ROUND-TRIP RESIDUAL UCB

Anonymous authors
Paper under double-blind review

ABSTRACT

We analyze the *round-trip residual* — the feature-space reconstruction error $R_t(x) = \|(C_t A_t - I)u_t(x)\|$ of a forward ridge head A_t and an inverse ridge head C_t trained on true observed pairs — as an exploration bonus for bandits with vector feedback. The analysis rests on a *store-once* design: every pair is embedded with the feature maps current at acquisition and never re-embedded, which preserves an exact algebraic identity, keeps the design matrix rank-one monotone (no feature-drift comparability conditions), and preserves the martingale structure needed under noise. Results: (i) the identity $R_t(x) = \lambda\|M_{t-1}G_t^{-1}u_t(x)\|$, $M_{t-1} \succeq I$, for arbitrary stored matrices; (ii) noiseless regret $2\|w\|S_A(1 + s_A^{-2})\sqrt{2\lambda T\gamma_T}$ with frozen features, and the same order under an ℓ^2-summable drift budget up to an $\widetilde{O}(E_\infty T^{3/4})$ term removable by horizon-calibrated drift; (iii) a tightness principle: the link $\omega_t \le \sqrt{L_U R_t/\lambda}$ between residual and elliptical width is saturated on every eigendirection, so within the pointwise-domination proof template no R-only bonus avoids the $T^{3/4}$ term when the error has a width-shaped component (formal lower bound open); (iv) observation noise makes that component unavoidable with coefficient $\Theta(\sigma\sqrt{\gamma_T})$, and — under a noisy residual-bound condition, certified only for noise amplitude $B_e = O(T^{-1/2})$ and otherwise assumed — round-trip UCB degrades to $\widetilde{O}(\gamma_T^{3/4}T^{3/4})$, or $\widetilde{O}(\gamma_T^{2/3}T^{2/3})$ with tuned λ, a polynomial gap from the $\widetilde{O}(\gamma_T\sqrt{T})$ of width-based bonuses. A finite-width NTK instantiation certifies the drift budget (trust-region schedule) and the design conditioning (Gram eigenvalue condition at polynomial width), and grounds the realizability norm in NTK-RKHS norms; the output-embedding margin s_A remains the design's one open structural assumption. All proofs are self-contained, modulo one imported self-normalized concentration inequality; auxiliary linear-algebra facts are proved in Appendix A.

1 INTRODUCTION

If a forward model maps inputs to predicted outcomes and an inverse model maps outcomes back, inputs on which the models are unreliable should fail to survive the round trip — an uncertainty signal computable by applying trained maps, with no confidence-matrix inversion. In ridge regression the intuition is exact: for paired matrices X, Z and ridge estimators $\widehat{W} = ZX^\top(XX^\top + \lambda I)^{-1}$, $\widehat{B} = XZ^\top(ZZ^\top + \lambda I)^{-1}$,

$$(\widehat{B}\widehat{W} - I)x = -\lambda\big[I + X(Z^\top Z + \lambda I)^{-1}X^\top\big](XX^\top + \lambda I)^{-1}x, \tag{1}$$

so reconstruction failure equals ridge *shrinkage* — the object controlling the forward prediction error — up to a bounded matrix factor. This paper develops the bandit theory of this mechanism in full: algorithm, identities, regret bounds in noiseless and noisy settings, and matching structural limits.[1]

Two design commitments shape everything. **Lifted outputs:** the inverse cannot be conditioned on raw outputs once the explored span exceeds the output dimension, so we lift, $v(y) = [\,y\,;\,\tau\phi_Y(y)\,]$, keeping y in the first block so the reward stays a fixed linear functional $w = [\mu; 0]$. **Store-once embedding:** with evolving (e.g., trained) feature maps, each pair is embedded at acquisition and

[1] This manuscript consolidates a sequence of working notes of the same project; companion notes develop the generalized-linear precursor identity and a width-based (identity-head) alternative.

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.

never re-embedded. Consequently the round-trip identity holds verbatim for the stored matrices (Section 3); the design gains one column per round, so the elliptical potential applies to mixed-time features with no spectral-comparability assumptions (Section 4); and the stored design is predictable for the noise filtration, so self-normalized concentration applies under noise (Section 6) — a property re-embedding destroys.

Contributions.

1. *Identity and sandwich* (Theorem 1): for arbitrary stored matrices, $\lambda\|G_t^{-1}u\| \le R_t(x) \le \bar{M}\sqrt{\lambda}\,\omega_t(x)$ under a margin condition, ω_t the elliptical width.
2. *Noiseless regret* (Theorems 2–3): $\widetilde{O}(\sqrt{\lambda T \gamma_T})$ with frozen features; with trainable features under an ℓ^2 drift budget, plus $\widetilde{O}(E_\infty T^{3/4})$, removable by horizon-calibrated drift.
3. *Tightness* (Proposition 1): the residual-to-width inequality is saturated direction-by-direction — R_t compresses the width by $\sqrt{\lambda/(\lambda+\mu)}$ in explored directions — so within the pointwise-domination template, any R-only bonus covering a width-shaped error of coefficient c pays $\Omega(cT^{3/4})$-type sums; no formal bandit lower bound is claimed.
4. *Noisy regret and its limit* (Theorem 5): noise injects a width-shaped error with non-tunable coefficient $\beta_T = \widetilde{O}(\sigma\sqrt{\gamma_T})$; under a noisy residual bound (certified only for noise amplitude $B_e = O(T^{-1/2})$, Proposition 4; otherwise assumed), round-trip UCB incurs $\widetilde{O}(\gamma_T^{3/4}T^{3/4})$, improvable only to $\widetilde{O}(\gamma_T^{2/3}T^{2/3})$ by tuning λ — versus $\widetilde{O}(\gamma_T\sqrt{T})$ for width-based bonuses.
5. *NTK instantiation* (Section 5): conditioning certified at polynomial width, drift certified by a trust-region schedule, S_A grounded in NTK-RKHS norms (Lemma K), with a finite-width regret theorem and a quantitative τ-tradeoff; the margin s_A is isolated as the single remaining structural assumption.
6. *Implementation* (Section 7): both heads computable by warm-started GD with weight decay, $O(\kappa_T \log T)$ steps per round ($\kappa_T = 1 + TL_U^2/\lambda$; $\sqrt{\kappa_T}$ with acceleration) for cumulative optimization error $O(T^{-1})$; exact kernelized reference.

2 Setup and Assumptions

2.1 Model and algorithm

At round $t \le T$ the learner sees a finite candidate set $\mathcal{X}_t \subset \mathbb{R}^d$, picks x_t, observes

$$y_t = F^\star(x_t) + e_t \in \mathbb{R}^q, \tag{2}$$

with $e_t \equiv 0$ in Sections 3–5 and $\mathbb{E}[e_t \mid \mathcal{F}_{t-1}, x_t] = 0$, $a^\top e_t$ conditionally σ-sub-Gaussian for every unit a, and $\|e_t\| \le B_e$ in Section 6.

Expected reward $r^\star(x) = \mu^\top F^\star(x)$. Regret

$$\mathrm{Reg}(T) = \sum_t [r^\star(x_t^\star) - r^\star(x_t)]$$

with $x_t^\star \in \arg\max_{x\in\mathcal{X}_t} r^\star(x)$.

At round t the learner holds maps $u_t : \mathbb{R}^d \to \mathbb{R}^p$ and $v_t : \mathbb{R}^q \to \mathbb{R}^{q'}$ where $v_t(y) = [\,y\,;\,\tau\phi_t(y)\,] \in \mathbb{R}^{q'}$. Let $w = [\mu; 0]$, so $r^\star(x) = w^\top v_t(F^\star(x))$ for all t.

A fixed *reference pair* $(\bar{u}, \bar{v})$ anchors the analysis (initial maps, or the limit of a settling trajectory).

Store-once: after observing y_s, store $u_s := u_s(x_s)$, $v_s := v_s(y_s)$; stack U_{t-1}, V_{t-1} and $G_t := \lambda I + U_{t-1}U_{t-1}^\top$. The heads are the ridge minimizers on stored pairs:

$$A_t = V_{t-1}U_{t-1}^\top G_t^{-1}, \qquad C_t = U_{t-1}V_{t-1}^\top(V_{t-1}V_{t-1}^\top + \lambda I)^{-1}. \tag{3}$$

With $u := u_t(x)$, the scores are $\widehat{r}_t(x) = w^\top A_t u$ and $R_t(x) = \|(C_tA_t - I)u\|$.

