\sqrt{(wI + w^2 H_I)\log(2n/\delta)}}{\kappa_{\mathrm{pe}} w\, m_I}.$$

Now substitute $m_I = (\tau + \tau_0^{\mathrm{pe}})/w$ and $wI = \tau$:

$$\frac{2Si_0^{\mathrm{pe}} + \sqrt{c^{(2)} i_0^{\mathrm{pe}}}}{m_I} = \frac{2S\tau_0^{\mathrm{pe}} + w\sqrt{c^{(2)} i_0^{\mathrm{pe}}}}{\tau + \tau_0^{\mathrm{pe}}} = \frac{2S\tau_0^{\mathrm{pe}} + \frac{2G\sqrt{\tau_0^{\mathrm{pe}}}}{\kappa_{\mathrm{pe}}}}{\tau + \tau_0^{\mathrm{pe}}},$$

$$\frac{8\sqrt{c^{(2)}}}{\sqrt{m_I}} = \frac{8\sqrt{c^{(2)} w}}{\sqrt{\tau + \tau_0^{\mathrm{pe}}}} = \frac{16G}{\kappa_{\mathrm{pe}}\sqrt{\tau + \tau_0^{\mathrm{pe}}}},$$

$$\frac{6c^{(3)}}{\sqrt{c^{(2)} i_0^{\mathrm{pe}}}\, m_I} = \frac{6c^{(3)} w}{\sqrt{c^{(2)} i_0^{\mathrm{pe}}}\,(\tau + \tau_0^{\mathrm{pe}})} = \frac{16L^2 G\, w^{3/2}}{\kappa_{\mathrm{pe}}\,\mu\,\sqrt{i_0^{\mathrm{pe}}}\,(\tau + \tau_0^{\mathrm{pe}})},$$

$$\frac{32G\sqrt{wI\log(2n/\delta)}}{\kappa_{\mathrm{pe}} w\, m_I} = \frac{32G\sqrt{\tau\log(2n/\delta)}}{\kappa_{\mathrm{pe}}\,(\tau + \tau_0^{\mathrm{pe}})} \le \frac{32G\sqrt{\log(2n/\delta)}}{\kappa_{\mathrm{pe}}\sqrt{\tau + \tau_0^{\mathrm{pe}}}},$$

$$\frac{32G\sqrt{w^2 H_I \log(2n/\delta)}}{\kappa_{\mathrm{pe}} w\, m_I} = \frac{32G\sqrt{H_I \log(2n/\delta)}}{\kappa_{\mathrm{pe}}\, m_I} \le \frac{32Gw\sqrt{\left(1 + \log\frac{n+\tau_0^{\mathrm{pe}}}{w}\right)\log(2n/\delta)}}{\kappa_{\mathrm{pe}}\,(\tau + \tau_0^{\mathrm{pe}})},$$

where the last line uses $H_I \le 1 + \log m_I = 1 + \log\frac{\tau+\tau_0^{\mathrm{pe}}}{w} \le 1 + \log\frac{n+\tau_0^{\mathrm{pe}}}{w}$. Note that the two leading $1/\sqrt{\tau + \tau_0^{\mathrm{pe}}}$ coefficients, $16G/\kappa_{\mathrm{pe}}$ and $32G\sqrt{\log(2n/\delta)}/\kappa_{\mathrm{pe}}$, are independent of w.

Step 5 (all local times). For $\tau \in (Iw, (I+1)w]$, the movement bound gives pathwise

$$\|e_\tau\|_2 \le s_I + w\eta_I G, \qquad w\eta_I G = \frac{2G}{\kappa_{\mathrm{pe}}\, m_I} = \frac{2Gw}{\kappa_{\mathrm{pe}}(Iw + \tau_0^{\mathrm{pe}})}.$$

The Step-4 bound at the window endpoint has the form $D_0/(Iw + \tau_0^{\mathrm{pe}}) + C_0/\sqrt{Iw + \tau_0^{\mathrm{pe}}}$ with $C_0 = \big(16 + 32\sqrt{\log(2n/\delta)}\big)G/\kappa_{\mathrm{pe}}$ and with

$$D_0 := 2S\tau_0^{\mathrm{pe}} + \frac{2G\sqrt{\tau_0^{\mathrm{pe}}}}{\kappa_{\mathrm{pe}}} + \frac{16L^2 G\, w^{3/2}}{\kappa_{\mathrm{pe}}\,\mu\,\sqrt{i_0^{\mathrm{pe}}}} + \frac{32Gw\sqrt{\left(1 + \log\frac{n+\tau_0^{\mathrm{pe}}}{w}\right)\log(2n/\delta)}}{\kappa_{\mathrm{pe}}},$$

the first four terms of (28), so that $D_0 + 2Gw/\kappa_{\mathrm{pe}} = A_w$ (the last term of (28) is exactly the movement allowance above). Hence

$$\|e_\tau\|_2 \;\le\; \frac{A_w}{Iw + \tau_0^{\mathrm{pe}}} + \frac{C_0}{\sqrt{Iw + \tau_0^{\mathrm{pe}}}}.$$

This window-start quantity exceeds its re-evaluation at τ; to re-express it at the argument τ we pay the ratio $c := \frac{\tau+\tau_0^{\mathrm{pe}}}{Iw+\tau_0^{\mathrm{pe}}} \in \big[1,\, 1 + w/\tau_0^{\mathrm{pe}}\big]$ (since $\tau \le Iw + w$), which gives

$$\frac{1}{Iw + \tau_0^{\mathrm{pe}}} = \frac{c}{\tau + \tau_0^{\mathrm{pe}}} \le \frac{1 + w/\tau_0^{\mathrm{pe}}}{\tau + \tau_0^{\mathrm{pe}}}, \qquad \frac{1}{\sqrt{Iw + \tau_0^{\mathrm{pe}}}} \le \frac{\sqrt{1 + w/\tau_0^{\mathrm{pe}}}}{\sqrt{\tau + \tau_0^{\mathrm{pe}}}}.$$

Collecting terms yields $\|e_\tau\|_2 \le \beta_\tau^{\mathrm{pe}}$ for every $\tau \in \{0, \dots, n\}$, with β_τ^{pe} as in (27)–(28): the $1/\sqrt{\tau + \tau_0^{\mathrm{pe}}}$ part contributes the coefficient $\sqrt{1 + w/\tau_0^{\mathrm{pe}}}\,\big(16 + 32\sqrt{\log(2n/\delta)}\big)G/\kappa_{\mathrm{pe}}$, and the $1/(\tau + \tau_0^{\mathrm{pe}})$ part is bounded by $\big(1 + w/\tau_0^{\mathrm{pe}}\big)A_w$. The prediction bound follows from $|y^\top e_\tau| \le \|y\|_2\|e_\tau\|_2$. The promotion factors cannot be reduced within this argument: at $(I, \tau) = (0, w)$ the chain holds with equality, since $1/\tau_0^{\mathrm{pe}} = \big(1 + w/\tau_0^{\mathrm{pe}}\big)/\big(w + \tau_0^{\mathrm{pe}}\big)$ identically (for every $\tau_0^{\mathrm{pe}} > 0$). □

Proof of Theorem 3. Apply Proposition 2 to each arm with $n = T$ and failure level δ/K and take a union bound over the K arms: with probability $1-\delta$, $|y_{k,t}^\top(z_{k,T_k(t-1)} - \theta_k)| \le \beta^{\mathrm{pe}}_{T_k(t-1)}\|y_{k,t}\|_2$ holds for all arms and rounds. The schedule β_τ^{pe} is deterministic and non-increasing in τ, so Proposition 3 (which is unchanged) gives

$$R_{\mathrm{loc}}(T) \le 2L \sum_{k=1}^{K} \sum_{\tau=0}^{T_k - 1} \beta_\tau^{\mathrm{pe}}.$$

The harmonic part is bounded by $3LKA'_w(1+\log(T+\tau_0^{\mathrm{pe}}))$ using $\sum_{\tau=0}^{T_k-1}\frac{1}{\tau+\tau_0^{\mathrm{pe}}} \le 1+\log(T+\tau_0^{\mathrm{pe}})$ and the promotion factor $1+w/\tau_0^{\mathrm{pe}} = 3/2$. The square-root part is bounded by

$$2L\,\frac{\sqrt{3/2}\,(16+32\sqrt{\log(2KT/\delta)})G}{\kappa_{\mathrm{pe}}}\sum_{k=1}^{K}2\sqrt{T_k} \le \frac{\sqrt{3/2}\,4LG(16+32\sqrt{\log(2KT/\delta)})}{\kappa_{\mathrm{pe}}}\sqrt{KT},$$

using $\sum_{\tau=0}^{T_k-1}(\tau+\tau_0^{\mathrm{pe}})^{-1/2} \le 2\sqrt{T_k}$ and $\sum_k\sqrt{T_k} \le \sqrt{KT}$; substituting $G = L/\bar{\gamma}$ gives (24). □

Proof of Theorem 8. Under only (23), Step 4 of Lemma 3 is modified at a single point: the pathwise lower bound on $\sum_j (y_{iw+j}^\top e_{iw})^2$ is replaced by

$$\sum_{j=1}^{w}(y_{iw+j}^\top e_{iw})^2 = \mathbb{E}\Big[\sum_{j=1}^{w}(y_{iw+j}^\top e_{iw})^2 \,\Big|\, \mathcal{F}_{iw}\Big] + \mathcal{X}_i \ge \mu w\, a_{iw} + \mathcal{X}_i,$$

where the inequality is (23) with $v = e_{iw}$, and $\mathcal{X}_i := \sum_{j=1}^{w}\big[(y_{iw+j}^\top e_{iw})^2 - \mathbb{E}[(y_{iw+j}^\top e_{iw})^2 \mid \mathcal{F}_{iw}]\big]$ satisfies $\mathbb{E}[\mathcal{X}_i \mid \mathcal{F}_{iw}] = 0$ and, writing $Y_j := (y_{iw+j}^\top e_{iw})^2$, the pathwise bound

$$|\mathcal{X}_i| \le \sum_{j=1}^{w} Y_j + \sum_{j=1}^{w}\mathbb{E}[Y_j \mid \mathcal{F}_{iw}] \le 2wL^2\|e_{iw}\|_2^2 \le 2wL^2 s_i^2,$$

where the second inequality uses $Y_j \le L^2\|e_{iw}\|_2^2$ almost surely (Cauchy–Schwarz and $\|y_{iw+j}\|_2 \le L$) and $\mathbb{E}[Y_j \mid \mathcal{F}_{iw}] \le L^2\|e_{iw}\|_2^2$ almost surely (since e_{iw} is $\mathcal{F}_{iw}$-measurable). The window recursion (29) therefore holds with $\mathcal{M}_i$ replaced by $\mathcal{M}_i - \eta_i C_{\mathrm{Cau}}\mathcal{X}_i$. In the q-dynamics this contributes the additional window-indexed martingale increment

$$\Delta_i^{\chi} := -\frac{(m_i+1)^2\eta_i C_{\mathrm{Cau}}}{2m_i s_i}\,\mathcal{X}_i, \qquad |\Delta_i^{\chi}| \le \frac{(m_i+1)^2\eta_i C_{\mathrm{Cau}}\,2wL^2 s_i^2}{2m_i s_i}$$
$$\le \frac{8L^2 s_i}{\mu} \le \frac{8L^2\big(2S+\sqrt{c^{(2)}}\big)}{\mu} =: b_\chi,$$

using $\frac{(m_i+1)^2}{2m_i} \le 2m_i$, $\eta_i = 2/(\kappa_{\mathrm{pe}} w m_i)$, $\kappa_{\mathrm{pe}} = C_{\mathrm{Cau}}\mu$, and $s_i \le \|e_{iw}\|_2 + \sqrt{c^{(2)}/m_i} \le 2S + \sqrt{c^{(2)}}$ (projection keeps $z_{iw} \in \mathcal{X}$). This bound is pathwise. We allocate failure probability $\delta/2$ to the step-level martingale event of Proposition 2 (re-derived with $\delta \leftarrow \delta/2$) and $\delta/2$ to the window-level event below; each is obtained by Azuma–Hoeffding with per-point failure level $\delta/(2n)$ and a union bound over at most n window endpoints, which is exactly what the shared $\log(2n/\delta)$ factors encode. With the required probability, simultaneously for all I with $Iw \le n$,

$$\sum_{i=0}^{I-1}\Delta_i^{\chi} \le b_\chi\sqrt{2I\log(2n/\delta)}, \qquad \frac{b_\chi\sqrt{2I\log(2n/\delta)}}{m_I} \le \frac{b_\chi\sqrt{2\log(2n/\delta)}}{\sqrt{m_I}}$$
$$= \frac{8\sqrt{2}\,L^2\big(2S+\sqrt{c^{(2)}}\big)\sqrt{w\log(2n/\delta)}}{\mu\sqrt{\tau+\tau_0^{\mathrm{pe}}}},$$

which at the window endpoints is precisely the additive term of (25) up to the promotion factor; promoting from window endpoints to all local times multiplies it, as in Step 5 of Proposition 2, by $\sqrt{1+w/\tau_0^{\mathrm{pe}}}$, giving (25) exactly. The regret statement follows by applying Proposition 3 as in the proof of Theorem 3, with the extra contribution $\sqrt{1+w/\tau_0^{\mathrm{pe}}}\cdot 2L\cdot\frac{8\sqrt{2}L^2(2S+\sqrt{c^{(2)}})\sqrt{w\log(2KT/\delta)}}{\mu}\cdot 2\sqrt{KT}$. □

E Proofs for the Ellipsoidal Variant

This section constructs and analyses the excitation-free variant of the direct policy: the ball-confidence schedule of Proposition 1 is replaced by an ellipsoidal confidence set, so that no excitation condition on the context process is needed at all — the contexts may be adversarial and predictable, subject only to $\|y_{k,t}\|_2 \le L$. The main result is Theorem 4. We use the notation of Section 2 and the

standing notation of Appendix A; the exact Cauchy drift identity of Lemma 1 is the only probabilistic ingredient. The plan has three steps: Lemma 4 is a pathwise linearized-regret bound for the direct update; Lemma 5 converts it, through the drift identity, into a cumulative prediction-error bound; and Lemma 6 completes the square against the regularized design matrix, turning that bound into an ellipsoidal confidence set with a closed-form UCB index. The policy (Algorithm 3) is given first, then the full form of Theorem 4 with all constants and the three accompanying remarks, and then the proofs.

Fix an arm k and index its samples locally by $s = 1, 2, \dots$; write $z_{k,s}$ for the iterate used to score the s-th pull (so $z_{k,1} = \mathbf{0}$). The update producing the next iterate is

$$z_{k,s+1} = \Pi_{\mathcal{X}}(z_{k,s} - \eta_s\, g_{k,s})\,, \qquad g_{k,s} = \varphi_{\bar\gamma}(y_{k,s}^\top z_{k,s} - r_{k,s})\, y_{k,s}, \qquad \eta_s = \frac{D}{G\sqrt{2s}}, \tag{33}$$

where $D = 2S$ is the diameter of $\mathcal{X}$ and $G = L/\bar\gamma$; this is exactly the update of Eq. (5) with the schedule η_s made explicit. The confidence index of Algorithm 3 uses the cumulative prediction-error budget $\mathcal{B}_\tau$ of Eq. (34) below, an explicit quantity of order $\frac{(B^2+(\Gamma)^2)SL}{\bar\gamma}\sqrt{\tau\log(K\tau/\delta)}$ with $\mathcal{B}_0 := 0$. The conversion argument of Appendix E shows that, with probability at least $1-\delta$, the cumulative squared prediction error of each arm's iterates is bounded by $\mathcal{B}_\tau$ simultaneously for all arms and local times (Lemma 5), and that completing the square against the regularized design matrix turns this confidence set into an ellipsoid with a closed-form UCB index (Lemma 6). Algorithm 3 implements the resulting policy; per arm it maintains the iterate $z_{k,\cdot}$, the design matrix $V_{k,\cdot}$ and its inverse, and the vector $b_{k,\cdot}$, with rank-one (Sherman–Morrison) updates.

Algorithm 3 Direct-Cauchy-Ellipsoidal-UCB

Require: Score scale $\bar\gamma > 0$, parameter radius $S > 0$, context bound L, regularizer $\lambda > 0$ (default $\lambda = L^2$), failure level $\delta \in (0,1)$; $B = 2LS$, $D = 2S$, $G = L/\bar\gamma$.

Initialize $T_k(0) = 0$, $z_{k,1} = \mathbf{0}$, $V_{k,0} = \lambda I_d$, $V_{k,0}^{-1} = \lambda^{-1} I_d$, $b_{k,0} = \mathbf{0}$ for all $k \in [K]$.
for round $t = 1, \dots, T$ **do**
  Observe contextual features $\{y_{1,t}, \dots, y_{K,t}\} \subset \mathbb{R}^d$.
  for each arm $k \in [K]$ **do**
    Set $\tau = T_k(t-1)$, $\hat z = V_{k,\tau}^{-1} b_{k,\tau}$, and compute the ellipsoidal index

$$\mathrm{UCB}_{k,t} = y_{k,t}^\top \hat z + \sqrt{\beta_\tau^{\mathrm{ell}}}\,\sqrt{y_{k,t}^\top V_{k,\tau}^{-1} y_{k,t}}, \qquad \beta_\tau^{\mathrm{ell}} = \mathcal{B}_\tau + \lambda S^2 \ \text{ with } \mathcal{B}_\tau \text{ from Eq. (34).}$$

  end for
  Select action $k_t \in \arg\max_{k\in[K]} \mathrm{UCB}_{k,t}$.
  Observe reward r_t; set $s = T_{k_t}(t-1) + 1$ and $y = y_{k_t,t}$.
  Score-update the played arm (Eq. (33)): $z_{k_t,s+1} = \Pi_{\mathcal{X}}\big(z_{k_t,s} - \eta_s\, \varphi_{\bar\gamma}(y^\top z_{k_t,s} - r_t)\, y\big)$ with $\eta_s = D/(G\sqrt{2s})$.
  Rank-one design updates for the played arm:

$$V_{k_t,s} = V_{k_t,s-1} + yy^\top, \qquad V_{k_t,s}^{-1} = V_{k_t,s-1}^{-1} - \frac{V_{k_t,s-1}^{-1} y\, y^\top V_{k_t,s-1}^{-1}}{1 + y^\top V_{k_t,s-1}^{-1} y},$$

$$b_{k_t,s} = b_{k_t,s-1} + (y^\top z_{k_t,s})\, y.$$

  Update pull counts: $T_{k_t}(t) = T_{k_t}(t-1) + 1$ and $T_k(t) = T_k(t-1)$ for $k \neq k_t$.
end for

For $\tau \geq 1$ the cumulative prediction-error budget of Theorem 4 is

$$\mathcal{B}_\tau \;:=\; \frac{\big(B^2 + (\Gamma)^2\big)SL}{\bar\gamma}\,\sqrt{2\tau}\,\Big(1 + 2\sqrt{\ell_\tau}\Big), \qquad \ell_\tau := \log\frac{\pi^2 K\tau^2}{6\delta}, \tag{34}$$

and $\mathcal{B}_0 := 0$.

Theorem 4 (full form). *Let $\delta \in (0,1)$. In the exact Cauchy linear model of Section 2 with adversarial predictable contexts, $\|y_{k,t}\|_2 \leq L$, execute Algorithm 3. Then, on a single event of probability at*

least $1-\delta$ on which the confidence statement of Lemma 5 holds simultaneously for all arms and all local times $\tau \le T$,

$$R_{\mathrm{loc}}(T) \;\le\; \sqrt{2dKT\log\Big(1+\frac{TL^2}{d\lambda}\Big)}\Bigg[B+2S\sqrt{\lambda} + 2\left(\frac{\big(B^2+(\Gamma)^2\big)SL\sqrt{2}}{\bar\gamma}\Big(1+2\sqrt{\ell_T}\Big)\right)^{1/2} T^{1/4}\Bigg], \tag{35}$$

where $\ell_T = \log(\pi^2 KT^2/(6\delta))$ and $B = 2LS$. In particular, since $\sqrt{T} \le T^{3/4}$ for $T \ge 1$, the dominant term gives $R_{\mathrm{loc}}(T) = \widetilde{\mathcal{O}}\big(\sqrt{dK}\,T^{3/4}\big)$, where $\widetilde{\mathcal{O}}$ hides polylogarithmic factors in $T, K, 1/\delta$ and constants depending on $L, S, \gamma_\star, \bar\gamma, \lambda$.

The additive $\sqrt{T}$-order terms of the display already carry the same $\sqrt{d}$ factor.

Three remarks accompany the theorem: the scale-misspecification reading (whose derivation is Appendix E.1), the memory and computation profile, and the positioning of the rate.

Remark 4 (Scale misspecification costs only constants)**.** *Theorem 4 never requires $\bar\gamma = \gamma_\star$: the drift identity of Lemma 1 is exact for every $\bar\gamma > 0$, and the scale parameters enter $\mathcal{B}_\tau$ only through the combination $B^2 + (\Gamma)^2$ and through $G = L/\bar\gamma$. The same construction therefore tolerates an arbitrary misspecified score scale, with constants degrading polynomially in the ratio $(\Gamma)/\bar\gamma$. The information-theoretic content of this statement is an identity that generalizes Lemma 2 and justifies viewing the conversion as an exact-KL argument: for $\xi \sim \mathrm{Cauchy}(0,\gamma_\star)$, any $\bar\gamma > 0$, and any $\Delta \in \mathbb{R}$,*

$$\mathbb{E}\big[\log\big(\bar\gamma^2+(\Delta-\xi)^2\big) - \log\big(\bar\gamma^2+\xi^2\big)\big] \;=\; \log\!\left(1+\frac{\Delta^2}{(\Gamma)^2}\right). \tag{36}$$

The derivation is in Appendix E.1.

Remark 5 (Memory, computation, and behavior under low-rank contexts)**.** *Algorithm 3 maintains, per arm, the iterate $z_{k,\cdot}$ and the vector $b_{k,\cdot}$ (d floats each) and the design matrix and its inverse $V_{k,\cdot}, V_{k,\cdot}^{-1}$ ($2d^2$ floats). Each pull of an arm costs $\mathcal{O}(d^2)$ work: one Sherman–Morrison rank-one update of V^{-1}, one matrix-vector product for the index, and the $\mathcal{O}(d)$ score update. The persistent memory footprint is therefore $\mathcal{O}(d^2)$ per arm, against the $\mathcal{O}(d)$ state and $\mathcal{O}(d)$ per-update cost of the excitation-based policy (Algorithm 1); this is the computational price of dropping Assumption 1. Statistically, the ellipsoid handles persistently low-rank context sequences gracefully: in a direction u that is never excited, $V_{k,\tau}u \approx \lambda u$, so $\|u\|_{V_{k,\tau}^{-1}} \approx \lambda^{-1/2}$ and the confidence width in that direction remains large, inducing continued exploration rather than spurious certainty; no excitation is needed for* validity*, only for fast regret decay. When the context sequence is well conditioned, $\sqrt{\beta_\tau^{\mathrm{ell}}} = \widetilde{\mathcal{O}}(\tau^{1/4})$ is the bottleneck that produces the $T^{3/4}$ rate, versus the $\widetilde{\mathcal{O}}(\tau^{-1/2})$ confidence radius of Proposition 1 under Assumption 1.*

Remark 6 (Positioning of the rate)**.** *Relative to Theorem 1, removing excitation degrades the rate from $\widetilde{\mathcal{O}}(\sqrt{KT})$ to $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ and introduces an explicit $\sqrt{d}$ factor; the analysis of Theorem 1 is dimension-free. Whether $T^{3/4}$ is minimax for exact-Cauchy contextual bandits with adversarial bounded contexts is open; the lower bound of Theorem 2 continues to apply and leaves a rate gap of $T^{1/4}$. The accompanying $\sqrt{d}$ factor is introduced by the ellipsoidal analysis and is not required by the (dimension-free) lower bound. We emphasize that Theorem 4 imposes no stochastic or deterministic design condition: contexts may be constant, rank-one, or adaptively degenerate.*

Setup and indexing. Fix an arm k. Index its samples locally by $s = 1, 2, \ldots$; write $y_{k,s}$, $r_{k,s}$ for the context and reward at the s-th pull, and $z_{k,s}$ for the iterate used to score that pull (so $z_{k,1} = \mathbf{0}$), updated by Eq. (33). We assume the local model of Section 2: $y_{k,s}$ is revealed at the s-th pull and satisfies $\|y_{k,s}\|_2 \le L$ (adversarial but predictable contexts), and the reward noise at the s-th pull satisfies $\xi_{k,s} \mid \mathcal{F}_{k,s-1} \sim \mathrm{Cauchy}(0,\gamma_\star)$. All local σ-fields below are understood to include the context about to be played, i.e. we work with $\mathcal{G}_{k,s-1} := \mathcal{F}_{k,s-1} \vee \sigma(y_{k,s})$; by the noise model of Section 2, $\xi_{k,s} \mid \mathcal{G}_{k,s-1} \sim \mathrm{Cauchy}(0,\gamma_\star)$ (smoothing of a constant conditional law), and $\Delta_{k,s}$ below is $\mathcal{G}_{k,s-1}$-measurable. Note that $z_{k,s}$ is a function of the first $s-1$ local samples, hence $\mathcal{F}_{k,s-1}$-measurable; likewise the local prediction error

$$\Delta_{k,s} := y_{k,s}^\top(z_{k,s} - \theta_k), \qquad |\Delta_{k,s}| \le \|y_{k,s}\|_2\,\|z_{k,s}-\theta_k\|_2 \le L\cdot 2S = B, \tag{37}$$

is $\mathcal{G}_{k,s-1}$-measurable. All boundedness statements below hold pathwise.

Lemma 4 (Deterministic linearized regret of the direct update)**.** *Fix an arm k and a local horizon $\tau \geq 1$. For* any *sequence of contexts and rewards $\{(y_{k,s}, r_{k,s})\}_{s=1}^{\tau}$ with $\|y_{k,s}\|_2 \leq L$ (no stochastic assumption), the iterates of Eq.* (33) *satisfy, for every $v \in \mathcal{X}$,*

$$\sum_{s=1}^{\tau} \langle g_{k,s},\, z_{k,s} - v \rangle \;\leq\; DG\sqrt{2\tau} \;=\; \frac{2SL}{\bar{\gamma}}\sqrt{2\tau}. \tag{38}$$

Proof. Since $|\varphi_{\bar{\gamma}}(u)| = 2|u|/(\bar{\gamma}^2 + u^2) \leq 1/\bar{\gamma}$ by AM–GM, we have $\|g_{k,s}\|_2 \leq L/\bar{\gamma} = G$ pathwise, and $\|z_{k,s} - v\|_2 \leq D$ since $z_{k,s}, v \in \mathcal{X}$. By nonexpansiveness of $\Pi_{\mathcal{X}}$,

$$\|z_{k,s+1} - v\|_2^2 \leq \|z_{k,s} - v\|_2^2 - 2\eta_s \langle g_{k,s}, z_{k,s} - v \rangle + \eta_s^2 G^2,$$

so, summing and telescoping with η_s nonincreasing (standard OGD algebra),

$$\sum_{s=1}^{\tau} \langle g_{k,s}, z_{k,s} - v \rangle \leq \frac{\|z_{k,1} - v\|_2^2}{2\eta_\tau} + \frac{G^2}{2} \sum_{s=1}^{\tau} \eta_s \leq \frac{D^2}{2\eta_\tau} + \frac{G^2}{2} \sum_{s=1}^{\tau} \eta_s.$$

With $\eta_s = D/(G\sqrt{2s})$ and $\sum_{s=1}^{\tau} s^{-1/2} \leq 2\sqrt{\tau}$ (integral comparison), both terms equal $\frac{1}{2} DG\sqrt{2\tau}$, giving Eq. (38). No convexity of the per-sample loss is used; the bound controls the linearized regret, which is all that the conversion below requires (see Remark 7). □

Lemma 5 (Cumulative prediction-error bound under adversarial contexts)**.** *Let $\delta \in (0,1)$ and let $\mathcal{B}_\tau$ be as in Eq.* (34)*. Under the local model above, with probability at least $1 - \delta$, simultaneously for all arms $k \in [K]$ and all local times $0 \leq \tau \leq T$,*

$$\sum_{s=1}^{\tau} \left(y_{k,s}^{\top} (z_{k,s} - \theta_k) \right)^2 \;\leq\; \mathcal{B}_\tau. \tag{39}$$

Proof. Fix an arm k and write $e_s = z_{k,s} - \theta_k$, $\Delta_s = \Delta_{k,s} = y_{k,s}^{\top} e_s$; both are $\mathcal{G}_{k,s-1}$-measurable. Since $r_{k,s} = y_{k,s}^{\top} \theta_k + \xi_{k,s}$, we have $y_{k,s}^{\top} z_{k,s} - r_{k,s} = \Delta_s - \xi_{k,s}$, hence

$$\langle g_{k,s}, e_s \rangle = \Delta_s \, \varphi_{\bar{\gamma}}(\Delta_s - \xi_{k,s}).$$

By Lemma 1 applied conditionally on $\mathcal{G}_{k,s-1}$ (the conditional law of $\xi_{k,s}$ given $\mathcal{G}_{k,s-1}$ is $\mathrm{Cauchy}(0, \gamma_\star)$ and Δ_s is $\mathcal{G}_{k,s-1}$-measurable),

$$\mathbb{E}[\langle g_{k,s}, e_s \rangle \mid \mathcal{G}_{k,s-1}] = \frac{2\Delta_s^2}{\Delta_s^2 + (\Gamma)^2} \;\geq\; \frac{2\Delta_s^2}{B^2 + (\Gamma)^2},$$

where the inequality uses $|\Delta_s| \leq B$ from Eq. (37) and the monotonicity of $\Delta^2/(\Delta^2 + c)$ in Δ^2. Define the martingale difference

$$X_s := \langle g_{k,s}, e_s \rangle - \mathbb{E}[\langle g_{k,s}, e_s \rangle \mid \mathcal{G}_{k,s-1}].$$

Pathwise, $|\langle g_{k,s}, e_s \rangle| \leq \|g_{k,s}\|_2 \|e_s\|_2 \leq GD$ and likewise for its conditional expectation, so $|X_s| \leq 2DG$ almost surely. Applying Lemma 4 with $v = \theta_k$ (which is valid pathwise, hence on every realization) and the *one-sided* Azuma–Hoeffding inequality, $\Pr\big(-\sum_{s=1}^{\tau} X_s \geq 2DG\sqrt{2\tau \log(1/\delta_\tau)}\,\big) \leq \delta_\tau$, at level $\delta_\tau := 6\delta/(K\pi^2\tau^2)$,

$$\sum_{s=1}^{\tau} \mathbb{E}[\langle g_{k,s}, e_s \rangle \mid \mathcal{G}_{k,s-1}] = \sum_{s=1}^{\tau} \langle g_{k,s}, e_s \rangle - \sum_{s=1}^{\tau} X_s \;\leq\; DG\sqrt{2\tau} + 2DG\sqrt{2\tau \log(1/\delta_\tau)},$$

with probability at least $1 - \delta_\tau$ for this fixed (k, τ). Chaining with the drift lower bound,

$$\sum_{s=1}^{\tau} \Delta_s^2 \;\leq\; \frac{B^2 + (\Gamma)^2}{2}\, DG\,\sqrt{2\tau}\left(1 + 2\sqrt{\log(1/\delta_\tau)}\right) \;=\; \mathcal{B}_\tau,$$

since $DG/2 = SL/\bar{\gamma}$ and $\log(1/\delta_\tau) = \ell_\tau$. A union bound over $k \in [K]$ and $\tau \geq 1$ uses $\sum_{k=1}^{K} \sum_{\tau=1}^{\infty} \delta_\tau = \sum_{k=1}^{K} \frac{6\delta}{K\pi^2} \cdot \frac{\pi^2}{6} = \delta$, and the case $\tau = 0$ holds vacuously. □

Lemma 5 says that, on a probability-$(1-\delta)$ event, θ_k belongs to the data-dependent set $\{v : \sum_{s\le\tau}(y_{k,s}^\top(v - z_{k,s}))^2 \le \mathcal{B}_\tau\}$ for every arm and local time. Completing the square turns this set into an ellipsoid in the metric of the regularized design matrix, which yields a closed-form UCB index.