Algorithm 1 Store-Once Round-Trip UCB

Require: regularization $\lambda \geq \max\{1, L_U^2\}$; bonus multipliers $b, \tilde{b} \geq 0$ (set per theorem; $\tilde{b} = 0$ in the frozen noiseless case); lift scale $\tau \geq 0$; initial feature maps $(u_1, v_1) = (u_{\theta_0}, v_{\psi_0})$ from (14); trust-region budgets $(m_s)_{s\geq 1}$; head tolerance $\varepsilon_{\text{head}}$ from Lemma 9; initial pool P_0 ($= \mathcal{X}_{T,K}$ in the known-pool setting of Lemma C; $= \emptyset$ online)

1: $U_0 \leftarrow [\,]$, $V_0 \leftarrow [\,]$
2: **for** $t = 1, 2, \ldots, T$ **do**
3: receive $\mathcal{X}_t$; $P_t \leftarrow P_{t-1} \cup \mathcal{X}_t$
4: freeze the round's maps (u_t, v_t)
5: $A_t, C_t \leftarrow$ warm-started GD with weight decay on the stored (U_{t-1}, V_{t-1}), run to tolerance $\varepsilon_{\text{head}}$ {exact ridge solutions (3) as reference}
6: **for** $x \in \mathcal{X}_t$ **do**
7: $u \leftarrow u_t(x)$; $\widehat{r}_t(x) \leftarrow w^\top A_t u$; $R_t(x) \leftarrow \|(C_t A_t - I)u\|$
8: **end for**
9: $x_t \leftarrow \arg\max_{x\in\mathcal{X}_t} \left[\widehat{r}_t(x) + b\, R_t(x) + \tilde{b}\sqrt{R_t(x)}\,\right]$; play x_t, observe y_t
10: **store once:** append column $u_t(x_t)$ to U_t and $v_t(y_t)$ to V_t
11: **if** representation training is enabled **then**
12: $\theta_{t+1} \leftarrow \textsc{RepUpdate}\big(\theta_t, \{(x_s, v_s)\}_{s\leq t}, P_t, m_t\big)$ (Algorithm 2); $u_{t+1} \leftarrow u_{\theta_{t+1}}$
13: **else**
14: $u_{t+1} \leftarrow u_t$
15: **end if**
16: $v_{t+1} \leftarrow v_t$ {output lift frozen by default; see Section 5.1}
17: **end for**

Round-Trip UCB selects

$$x_t \in \arg\max_{x\in\mathcal{X}_t} \left[\widehat{r}_t(x) + b\, R_t(x) + \tilde{b}\,\sqrt{R_t(x)}\right], \tag{4}$$

with $b, \tilde{b}$ set per theorem ($\tilde{b} = 0$ in the frozen noiseless case).

Denote $\omega_t(x) := \sqrt{u^\top G_t^{-1} u}$ and $\gamma_T := \log\det(I + \lambda^{-1} U_T^\top U_T)$. γ_T^{ref} for the reference features.

The full procedure is Algorithm 1. Line 10 (the stored vector is the scored vector, by construction) is what Lemma 4(a)'s monotone-potential argument consumes; the placement of line 12, after scoring and storing, constrained on the pool that already contains $\mathcal{X}_t$ (line 3), is what the pool-relative drift lemma of Section 5.4 consumes.

2.2 Assumptions

A1 (Reference realizability). $\exists A^\star$, $\|A^\star\|_{\text{op}} \leq S_A$, with $\bar{v}(F^\star(x)) = A^\star \bar{u}(x)$ for all candidates of all rounds.

A2 (Drift budget, pool-relative, budget form). For every s and every $x \in \mathcal{X}_s \cup \{x_1, \ldots, x_{s-1}\}$ (all the analysis evaluates at round s):

$$\|u_s(x) - \bar{u}(x)\| \leq \varepsilon_{X,s}$$

and

$$\|v_s(F^\star(x)) - \bar{v}(F^\star(x))\| \leq \tau\varepsilon_{Y,s}.$$

Denote $\Delta_s := \tau\varepsilon_{Y,s} + S_A\varepsilon_{X,s}$, $\varepsilon_{X,s} \leq c_\varepsilon$, and finite aggregate budgets

$$E_U^2 := \sum_s \varepsilon_{X,s}^2, \qquad E_\infty^2 := \sum_s \Delta_s^2, \qquad \Sigma_\Delta := \sum_s \Delta_s. \tag{5}$$

The analysis consumes only these aggregates (and Δ_t at score time); no pointwise decay is required. The decaying schedule $\Delta_s \leq c_\Delta s^{-(1/2+\delta)}$ is one sufficient instance, giving $E_U, E_\infty = O(1)$ and $\Sigma_\Delta = O(T^{1/2-\delta})$ for $\delta > 0$; a uniform finite-horizon bound $\Delta_s \leq \bar{\Delta}$ is another, giving $E_\infty \leq \sqrt{T}\bar{\Delta}$, $\Sigma_\Delta \leq T\bar{\Delta}$. Frozen case: $\varepsilon \equiv 0$.

A3 (Reference margin). $\|A^\star z\| \geq s_A \|z\|$ on $\mathrm{range}(\bar{U}_{t-1})$, $\forall t$.

A4 (Reference conditioning). $\lambda_{\min}(\bar{U}_{t-1}^\top \bar{U}_{t-1}) \geq \kappa_0^2 > 0$ for $t > 1$, and the smallness condition

$$\frac{(s_A + S_A)E_U + E_\infty}{\kappa_0 - E_U} \leq \frac{s_A}{2}, \qquad E_U < \kappa_0. \tag{6}$$

A4 is restrictive: it excludes repeated contexts (duplicates give $\lambda_{\min} = 0$) and histories beyond the effective feature rank. In the NTK instantiation, it follows from the pool eigenvalue condition by interlacing (Fact A.5). Accordingly, *all theorems that use A4 are stated for distinct selected contexts only.*

A5 (Boundedness). $\sup_x \|\bar{u}(x)\| \leq L_u$, hence $\|u\| \leq L_U := L_u + c_\varepsilon$ everywhere relevant; $\lambda \geq \max\{1, L_U^2\}$; the lift is L_ϕ-Lipschitz on the output region (noisy section only).

Remarks. A1 is automatic on a finite pool with nonsingular reference Gram; its content is the norm S_A (grounded in Section 5.3). A3 is the structural heart: *the lifted output embedding must not collapse input directions the learner can distinguish.*

The tiers are bookkeeping regimes for *how much representation movement is allowed, and what parts of the theory are certified versus assumed.*

Frozen: *No feature updates: $m_s \equiv 0$. Then*

$$E_U = E_\infty = \Sigma_\Delta = 0.$$

This is the cleanest case: the regret is $\widetilde{O}(\sqrt{T\gamma_T})$, and there is no drift penalty.

Tier C, certified: *The reference is initialization:*

$$(\bar{u}, \bar{v}) = (u_0, v_0).$$

The total feature movement is small:

$$B_X, B_Y \leq c_b T^{-1/2}.$$

This certifies A2/A4/A5 from initialization events, but it gives

$$E_\infty = O(1), \qquad \Sigma_\Delta = O(\sqrt{T}).$$

*So for the pure R_t-based score, the drift-width term produces a $\widetilde{O}(T^{3/4})$ contribution. Tier C is "certified" but does **not** preserve the optimal $\sqrt{T}$ rate for the pure round-trip score.*

Tier C$^+$, strengthened certified: *Same initialization-reference logic, but with smaller movement:*

$$B_X, B_Y \leq c_b T^{-3/4}.$$

Then

$$E_\infty = O(T^{-1/4}), \qquad \Sigma_\Delta = O(T^{1/4}).$$

This makes the problematic $E_\infty T^{3/4}$ term become $\widetilde{O}(\sqrt{T})$. So Tier C$^+$ restores the $\sqrt{T}$-order regret, but permits less representation learning.

Tier L, assumed-limit *Here the reference is not initialization. It is the limiting learned representation:*

$$\bar{u} = \lim_{t \to \infty} u_t.$$

The total movement can be $O(1)$, so this allows meaningful representation learning. The tail movement must decay:

$$r_t = \sum_{s \geq t} m_s \lesssim t^{-(1/2+\delta)}.$$

This gives A2-style drift control, but A1/A3/A4 at the learned limit are not certified from NTK initialization. They are assumed, though the draft suggests monitoring them empirically.