Lemma 6 (Ellipsoidal form and closed-form index)**.** *Fix an arm k, a regularizer $\lambda > 0$, and a local time τ. Define*

$$\begin{aligned} V_{k,\tau} &= \lambda I_d + \sum_{s=1}^{\tau} y_{k,s} y_{k,s}^\top, \qquad b_{k,\tau} = \sum_{s=1}^{\tau} (y_{k,s}^\top z_{k,s})\, y_{k,s}, \\ \hat{z}_{k,\tau} &= V_{k,\tau}^{-1} b_{k,\tau}, \qquad \beta_\tau^{\mathrm{ell}} = \mathcal{B}_\tau + \lambda S^2. \end{aligned} \tag{40}$$

(Unlike the decreasing scalar confidence radius β_τ of Proposition 1, $\beta_\tau^{\mathrm{ell}}$ is a growing squared *ellipsoid radius, in units of squared prediction error.) Then the set $\Theta_{k,\tau} = \{v \in \mathcal{X} : \sum_{s\le\tau}(y_{k,s}^\top(v - z_{k,s}))^2 \le \mathcal{B}_\tau\}$ is contained in the ellipsoid*

$$\mathcal{E}_{k,\tau} := \left\{ v \in \mathbb{R}^d : (v - \hat{z}_{k,\tau})^\top V_{k,\tau}\, (v - \hat{z}_{k,\tau}) \le \beta_\tau^{\mathrm{ell}} \right\}, \tag{41}$$

and for every context $y \in \mathbb{R}^d$,

$$\max_{v \in \Theta_{k,\tau}} y^\top v \;\le\; \max_{v \in \mathcal{E}_{k,\tau}} y^\top v \;=\; y^\top \hat{z}_{k,\tau} + \sqrt{\beta_\tau^{\mathrm{ell}}}\, \|y\|_{V_{k,\tau}^{-1}}. \tag{42}$$

In particular, on the event of Lemma 5, the right-hand side of Eq. (42) *is a valid upper confidence bound on $y^\top \theta_k$, simultaneously for all arms, local times, and contexts.*

Proof. Expand the quadratic defining $\Theta_{k,\tau}$: with $\alpha_s := y_{k,s}^\top z_{k,s}$ and $c_\tau := \sum_{s\le\tau} \alpha_s^2$,

$$\sum_{s\le\tau} (y_{k,s}^\top v - \alpha_s)^2 = v^\top (V_{k,\tau} - \lambda I_d) v - 2 b_{k,\tau}^\top v + c_\tau \le \mathcal{B}_\tau.$$

For $v \in \mathcal{X}$ we have $\lambda \|v\|_2^2 \le \lambda S^2$, so every $v \in \Theta_{k,\tau}$ satisfies $v^\top V_{k,\tau} v - 2 b_{k,\tau}^\top v \le \mathcal{B}_\tau + \lambda S^2 - c_\tau$. Completing the square,

$$(v - \hat{z}_{k,\tau})^\top V_{k,\tau} (v - \hat{z}_{k,\tau}) \le \mathcal{B}_\tau + \lambda S^2 - c_\tau + b_{k,\tau}^\top V_{k,\tau}^{-1} b_{k,\tau}.$$

Write Y for the $\tau \times d$ matrix with rows $y_{k,s}^\top$ and $\alpha \in \mathbb{R}^\tau$ for the vector of α_s, so that $b_{k,\tau} = Y^\top \alpha$, $V_{k,\tau} = \lambda I_d + Y^\top Y$, and $c_\tau = \|\alpha\|_2^2$. By the push-through identity, $Y(\lambda I_d + Y^\top Y)^{-1} Y^\top = YY^\top (\lambda I_\tau + YY^\top)^{-1} \preceq I_\tau$, hence

$$b_{k,\tau}^\top V_{k,\tau}^{-1} b_{k,\tau} = \alpha^\top Y V_{k,\tau}^{-1} Y^\top \alpha \;\le\; \|\alpha\|_2^2 = c_\tau,$$

so $-c_\tau + b_{k,\tau}^\top V_{k,\tau}^{-1} b_{k,\tau} \le 0$ and every $v \in \Theta_{k,\tau}$ lies in $\mathcal{E}_{k,\tau}$. (The tighter radius $\mathcal{B}_\tau + \lambda S^2 - c_\tau + b_{k,\tau}^\top V_{k,\tau}^{-1} b_{k,\tau}$ established in the display above is nonnegative on the good event: $\theta_k \in \Theta_{k,\tau} \subseteq \mathcal{X}$ then satisfies the quadratic inequality, so the radius is at least $(\theta_k - \hat{z}_{k,\tau})^\top V_{k,\tau} (\theta_k - \hat{z}_{k,\tau}) \ge 0$; we relax it to the manifestly nonnegative $\beta_\tau^{\mathrm{ell}}$ to obtain a closed-form index.) The first inequality in Eq. (42) is the containment just proved — intersecting with the ball $\mathcal{X}$ only shrinks the feasible set, so the unconstrained ellipsoid maximum remains a valid UCB — and the equality is Cauchy–Schwarz in the $V_{k,\tau}$-metric:

$$\max_{v \in \mathcal{E}_{k,\tau}} y^\top (v - \hat{z}_{k,\tau}) = \sqrt{\beta_\tau^{\mathrm{ell}}}\, \sqrt{y^\top V_{k,\tau}^{-1} y}. \qquad \square$$

Proof of Theorem 4. Work on the probability-$(1-\delta)$ event of Lemma 5, on which $\theta_k \in \Theta_{k,\tau} \subseteq \mathcal{E}_{k,\tau}$ for every arm k and every $\tau \le T$ (Lemma 6). Fix a round t and write $\tau_k = T_k(t-1)$, $y_t = y_{k_t,t}$, $w_{k,\tau} = \|y_{k,\tau+1}\|_{V_{k,\tau}^{-1}}$. By optimism: since $\theta_{k_t^\star} \in \mathcal{E}_{k_t^\star,\tau_{k_t^\star}}$ and k_t maximizes the index,

$$y_{k_t^\star,t}^\top \theta_{k_t^\star} \;\le\; \mathrm{UCB}_{k_t^\star,t} \;\le\; \mathrm{UCB}_{k_t,t} \;=\; y_t^\top \hat{z}_{k_t,\tau_{k_t}} + \sqrt{\beta_{\tau_{k_t}}^{\mathrm{ell}}}\, \|y_t\|_{V_{k_t,\tau_{k_t}}^{-1}}.$$

Since $\theta_{k_t} \in \mathcal{E}_{k_t,\tau_{k_t}}$, Cauchy–Schwarz in the V-metric gives $y_t^\top \theta_{k_t} \geq y_t^\top \hat{z}_{k_t,\tau_{k_t}} - \sqrt{\beta^{\mathrm{ell}}_{\tau_{k_t}}}\, \|y_t\|_{V^{-1}_{k_t,\tau_{k_t}}}$. Subtracting, the instantaneous location regret satisfies

$$\rho_t := y_{k_t^\star,t}^\top \theta_{k_t^\star} - y_t^\top \theta_{k_t} \;\leq\; 2\sqrt{\beta^{\mathrm{ell}}_{\tau_{k_t}}}\, \|y_t\|_{V^{-1}_{k_t,\tau_{k_t}}}, \qquad \text{and also} \qquad \rho_t \leq 2LS = B,$$

the latter since $|y^\top \theta| \leq LS$ for all y, θ involved. Since $\beta^{\mathrm{ell}}_\tau$ is nondecreasing in τ, for every pull of arm k at local time $\tau + 1 \leq T_k$,

$$\rho \;\leq\; \min\bigl(B,\, 2\sqrt{\beta^{\mathrm{ell}}_\tau} w_{k,\tau}\bigr) \;\leq\; \min\bigl(B,\, 2\sqrt{\beta^{\mathrm{ell}}_{T_k}} w_{k,\tau}\bigr) \;\leq\; \bigl(B + 2\sqrt{\beta^{\mathrm{ell}}_{T_k}}\bigr) \min(1, w_{k,\tau}),$$

where the last step checks the two cases $w_{k,\tau} \leq 1$ and $w_{k,\tau} > 1$ separately. Summing over the pulls of arm k and applying Cauchy–Schwarz,

$$R^{(k)}_{\mathrm{loc}}(T) \;\leq\; \bigl(B + 2\sqrt{\beta^{\mathrm{ell}}_{T_k}}\bigr)\, \sqrt{T_k}\, \Bigl(\sum_{\tau=0}^{T_k-1} \min(1, w_{k,\tau}^2)\Bigr)^{1/2}.$$

By the elliptical potential lemma: $\min(1,u) \leq 2\log(1+u)$ for $u \geq 0$, and the determinant identity $\det V_{k,\tau} = \det V_{k,\tau-1}\,(1 + w^2_{k,\tau-1})$ telescopes, while $\det V_{k,T_k} \leq (\lambda + T_k L^2/d)^d$ by AM–GM on the eigenvalues ($\mathrm{tr}\, V_{k,T_k} \leq \lambda d + T_k L^2$); hence

$$\sum_{\tau=0}^{T_k-1} \min(1, w_{k,\tau}^2) \;\leq\; 2\log \frac{\det V_{k,T_k}}{\lambda^d} \;\leq\; 2d \log\Bigl(1 + \frac{T_k L^2}{d\lambda}\Bigr).$$

Summing over arms with $\sum_k \sqrt{T_k} \leq \sqrt{KT}$ (Cauchy–Schwarz, $\sum_k T_k = T$) and $T_k \leq T$ inside the logarithm,

$$R_{\mathrm{loc}}(T) \;\leq\; \sqrt{2dKT \log\Bigl(1 + \frac{TL^2}{d\lambda}\Bigr)}\, \bigl(B + 2\sqrt{\beta^{\mathrm{ell}}_T}\bigr).$$

Finally $\sqrt{\beta^{\mathrm{ell}}_T} = \sqrt{\mathcal{B}_T + \lambda S^2} \leq \sqrt{\mathcal{B}_T} + S\sqrt{\lambda}$, and substituting $\mathcal{B}_T$ from Eq. (34) gives Eq. (35). The dominant term scales as $T^{1/2} \cdot T^{1/4}\sqrt{\log T} = \widetilde{\mathcal{O}}(T^{3/4})$, with leading constant $2\bigl((B^2 + (\Gamma)^2) SL\sqrt{2}/\bar{\gamma}\bigr)^{1/2}\sqrt{2dK}$ up to logarithmic factors. □

Remark 7 (Why the conversion uses the linearized regret, not the NLL regret)**.** *The per-sample Cauchy NLL $f_s(z) = \log(\bar{\gamma}^2 + (y_s^\top z - r_s)^2)$ is* not convex *in z: the scalar map $u \mapsto \log(\bar{\gamma}^2 + u^2)$ has second derivative $2(\bar{\gamma}^2 - u^2)/(\bar{\gamma}^2 + u^2)^2$, negative for $|u| > \bar{\gamma}$, and the residual $y_s^\top z - r_s$ is unbounded under Cauchy noise. The classical convex OGD bound $\sum_s [f_s(z_s) - f_s(\theta)] \leq DG\sqrt{T}$ is therefore unavailable, and an online-to-confidence-set conversion based on exp-concavity (as in the squared-loss analysis of Abbasi-Yadkori et al., 2012) fails as well, since $u \mapsto (\bar{\gamma}^2 + u^2)^{-\eta}$ is concave only on $|u| \leq \bar{\gamma}/\sqrt{2\eta+1}$. Lemma 4 sidesteps this by controlling the* linearized *regret $\sum_s \langle g_s, z_s - \theta\rangle$, a purely geometric property of the projected update that holds pathwise for any loss sequence; the exact drift identity of Lemma 1 then converts linearized regret into the quadratic prediction-error bound of Lemma 5. The KL identity of Eq.* (36) *shows that this route loses nothing statistically relative to an NLL-regret conversion: the expected per-sample excess loss is exactly $\log(1 + \Delta_s^2/(\Gamma)^2)$, which is $\Theta\bigl(\Delta_s^2/(\Gamma)^2\bigr)$ on $|\Delta_s| \leq \Gamma$, since there $\frac{1}{2}\Delta_s^2/(\Gamma)^2 \leq \log(1 + \Delta_s^2/(\Gamma)^2) \leq \Delta_s^2/(\Gamma)^2$.*

E.1 Scale-misspecification identities

Indeed, both individual expectations are finite: $|\log(\bar{\gamma}^2 + x^2)| \leq 2|\log\bar{\gamma}| + \log(1 + x^2)$, and $\mathbb{E}[\log(1+\xi^2)] < \infty$ since the Cauchy density decays like x^{-2} while $\log(1+x^2) = o(x^\varepsilon)$ for every $\varepsilon > 0$. Differentiating the left side of Eq. (36) in Δ (justified by the boundedness of $\varphi_{\bar{\gamma}}$) gives $\mathbb{E}[\varphi_{\bar{\gamma}}(\Delta - \xi)] = 2\Delta/(\Delta^2 + (\Gamma)^2)$ by Lemma 1, which is exactly the derivative of the right side; both sides vanish at $\Delta = 0$. For $\bar{\gamma} = \gamma_\star$ this recovers the equal-scale KL of Lemma 2, $\log(1 + \Delta^2/(4\gamma_\star^2))$. For completeness we record the general cross-scale formula, which we verify independently by the same differentiation-under-integral technique as in the proof of Lemma 2 (the verification is the display immediately below):

$$\mathrm{KL}\bigl(\mathrm{Cauchy}(\mu_1, g_1) \,\|\, \mathrm{Cauchy}(\mu_2, g_2)\bigr) = \log \frac{(g_1 + g_2)^2 + (\mu_1 - \mu_2)^2}{4 g_1 g_2}. \tag{43}$$

To verify Eq. (43), write $\delta = \mu_1 - \mu_2$ and $u \sim \mathrm{Cauchy}(0, g_1)$, so the KL equals $\log(g_1/g_2) + A(g_2, \delta) - A(g_1, 0)$ with $A(a, \delta) := \mathbb{E}[\log(a^2 + (u+\delta)^2)]$. Differentiating under the integral and using $1/(a^2+x^2) = \pi p_a(x)/a$ together with the Poisson-kernel convolution $p_a * p_{g_1} = p_{a+g_1}$ (as in Lemma 1),

$$\frac{\partial A}{\partial \delta} = \mathbb{E}[\varphi_a(\delta - u)] = \frac{2\delta}{\delta^2 + (g_1+a)^2}, \qquad \frac{\partial A}{\partial a} = \mathbb{E}\left[\frac{2a}{a^2 + (u+\delta)^2}\right] = \frac{2(g_1+a)}{(g_1+a)^2 + \delta^2},$$

which match the partial derivatives of $\log(((g_1+g_2)^2 + \delta^2)/(4g_1g_2))$ in (δ, g_2), and both sides vanish at $(g_2, \delta) = (g_1, 0)$. Comparing with Eq. (36) shows the location-shift NLL gap of Eq. (36) equals the cross-scale KL up to the Δ-independent constant $\log((\Gamma)^2/(4\gamma_\star\bar{\gamma})) = \mathrm{KL}\big(\mathrm{Cauchy}(0, \gamma_\star) \,\|\, \mathrm{Cauchy}(0, \bar{\gamma})\big) \geq 0$; under constant-factor misspecification this constant is $\mathcal{O}(1)$.

F Proofs Beyond Exact Cauchy

This section proves the results of Section 4.3: the sign-update guarantees (Theorem 5) through the sign drift identity (Lemma 7), a mean-square contraction (Proposition 8), and a uniform confidence sequence (Proposition 9), the perturbation-stability result (Proposition 7), and the matching floor for the per-arm perturbation class (Proposition 10).

The sign-update variant summarized in the body rests on the following assumption, update rule, and constants.

Assumption 4 (Symmetric median-growth noise)**.** *There exists a deterministic constant $p_0 > 0$ such that, for every round t, conditionally on $\mathcal{F}_{t-1}$ and the round-t contexts $\{y_{j,t}\}_{j=1}^K$, the noise ξ_t is symmetric about 0 and its conditional cumulative distribution function F_t satisfies, almost surely,*

$$F_t(u) - \tfrac{1}{2} \;\geq\; p_0\, u \qquad \textit{for all } 0 \leq u \leq B, \quad B = 2LS. \tag{44}$$

Assumption 4 imposes no moment condition of any order. It holds, for instance, for every symmetric α-stable law with any $\alpha \in (0, 2]$ (including Cauchy), for Student laws with any degrees of freedom, for Laplace and Gaussian laws, and for arbitrary symmetric mixtures of these with an atom at 0: in each case the density (or sub-density) is positive and non-increasing on $[0, \infty)$, so one may take $p_0 = p(B) > 0$, and an atom of mass a_0 at 0 only helps ($p_0 = a_0/(2B)$).

We replace the Cauchy score in Eq. (5) by the sign map

$$z_{k_t,\tau+1} = \Pi_{\mathcal{X}}\big(z_{k_t,\tau} - \eta_\tau\, \psi\big(y_{k_t,t}^\top z_{k_t,\tau} - r_t\big)\, y_{k_t,t}\big), \qquad \psi(u) = 2\mathbf{1}\{u \geq 0\} - 1. \tag{45}$$

The map ψ is the sign function with the tie-breaking convention $\psi(0) = +1$; the convention only matters for noise distributions with atoms and is irrelevant for continuous laws (see the proof of Lemma 7). Since $|\psi| \leq 1$, the stochastic gradient is bounded pathwise by $L =: G_s$ (Eq. (48)), and (45) is projected OGD on the absolute-deviation (median) loss $|y^\top z - r|$, which needs no moment of the noise to exist. The contraction modulus, gradient bound, and burn-in of the sign update are

$$\kappa_s := 2p_0\, \mu_{\mathrm{dir}}, \qquad G_s := L, \qquad \tau_0^{(s)} = \left\lceil \max\left\{1,\; \frac{2L^2}{\kappa_s},\; \frac{4L^4}{\kappa_s^2}\right\}\right\rceil, \tag{46}$$

the exact analogues of $C_{\mathrm{Cau}}\mu_{\mathrm{dir}}$, $L/\bar{\gamma}$, and τ_0 in Section 4 (with $\ell_s := 2L^2$ in place of $\ell = 2L^2/(\gamma_\star + \bar{\gamma})^2$); as in the main development, the proofs only use $\tau_0^{(s)} \geq 1$.

Theorem 5 (full form). *Let $\delta \in (0, 1)$. Run Algorithm 1 with the update* (5) *replaced by the sign update* (45)*, step sizes $\eta_\tau = \frac{1}{\kappa_s(\tau + \tau_0^{(s)} + 1)}$, and confidence schedule $\beta_\tau^{(s)}$ from Proposition 9 instantiated at failure level δ/K, under Assumptions 1 and 4. Then, with probability at least $1 - \delta$,*

$$\begin{aligned} R_{\mathrm{loc}}(T) \leq 2L\left(2S\tau_0^{(s)} + \frac{L}{\kappa_s}\sqrt{\tau_0^{(s)}}\right) K\big(1 + \log(T + \tau_0^{(s)})\big) \\ + \frac{8L^2}{\kappa_s}\sqrt{KT} + \frac{16\sqrt{2}\,L^2}{\kappa_s}\sqrt{KT\log(2KT/\delta)}. \end{aligned} \tag{47}$$

The bound implies the stated $\widetilde{\mathcal{O}}\big(K\log T + \sqrt{KT\log(KT/\delta)}\big)$ form.

The pathwise gradient bound of the sign update, used by the two proofs below, is

$$\left\|\psi(y_{k,\tau+1}^{\top} z_{k,\tau} - r_{k,\tau+1})\, y_{k,\tau+1}\right\|_2 \le L =: G_s. \tag{48}$$

We use the perturbed drift constant and the resulting error floor

$$\Gamma := \gamma_\star + \bar{\gamma}, \qquad \lambda := \frac{2\rho L}{\Gamma^2}, \qquad \Phi := \frac{\lambda}{\kappa_{\mathrm{dir}}} = \frac{2\rho L}{\Gamma^2 \kappa_{\mathrm{dir}}}.$$

Here λ is the perturbation constant of Eq. (49), not the regularizer λ of Algorithm 3 or the design eigenvalue $\lambda_{\min}$ of Assumption 2. The body summarizes the perturbation-stability result by its regret part:

Proposition 7 (Regret under bounded deviations from Cauchy)**.** *Fix an arm k and run the Direct-Cauchy update* (5) *with the exact-Cauchy step size $\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau+\tau_0+1)}$ under the model* (11) *and Assumption 1.*

Run Algorithm 1 with the confidence schedule $\beta_\tau^D := \beta_\tau^{\mathrm{Cau}} + \Phi\big(1 + \frac{1}{e\tau_0}\big)$ instantiated at failure level δ/K. With probability at least $1-\delta$,

$$R_{\mathrm{loc}}(T) \;=\; \widetilde{\mathcal{O}}\Big(K \log T + \sqrt{KT \log(2KT/\delta)}\Big) \;+\; \frac{4\rho L^2}{\Gamma^2 \kappa_{\mathrm{dir}}}\Big(1 + \frac{1}{e\tau_0}\Big)\, T,$$

i.e. the bound of Theorem 1 plus a term linear in ρT whose explicit constant is $\frac{4L^2}{\Gamma^2\kappa_{\mathrm{dir}}}\big(1+\frac{1}{e\tau_0}\big) = \frac{2L^2\left(B^2+\Gamma^2\right)}{\mu_{\mathrm{dir}}\Gamma^2}\big(1+O(1/\tau_0)\big)$ (Eq. (51)*).*

The full form, with the drift and confidence-sequence parts and all constants, is given below. The floor Φ there is negligible relative to the terminal stochastic term whenever $\rho \lesssim \Gamma^2\sqrt{\log(2T/\delta)}/\big(\bar{\gamma}\sqrt{T+\tau_0}\big)$, and is informative only when $\Phi < 2S$ (the diameter of $\mathcal{X}$; otherwise the trivial bound $\|e_{k,\tau}\|_2 \le 2S$ dominates) and when $\tau_0 \lesssim T$.

Proposition 7 (full form). *Fix an arm k and run the Direct-Cauchy update* (5) *with the exact-Cauchy step size $\eta_\tau = \frac{1}{\kappa_{\mathrm{dir}}(\tau+\tau_0+1)}$ under the model* (11) *and Assumption 1.*

(i) Perturbed drift. *For $\Delta_{k,\tau+1} = y_{k,\tau+1}^{\top} e_{k,\tau}$ and the conditional mean map $h_{k,\tau}$ of the Cauchy score,*

$$\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \ge \kappa_{\mathrm{dir}}\, \|e_{k,\tau}\|_2^2 - \lambda\, \|e_{k,\tau}\|_2. \tag{49}$$

(ii) Confidence sequence with an error floor. *Let $\delta \in (0,1)$ and let $\beta_\tau^{\mathrm{Cau}}$ denote the radius* (13) *of Proposition 1. With probability at least $1-\delta$, simultaneously for all $\tau \in \{0,\dots,n\}$,*

$$\|z_{k,\tau} - \theta_k\|_2 \le \beta_\tau^{\mathrm{Cau}} + \Phi\,\frac{\tau + 1 + \log(1+\tau/\tau_0)}{\tau+\tau_0} \le \beta_\tau^{\mathrm{Cau}} + \Phi\Big(1 + \frac{1}{e\,\tau_0}\Big). \tag{50}$$

Thus the estimation error decays at the usual $O(1/\sqrt{\tau})$ rate until it reaches the floor $\Phi = \frac{2\rho L}{\Gamma^2\kappa_{\mathrm{dir}}}$, which is linear in ρ.

(iii) Regret. *Run Algorithm 1 with the confidence schedule $\beta_\tau^D := \beta_\tau^{\mathrm{Cau}} + \Phi\big(1+\frac{1}{e\tau_0}\big)$ instantiated at failure level δ/K. With probability at least $1-\delta$,*

$$\begin{aligned} R_{\mathrm{loc}}(T) \le 2L\left(2S\tau_0 + \frac{L}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{\tau_0}\right) K(1+\log(T+\tau_0)) + \frac{8L^2}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{KT} \\ + \frac{16\sqrt{2}\,L^2}{\kappa_{\mathrm{dir}}\bar{\gamma}}\sqrt{KT\log(2KT/\delta)} + \frac{4\rho L^2}{\Gamma^2\kappa_{\mathrm{dir}}}\Big(1+\frac{1}{e\tau_0}\Big)\,T, \end{aligned} \tag{51}$$

i.e. the bound of Theorem 1 plus a term linear in ρT with explicit constant $\frac{4L^2}{\Gamma^2\kappa_{\mathrm{dir}}}\big(1+\frac{1}{e\tau_0}\big) = \frac{2L^2\left(B^2+\Gamma^2\right)}{\mu_{\mathrm{dir}}\Gamma^2}\big(1+O(1/\tau_0)\big)$.

Lemma 7 (Sign drift)**.** *Let ξ be symmetric about 0 with CDF F satisfying* (44)*. Then, with ψ as in* (45)*, for every deterministic $u \in [-B, B]$,*

$$u\,\mathbb{E}[\psi(u-\xi)] = u\,(2F(u)-1) \;\ge\; 2p_0\,u^2. \tag{52}$$

Proof. Since $\psi(u-\xi) = +1$ on $\{\xi \le u\}$ and $\psi(u-\xi) = -1$ on $\{\xi > u\}$ (the tie-breaking convention assigns $\{\xi = u\}$ to the first event),

$$\mathbb{E}[\psi(u-\xi)] = F(u) - (1-F(u)) = 2F(u) - 1, \qquad \text{for every } u \in \mathbb{R},$$

including at atoms of F. For $0 < u \le B$ the bound $u(2F(u)-1) \ge 2p_0u^2$ is exactly (44). For $u = 0$ both sides vanish. For $-B \le u < 0$, symmetry of ξ gives $F(u) = \Pr(\xi \le u) = \Pr(\xi \ge -u) = 1 - F((-u)-)$, where $F(v-)$ denotes the left limit; taking left limits in (44) (apply it at $w \uparrow v$) yields $F(v-) - \frac{1}{2} \ge p_0 v$ for $0 < v \le B$, hence

$$1 - 2F(u) = 2F((-u)-) - 1 \ge 2p_0(-u),$$

and multiplying by $u < 0$ gives $u(2F(u)-1) \ge 2p_0u^2$. □

Proposition 8 (One-arm sign contraction)**.** *Fix an arm k and let $e_{k,\tau} = z_{k,\tau} - \theta_k$. Under Assumptions 1 and 4, the update* (45) *with step size $\eta_\tau = \frac{1}{\kappa_s(\tau+\tau_0^{(s)}+1)}$ satisfies*

$$\mathbb{E}[\|e_{k,\tau+1}\|_2^2 \mid \mathcal{F}_{k,\tau}] \le \left(1 - \frac{2}{\tau+\tau_0^{(s)}+1}\right)\|e_{k,\tau}\|_2^2 + \frac{L^2}{\kappa_s^2(\tau+\tau_0^{(s)}+1)^2}, \tag{53}$$

and consequently $\mathbb{E}\|e_{k,\tau}\|_2^2 \le \frac{D_s}{\tau+\tau_0^{(s)}}$ with $D_s = \max\Big\{4S^2\tau_0^{(s)},\ \frac{L^2}{\kappa_s^2}\Big\}$.

Proof. Write $\Delta_{k,\tau+1} := y_{k,\tau+1}^\top e_{k,\tau}$ and define the conditional mean map

$$h_{k,\tau}(x) = \mathbb{E}\big[\psi(y_{k,\tau+1}^\top(x-\theta_k) - \xi_{\tau+1})\, y_{k,\tau+1} \mid \mathcal{F}_{k,\tau}\big].$$

Condition first on $\mathcal{F}_{k,\tau}$ and $y_{k,\tau+1}$, so that $y_{k,\tau+1}^\top(x-\theta_k) =: u$ is frozen. Since $z_{k,\tau}$ is $\mathcal{F}_{k,\tau}$-measurable and $\|z_{k,\tau}-\theta_k\|_2 \le 2S$, $\|y_{k,\tau+1}\|_2 \le L$, we have $|u| \le B$ pathwise, so Lemma 7 applied to the conditional law of $\xi_{\tau+1}$ given $(\mathcal{F}_{k,\tau}, \{y_{j,\tau+1}\})$ gives $u\,\mathbb{E}[\psi(u-\xi_{\tau+1}) \mid \mathcal{F}_{k,\tau}, y_{k,\tau+1}] \ge 2p_0u^2$ pathwise. Here we use that Assumption 4, stated given $(\mathcal{F}_{t-1}, \{y_{j,t}\})$, transfers to this coarser σ-field (at the global round of the $(\tau+1)$-th pull, $\mathcal{F}_{k,\tau} \vee \sigma(\text{selection}) \subseteq \mathcal{F}_{t-1} \vee \sigma(\{y_{j,t}\})$): the conditional law under the coarser σ-field is a mixture of the finer conditional laws, and both required properties—symmetry about 0 and the growth bound (44)—are linear in F, hence preserved under mixing, so the conditional law given $(\mathcal{F}_{k,\tau}, \{y_{j,\tau+1}\})$ satisfies Assumption 4 almost surely. Taking the expectation over $y_{k,\tau+1}$ and applying Assumption 1 with the $\mathcal{F}_{k,\tau}$-measurable vector $v = e_{k,\tau}$,

$$\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \ge 2p_0\,\mathbb{E}[\Delta_{k,\tau+1}^2 \mid \mathcal{F}_{k,\tau}] \ge 2p_0\,\mu_{\mathrm{dir}}\|e_{k,\tau}\|_2^2 = \kappa_s\|e_{k,\tau}\|_2^2. \tag{54}$$

This is the precise analogue of the Cauchy-score drift step of Proposition 5: the inner noise expectation is taken with the context frozen, and Assumption 1 is applied only to the outer context expectation. Projection nonexpansiveness gives, pathwise,

$$\|e_{k,\tau+1}\|_2^2 \le \|e_{k,\tau}\|_2^2 - 2\eta_\tau\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle - 2\eta_\tau\langle e_{k,\tau}, M_{k,\tau+1}\rangle + \eta_\tau^2\|g_{k,\tau+1}\|_2^2,$$

where $g_{k,\tau+1} = \psi(\cdot)y_{k,\tau+1}$ and $M_{k,\tau+1} = g_{k,\tau+1} - h_{k,\tau}(z_{k,\tau})$. Taking the conditional expectation kills the martingale term; inserting (54) and the pathwise bound $\|g_{k,\tau+1}\|_2 \le L$ from (48),

$$\mathbb{E}[\|e_{k,\tau+1}\|_2^2 \mid \mathcal{F}_{k,\tau}] \le (1 - 2\kappa_s\eta_\tau)\|e_{k,\tau}\|_2^2 + \eta_\tau^2 L^2.$$

Substituting $\eta_\tau = 1/(\kappa_s(\tau+\tau_0^{(s)}+1))$ gives (53). The mean-square bound follows by the induction of Proposition 5 verbatim: the base case uses $\|e_{k,0}\|_2^2 \le 4S^2 \le D_s/\tau_0^{(s)}$, and the inductive step uses only $D_s \ge L^2/\kappa_s^2$ and the displayed algebra of Proposition 5, which depends on the score only through the constants (κ_s, L). □

Proposition 9 (One-arm sign confidence sequence)**.** *Let $\delta \in (0,1)$. Fix an arm k and run the update* (45) *with $\eta_\tau = \frac{1}{\kappa_s(\tau+\tau_0^{(s)}+1)}$ under Assumptions 1 and 4. With probability at least $1-\delta$, simultaneously for all local times $\tau \in \{0,\dots,n\}$,*

$$\|z_{k,\tau} - \theta_k\|_2 \le \beta_\tau^{(s)} := \frac{2S\tau_0^{(s)} + \frac{L}{\kappa_s}\sqrt{\tau_0^{(s)}}}{\tau+\tau_0^{(s)}} + \frac{2L}{\kappa_s\sqrt{\tau+\tau_0^{(s)}}} + \frac{4\sqrt{2}\,L}{\kappa_s}\sqrt{\frac{\log(2n/\delta)}{\tau+\tau_0^{(s)}}}, \tag{55}$$

and consequently $|y^\top(z_{k,\tau}-\theta_k)| \le \beta_\tau^{(s)}\|y\|_2$ for every context y.

Proof. The proof is the variance-smoothed martingale argument of Proposition 1 (Appendix A.4) with the substitutions $\kappa_{\mathrm{dir}} \to \kappa_s$ and $L/\bar{\gamma} \to L$; we re-derive each step so that every inequality is justified for the sign score. Set

$$c_\star = \frac{L^2}{\kappa_s^2}, \qquad m_\tau = \tau + \tau_0^{(s)}, \qquad a_{k,\tau} = \|e_{k,\tau}\|_2^2, \qquad \xi_{k,\tau+1} = \langle e_{k,\tau}, M_{k,\tau+1}\rangle.$$

The martingale difference $M_{k,\tau+1}$ is bounded pathwise: $\|M_{k,\tau+1}\|_2 \le \|g_{k,\tau+1}\|_2 + \|h_{k,\tau}(z_{k,\tau})\|_2 \le 2L$, using (48) and Jensen's inequality for h. From the pathwise expansion in the proof of Proposition 8 and the drift (54),

$$a_{k,\tau+1} \le (1 - 2\kappa_s\eta_\tau)a_{k,\tau} - 2\eta_\tau\xi_{k,\tau+1} + \eta_\tau^2 L^2,$$

and substituting $\eta_\tau = 1/(\kappa_s(m_\tau+1))$ gives the pathwise recursion

$$a_{k,\tau+1} \le \left(1 - \frac{2}{m_\tau+1}\right) a_{k,\tau} - \frac{2}{\kappa_s(m_\tau+1)}\xi_{k,\tau+1} + \frac{c_\star}{(m_\tau+1)^2}.$$

Define $s_{k,\tau} = \sqrt{a_{k,\tau} + c_\star/m_\tau}$. The algebra converting this into

$$s_{k,\tau+1}^2 \le \frac{m_\tau^2}{(m_\tau+1)^2} s_{k,\tau}^2 - \frac{2}{\kappa_s(m_\tau+1)}\xi_{k,\tau+1} + \frac{2c_\star}{(m_\tau+1)^2}$$

is identical to Appendix A.4 (it uses only $1 - \frac{2}{m+1} = \frac{m^2-1}{(m+1)^2}$ and $s_{k,\tau}^2 \ge c_\star/m_\tau$, both score-independent). The concave square-root tangent bound $\sqrt{x+r} \le \sqrt{x} + r/(2\sqrt{x})$ at $x = m_\tau^2 s_{k,\tau}^2/(m_\tau+1)^2$ then yields

$$s_{k,\tau+1} \le \frac{m_\tau}{m_\tau+1} s_{k,\tau} - \frac{\xi_{k,\tau+1}}{\kappa_s m_\tau s_{k,\tau}} + \frac{c_\star}{m_\tau(m_\tau+1)s_{k,\tau}}.$$

Multiplying by $m_\tau + 1$ and writing $q_{k,\tau} = m_\tau s_{k,\tau}$, $u_{k,\tau} = e_{k,\tau}/s_{k,\tau}$ (an $\mathcal{F}_{k,\tau}$-measurable vector of norm at most 1, since $s_{k,\tau} \ge \|e_{k,\tau}\|_2$),

$$q_{k,\tau+1} \le q_{k,\tau} - \frac{m_\tau+1}{\kappa_s m_\tau}\langle u_{k,\tau}, M_{k,\tau+1}\rangle + \frac{\sqrt{c_\star}}{\sqrt{m_\tau}},$$

using $c_\star/(m_\tau s_{k,\tau}) \le \sqrt{c_\star/m_\tau}$. The increments form a martingale difference sequence with the pathwise bound

$$\left|\frac{m_\tau+1}{\kappa_s m_\tau}\langle u_{k,\tau}, M_{k,\tau+1}\rangle\right| \le \frac{m_\tau+1}{\kappa_s m_\tau}\, 2L \le \frac{4L}{\kappa_s} = 4\sqrt{c_\star},$$

since $m_\tau \ge \tau_0^{(s)} \ge 1$. Azuma–Hoeffding and a union bound over $\tau \le n$ give, with probability $1-\delta$, simultaneously for all τ,

$$q_{k,\tau} \le q_{k,0} + 4\sqrt{2c_\star\,\tau\log(2n/\delta)} + \sqrt{c_\star}\sum_{s=0}^{\tau-1}\frac{1}{\sqrt{m_s}}.$$