So the intuitive hierarchy is:

$$\textit{Frozen} \subset \textit{Tier } C^+ \subset \textit{Tier } C \subset \textit{Tier } L.$$

But the tradeoff is reversed for rigor:

$$\textit{Frozen and Tier } C^+ : \textit{ best regret, least learning;}$$

$$\textit{Tier } L : \textit{ most learning, weakest certification.}$$

One important caveat: all tiers still assume the output-lift margin $s_A > 0$. The tiers certify drift and conditioning, not the structural identifiability margin.

3 The Exact Identity

Theorem 1. *For every t and $z \in \mathbb{R}^p$,*

$$(C_t A_t - I)z = -\lambda M_{t-1} G_t^{-1} z, \qquad M_{t-1} := I + U_{t-1}(V_{t-1}^\top V_{t-1} + \lambda I)^{-1} U_{t-1}^\top \succeq I, \tag{7}$$

hence $\lambda \|G_t^{-1} u_t(x)\| \le R_t(x)$.

Moreover, if $V_{t-1}^\top V_{t-1} \succeq s^2 U_{t-1}^\top U_{t-1}$ then $\|M_{t-1}\|_{\mathrm{op}} \le 1 + s^{-2}$ and

$$\lambda \|G_t^{-1} u\| \;\le\; R_t(x) \;\le\; (1 + s^{-2})\sqrt{\lambda}\, \omega_t(x). \tag{8}$$

Proof. Abbreviate $U = U_{t-1}$, $V = V_{t-1}$, $G = G_t$.

Step 1 (composition). By the push-through identity (Fact A.1, applied to V),

$$V^\top (VV^\top + \lambda I)^{-1} V = I - \lambda (V^\top V + \lambda I)^{-1},$$

so, substituting the closed forms (3),

$$\begin{aligned} C_t A_t &= UV^\top (VV^\top + \lambda I)^{-1}\, VU^\top G^{-1} = U\big[I - \lambda (V^\top V + \lambda I)^{-1}\big] U^\top G^{-1} \\ &= UU^\top G^{-1} - \lambda\, U (V^\top V + \lambda I)^{-1} U^\top G^{-1}. \end{aligned}$$

Step 2 (subtract the identity). Since $UU^\top = G - \lambda I$, we have $UU^\top G^{-1} = I - \lambda G^{-1}$, hence

$$\begin{aligned} C_t A_t - I &= -\lambda G^{-1} - \lambda\, U (V^\top V + \lambda I)^{-1} U^\top G^{-1} \\ &= -\lambda \big[I + U (V^\top V + \lambda I)^{-1} U^\top\big] G^{-1} = -\lambda M_{t-1} G^{-1}, \end{aligned}$$

which is (7). M_{t-1} is symmetric, and $M_{t-1} \succeq I$ because $U(V^\top V + \lambda I)^{-1} U^\top \succeq 0$ (it is of the form $UBU^\top$ with $B \succ 0$).

Step 3 (lower bound). For symmetric $M \succeq I$ and any z, $\|Mz\| \ge \|z\|$ (Fact A.3). Applying this with $z = G^{-1}u$ in (7) gives

$$R_t(x) = \lambda \|M_{t-1} G^{-1} u\| \ge \lambda \|G^{-1} u\|.$$

Step 4 (upper bound). Assume $V^\top V \succeq s^2 U^\top U$. Then

$$V^\top V + \lambda I \succeq s^2 U^\top U + \lambda I,$$

and by inverse monotonicity (Fact A.2),

$$(V^\top V + \lambda I)^{-1} \preceq (s^2 U^\top U + \lambda I)^{-1}.$$

Thus,

$$\big\| U (V^\top V + \lambda I)^{-1} U^\top \big\|_{\mathrm{op}} \le \big\| U (s^2 U^\top U + \lambda I)^{-1} U^\top \big\|_{\mathrm{op}}.$$

Let $U = P\Sigma Q^\top$ be an SVD with singular values σ_i. Then

$$U(s^2U^\top U + \lambda I)^{-1}U^\top = P\,\Sigma(s^2\Sigma^2 + \lambda I)^{-1}\Sigma\,P^\top,$$

whose eigenvalues are $\sigma_i^2/(s^2\sigma_i^2 + \lambda) \le 1/s^2$. Hence

$$\|M_{t-1}\|_{\mathrm{op}} \le 1 + s^{-2}$$

and

$$R_t(x) \le \|M_{t-1}\|_{\mathrm{op}} \cdot \lambda\|G^{-1}u\| \le (1 + s^{-2})\,\lambda\|G^{-1}u\|.$$

Finally, $\lambda^2\, u^\top G^{-2}u \le \lambda\, u^\top G^{-1}u$ because $G \succeq \lambda I$ implies $G^{-1} \preceq \lambda^{-1}I$, hence

$$G^{-2} = G^{-1/2}G^{-1}G^{-1/2} \preceq \lambda^{-1}G^{-1}.$$

Taking square roots, $\lambda\|G^{-1}u\| \le \sqrt{\lambda}\,\omega_t(x)$, which yields Equation (8). □

The identity is unconditional for any stored matrices, any noise, any misspecification. It settles what R_t is: ridge shrinkage $\lambda\|G_t^{-1}u\|$ modulated by a factor M_{t-1} depending only on the output spectrum $V^\top V$. The round trip is geometry: it never sees fit quality, noise realizations, or misspecification. This generates both the positive results (geometry is exactly what optimism needs when errors are shrinkage-shaped) and the limits (geometry cannot cover width-shaped errors).

4 Noiseless Analysis

Throughout this section and the next, $e_t \equiv 0$.

Lemma 1 (Decomposition)**.** *Let*

$$U_{t-1} = [u_1, \dots, u_{t-1}], \qquad V_{t-1} = [v_1, \dots, v_{t-1}].$$

Define

$$E_{t-1} := [\epsilon_1, \dots, \epsilon_{t-1}],$$

where each column is

$$\epsilon_s := v_s(F^\star(x_s)) - A^\star u_s(x_s).$$

Then, under A1-A2, $\|\epsilon_s\| \le \Delta_s$, and $\|E_{t-1}\|_{\mathrm{op}} \le E_\infty$.

At score time, $r^\star(x) = w^\top[A^\star u_t(x) + \epsilon_t^{\mathrm{sc}}(x)]$ with $\|\epsilon_t^{\mathrm{sc}}(x)\| \le \Delta_t$.

Proof. Using the reference identity

$$\bar v(F^\star(x)) = A^\star \bar u(x),$$

we have

$$\epsilon_s = v_s(F^\star(x_s)) - A^\star u_s(x_s) = \underbrace{\big[v_s(F^\star(x_s)) - \bar v(F^\star(x_s))\big]}_{\text{norm}\le\tau\varepsilon_{Y,s}\ \text{(A2)}} - A^\star \underbrace{\big[u_s(x_s) - \bar u(x_s)\big]}_{\text{norm}\le\varepsilon_{X,s}\ \text{(A2)}}.$$

Hence $\|\epsilon_s\| \le \tau\varepsilon_{Y,s} + \|A^\star\|_{\mathrm{op}}\,\varepsilon_{X,s} \le \Delta_s$.

Note that each ϵ_s carries the feature drift relative to the reference maps at round s. Therefore,

$$\|E_{t-1}\|_{\mathrm{op}} \le \|E_{t-1}\|_F = \left(\sum_{s<t}\|\epsilon_s\|^2\right)^{1/2} \le E_\infty.$$

The score-time claim is the same computation at a candidate $x \in \mathcal{X}_t$ with the round-t maps:

$$r^\star(x) = w^\top v_t(F^\star(x)) = w^\top[A^\star u_t(x) + \epsilon_t^{\mathrm{sc}}(x)]$$

with

$$\epsilon_t^{\mathrm{sc}}(x) = [v_t(F^\star(x)) - \bar v(F^\star(x))] - A^\star[u_t(x) - \bar u(x)].$$

Thus, $\|\epsilon_t^{\mathrm{sc}}(x)\| \le \Delta_t$ by A2 at round t. □

Lemma 2 (Forward error, sharpened). *Under A1-A2, for every candidate x with $u = u_t(x)$,*

$$\begin{aligned} &\|A_t u - A^\star u\| \le S_A\, R_t(x) + E_\infty\, \omega_t(x), \\ \textit{hence} \qquad &|\widehat{r}_t(x) - r^\star(x)| \le \|w\| \big[S_A R_t(x) + E_\infty \omega_t(x) + \Delta_t\big]. \end{aligned} \tag{9}$$

Proof. Step 1 (ridge error form). Substituting Lemma 1 into Equation (3),

$$A_t = (A^\star U + E)U^\top G^{-1} = A^\star(G - \lambda I)G^{-1} + EU^\top G^{-1} = A^\star - \lambda A^\star G^{-1} + EU^\top G^{-1}.$$

Thus, $A_t - A^\star = \big(EU^\top - \lambda A^\star\big)G^{-1}$.