Finally $q_{k,0} \le 2S\tau_0^{(s)} + \sqrt{c_\star\tau_0^{(s)}}$ (from $\|e_{k,0}\|_2 \le 2S$ and $\sqrt{a+b} \le \sqrt{a}+\sqrt{b}$) and $\sum_{s=0}^{\tau-1} m_s^{-1/2} \le 2\sqrt{\tau+\tau_0^{(s)}}$; dividing by $m_\tau = \tau + \tau_0^{(s)}$ and using $\tau/(\tau+\tau_0^{(s)}) \le 1$ inside the square-root terms gives (55). The prediction bound is Cauchy–Schwarz. □

Proof of Theorem 5. Apply Proposition 9 to each arm with local horizon $n = T$ and failure level δ/K, and take a union bound over the K arms. The schedule $\beta_\tau^{(s)}$ is deterministic and non-increasing, and on the intersection event it provides the simultaneous confidence bounds required by Proposition 3, whose reduction is score-independent; hence $R_{\mathrm{loc}}(T) \le 2L\sum_{k=1}^K\sum_{\tau=0}^{T_k-1}\beta_\tau^{(s)}$. The harmonic term is bounded by $2L\left(2S\tau_0^{(s)} + \frac{L}{\kappa_s}\sqrt{\tau_0^{(s)}}\right)K(1+\log(T+\tau_0^{(s)}))$ using $\sum_{\tau=0}^{T_k-1}\frac{1}{\tau+\tau_0^{(s)}} \le 1 + \log(T+\tau_0^{(s)})$, and the two square-root terms are bounded using $\sum_{\tau=0}^{T_k-1}\frac{1}{\sqrt{\tau+\tau_0^{(s)}}} \le 2\sqrt{T_k}$ and $\sum_k\sqrt{T_k} \le \sqrt{KT}$, exactly as in Theorem 1. □

Proof of Proposition 7. (i) Let $m_{\mathrm{Cau}}(u) = \frac{2u}{u^2+\Gamma^2}$. Condition on $(\mathcal{F}_{k,\tau}, y_{k,\tau+1}, \zeta_{k,\tau+1})$; with $u = \Delta_{k,\tau+1}$ frozen, Lemma 1 applied to the inner expectation over $\xi_{C,k,\tau+1}$ (with $\Delta = u - \zeta_{k,\tau+1}$, valid for every real Δ) gives

$$\mathbb{E}\big[\varphi_{\bar\gamma}(u - \xi_{C,k,\tau+1} - \zeta_{k,\tau+1}) \,\big|\, \mathcal{F}_{k,\tau}, y_{k,\tau+1}, \zeta_{k,\tau+1}\big] = m_{\mathrm{Cau}}(u - \zeta_{k,\tau+1}) \qquad \text{pathwise.}$$

Here the transfer of the model (11) from the conditioning $(\mathcal{F}_{t-1}, \{y_{j,t}\})$ to the coarser local conditioning is the same mixture/coarsening argument as in Proposition 8 (in fact simpler, since the Cauchy kernel and the bound $|\zeta| \le \rho$ are preserved under mixing), and conditioning additionally on $\zeta_{k,\tau+1}$ inside the expectation over $\xi_{C,k,\tau+1}$ is licensed by the stated conditional independence $\xi_{C,k,t} \perp \zeta_{k,t}$. From the derivative computed in the proof of Lemma 1,

$$|m'_{\mathrm{Cau}}(\Delta)| = \frac{2|\Gamma^2 - \Delta^2|}{(\Delta^2+\Gamma^2)^2} \le \frac{2(\Gamma^2+\Delta^2)}{(\Delta^2+\Gamma^2)^2} = \frac{2}{\Delta^2+\Gamma^2} \le \frac{2}{\Gamma^2},$$

so m_{Cau} is $(2/\Gamma^2)$-Lipschitz and, since $|\zeta_{k,\tau+1}| \le \rho$ pathwise, $m_{\mathrm{Cau}}(u - \zeta_{k,\tau+1}) \ge m_{\mathrm{Cau}}(u) - 2\rho/\Gamma^2$. Multiplying by u (the perturbation contributes $-\frac{2\rho}{\Gamma^2}|u|$ regardless of the sign of u), averaging over $\zeta_{k,\tau+1}$, and using $|u| \le B$ pathwise,

$$u\,\mathbb{E}[\varphi_{\bar\gamma}(u - \xi_{\tau+1}) \mid \mathcal{F}_{k,\tau}, y_{k,\tau+1}] \ge \frac{2u^2}{u^2+\Gamma^2} - \frac{2\rho}{\Gamma^2}|u| \ge C_{\mathrm{Cau}}\,u^2 - \frac{2\rho}{\Gamma^2}|u| \qquad \text{pathwise.}$$

Taking the outer expectation over $y_{k,\tau+1}$, applying Assumption 1 with $v = e_{k,\tau}$ to the quadratic term, and using $|u| = |y_{k,\tau+1}^\top e_{k,\tau}| \le L\|e_{k,\tau}\|_2$ pathwise for the linear term gives (49): $\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \ge C_{\mathrm{Cau}}\mu_{\mathrm{dir}}\|e_{k,\tau}\|_2^2 - \frac{2\rho L}{\Gamma^2}\|e_{k,\tau}\|_2$.

(ii) Write $a_{k,\tau} = \|e_{k,\tau}\|_2^2$ and follow Appendix A.4 verbatim except for the drift step: projection nonexpansiveness and the decomposition $g_{k,\tau+1} = h_{k,\tau}(z_{k,\tau}) + M_{k,\tau+1}$ give, pathwise,

$$a_{k,\tau+1} \le a_{k,\tau} - 2\eta_\tau\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle - 2\eta_\tau\xi_{k,\tau+1} + \eta_\tau^2\|g_{k,\tau+1}\|_2^2,$$

with $\xi_{k,\tau+1} = \langle e_{k,\tau}, M_{k,\tau+1}\rangle$, $\|M_{k,\tau+1}\|_2 \le 2L/\bar\gamma$ and $\|g_{k,\tau+1}\|_2 \le L/\bar\gamma$ (both pathwise, unchanged since they use only $|\varphi_{\bar\gamma}| \le 1/\bar\gamma$). Inserting (49) with $\sqrt{a_{k,\tau}} \le s_{k,\tau}$, where $s_{k,\tau} = \sqrt{a_{k,\tau} + c_\star/m_\tau}$, $c_\star = \frac{L^2}{\kappa_{\mathrm{dir}}^2\bar\gamma^2}$, $m_\tau = \tau + \tau_0$,

$$a_{k,\tau+1} \le (1 - 2\kappa_{\mathrm{dir}}\eta_\tau)a_{k,\tau} + 2\lambda\eta_\tau s_{k,\tau} - 2\eta_\tau\xi_{k,\tau+1} + \eta_\tau^2\frac{L^2}{\bar\gamma^2}.$$

Substituting $\eta_\tau = 1/(\kappa_{\mathrm{dir}}(m_\tau+1))$ and adding $c_\star/(m_\tau+1)$ to both sides, the exact-Cauchy algebra of Appendix A.4 (which is unchanged) gives

$$s_{k,\tau+1}^2 \le \frac{m_\tau^2}{(m_\tau+1)^2}s_{k,\tau}^2 - \frac{2}{\kappa_{\mathrm{dir}}(m_\tau+1)}\xi_{k,\tau+1} + \frac{2c_\star}{(m_\tau+1)^2} + \frac{2\lambda}{\kappa_{\mathrm{dir}}(m_\tau+1)}\,s_{k,\tau}.$$

The concave tangent bound $\sqrt{x+r} \le \sqrt{x} + r/(2\sqrt{x})$ at $x = m_\tau^2 s_{k,\tau}^2/(m_\tau+1)^2$ contributes, from the new term, $\frac{2\lambda s_{k,\tau}}{\kappa_{\mathrm{dir}}(m_\tau+1)} \cdot \frac{1}{2\sqrt{x}} = \frac{\lambda}{\kappa_{\mathrm{dir}}m_\tau}$, so

$$s_{k,\tau+1} \le \frac{m_\tau}{m_\tau+1}s_{k,\tau} - \frac{\xi_{k,\tau+1}}{\kappa_{\mathrm{dir}}m_\tau s_{k,\tau}} + \frac{c_\star}{m_\tau(m_\tau+1)s_{k,\tau}} + \frac{\lambda}{\kappa_{\mathrm{dir}}m_\tau}.$$

Multiplying by $m_\tau + 1$, with $q_{k,\tau} = m_\tau s_{k,\tau}$ and $u_{k,\tau} = e_{k,\tau}/s_{k,\tau}$,

$$q_{k,\tau+1} \le q_{k,\tau} - \frac{m_\tau+1}{\kappa_{\mathrm{dir}}m_\tau}\langle u_{k,\tau}, M_{k,\tau+1}\rangle + \frac{\sqrt{c_\star}}{\sqrt{m_\tau}} + \frac{\lambda}{\kappa_{\mathrm{dir}}} \cdot \frac{m_\tau+1}{m_\tau}.$$

The martingale increments are unchanged and bounded pathwise by $\frac{m_\tau+1}{\kappa_{\mathrm{dir}}m_\tau} \cdot \frac{2L}{\bar\gamma} \le 4\sqrt{c_\star}$; the perturbation adds only the deterministic increment $\frac{\lambda}{\kappa_{\mathrm{dir}}} \cdot \frac{m_\tau+1}{m_\tau} = \frac{\lambda}{\kappa_{\mathrm{dir}}}\big(1 + \frac{1}{m_\tau}\big)$, which accumulates to exactly $\frac{\lambda}{\kappa_{\mathrm{dir}}}\sum_{s=0}^{\tau-1}\frac{m_s+1}{m_s} = \frac{\lambda}{\kappa_{\mathrm{dir}}}\big(\tau + \sum_{s=0}^{\tau-1}\frac{1}{m_s}\big) \le \frac{\lambda}{\kappa_{\mathrm{dir}}}\big(\tau + 1 + \log(1+\tau/\tau_0)\big)$ over τ steps, using $\sum_{s=0}^{\tau-1}\frac{1}{s+\tau_0} \le \frac{1}{\tau_0} + \log\frac{\tau-1+\tau_0}{\tau_0} \le 1 + \log(1+\tau/\tau_0)$ (since $\tau_0 \ge 1$). Azuma–Hoeffding with a union bound over $\tau \le n$ therefore gives, with probability $1-\delta$, simultaneously for all τ,

$$q_{k,\tau} \le q_{k,0} + 4\sqrt{2c_\star\tau\log(2n/\delta)} + 2\sqrt{c_\star(\tau+\tau_0)} + \frac{\lambda}{\kappa_{\mathrm{dir}}}\big(\tau + 1 + \log(1+\tau/\tau_0)\big),$$

using $q_{k,0} \le 2S\tau_0 + \sqrt{c_\star \tau_0}$ and $\sum_{s=0}^{\tau-1} m_s^{-1/2} \le 2\sqrt{\tau+\tau_0}$ as in Appendix A.4. Dividing by $m_\tau = \tau + \tau_0$ yields (50), since $\|e_{k,\tau}\|_2 \le s_{k,\tau} = q_{k,\tau}/m_\tau$ and $\frac{\lambda}{\kappa_{\mathrm{dir}}} = \Phi$; the uniform form of the floor uses $\frac{\tau+1+\log(1+\tau/\tau_0)}{\tau+\tau_0} \le 1 + \frac{1}{e\tau_0}$, which follows from $\log x \le x/e$ for $x = (\tau+\tau_0)/\tau_0 \ge 1$ (and $\tau_0 \ge 1$). The stated threshold on ρ equates Φ with the leading $O(1/\sqrt{T+\tau_0})$ term of β_T^{Cau}.

(iii) The schedule $\beta_\tau^D = \beta_\tau^{\mathrm{Cau}} + \Phi\big(1 + \frac{1}{e\tau_0}\big)$ is deterministic and non-increasing (a non-increasing sequence plus a constant), and by part (ii) with failure level δ/K per arm and a union bound over arms it provides simultaneous valid confidence bounds with probability $1-\delta$. Proposition 3 then gives $R_{\mathrm{loc}}(T) \le 2L \sum_k \sum_{\tau<T_k} \beta_\tau^D$. The sum of $\beta_\tau^{\mathrm{Cau}}$ is bounded by the first three terms of (51) exactly as in Theorem 1, and the floor contributes $2L\Phi\big(1+\frac{1}{e\tau_0}\big)\sum_k T_k = \frac{4\rho L^2}{\Gamma^2 \kappa_{\mathrm{dir}}}\big(1+\frac{1}{e\tau_0}\big)T$. □

F.1 Comparison of the sign and exact-Cauchy constant modules

This subsection collects the constant-level comparison of the sign and exact-Cauchy updates that the body summarizes; the remark it supports is stated first.

Remark 8 (Exact Cauchy score as the statistically efficient special case)**.** *For $\xi \sim \mathrm{Cauchy}(0,\gamma_\star)$ the median-growth constants are explicit: $p_0 = p(B) = 1/\big(\pi\gamma_\star(1+(B/\gamma_\star)^2)\big)$ is an admissible conservative choice, and the largest admissible constant is the strictly larger $p_0^\star = \arctan(B/\gamma_\star)/(\pi B)$. Under exact Cauchy noise the sign update is therefore valid with $\kappa_s = \frac{2\mu_{\mathrm{dir}}\gamma_\star}{\pi(B^2+\gamma_\star^2)}$ under the conservative tuning and with $\kappa_s^\star = \frac{2\mu_{\mathrm{dir}}\arctan(B/\gamma_\star)}{\pi B}$ under the optimal tuning, whereas the exact-score update achieves $\kappa_{\mathrm{dir}}\bar\gamma = \frac{2\mu_{\mathrm{dir}}\bar\gamma}{B^2+(\gamma_\star+\bar\gamma)^2}$. The two branches compare in the dimensionally matching combination L^2/κ_s versus $L^2/(\kappa_{\mathrm{dir}}\bar\gamma)$, and at the matched choice $\bar\gamma = \gamma_\star$ the conservatively tuned ratio lies in $[\pi/4, \pi]$, so the exact Cauchy score improves the leading regret constant by up to a factor π and is never worse by more than $4/\pi$. With the optimal tuning the* proved *sign bound becomes the tighter of the two — a reversal that is an artifact of bounding rather than of the score, since κ_{dir} is a uniform lower bound on the exact-score drift while $p_0^\star$ is the exact optimum of the sign-drift constraint. The structural conclusion is unaffected: the sign map is the likelihood score of the Laplace law and is the distribution-free robust choice, valid under Assumption 4 alone, while the Cauchy score exploits the exact noise law — its drift is exactly proportional to $u/(u^2+\Gamma^2)$ rather than merely bounded below — and is the statistically efficient special case. The derivations behind these statements, including the optimal-tuning ratio and the Laplace-noise comparison, are in Appendix F.1.*

The computations below support Remark 8.

For $\xi \sim \mathrm{Cauchy}(0,\gamma_\star)$ the density $p(x) = \frac{1}{\pi\gamma_\star(1+(x/\gamma_\star)^2)}$ is strictly decreasing on $[0,\infty)$, so for $0 \le u \le B$, $F(u) - \frac{1}{2} = \int_0^u p(x)\,dx \ge u\,p(B)$, and

$$p_0 = p(B) = \frac{1}{\pi\gamma_\star\big(1+(B/\gamma_\star)^2\big)}, \qquad 2p_0 = \frac{2}{\pi\gamma_\star\big(1+(B/\gamma_\star)^2\big)},$$

is an admissible (conservative) choice. It is not the optimal one: (44) requires $p_0 \le \big(F(u)-\frac{1}{2}\big)/u$ for all $u \in (0,B]$, and $u \mapsto \frac{1}{u}\int_0^u p$ is decreasing for a decreasing density (equivalently $\arctan(u/\gamma_\star) > u/(\gamma_\star + u^2/\gamma_\star)$ for $u>0$), so the constraint binds at $u=B$ and the largest admissible constant is

$$p_0^\star = \frac{F(B)-\frac{1}{2}}{B} = \frac{\arctan(B/\gamma_\star)}{\pi B} \;>\; p(B),$$

the inequality being strict because the average of a strictly decreasing function exceeds its endpoint value (e.g. $B=\gamma_\star$ gives $p_0^\star = 1/4$ vs. $p(B) = 1/(2\pi)$). Hence under exact Cauchy noise the sign update is valid with $\kappa_s = \frac{2\mu_{\mathrm{dir}}\gamma_\star}{\pi(B^2+\gamma_\star^2)}$ (conservative tuning) or $\kappa_s^\star = \frac{2\mu_{\mathrm{dir}}\arctan(B/\gamma_\star)}{\pi B}$ (optimal tuning), while the exact-score update of Eq. (5) achieves $\kappa_{\mathrm{dir}}\bar\gamma = \frac{2\mu_{\mathrm{dir}}\bar\gamma}{B^2+(\gamma_\star+\bar\gamma)^2}$; the regret constants compare in the dimensionally matching combination L^2/κ_s vs. $L^2/(\kappa_{\mathrm{dir}}\bar\gamma)$ (this is a comparison of the leading L^2/κ constants of the $\sqrt{KT}$ terms only; the burn-in terms of the two branches scale differently with the tuning, through $\tau_0^{(s)}$ versus τ_0). At the matched choice $\bar\gamma = \gamma_\star$ and the conservative tuning $p_0 = p(B)$,

$$\frac{L^2/\kappa_s}{L^2/(\kappa_{\mathrm{dir}}\bar\gamma)} = \frac{\kappa_{\mathrm{dir}}\bar\gamma}{\kappa_s} = \frac{\pi\,(B^2+\gamma_\star^2)}{B^2+4\gamma_\star^2} \;\in\; \big[\tfrac{\pi}{4},\,\pi\big],$$

so the exact Cauchy score improves the leading regret constant by up to a factor π (approached as $B/\gamma_\star \to \infty$) and is never worse by more than the factor $4/\pi$ (approached as $B/\gamma_\star \to 0$). With the optimal tuning $p_0^\star$ the ratio becomes

$$\frac{L^2/\kappa_s^\star}{L^2/(\kappa_{\mathrm{dir}}\bar{\gamma})} = \frac{\pi\, B\, \gamma_\star}{\left(B^2 + 4\gamma_\star^2\right)\arctan(B/\gamma_\star)} \xrightarrow{B/\gamma_\star \to 0} \frac{\pi}{4}, \quad \frac{L^2/\kappa_s^\star}{L^2/(\kappa_{\mathrm{dir}}\bar{\gamma})} \sim \frac{2\gamma_\star}{B} \xrightarrow{B/\gamma_\star \to \infty} 0, \tag{56}$$

i.e. the *proved* sign bound is then tighter than the proved Cauchy bound by a factor growing like $B/(2\gamma_\star)$. This reversal is an artifact of bounding, not of the score: $C_{\mathrm{Cau}} = 2/(B^2+\Gamma^2)$ is a uniform lower bound on the exact-score drift $u\, m_{\mathrm{Cau}}(u) = 2u^2/(u^2+\Gamma^2)$; the quadratic coefficient of this drift at $u = 0$ is $2/\Gamma^2$, a factor $(B^2+\Gamma^2)/\Gamma^2$ larger than the uniform bound; the sign-drift lemma, by contrast, is essentially tight since $p_0^\star$ is the exact optimum. The structural conclusion is unaffected: the sign map is the likelihood score of the Laplace law and is the distribution-free robust choice (valid under Assumption 4 alone), while the Cauchy score exploits the exact noise law—its drift is exactly proportional to $u/(u^2+\Gamma^2)$ rather than merely bounded below—and is the statistically efficient special case. For exact Laplace noise (of scale b; we adopt the convention $\gamma_\star = \bar{\gamma} = b$, $\Gamma = 2b$ for the Cauchy-score constants, since the exact-score theory does not fix $\gamma_\star$ there) the comparison is tuning-dependent: with the conservative tuning $p_0 = p(B)$ the sign bound's leading constant is tighter only for $B/b \lesssim 0.864$, with the optimal tuning $p_0^\star$ it is tighter for all B/b, while a tuned Cauchy score beats the conservatively tuned sign branch for $B/b \gtrsim 0.834$.

F.2 Step-size adaptation over unknown p_0

The body summarizes how the sign update behaves when the median-growth constant p_0 is unknown, including the open screening point; the full remark is stated here.

Remark 9 (Step-size adaptation over unknown p_0)**.** *The step size η_τ and the radius $\beta_\tau^{(s)}$ depend on the noise law only through $\kappa_s = 2p_0\mu_{\mathrm{dir}}$. Underestimating p_0 is safe: if the algorithm is run with any $\hat{\kappa} \in (0, \kappa_s]$, then the drift inequality $\langle e_{k,\tau}, h_{k,\tau}(z_{k,\tau})\rangle \geq \kappa_s\|e_{k,\tau}\|_2^2 \geq \hat{\kappa}\|e_{k,\tau}\|_2^2$ still holds, so every step of Propositions 8 and 9 applies with κ_s replaced by $\hat{\kappa}$, yielding a valid (conservative) confidence sequence. Its leading constants inflate by the factor $\kappa_s/\hat{\kappa}$; the burn-in $\tau_0^{(s)}$ may grow by at most its square, $(\kappa_s/\hat{\kappa})^2$, and at leading order this inflation affects only the $K\log T$ burn-in term, since the $\tau + \tau_0^{(s)}$ shifts in the denominators of $\beta_\tau^{(s)}$ make the $\sqrt{KT}$-order terms smaller, not larger. Moreover* (44) *forces $p_0 \leq \frac{1}{2B}$, since $F(B) \leq 1$, so a dyadic grid $\{\hat{p}_m = 2^{-m}/(2B)\}$ over a known range $[p_{\min}, \frac{1}{2B}]$ contains a branch with $\hat{\kappa} \in [\kappa_s/2, \kappa_s]$, at the price of $M = \lceil \log_2(1/(2Bp_{\min}))\rceil + 1$ branches and a factor $\log M$ inside the logarithmic term, mirroring the dyadic scale adaptation of Theorem 6. One caveat differs from the scale grid: there every branch is exactly valid for any $\bar{\gamma}_m$, whereas here a branch with $\hat{p} > p_0$ is not covered by Proposition 9. The fully rigorous instantiation is therefore either a single conservative branch $\hat{p} = p_{\min}$, losing the factor $p_0/p_{\min}$ in constants, or the grid with a branch-validity screening step, which we leave as a technical open point. Finally, the $\widetilde{\mathcal{O}}(\sqrt{KT})$ shape of Theorem 5 is visible once $T \gtrsim \tau_0^{(s)}$: for $T \ll \tau_0^{(s)}$ the harmonic burn-in term dominates the bound, which is why conservative underestimates of p_0 are doubly expensive — they inflate $\tau_0^{(s)}$ quadratically in $\kappa_s/\hat{\kappa}$.*

F.3 The per-arm perturbation floor, and its proof

The body cites this floor as the matching lower bound for the per-arm perturbation class; the statement and the scope remark are repeated here before the proof.

Proposition 10 (Linear location-regret floor under per-arm perturbations)**.** *Fix $\gamma_\star > 0$, $L > 0$, $d \geq 1$, $K \geq 2$, $S > 0$, and $0 < \rho \leq 2LS$. There exist two instances A, B of the per-arm perturbation model* (11) *with exogenous i.i.d. contexts $y_{k,t} = L\,U_t$ (U_t uniform on the unit sphere, shared across arms) such that, for every bandit policy,*

$$\max(\mathbb{E}_A[R_{\mathrm{loc}}(T)],\ \mathbb{E}_B[R_{\mathrm{loc}}(T)]) \ \geq\ \frac{\rho T}{4}\,\mathbb{E}|U_1| \ \geq\ \frac{\rho T}{8\sqrt{d}}, \tag{57}$$

where U_1 denotes the first coordinate of U.

Remark 10 (The linear-in-ρ regret floor is genuine for the per-arm class)**.** *Proposition 10 shows that the linear dependence on ρT in* (51) *cannot be removed uniformly over the per-arm class* (11)*: the*

two instances are observationally identical in the strongest sense, an exact pathwise coupling, while ranking the arms oppositely, so the floor is the price of non-identifiability of the θ-based location benchmark, not a statistical difficulty of learning the reward laws. In TV terms, the shifted arm law is at distance $\mathrm{TV}(\mathrm{Cauchy}(0,\gamma_\star),\mathrm{Cauchy}(\rho,\gamma_\star)) = \frac{2}{\pi}\arctan\frac{\rho}{2\gamma_\star} \le \frac{\rho}{\pi\gamma_\star}$ *from the exact-Cauchy law, so the bounded-shift class nests inside the corresponding TV ball. For the arm-shared (single-law) subclass, where all arms share the same perturbation, no exactly indistinguishable flipped pair exists: observational identity forces the projected gaps to coincide, and the arm-common part of a perturbation cancels in the inter-arm differences that drive the UCB comparisons. The true floor may therefore be smaller; settling it requires a non-exact (Le Cam type) argument, which we leave open. Our bound covers the larger per-arm class, for which the linear floor is genuine.*

Proof of Proposition 10. Take $\Delta = \rho/2$ and $v = (\rho/L)e_1$. Instance A: $\theta_1^A = 0$, $\theta_2^A = -(\Delta/L)e_1$, exact $\mathrm{Cauchy}(0,\gamma_\star)$ noise on both arms. Instance B: $\theta_1^B = 0$, $\theta_2^B = (\Delta/L)e_1$; arm 1 has exact Cauchy noise, and arm 2 has noise $\mathrm{Cauchy}(-s_t,\gamma_\star)$ with $s_t = y_t^\top v$ — a per-arm bounded shift, $|s_t| \le L\|v\|_2 = \rho$, admissible in (11) (deterministic given the round-t contexts). Realize both instances from the same driver: arm 2's noise is $\xi_t \sim \mathrm{Cauchy}(0,\gamma_\star)$ in A and $\xi_t - s_t$ in B; then arm 2's reward in B equals $-\Delta\, x_t + \xi_t$ with $x_t = U_t^\top e_1$, pathwise equal to its reward in A, and every other arm coincides trivially (if $K > 2$, arms $3,\dots,K$ duplicate arm 1 exactly). Hence the realized histories of any policy coincide pathwise in the two instances, and $p_t = \Pr(k_t = 2 \mid y_t)$ is the same in both. In A, arm 1 is optimal iff $x_t > 0$, with gap $\Delta|x_t|$; in B, arm 2 is optimal iff $x_t > 0$, with the same gap. Therefore $\mathbb{E}_A[R_{\mathrm{loc}}(T)] + \mathbb{E}_B[R_{\mathrm{loc}}(T)] = \Delta\sum_{t=1}^T \mathbb{E}|x_t| = \Delta\, T\,\mathbb{E}|U_1|$ for every policy, so $\max(\mathbb{E}_A,\mathbb{E}_B) \ge \frac{\Delta}{2}\,T\,\mathbb{E}|U_1| = \frac{\rho T}{4}\mathbb{E}|U_1|$. Finally, $U_1^2 \sim \mathrm{Beta}\big(\frac{1}{2},\frac{d-1}{2}\big)$ gives $\mathbb{E}U_1^2 = 1/d$ and $\mathbb{E}U_1^4 = 3/(d(d+2))$, and Hölder's inequality yields $\mathbb{E}|U_1| \ge (\mathbb{E}U_1^2)^{3/2}/(\mathbb{E}U_1^4)^{1/2} = \frac{1}{d}\sqrt{\frac{d+2}{3}} \ge \frac{1}{2\sqrt{d}}$ (for $d = 1$ read $U_1 = \pm 1$, so $\mathbb{E}|U_1| = 1$ directly). Finally, both instances satisfy Assumption 1 with $\mu_{\mathrm{dir}} = L^2/d$: the contexts are exogenous i.i.d. uniform on the sphere of radius L, so $\mathbb{E}[y_{k,t}y_{k,t}^\top] = (L^2/d)I_d$ independently of the history (cf. Proposition 4); the pair therefore lies in the class covered by the upper bound of Proposition 7. □

G Experimental details and reproducibility

This appendix reports the five studies behind Section 5: the flagship exact-Cauchy benchmark, in which every constant of the comparison is fixed once; the noise-scale sweep, which moves the parameter the method depends on most; the score-scale misspecification study, which tests Corollary 1; the nearly-Cauchy ablation, which varies the noise law instead of its scale; and a real-data experiment on six decades of daily industry-portfolio returns.

G.1 Configurations of the direct policy

Four configurations of the direct policy appear in these runs, together with the sign and ellipsoidal branches of the same update.

G.1.1 Direct-Cauchy-Tuned

The practical configuration: a constant step size $\eta = 0.05$, a proxy confidence bonus with multiplier $\alpha = 0.08$, an online-Newton preconditioner and a box clip. The two tuned constants are chosen once, in the flagship study, and never changed afterwards, so every later run is a pure check. The exact values, and how they relate to the analysed schedules, are listed as deviations below.

G.1.2 Direct-Cauchy-Scheduled

The analysed schedule of Proposition 1 — the harmonic step size and the explicit confidence radius, with κ_{dir} and ℓ computed from the known instance quantities. The burn-in is capped at a single numerically chosen value, selected once on a dedicated sweep over $\tau_0 \in \{1, 3, 10, 30, 100, 300, 1000, 10^4\}$ and frozen afterwards. The selection rule takes the argmin in the regime where the analysed burn-in fits the horizon and the regret curve has an interior minimum, requiring dominance over all smaller values in the other two regimes; a plain argmin on the flagship geometry is degenerate, since there the regret decreases along the whole grid as the step vanishes. The rule gives $\tau_0 = 1000$.

Table 1: Benchmark suite used in the exact-Cauchy experiments. The critical comparison is between the direct Cauchy score and strong finite-mean robustifications of linear UCB.

Method	Estimation style	Robustness device
Direct-Cauchy-Tuned	One-pass current iterate	Exact Cauchy likelihood score
Huber-LinUCB	Ridge-style linear bandit	Huber loss / clipped residuals
Trunc-LinUCB	Ridge-style linear bandit	Reward truncation
Catoni-LinUCB	Ridge-style linear bandit	Catoni-style robust mean control
MoM-SGD-UCB	One-pass linear update	Median-of-means aggregation
Huber-OnePass-UCB	One-pass current iterate	Huberized score update
LinUCB	Classical ridge-UCB	None

G.1.3 Direct-Cauchy-Grid and Direct-Cauchy-Doubling

The dyadic-bracket adaptation of Theorem 6 and its restarting variant of Remark 1. The grid runs branches over $[0.5, 2]\hat{\gamma}$ ($\hat{\gamma}$ being the true scale in the synthetic studies and the MAD of the training residuals on real data), with surrogate schedules of the theorem and branch failure level $\delta/(K(M+1))$. The doubling row restarts the grid on phases $[\gamma_{\min}, \Gamma_0 2^i]$ with $\Gamma_0 = \gamma_{\min}$, phase lengths $n_i = n_0 2^i$ with $n_0 = T/10$, and phase failure levels $\delta_i = 6\delta/(\pi^2(i+1)^2)$.

G.1.4 The sign and ellipsoidal rows

The sign branch of Theorem 5 uses $\kappa_s = 2p_0\mu_{\mathrm{dir}}$ with p_0 the tightest admissible median-growth constant $(F(B) - 1/2)/B$ of each noise family: in closed form for the contaminated and truncated ladders, by quadrature of the density for the Student-t ladder, by Monte Carlo with a 0.9 safety factor for the α-stable ladder, and as the empirical minimum on the training residuals on real data, discounted by 0.9. The ellipsoidal variant of Theorem 4 uses regularizer $\lambda = L^2$ and, on real data, the safe upper bound $2\hat{\gamma}$ for the scale inside its budget.

The theorem guarantees behind the non-tuned rows are conservative. In the flagship regime the explicit radius, rather than the burn-in, keeps the scheduled row well behind the tuned one; the grid and doubling burn-ins exceed the horizons by five to eight orders of magnitude (measured on the real-data folds below), so those rows quantify the price of theorem-exact constants rather than the method's practical behaviour. The sign branch is the one analysed branch whose burn-in is small out of the box (about 10^{-2} of a real-data fold).

What is reported. Every study plays the same suite — the configurations above against the baselines of Table 1, whose two clipping baselines are handed the true noise scale — and is scored by the latent location regret of Section 4 (on real data it is a frozen-model plug-in, defined in that subsection). Per environment, statistics are the mean over the evaluation seeds; a method *wins* an environment when its environment-mean regret is the lowest of the suite, and win counts are integers by construction with no reported uncertainty measure. Error bars on environment-level figures are standard deviations across environments, i.e. between-environment spread, and the per-step curves carry ± 1 standard-error bands (the real-data curves pool the five folds); the intervals quoted for the real data are 95% moving-block bootstrap intervals (blocks of 21 trading days, 2000 resamples). All figures are redrawn from the saved per-step series or summary files by the experiment entry points (`-only_draw`), so a figure can be restyled without recomputing anything.

G.2 The exact-Cauchy benchmark

The flagship study fixes the constants used everywhere below and establishes the reference comparison against the robust suite. Its instances are stationary linear contextual bandits with $K = 4$ arms, $d = 4$ features and horizon $T = 3000$: at every round a context is drawn from the chosen isotropic family (Gaussian or sphere) and the played arm's reward is its noise-free location plus independent Cauchy noise of scale $\gamma_\star = 0.12$; the latent parameters are drawn once per parameter seed (normal entries with a displacement on the first coordinate), giving six environments (seeds $2026, \ldots, 2031$) per family, each averaged over eight evaluation seeds.