Step 2 ($A^\star$-piece). $\|\lambda A^\star G^{-1} u\| \le \|A^\star\|_{\mathrm{op}}\, \lambda \|G^{-1}u\| \le S_A\, R_t(x)$ by Theorem 1's lower bound.

Step 3 (E-piece). $\|EU^\top G^{-1}u\| \le \|E\|_{\mathrm{op}}\, \|U^\top G^{-1}u\|$, and

$$\|U^\top G^{-1}u\|^2 = u^\top G^{-1} U U^\top G^{-1} u = u^\top G^{-1}(G - \lambda I)G^{-1}u = \omega_t(x)^2 - \lambda \|G^{-1}u\|^2 \le \omega_t(x)^2,$$

so $\|EU^\top G^{-1}u\| \le E_\infty\, \omega_t(x)$ by Lemma 1.

The triangle inequality gives the first claim.

Step 4 (reward). By Lemma 1's score-time part, $r^\star(x) = w^\top A^\star u + w^\top \epsilon_t^{\mathrm{sc}}(x)$ and $\widehat{r}_t(x) = w^\top A_t u$. Hence,

$$|\widehat{r}_t(x) - r^\star(x)| \le \|w\|\, \|A_t u - A^\star u\| + \|w\|\, \|\epsilon_t^{\mathrm{sc}}(x)\| \le \|w\| \big[S_A R_t(x) + E_\infty \omega_t(x) + \Delta_t\big]. \quad \square$$

This decomposition is the paper's pivot: the realizable error is shrinkage-shaped *and certified by R_t with constant S_A; the drift error is* width-shaped *with coefficient E_∞. The design's fate in every regime is decided by the size and tunability of the width-shaped coefficient.*

Lemma 3 (Margin transfer). *Under A1–A4 (in the frozen case ($\varepsilon \equiv 0$), under A1/A3 alone), $\|V_{t-1}a\| \ge \frac{s_A}{2}\|U_{t-1}a\|$ for all a, so Equation* (8) *holds with $\bar{M} := 1 + 4/s_A^2$:*

$$R_t(x) \le \bar{M}\sqrt{\lambda}\, \omega_t(x). \tag{10}$$

Proof. Frozen case. $V = A^\star \bar{U}$ exactly (Lemma 1 with $E = 0$, $\bar{U} = U$). Thus, for any a,

$$\|Va\| = \|A^\star \bar{U} a\| \ge s_A \|\bar{U}a\|$$

by A3 ($\bar{U}a \in \mathrm{range}(\bar{U})$).

Equivalently, $V^\top V \succeq s_A^2 U^\top U$, and Theorem 1 applies with $s = s_A$, $\bar{M} = 1 + s_A^{-2}$.

General case. Write $\Delta U := U_{t-1} - \bar{U}_{t-1}$. Its columns are $u_s(x_s) - \bar{u}(x_s)$. Therefore, $\|\Delta U\|_{\mathrm{op}} \le \|\Delta U\|_F \le E_U$ by A2. By the singular-value perturbation bound (Fact A.4) and A4,

$$\sigma_{\min}(U_{t-1}) \ge \sigma_{\min}(\bar{U}_{t-1}) - \|\Delta U\|_{\mathrm{op}} \ge \kappa_0 - E_U =: \kappa > 0.$$

Fix a. Then, using A3 on $\bar{U}a$, the reverse triangle inequality twice, and $\|a\| \le \|Ua\|/\sigma_{\min}(U) \le \|Ua\|/\kappa$:

$$\begin{aligned} \|A^\star U a\| &\ge \|A^\star \bar{U} a\| - \|A^\star \Delta U a\| \ge s_A \|\bar{U}a\| - S_A \|\Delta U a\| \\ &\ge s_A \|Ua\| - (s_A + S_A)\|\Delta U a\| \ge \Big[s_A - \frac{(s_A + S_A)E_U}{\kappa}\Big]\|Ua\|, \end{aligned}$$

where the third inequality used $\|\bar{U}a\| \ge \|Ua\| - \|\Delta U a\|$ and the fourth $\|\Delta U a\| \le E_U \|a\| \le (E_U/\kappa)\|Ua\|$.

Subtracting the misspecification,

$$\|Va\| \ge \|A^\star U a\| - \|E_{t-1}a\| \ge \big[s_A - ((s_A + S_A)E_U + E_\infty)/\kappa\big]\|Ua\| \ge \frac{s_A}{2}\|Ua\|$$

by the smallness condition (6).

Theorem 1 with $s = s_A/2$ gives $\|M_{t-1}\|_{\mathrm{op}} \le 1 + 4/s_A^2 = \bar{M}$ and hence Equation (10) via $R_t \le \|M\|_{\mathrm{op}}\, \lambda \|G^{-1}u\| \le \bar{M}\sqrt{\lambda}\, \omega_t$. $\square$

Lemma 4 (Mixed-feature potential and information-gain stability)**.** *Under Algorithm 1, we have*

(a) Under A5, $\sum_{t\le T}\omega_t(x_t) \le \sqrt{2T\gamma_T}$.

(b) $\gamma_T \le \gamma_T^{\mathrm{ref}} + 2E_U\sqrt{\gamma_T^{\mathrm{ref}}/\lambda} + E_U^2/\lambda$.

Proof. (a) Store-once makes $G_{t+1} = G_t + u_t(x_t)u_t(x_t)^\top$, where $u_t(x_t)$ is exactly the vector scored at round t (line 10 of Algorithm 1). By the matrix determinant lemma (Fact A.6),

$$\det G_{t+1} = \det G_t\left(1 + u_t(x_t)^\top G_t^{-1}u_t(x_t)\right) = \det G_t\,(1+\omega_t(x_t)^2).$$

Telescoping from $G_1 = \lambda I$ and using Sylvester's identity $\det(I + \lambda^{-1}U_TU_T^\top) = \det(I + \lambda^{-1}U_T^\top U_T)$ (Fact A.6),

$$\sum_{t=1}^{T}\log\left(1+\omega_t(x_t)^2\right) = \log\frac{\det G_{T+1}}{\det(\lambda I)} = \log\det\left(I+\lambda^{-1}U_T^\top U_T\right) = \gamma_T.$$

By A5,

$$\omega_t(x_t)^2 \le \|u_t(x_t)\|^2\left\|G_t^{-1}\right\|_{\mathrm{op}} \le L_U^2/\lambda \le 1,$$

and $s \le 2\log(1+s)$ for $s \in [0,1]$ (Fact A.7(i)). Therefore, $\sum_t \omega_t(x_t)^2 \le 2\gamma_T$.

Cauchy-Schwarz gives $\sum_t \omega_t(x_t) \le \sqrt{T\sum_t \omega_t(x_t)^2} \le \sqrt{2T\gamma_T}$. We note that the argument does not require that the columns come from a single feature map.

(b) Let $\bar K := \bar U_T^\top \bar U_T$, $K := U_T^\top U_T$, $\Delta K := K - \bar K$, and $S := (\lambda I + \bar K)^{-1/2}\Delta K(\lambda I + \bar K)^{-1/2}$. Since $\lambda I + K \succ 0$, $I + S = (\lambda I + \bar K)^{-1/2}(\lambda I + K)(\lambda I + \bar K)^{-1/2} \succ 0$, i.e., $S \succ -I$. Then

$$\gamma_T - \gamma_T^{\mathrm{ref}} = \log\frac{\det(\lambda I + K)}{\det(\lambda I + \bar K)} = \log\det(I+S) = \sum_i \log(1+\mu_i(S)) \le \sum_i \mu_i(S) = \operatorname{tr} S,$$

using $\log(1+x) \le x$ for $x > -1$ (Fact A.7(ii)).