Table 2 reports the reference comparison: Direct-Cauchy-Tuned attains the lowest environment-mean location regret on both families and leads in five of the six environments, its closest competitor Huber-

Table 2: Exact Cauchy noise at scale 0.12. Entries designate environment-sweep location-regret means; wins count how many of the six latent environments each method successfully leads after averaging over eight run seeds per environment.

	Gaussian		Sphere	
Algorithm	Regret	Wins	Regret	Wins
Direct-Cauchy-Tuned	30.00	5/6	30.21	5/6
Huber-LinUCB	40.30	1/6	36.63	1/6
Trunc-LinUCB	40.30	0/6	36.63	0/6
Catoni-LinUCB	99.66	0/6	75.89	0/6
MoM-SGD-UCB	104.40	0/6	101.71	0/6
Huber-OnePass-UCB	119.11	0/6	105.32	0/6
LinUCB	727.88	0/6	809.08	0/6

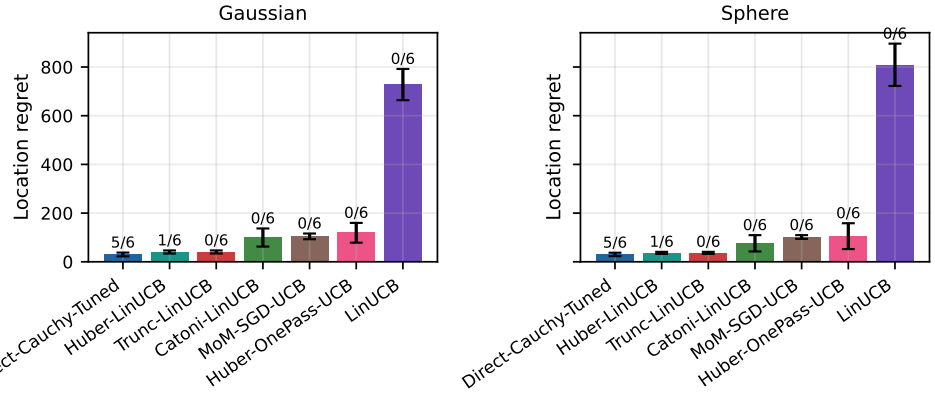


Figure 2: Flagship exact-Cauchy benchmark. Bars are environment-sweep location regret, error bars the spread across the six environments, and the annotation above each bar the number of environments the method leads.

LinUCB taking the remaining one, while classical mean-oriented LinUCB degrades by roughly an order of magnitude and never wins. Figure 2 shows the same hierarchy with the environment-level spread: the margin over the runner-up is larger than the between-environment spread, and the spread of LinUCB, the largest of the suite, is what makes its average an order of magnitude worse.

Two properties of the suite are worth recording here. Huber-LinUCB and Trunc-LinUCB coincide at this configuration: both accumulate the same per-arm ridge statistics, with the same ridge parameters and with the residual clipped to the same threshold $\tau_t = 2.5\,\gamma_\star\sqrt{\log(t+2)}\,(n_a+1)^{1/4}$, which is why their regret entries in Table 2 are identical. Their win counts nevertheless differ ($1/6$ against $0/6$ on both context families); the tie-breaking rule behind the win counts is not documented in the saved run artifacts. Under exact Cauchy noise the truncation bias of both does not vanish, which is the conceptual reason the paper replaces truncation by the bounded Cauchy score. Finally, MoM-SGD-UCB aggregates per-block gradient means with five blocks, fewer than the $8\log(1/\delta)$ a genuine $1-\delta$ median-of-means guarantee requires, so it carries a weak confidence by construction.

G.3 The noise-scale sweep

The sweep moves the one parameter the method depends on most, the noise scale, to locate where its advantage ends. The instances are those of the flagship study with $\gamma_\star \in \{0.03, 0.06, 0.12, 0.24, 0.48\}$ and four evaluation seeds per environment, as the caption of Figure 1 records; everything else is unchanged. Figure 1 of the main text traces the crossover: the direct update and the clipping baselines stay close at the small scales, Catoni-LinUCB falls steeply from its poor small-scale behaviour, and by scale 0.48 it is the best of the five rows shown on both families while LinUCB has moved an order of magnitude away from every other method. The direct implementation leads both families from 0.03 through 0.12; at 0.24 it remains best on Gaussian contexts while Catoni-LinUCB overtakes

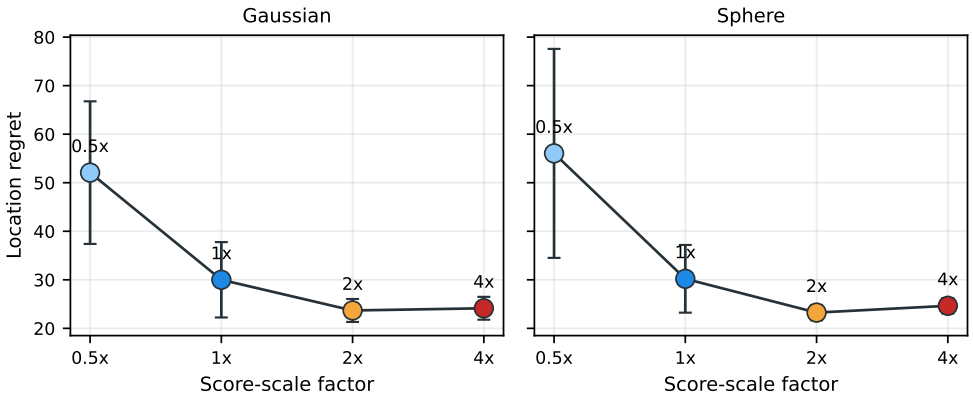


Figure 3: Score-scale misspecification: regret against the multiplier applied to the learner's score scale, the true Cauchy scale being 0.12.

it on the sphere; at 0.48 Catoni-LinUCB leads on both, consistently with the contraction modulus degrading quadratically in the scale-to-signal ratio.

G.4 The score-scale misspecification study

The misspecification study runs the tuned configuration and tests the constant-factor regime of Corollary 1: the noise is held at the reference scale 0.12 and the learner's score scale is multiplied by $0.5, 1, 2, 4$ around it, everything else staying at the flagship values, with eight evaluation seeds per environment. Figure 3 shows the asymmetric sensitivity: the regime below the matched scale is both worse and far more variable across environments than the regime above it, and at this horizon doubling or quadrupling the scale is slightly better than the matched one — the finite-sample reading of the constant-factor claim.

G.5 The nearly-Cauchy ablation

The ablation keeps the noise scale matched and varies the noise law instead, to show where the exact-Cauchy model stops describing the environment and which method should be used there. Its instances use $K = d = 4$ arms and features, horizon $T = 2000$, four latent environments (the parameter seeds $2026, 2027$ crossed with the two context families), context scale $c = 0.25$, so $L = 0.5$ and $\mu_{\mathrm{dir}} = c^2$ (the smallest eigenvalue of the design's second moment), and parameter radius $S = 1.5$, hence $B = 1.5$ and $\Gamma = 2\gamma_\star = 0.8$ — the viable regime in which the analysed schedules are alive; each cell averages six run seeds. Four ladders move the shape of the tail: α-stable, Student-t, contaminated Cauchy, and truncated Cauchy; each rung is rescaled so that its median absolute deviation equals that of $\mathrm{Cauchy}(0, 0.4)$, so only the shape changes, except the contaminated ladder, which keeps the core scale at 0.4 and varies the tail mass instead (matching its MAD would shrink the core scale as the mass grows and conflate the two effects, so its measured MAD is recorded with the run). The measured tail index of each rung — a Hill estimate on the largest 2% of the noise samples, 1 for Cauchy and 2 for Gaussian noise — is the natural axis of the comparison.

Four Cauchy-score rows play the suite, configured as in Appendix G.1: the tuned configuration, the scheduled configuration at failure level δ/K, the sign branch, and the grid at branch failure level $\delta/(K(M+1))$. The baselines are the same suite.

On Cauchy-like tails (measured index up to about 1.5) Direct-Cauchy-Tuned attains the lowest regret of the suite, and the classical ridge baselines are not usable: Huber-LinUCB and LinUCB pay two to four times as much regret as the Cauchy-score rows. The scheduled configuration is second of the eleven rows under exact Cauchy noise, ahead of every ridge baseline, so the ordering is not an artefact of tuning. As the tail lightens the ordering turns over: beyond index ≈ 1.5 Catoni-LinUCB pulls ahead, and with bounded (truncated) noise the ridge baselines are ahead of every Cauchy-score method — the expected behaviour of a score calibrated for infinite-variance noise. The contaminated ladder, which keeps the Cauchy tail index but adds a four-times-wider component, is the sharpest

trade-off: the tuned configuration wins at every weight up to 0.25 and ties Catoni-LinUCB at 0.5, while Huber-LinUCB and LinUCB are two-and-a-half to four times worse. Figure 4 locates the turnover within each ladder, and Figures 5 and 6 give the same runs as functions of the round for three representative rungs (exact Cauchy, the turnover rung $\alpha = 1.5$, and the bounded rung): on the exact-Cauchy rung the per-round regret of the tuned configuration falls by a factor of five between the first few hundred rounds and the end of the horizon, the analysed schedule converges more slowly, Huber-LinUCB stalls at a level several times higher, and uniform sampling never improves at all; on the bounded rung the ridge baselines converge twice as far as any Cauchy-score row, which is the turnover seen as a crossing of curves. The branches that avoid knowing $\gamma_\star$ trade regret level for stability: the grid, which needs only a bracket, is behind Direct-Cauchy-Tuned on all seventeen rungs but varies least across the four noise families, whereas the ridge baselines swing most.

G.6 The real-data experiment

The real-data experiment replaces the synthetic instance by six decades of daily industry-portfolio returns, to show that the conclusions are not an artefact of the generative model. The data are the daily value-weighted returns of the ten US industry portfolios of the Kenneth R. French Data Library (26 317 daily rows, 1926-07-01 to 2026-08-31; the SHA-256 of the normalized CSV is recorded with the run). Their noise is heavy-tailed but not Cauchy: the measured tail index is 2.54 for the returns and 1.98 for the residuals of a linear location model (Hill estimates on the largest 5% of $|x|$), and the empirical survival function falls below a MAD-matched Cauchy reference in the far tail (Figure 7); the instances are non-stationary in the usual market sense, with volatility clustering.

The evaluation is walk-forward with expanding training windows, so that no quantity used inside a fold is computed from that fold's or any later data. The first 40% of the rows form the initial training window; the remainder splits into five contiguous evaluation folds of about $3\,155$ trading days each ($15\,779$ in total, 1963–2026). Fold i is scored with all rows strictly before its start as the training window, and once the fold is scored its rows join the training window of fold $i + 1$. The whitening statistics, the frozen location model below, and every algorithm parameter are fit on the training window only. The context is one market-state vector per day, shared by all arms and computed from data up to the previous day (5-day market momentum, 20-day market volatility, the previous day's return and cross-industry dispersion), standardized and clipped using training statistics only, which fixes $L = 2$; the measured directional-excitation constant on the training designs is 0.15–0.17. On real data the noise-free location of the location regret is not observable, so it is replaced by a frozen Huber (IRLS) fit of each arm's return on the training window, never refitted on the evaluation rows; the model-free complement is the full-feedback realized regret $\max_k r_t(k) - r_t(k_t)$, computable because every portfolio's return is on record each day.

The configurations are those of Appendix G.1, with every quantity computed on training rows only: the Cauchy-score rows are told $\hat{\gamma}$, the MAD of the training residuals (which equals the scale for an exact Cauchy law); the parameter radius is $S = 1.01 \max_k \|\hat{\theta}_k\|$, the tightest value containing the fitted locations; and the clipped baselines receive the same known scale $\hat{\gamma}$. The tuned configuration is unchanged.

The tuned configuration attains the lowest regret of the suite on both metrics, and its bootstrap intervals are disjoint from the closest baselines': on the frozen-location metric it leads by 30–34% per day and wins four of the five folds (the grid taking the 2014–2026 fold), and on the realized metric it leads by about 13%. Figure 8 shows the cumulative location regret inside each fold: the ordering is stable across the five eras. With the frozen burn-in the analysed schedule is alive here, at the level of the sign branch; the grid and doubling rows accumulate regret at roughly the uniform-sampling rate, since their analysed burn-ins exceed the fold length by a factor of 3.8×10^5–1.4×10^8, and the doubling restarts do not pay off within a fold. Figure 9 reports the model-free complement with its 95% moving-block bootstrap intervals: the ordering matches the location metric, so the conclusion does not depend on the frozen location model. Averaged over the five folds, the average regret of the tuned configuration falls from about 0.0017 at day 250 to 0.0014 at the end of a fold (it is still falling), while LinUCB and the ridge baselines level off near 0.0018 after an early dip, the sign branch stays flat at about 0.0019, the analysed schedule holds at 0.0020, and the grid and doubling rows rise to 0.0027; the full-feedback metric orders the rows identically. The real data therefore supports the moderate-noise conclusion of the synthetic studies. The remaining gap of the analysed schedule is

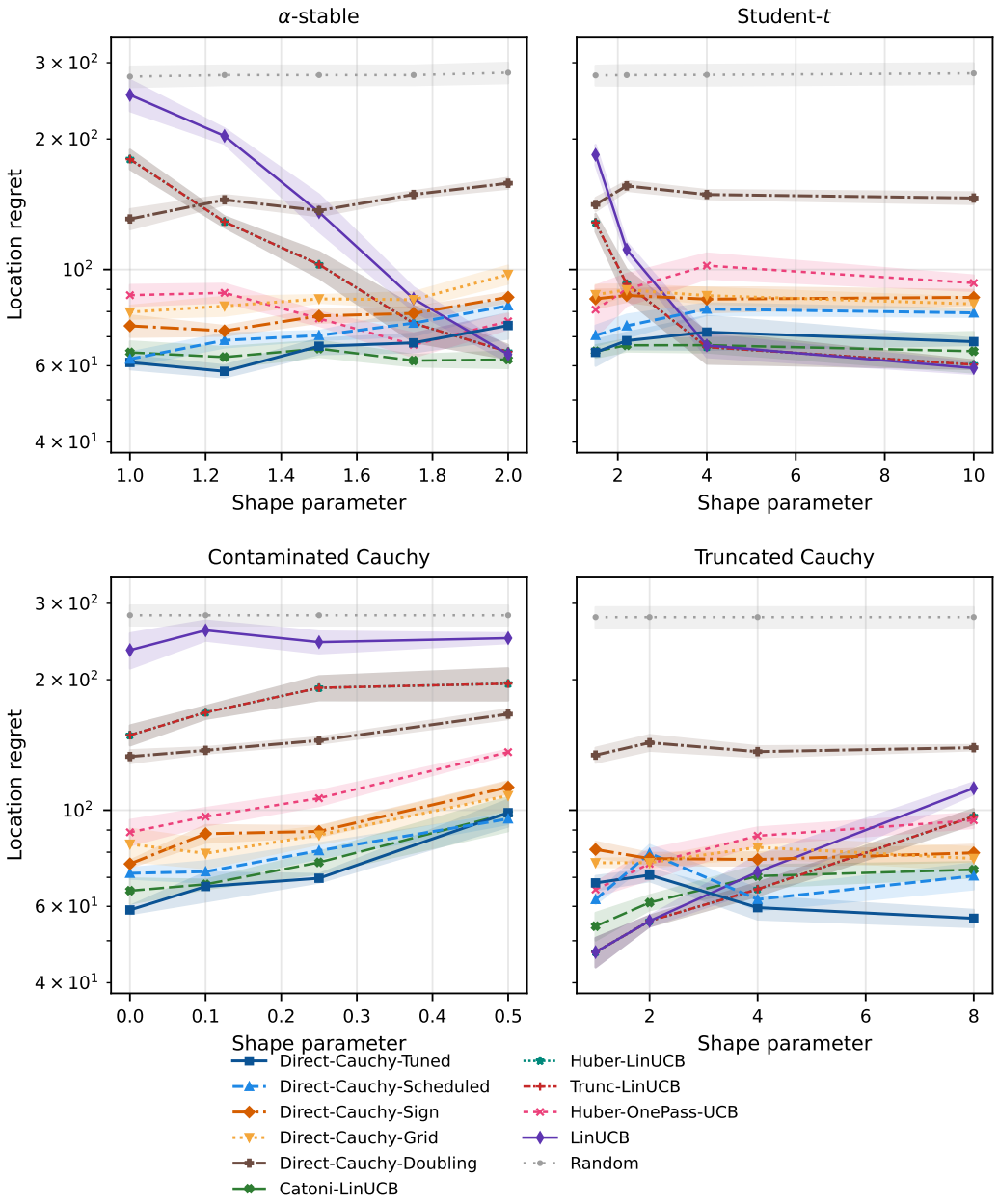


Figure 4: Nearly-Cauchy ablation, one panel per ladder, with the family's shape parameter on the horizontal axis, on a logarithmic regret axis.

explained by the scale of the theorem-exact radius rather than the burn-in, and at this signal-to-noise ratio the tuned configuration is the row to use.

G.7 Constants and deviations from the analysed schedules

Most constants of the comparison are fixed once in the flagship study and never changed, as stated in the opening; the instance-dependent constants follow the rules stated with each experiment, and on real data the feature clip is 1 and the Huber threshold one residual MAD. The list below records the seven points at which the implementation used in our experiments departs from the analysed schedules or assumptions: four deviations that affect the tuned direct-policy row, one scope note on

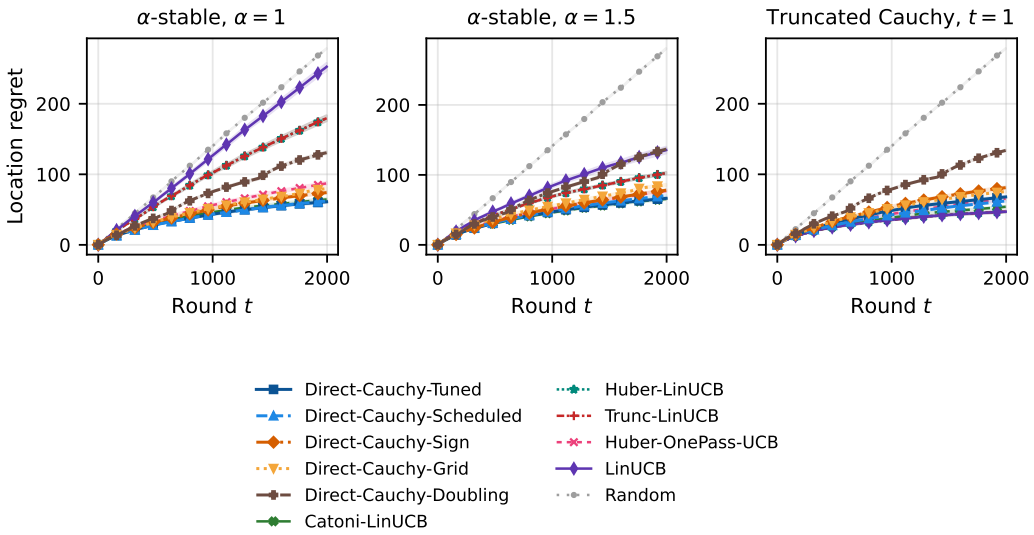


Figure 5: Nearly-Cauchy ablation: cumulative location regret against the round for three rungs, all rows.

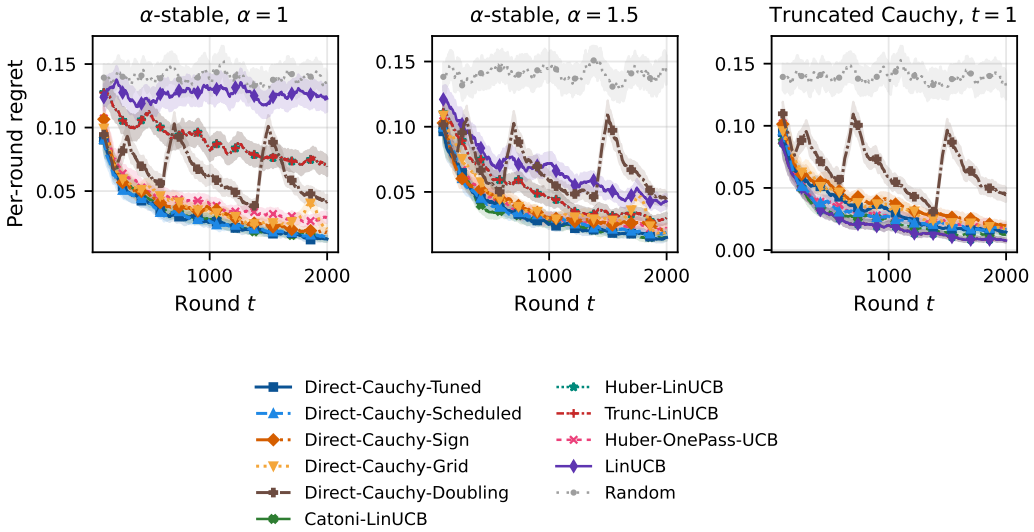


Figure 6: Nearly-Cauchy ablation: per-round location regret (100-round mean) for the same three rungs; the end-of-horizon level of each curve is its convergence speed.

the adaptation mechanism, one protocol difference concerning the baselines, and one scope caveat about coverage rather than deviations.

1. *Step size.* The theorems assume the harmonic schedule $\eta_\tau = 1/(\kappa_{\mathrm{dir}}(\tau + \tau_0 + 1))$ of Proposition 1; the tuned configuration uses a constant step size, $\eta = 0.05$, tuned. The scheduled row runs the harmonic schedule (Appendix G.1).

2. *Confidence radius.* The optimistic bonus is a proxy $\alpha\|y\|_2\sqrt{\log(t+2)}/(n_a+1)^{1/2}$ with $\alpha = 0.08$: it keeps the leading $(\tau+\tau_0)^{-1/2}$ decay of the radius β_τ of Proposition 1, but folds A_{Cau}, B_{Cau} and δ into that tuned multiplier and omits the $1/(\tau+\tau_0)$ term of A_{Cau}.

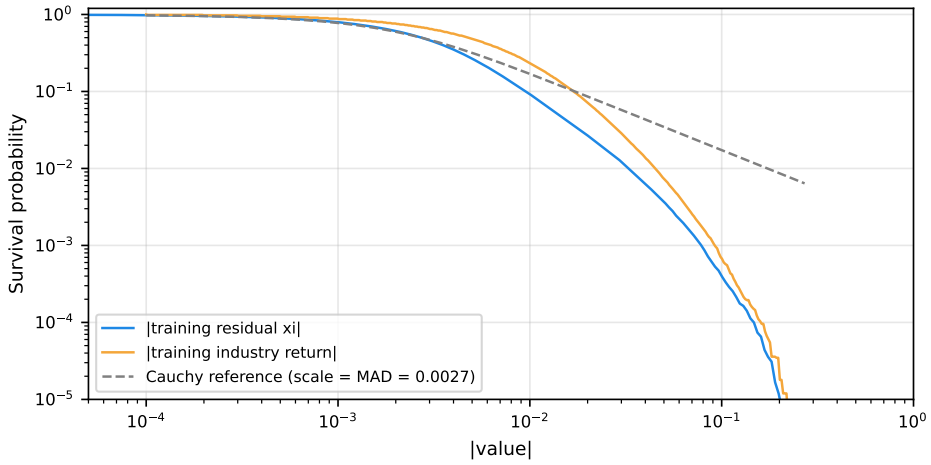


Figure 7: Real data: survival functions of the daily returns and of the frozen-location residuals, against a Cauchy reference with the MAD-matched scale.

The tuned values quoted here are those recorded in the run configuration committed with the code.

3. *Projection.* The code clips to the box $\{x : \|x\|_\infty \le S\}$ rather than to the Euclidean ball $\mathcal{X} = \{x : \|x\|_2 \le S\}$ used in Section 3; the box contains the ball, so the iterates remain bounded, but its diameter is $\sqrt{d}$ times larger.
4. *Preconditioning.* The direct-policy runs use an online-Newton preconditioner in place of the plain projected gradient step analysed in Section 3.
5. *Grid and doubling.* The dyadic-grid adaptation of Theorem 6 and the restarting doubling schedule of Remark 1 are implemented as the Direct-Cauchy-Grid and Direct-Cauchy-Doubling rows, parameterized per the theorem and the remark. The misspecification study varies a single policy's score scale by fixed multipliers, which tests the constant-factor regime of Corollary 1 rather than the model-selection mechanism of Theorem 6.
6. *Baseline knowledge.* The two baselines that clip or truncate the reward are given the true noise scale, an oracle quantity the direct policy does not require; Corollary 1 instead covers a constant-factor mismatch of the score scale.
7. *Context bounds.* The sphere design satisfies the bounded-context assumption exactly; in the Gaussian design the contexts have unbounded support, so L is a high-probability surrogate there.

These deviations affect how tightly the measured curves should be compared with the constants of the bounds; they do not change the metric (location regret), the environments or the reported values.

G.8 Reproducibility

Commands. Each run is driven by a configuration file committed with the code, so each of the five studies is reproduced by a single command, followed by the figure builder for Figures 2, 1 and 3; the newer experiments, including the burn-in selection sweep, have their own entry points and configuration files:

```
# entry points live in experiments/, configurations in configs/
run_paper_experiments.py \
    --config configs/paper_experiments.yaml     # flagship
run_paper_experiments.py \
    --config configs/paper_scale_sweep.yaml     # scale sweep
run_cauchy_misspec_experiment.py \
    --config configs/cauchy_misspec.yaml        # misspecification
```

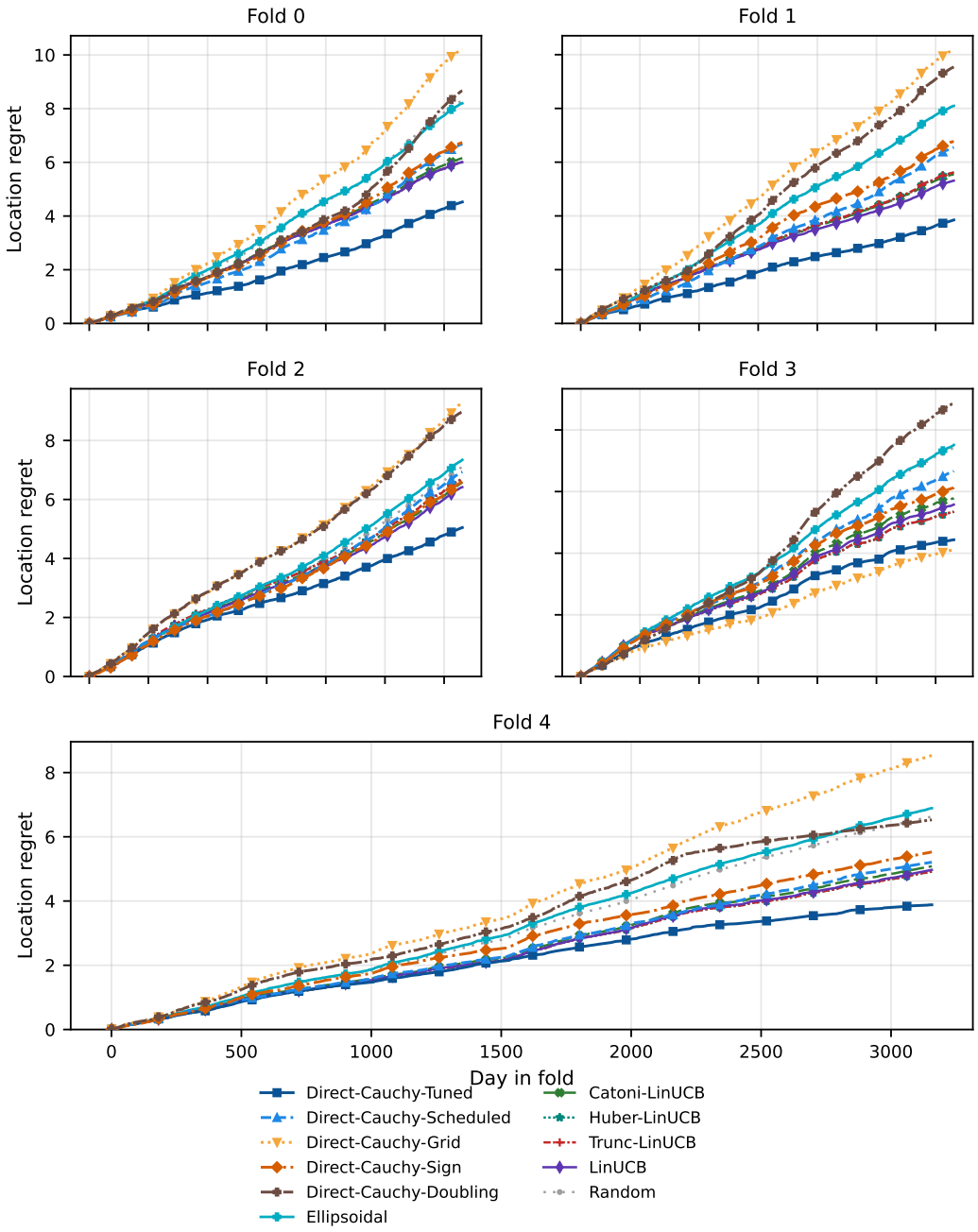


Figure 8: Real data: cumulative location regret inside each walk-forward fold, all rows.

```
run_nearly_cauchy_experiment.py \
    --config configs/nearly_cauchy.yaml          # nearly Cauchy
run_real_data_experiment.py \
    --config configs/real_data.yaml              # real data
run_tau0_sweep_experiment.py \
    --config configs/tau0_sweep.yaml             # burn-in selection sweep
make_paper_benchmark_figures.py                  # the exact-Cauchy figures
```

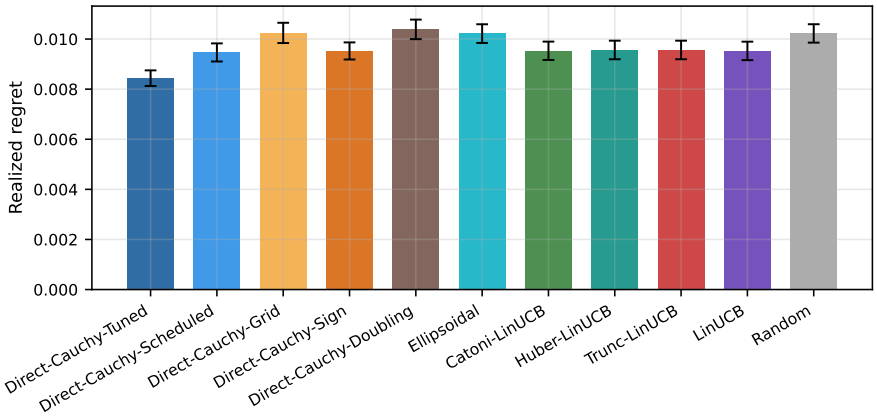


Figure 9: Real data: full-feedback realized regret pooled over the folds, with 95% moving-block bootstrap intervals.

What reproduces. Re-running the flagship and the scale sweep from the committed configurations reproduces the published entries of Table 2 to the printed precision (the largest deviation over the fourteen numbers is 0.005, at LinUCB's 809) and the leadership pattern of Figure 1 cell by cell, so the numbers of this section are reproducible from the repository as published. Every run writes its fully resolved configuration, environment seeds, command line, package versions and the policy constructor arguments into a `metadata/` directory, together with the raw metrics, the table rows and the per-step curves; a top-level `summary.csv` holds the paper-ready table, and the figures are drawn from those artifacts rather than re-simulated. The pre-existing `results/paper_experiments_*` directories are deliberately reduced verification runs (two run seeds and a horizon of 30 instead of 3000), so they are not the record of the runs behind the reported numbers; the reproductions live in `results/paper_reproduction/` and `results/paper_reproduction_scale/`, each a fresh run of the current code. A fast regression suite guards the policies and environments against silent breakage.

Compute. The experiments were conducted on a MacBook Pro with an M1 Pro chip; no GPU is required. Every configuration is small: six environments, at most eight seeds per environment and 3000 rounds per run, across seven methods. A full sweep of this size completes in minutes, and the repository's reduced trial configuration of the same sweeps runs in seconds.

H Limitations and Discussion

This section collects the limitations stated in the body and the open questions that accompany the results.

Model scope. The theoretical scope is symmetric noise with a median-growth condition, with the sharpest results for exact symmetric Cauchy noise; broader bounded-score families are left for future work. The perturbation floor of Proposition 10 is proved for the per-arm deviation class; the arm-shared case, where the common part cancels in the inter-arm comparisons, requires a non-exact (Le Cam type) argument that is left open (Remark 10).

Excitation. The main regret guarantees rely on excitation conditions. Windowed persistent excitation relaxes Assumption 1 to a cumulative window condition (Section 4.2), but persistently degenerate stochastic context processes can violate it. Dropping excitation entirely is possible only through the ellipsoidal variant, at $\mathcal{O}(d^2)$ per-arm memory and a degraded $\widetilde{\mathcal{O}}(\sqrt{dK}\,T^{3/4})$ rate with a $T^{1/4}$ gap to the lower bound; whether that rate is minimax is open (Remark 6), and the accompanying $\sqrt{d}$ comes from the ellipsoidal analysis. Whether the $\sqrt{w}$ factor of the stochastic windowed variant can be removed is likewise open (Remark 3).

Knowledge requirements. The sign update requires a known lower bound on the median-growth constant p_0 or a conservative underestimate, and grid adaptation over p_0 leaves branch-validity screening open (Remark 9). The scale adaptation of Theorem 6 comes with conservative constants: the analysed burn-ins exceed the horizons of our experiments by orders of magnitude, which is why the reported row is the tuned configuration and the analysed schedule is evaluated separately with a numerically chosen burn-in (Appendix G).

Empirical scope. The benchmarks cover synthetic exact-Cauchy and nearly-Cauchy environments and one real return dataset; they are aligned with the location-regret setting and should not be read as evidence for all real-world heavy-tailed problems. In particular, the adversarial-context results of Section 4.2 and the perturbation guarantees of Section 4.3 have no experimental counterpart yet, and ablations in those regimes are future work.