With $\Delta U := U_T - \bar U_T$ ($\|\Delta U\|_F \le E_U$ by A2),

$$\Delta K = \bar U^\top \Delta U + \Delta U^\top \bar U + \Delta U^\top \Delta U, \qquad \operatorname{tr} S = \operatorname{tr}\left((\lambda I + \bar K)^{-1}\Delta K\right),$$

by cyclicity.

For the cross terms, the trace Cauchy-Schwarz inequality (Fact A.8) gives

$$\left|\operatorname{tr}\left((\lambda I + \bar K)^{-1}\bar U^\top \Delta U\right)\right| = \left|\left\langle \bar U(\lambda I + \bar K)^{-1},\, \Delta U\right\rangle_F\right| \le \left\|\bar U(\lambda I + \bar K)^{-1}\right\|_F \|\Delta U\|_F,$$

and, writing μ_i for the eigenvalues of $\bar K$,

$$\left\|\bar U(\lambda I + \bar K)^{-1}\right\|_F^2 = \operatorname{tr}\left((\lambda I + \bar K)^{-2}\bar K\right) = \sum_i \frac{\mu_i}{(\lambda+\mu_i)^2} \le \frac{1}{\lambda}\sum_i \frac{\mu_i}{\lambda+\mu_i} = \frac{d_{\mathrm{eff}}}{\lambda},$$

where $d_{\mathrm{eff}} \le \sum_i \log(1+\mu_i/\lambda) = \gamma_T^{\mathrm{ref}}$ by $x/(1+x) \le \log(1+x)$ (Fact A.7(iii)) applied to $x = \mu_i/\lambda$.

The quadratic term: $\operatorname{tr}((\lambda I + \bar K)^{-1}\Delta U^\top \Delta U) \le \left\|(\lambda I + \bar K)^{-1}\right\|_{\mathrm{op}}\|\Delta U\|_F^2 \le E_U^2/\lambda$. Collecting, $\operatorname{tr} S \le 2E_U\sqrt{\gamma_T^{\mathrm{ref}}/\lambda} + E_U^2/\lambda$. □

Theorem 2 (Frozen features: pure round-trip UCB)**.** *With* $\varepsilon \equiv 0$*, A1, A3, A5, run Equation* (4) *with* $b = \|w\|S_A$, $\tilde b = 0$*. Then*

$$\mathrm{Reg}(T) \le 2\|w\|S_A\left(1 + \frac{1}{s_A^2}\right)\sqrt{2\lambda T\gamma_T} + \mathcal{E}_{\mathrm{head}}. \tag{11}$$

Proof. By (9) with $E_\infty = \Delta_t = 0$, for every candidate $x \in \mathcal{X}_t$,

$$r^\star(x) \le \widehat{r}_t(x) + \|w\| S_A \, R_t(x) = \widehat{r}_t(x) + b \, R_t(x) \quad \text{and} \quad \widehat{r}_t(x_t) \le r^\star(x_t) + b \, R_t(x_t).$$

Since x_t maximizes the score (Equation (4)) (with $\tilde{b} = 0$) over $\mathcal{X}_t \ni x_t^\star$,

$$r^\star(x_t^\star) - r^\star(x_t) \le \big[\widehat{r}_t(x_t^\star) + bR_t(x_t^\star)\big] - r^\star(x_t) \le \big[\widehat{r}_t(x_t) + bR_t(x_t)\big] - r^\star(x_t) \le 2b \, R_t(x_t).$$

Summing over t and applying the frozen-case part of Lemma 3 ($R_t(x_t) \le (1 + s_A^{-2})\sqrt{\lambda}\,\omega_t(x_t)$) and Lemma 4(a),

$$\mathrm{Reg}(T) \le 2b \sum_t R_t(x_t) \le 2\|w\| S_A \Big(1 + \frac{1}{s_A^2}\Big)\sqrt{\lambda} \sum_t \omega_t(x_t) \le 2\|w\| S_A \Big(1 + \frac{1}{s_A^2}\Big)\sqrt{2\lambda T \gamma_T}.$$

Head-optimization error adds $\mathcal{E}_{\mathrm{head}}$ by Lemma 9. □

This is the design's clean positive result: an inverse-free score whose residual itself *certifies the forward error — no confidence set is constructed — at the standard $\sqrt{\lambda T \gamma_T}$ order.*

Theorem 3 (Trainable features)**.** *Under A1-A5 (distinct selected contexts only; see A4), run Equation* (4) *with $b = \|w\| S_A$, $\tilde{b} = \|w\| E_\infty \sqrt{L_U/\lambda}$. Then*

$$\mathrm{Reg}(T) \le 2\|w\| S_A \bar{M} \sqrt{2\lambda T \gamma_T} + 2^{5/4} \|w\| E_\infty \sqrt{\tfrac{L_U}{\lambda}} \bar{M}^{1/2} \lambda^{1/4} \gamma_T^{1/4} T^{3/4} + 2\|w\| \Sigma_\Delta + \mathcal{E}_{\mathrm{head}}. \quad (12)$$

Proof. Step 1 (the $\sqrt{R}$ bonus dominates the width term). By Cauchy–Schwarz and Theorem 1's lower bound,

$$\omega_t(x)^2 = u^\top G_t^{-1} u \le \|u\| \, \|G_t^{-1} u\| \le L_U \cdot \frac{R_t(x)}{\lambda},$$

$$\text{so} \qquad E_\infty \, \omega_t(x) \le E_\infty \sqrt{\tfrac{L_U}{\lambda}} \sqrt{R_t(x)} = \frac{\tilde{b}}{\|w\|} \sqrt{R_t(x)}.$$

Step 2 (optimism up to Δ_t). By Equation (9) and Step 1, for every $x \in \mathcal{X}_t$,

$$r^\star(x) \le \widehat{r}_t(x) + bR_t(x) + \tilde{b}\sqrt{R_t(x)} + \|w\|\Delta_t, \qquad \widehat{r}_t(x_t) \le r^\star(x_t) + bR_t(x_t) + \tilde{b}\sqrt{R_t(x_t)} + \|w\|\Delta_t.$$

Since x_t maximizes (4),

$$r^\star(x_t^\star) - r^\star(x_t) \le 2bR_t(x_t) + 2\tilde{b}\sqrt{R_t(x_t)} + 2\|w\|\Delta_t.$$

Step 3 (summation). By Lemma 3 and Lemma 4(a),

$$\sum_t R_t(x_t) \le \bar{M}\sqrt{\lambda} \sum_t \omega_t(x_t) \le \bar{M}\sqrt{\lambda}\sqrt{2T\gamma_T}.$$

By Cauchy-Schwarz,

$$\sum_t \sqrt{R_t(x_t)} \le \sqrt{T \sum_t R_t(x_t)} \le \sqrt{T}\,\big(\bar{M}\sqrt{\lambda}\sqrt{2T\gamma_T}\big)^{1/2} = 2^{1/4}\,\bar{M}^{1/2}\lambda^{1/4}\gamma_T^{1/4}\,T^{3/4}.$$

Substituting into Step 2 and summing, the first term is

$$2b\bar{M}\sqrt{\lambda}\sqrt{2T\gamma_T} = 2\|w\| S_A \bar{M}\sqrt{2\lambda T\gamma_T}.$$

The second is $2\tilde{b} \cdot 2^{1/4}\bar{M}^{1/2}\lambda^{1/4}\gamma_T^{1/4}T^{3/4}$, which equals the middle term of (12). The third is $2\|w\|\Sigma_\Delta$. Heads add $\mathcal{E}_{\mathrm{head}}$ (Lemma 9). □

Corollary 1. *If $\Delta_s \le c_\Delta s^{-(1/2+\delta)}$ (the decaying instance described after A2) with $c_\Delta = \Theta(T^{-1/4})$ and $\delta > 0$, then*

$$E_\infty^2 = \sum_s \Delta_s^2 \le c_\Delta^2 \sum_{s \ge 1} s^{-1-2\delta} \le c_\Delta^2 \Big(1 + \frac{1}{2\delta}\Big) \quad \Rightarrow \quad E_\infty = O(T^{-1/4}),$$

so the middle term of Equation (12) *is $O(T^{-1/4}) \cdot O(T^{3/4}) = O(\sqrt{T})$ (up to the $\gamma_T^{1/4}$ and constant factors), and $\Sigma_\Delta \le c_\Delta \sum_{s \le T} s^{-1/2-\delta} \le c_\Delta \cdot \frac{T^{1/2-\delta}}{1/2-\delta} = O(T^{1/4-\delta})$ for $\delta < 1/2$.*

Hence Equation (12) *is $\widetilde{O}(\|w\| S_A \bar{M}\sqrt{\lambda T \gamma_T})$; the per-round drift allowance ranges from $T^{-1/4}$ (round 1) to $T^{-3/4}$ (round T): front-loaded representation movement.*

Proposition 1 (Tightness). *For $u = \|u\|e$ with e a unit eigenvector of $U_{t-1}U_{t-1}^\top$ of eigenvalue μ:*

$$\omega_t = \frac{\|u\|}{\sqrt{\lambda+\mu}}, \qquad \lambda\|G_t^{-1}u\| = \frac{\lambda\|u\|}{\lambda+\mu}, \qquad R_t \in \Big[\frac{\lambda\|u\|}{\lambda+\mu},\ \bar{M}\frac{\lambda\|u\|}{\lambda+\mu}\Big],$$

so $\omega_t = \sqrt{\|u\|R_t/\lambda}\cdot\Theta(1)$: the inequality of Theorem 3, Step 1, is an equality family, *saturated at every exploration level μ. Hence any bonus $g(R_t(x))$ dominating a width-shaped error $c\,\omega_t(x)$ at all candidates obeys $g(R) \geq c\sqrt{\|u\|R/\lambda}\cdot\Theta(1)$ along this family, and on a trajectory with $R_t(x_t) \asymp \sqrt{\lambda}\,\omega_t(x_t)$ and $\omega_t(x_t) \asymp \sqrt{\gamma/t}$,*

$$\sum_{t\leq T} g(R_t(x_t)) \gtrsim c\sqrt{\tfrac{1}{\lambda}}\sum_{t\leq T}\sqrt{\|u\|\,R_t(x_t)} \asymp c\,\lambda^{1/4}\gamma^{1/4}\sum_{t\leq T} t^{-1/4}\cdot\Theta(1) = \Theta\big(cT^{3/4}\big).$$

In explored directions ($\mu \gg \lambda$), R_t compresses the width by $\sqrt{\lambda/(\lambda+\mu)}$: the information is absent from the score, not the analysis. This is tightness of the proof template; constructing a formal bandit lower-bound instance is open.

Proof. For the eigenvector u: $G_t u = (\lambda+\mu)u$, so $G_t^{-1}u = u/(\lambda+\mu)$, giving the displayed ω_t and $\lambda\|G_t^{-1}u\|$ by direct computation; the bracket for R_t is Theorem 1's sandwich ($R_t = \lambda\|M_{t-1}G_t^{-1}u\|$ with $I \preceq M_{t-1} \preceq \bar{M}I$). The saturation claim is the identity $\omega_t^2 = \|u\|\cdot\lambda\|G_t^{-1}u\|/\lambda$ valid on every eigendirection. The trajectory computation substitutes $R_t \asymp \sqrt{\lambda}\,\omega_t \asymp \sqrt{\lambda\gamma/t}$, so $\sqrt{R_t} \asymp \lambda^{1/4}\gamma^{1/4}t^{-1/4}$, and the integral comparison $\sum_{t\leq T} t^{-1/4} \geq \int_1^{T+1} t^{-1/4}dt = \frac{4}{3}((T+1)^{3/4}-1)$. □

In the noiseless case $c = E_\infty$ is a controllable budget (Corollary 1). Under noise, c is set by nature.

5 NTK Instantiation

5.1 Architecture

Two scalar-output, depth-L, width-m (m even), fully connected networks with NTK scaling generate the feature maps; their forward outputs are never used by the algorithm. For $a \in \{X, Y\}$ let the input dimension be $d_X = d$ (actions, $\|x\| \leq 1$) and $d_Y = q$ (observations, $\|y\| \leq B_y$). For an input $s \in \mathbb{R}^{d_a}$:

$$h_a^0(s) = s, \qquad h_a^\ell(s) = \sqrt{\tfrac{2}{m}}\,\sigma\big(W_a^\ell\, h_a^{\ell-1}(s)\big)\ \ (\ell = 1,\dots,L), \qquad f_a(s;\vartheta_a) = \tfrac{1}{\sqrt{m}}\, b_a^\top h_a^L(s), \tag{13}$$

with $W_a^1 \in \mathbb{R}^{m\times d_a}$, $W_a^\ell \in \mathbb{R}^{m\times m}$ for $2 \leq \ell \leq L$, $b_a \in \mathbb{R}^m$, activation σ (ReLU by default; smooth σ in the parameter-surrogate branch of Lemma C). The parameter vectors and their dimensions are

$$\theta = \mathrm{vec}\,\big(W_X^1,\dots,W_X^L,b_X\big) \in \mathbb{R}^p, \quad \psi = \mathrm{vec}\,\big(W_Y^1,\dots,W_Y^L,b_Y\big) \in \mathbb{R}^{p_Y},$$
$$p = md + (L-1)m^2 + m, \qquad p_Y = mq + (L-1)m^2 + m.$$

Initialization. *Entries of each W_a^ℓ are i.i.d. $\mathcal{N}(0,1)$ (the $\sqrt{2/m}$ He scaling is factored into (13)) and entries of b_a are i.i.d. Rademacher. We then apply the standard* symmetric initialization *(Zhou et al., 2020): neurons are duplicated in pairs with identical hidden weights and opposite output signs ($W_a^\ell[r+\frac{m}{2},:] = W_a^\ell[r,:]$, $b_a[r+\frac{m}{2}] = -b_a[r]$ for $r \leq m/2$), so that $f_X(\cdot;\theta_0) \equiv 0$ and $f_Y(\cdot;\psi_0) \equiv 0$ while the tangent features below are unaffected in law (up to a factor $\sqrt{2}$ absorbed into constants). The zero-output property is not needed by the round-trip identity; it makes the discarded forward outputs carry literally no information and lets the NeuralUCB concentration lemmas import verbatim.*

Tangent features. *The algorithm uses only the parameter Jacobians*

$$u_\theta(x) := \nabla_\theta f_X(x;\theta) \in \mathbb{R}^p, \qquad \phi_\psi(y) := \nabla_\psi f_Y(y;\psi) \in \mathbb{R}^{p_Y}. \tag{14}$$

Blockwise, with pre-activations $z_a^\ell = W_a^\ell h_a^{\ell-1}$ and backpropagated vectors

$$\delta_a^L = \tfrac{1}{\sqrt{m}}\sqrt{\tfrac{2}{m}}\ \mathrm{diag}\big(\sigma'(z_a^L)\big)\, b_a, \qquad \delta_a^\ell = \sqrt{\tfrac{2}{m}}\ \mathrm{diag}\big(\sigma'(z_a^\ell)\big)\,(W_a^{\ell+1})^\top\delta_a^{\ell+1}\ \ (\ell < L),$$

the Jacobian blocks are $\partial f_a/\partial W_a^\ell = \delta_a^\ell\,(h_a^{\ell-1})^\top$ *and* $\partial f_a/\partial b_a = \frac{1}{\sqrt{m}} h_a^L$ *(a.e. for ReLU). The induced kernels are the empirical NTKs and the lifted output kernel:*

$$k_X(x,x') = u_\theta(x)^\top u_\theta(x'), \quad k_Y(y,y') = \phi_\psi(y)^\top \phi_\psi(y'), \quad k_V(y,y') = y^\top y' + \tau^2 k_Y(y,y'), \tag{15}$$

the last realized by the lift $v_\psi(y) = [\,y\,;\,\tau\phi_\psi(y)\,] \in \mathbb{R}^{q'}$, $q' = q + p_Y$, *with reward read-off* $\Pi_y v_\psi(y) = y$ *for* $\Pi_y = [\,I_q\ 0\,]$. *The output network is frozen by default* ($\psi \equiv \psi_0$)*; only* θ *may move, under the Section 5.4 budget. Neither* u_θ *nor* ϕ_ψ *need be materialized: all quantities in Sections 3-7 reduce to the kernels (15), computable by Jacobian-vector products at the stored parameters.*

5.2 Certifying A4 and A5

Let $\mathcal{X}_{T,K}$ *be the pool of* $N = TK$ distinct *candidates (the distinct-contexts scope of A4) and assume* $\lambda_{\min}(H_X^\infty) \geq \lambda_0 > 0$ *for the infinite-width NTK Gram on the pool. For width* $m \geq \mathrm{poly}(T, K, L, \lambda_0^{-1}, \log(1/\delta))$, $\left\|K_X^0 - H_X^\infty\right\|_{\mathrm{op}} \leq \lambda_0/4$ *w.p.* $\geq 1-\delta$ *(Jacot et al., 2018; Zhou et al., 2020), so*

$$\lambda_{\min}(K_X^0) \geq \lambda_{\min}(H_X^\infty) - \left\|K_X^0 - H_X^\infty\right\|_{\mathrm{op}} \geq \lambda_0 - \tfrac{\lambda_0}{4} = \tfrac{3\lambda_0}{4}.$$

Every selected history Gram $\bar{U}_{t-1}^\top \bar{U}_{t-1}$ *is a principal submatrix of* K_X^0 *(distinct selected contexts index distinct pool points), so by eigenvalue interlacing (Fact A.5) A4 holds at the initialization reference with* $\kappa_0 = \sqrt{3\lambda_0/4}$. *Also* $L_u^2 = \sup_x k_X(x,x) = O(L)$ *w.h.p. at this width, fixing the scale of* $\lambda \geq \max\{1, L_U^2\}$.

5.3 Grounding S_A; the τ-tradeoff

On a finite pool with nonsingular Gram, an interpolating $A^\star$ *exists; A1's content is its norm. Assume each coordinate* j *of* $x \mapsto \bar{v}(F^\star(x))$ *lies in the* infinite-width *input NTK RKHS with norm* B_j, *and set*

$$B_{\mathrm{lift}}^2(\tau) := \sum_{j \leq q} B_j^2 + \tau^2 \sum_{j > q} B_j^2.$$

This is the analogue of NeuralUCB's bounded-RKHS-norm assumption (Zhou et al., 2020), applied to the lifted observation map. The finite-width transfer is explicit:

Lemma 5 (Finite-width RKHS transfer)**.** *On the event* $\left\|K_X^0 - H_X^\infty\right\|_{\mathrm{op}} \leq \lambda_0/4$ *of Section 5.2, the minimum-norm empirical-feature interpolant* $A^\star$ *(rows* $a_j = \arg\min \|a\|$ *s.t.* $a^\top u_0(x_i) = [\bar{v}(F^\star(x_i))]_j$ *on the pool) satisfies*

$$S_A \leq \|A^\star\|_F \leq \frac{2}{\sqrt{3}}\, B_{\mathrm{lift}}(\tau). \tag{16}$$

Proof. Write $U_0 \in \mathbb{R}^{p \times N}$ for the matrix of pool features $u_0(x_i)$, so $K_X^0 = U_0^\top U_0$, and let $g_j := ([\bar{v}(F^\star(x_i))]_j)_{i \leq N} \in \mathbb{R}^N$ be the j-th target vector. The RKHS-norm assumption restricted to the finite pool states $g_j^\top (H_X^\infty)^{-1} g_j \leq B_j^2$ (the RKHS norm of any function dominates the H^{-1}-quadratic form of its values on any finite set). The minimum-norm solution of $U_0^\top a = g_j$ is $a_j = U_0 (K_X^0)^{-1} g_j$ with $\|a_j\|^2 = g_j^\top (K_X^0)^{-1} g_j$ (Fact A.9; K_X^0 is invertible on the event by Section 5.2). Since $H_X^\infty \succeq \lambda_0 I$, the event gives

$$K_X^0 \succeq H_X^\infty - \tfrac{\lambda_0}{4} I \succeq H_X^\infty - \tfrac{1}{4} H_X^\infty = \tfrac{3}{4} H_X^\infty,$$

hence $(K_X^0)^{-1} \preceq \frac{4}{3}(H_X^\infty)^{-1}$ by inverse monotonicity (Fact A.2), and $\|a_j\|^2 \leq \frac{4}{3} g_j^\top (H_X^\infty)^{-1} g_j \leq \frac{4}{3} B_j^2$.

Summing over the rows $j \leq q'$ (with the τ weights on $j > q$ included in the definition of g_j for the lifted block), we have $\|A^\star\|_F^2 = \sum_j \|a_j\|^2 \leq \frac{4}{3} B_{\mathrm{lift}}^2(\tau)$, and $\|A^\star\|_{\mathrm{op}} \leq \|A^\star\|_F$. □

The margin side is genuinely structural: with $\tau = 0$, *A3* fails *once the explored span exceeds* q *(*$s_A = 0$*, since a rank-*q *map cannot be injective on a higher-dimensional span); a nondegenerate lift makes a margin possible, while* $\tau \uparrow$ *inflates* $B_{\mathrm{lift}}(\tau)$ *and the lift-noise scale* $\sigma_v = \sigma\sqrt{1 + \tau^2 L_\phi^2}$. *Exhibiting a model class with* s_A *bounded below is the design's key open structural problem.*

5.4 CERTIFYING A2: FEATURE-SPACE TRUST REGION

A2 is enforced, not assumed. The representation update at round s (any optimizer, any objective) is backtracked until

$$\sup_{x\in P_s} \|u_{\theta_{s+1}}(x) - u_{\theta_s}(x)\| \le m_s, \qquad P_s = \bigcup_{\tau\le s} \mathcal{X}_\tau, \tag{17}$$

with budgets fixed in advance, executed at line 12 of Algorithm 1 (after scoring and storing, on the pool that already contains $\mathcal{X}_t$).

Lemma 6 (Pool-relative drift)**.** *Run Algorithm 1 with rule (17) and $\sum_s m_s < \infty$, and set $r_t := \sum_{s\ge t} m_s$. Then (i) for every x that ever appears in a pool, $(u_s(x))_{s\ge t(x)}$ is Cauchy ($t(x)$ its first appearance), so a limit $\bar u(x)$ exists on $\bigcup_t P_t$; (ii) for every t and every $x \in \mathcal{X}_t \cup \{x_1,\dots,x_{t-1}\}$, $\|u_t(x) - \bar u(x)\| \le r_t$, i.e., A2 holds with reference $\bar u$ and $\varepsilon_{X,t} = r_t$. The schedule $m_s \propto s^{-(3/2+\delta)}$ gives $r_t \lesssim t^{-(1/2+\delta)}$, hence the decaying-budget instance described after A2 ($E_U, E_\infty = O(1)$, $\Sigma_\Delta = O(T^{1/2-\delta})$).*

Proof. (i) Fix x with first appearance $t(x)$. By line 3 of Algorithm 1, $x \in P_s$ for all $s \ge t(x)$, and the update between rounds s and $s+1$ obeys (17) at x (line 12 runs after the pool update). Hence for $t(x) \le t < t'$,

$$\|u_{t'}(x) - u_t(x)\| \le \sum_{s=t}^{t'-1} \|u_{s+1}(x) - u_s(x)\| \le \sum_{s=t}^{t'-1} m_s \le r_t \xrightarrow[t\to\infty]{} 0,$$

so the sequence is Cauchy in $\mathbb{R}^p$ and converges to some $\bar u(x)$.

(ii) For $x \in \mathcal{X}_t \cup \{x_1,\dots,x_{t-1}\}$ we have $t(x) \le t$, so letting $t' \to \infty$ above, $\|u_t(x) - \bar u(x)\| \le r_t$. Updates performed before x first appears do not enter the telescope, so unseen future candidates pose no gap; points never in any pool are never evaluated by the algorithm or the analysis. For the schedule:

$$r_t = \sum_{s\ge t} cs^{-(3/2+\delta)} \le c\int_{t-1}^{\infty} s^{-(3/2+\delta)} ds = \tfrac{c}{1/2+\delta}(t-1)^{-(1/2+\delta)}.$$

The aggregate behavior follows as in the discussion after A2. □

Two budget tiers, with the transfer to the abstract assumptions made explicit:

Lemma 7 (Tier-C transfer)**.** *Take the reference to be the* initialization $(\bar u, \bar v) = (u_0, v_0)$*, impose total budgets $\sum_s m_s \le B_X$ (input) and, if the output net trains, $\sum_s m_{Y,s} \le B_Y$, of arbitrary shape,* ***and assume one of the two settings under which the budget controls every evaluated point from time zero****:*

- ***(known pool)*** *$P_0 := \mathcal{X}_{T,K}$: the full candidate pool is available in advance and the trust region (17) is enforced on it from the first update (the natural setting here, since the NTK event of Section 5.2 is already an assumption on this same pool); or*
- ***(parameter surrogate)*** *with a smooth activation, the budget is enforced in parameter space: constrain each update by $L_{\phi,X}\|\theta_{s+1} - \theta_s\| \le m_s$, where $L_{\phi,X}$ is the feature-Lipschitz modulus of (18), i.e., m_s is the* induced feature budget*, valid over* all *of $\mathcal{X}$ uniformly, seen or unseen.*

Then, on the initialization event of Section 5.2, ***A2 (budget form), A4, and A5 hold for the stored dynamic design****, with*

$$\varepsilon_{X,s} \le B_X, \qquad \Delta_s \le \tau B_Y + S_A B_X, \qquad E_U \le \sqrt{T} B_X,$$
$$E_\infty \le \sqrt{T}(\tau B_Y + S_A B_X), \qquad \Sigma_\Delta \le T(\tau B_Y + S_A B_X),$$

and the smallness condition (6) evaluated at these constants: a constraint on (B_X, B_Y) only. A1 holds given *the RKHS realizability assumption of Section 5.3 (with $S_A \le \frac{2}{\sqrt{3}} B_{\mathrm{lift}}(\tau)$ by Lemma K); A3 is not supplied: the margin $s_A > 0$ remains the separate structural assumption that every theorem states explicitly. In particular,* ***Tier C*** *($B_X, B_Y \le c_b T^{-1/2}$) gives $E_U \le c_b$, $E_\infty \le (\tau + S_A)c_b = O(1)$, $\Sigma_\Delta \le (\tau + S_A)c_b\sqrt{T}$; the* ***strengthened Tier C$^+$*** *($B_X, B_Y \le c_b T^{-3/4}$) gives $E_\infty \le (\tau + S_A)c_b T^{-1/4}$, $\Sigma_\Delta \le (\tau + S_A)c_b T^{1/4}$.*

Proof. In either setting the cumulative-movement bound $\|u_s(x) - u_0(x)\| \le \sum_{\tau<s} m_\tau \le B_X$ holds for every x the analysis evaluates: under (known pool) because $x \in P_\tau$ for *all* τ, so every update was constrained at x and the telescope runs from time zero; under (parameter surrogate) because

$$\|u_s(x) - u_0(x)\| \le L_{\phi,X}\|\theta_s - \theta_0\| \le L_{\phi,X} \sum_{\tau<s} \|\theta_{\tau+1} - \theta_\tau\| \le \sum_{\tau<s} m_\tau,$$

pointwise in x, by the per-step constraint $L_{\phi,X}\|\theta_{\tau+1} - \theta_\tau\| \le m_\tau$ and the validity of the modulus over all of $\mathcal{X}$. This gives $\varepsilon_{X,s} \le B_X$ uniformly (similarly $\tau\varepsilon_{Y,s} \le \tau B_Y$ on the output side), hence $\Delta_s \le \tau B_Y + S_A B_X$ by definition, and the aggregates by summing $s \le T$: $E_U \le \sqrt{T} B_X$ etc. A4 and A5 at the initialization reference hold by Section 5.2 on the certified event; the Tier computations substitute B_X, B_Y. □

Remark (online candidate arrival). *If candidates arrive online and only the running pool is constrained, the initialization reference is* not *valid: a point first appearing at round $t(x)$ was unconstrained before $t(x)$ and may already be far from $u_0(x)$. Lemma 6 still applies, but its reference is the constructive limit $\bar{u}$, i.e., the online-arrival setting forces* ***Tier L****: $O(1)$ budgets with tail $r_t \lesssim t^{-(1/2+\delta)}$, while A1/A3/A4 at $\bar{u}$ remain conditional and monitorable (head-fit residual; the generalized singular value $\min_a \|V_{t-1}a\|/\|U_{t-1}a\|$), with freeze-on-breach fallback. This reference-consistency requirement is the precise boundary between the certified and conditional regimes.*

A parameter-norm surrogate *can certify the feature budget at constant monitoring cost for smooth activations (the full pool rule (Equation* (17)*) costs $O(|P_s|)$ feature evaluations per update): for the two-layer instance $f = m^{-1/2}\sum_r a_r\sigma(w_r^\top x)$, $|\sigma'| \le B_1$, $|\sigma''| \le B_2$, $a_r \in \{\pm 1\}$ fixed, $u_W(x) = \nabla_W f(x;W)$ has blocks $m^{-1/2}a_r\sigma'(w_r^\top x)x$, so deterministically for all W, W' and $\|x\| \le 1$,*

$$\|u_W(x)\|^2 = \frac{1}{m}\sum_r \sigma'(w_r^\top x)^2\|x\|^2 \le B_1^2, \qquad \|u_W(x) - u_{W'}(x)\| \le \frac{B_2}{\sqrt{m}}\|W - W'\|_F, \quad (18)$$

the second by $|\sigma'(w_r^\top x) - \sigma'(w_r'^\top x)| \le B_2|(w_r - w_r')^\top x|$ and $\sum_r |(w_r - w_r')^\top x|^2 \le \|W - W'\|_F^2\|x\|^2$. A feature budget m_s thus permits parameter movement $\Theta(\sqrt{m}\, m_s)$: large function change at large width: the theory taxes tangent-feature geometry, not the network function, because heads are refit every round. For ReLU, known perturbation lemmas give only ball-radius (Hölder-$\omega^{1/3}$) moduli around init: enough for Tier C with recalibrated radii, not for trajectory settling; the feature-space rule (Equation (17)*) is activation-agnostic and recommended.*

5.5 Representation training: objective, optimizer, schedule

Algorithm 1, line 12, delegates the representation update to the subroutine Algorithm 2; here is its specification. We emphasize the division of labor up front: the regret guarantees consume only the budget (m_s), through Lemmas P and C they are agnostic to the loss and optimizer below. *The objective influences only* where *the features settle, i.e., the constants (S_A, s_A) that A1/A3 assess at the reference. No claim is made (or needed) that the optimizer decreases the loss, converges, or improves the margin; making that link rigorous is part of open problems (2) and (6) of Section 8.*

Objective. *The default is anchored forward–inverse consistency on the* true stored pairs*, with auxiliary linear heads $\widetilde{A} \in \mathbb{R}^{q' \times p}$, $\widetilde{C} \in \mathbb{R}^{p \times q'}$:*

$$\mathcal{L}_{\mathrm{rep},t}(\theta; \widetilde{A}, \widetilde{C}) := \frac{1}{t}\sum_{s \le t}\left[\tfrac{1}{2}\left\|\widetilde{A}\,u_\theta(x_s) - v_s\right\|^2 + \tfrac{\alpha_{\mathrm{inv}}}{2}\left\|\widetilde{C}\,v_s - u_\theta(x_s)\right\|^2\right] + \tfrac{\nu}{2}\|\theta - \theta_0\|^2, \quad (19)$$

with inverse weight $\alpha_{\mathrm{inv}} \ge 0$ and anchoring strength $\nu \ge 0$. Three deliberate choices: (i) the targets are the stored *columns v_s — true observed outputs under the frozen lift, never synthetic predictions, so the inverse term cannot co-adapt to undo forward errors; (ii) the data-fit is* averaged *$(1/t)$, keeping the gradient scale t-independent so a fixed step size remains stable as the history grows; (iii) the anchor $\nu\|\theta - \theta_0\|^2$ complements — but does not replace — the trust region (17), which alone carries the proof obligations. Heuristically, the forward term pushes θ toward features in which the lifted observation map is linearly realizable with small norm (the S_A side of A1), and the inverse term penalizes collapsing input directions that remain distinguishable in outcome space (the s_A side of A3) while giving no gradient credit to feature directions unrecoverable from outcomes.*

Algorithm 2 REPUPDATE$\big(\theta_t, \{(x_s, v_s)\}_{s\le t}, P_t, m_t\big)$ — anchored round-trip representation step

Require: inverse weight α_{inv}; anchor ν; steps J_{rep}; minibatch size B; optimizer $\mathcal{O} \in \{$minibatch AdamW-style (default), SGD, full-batch GD/Nesterov$\}$ with step size η_{rep} (for full-batch GD: $\eta_{\mathrm{rep}} \le 1/(L_{\mathrm{rep}} + \nu)$, L_{rep} a local smoothness bound of the averaged data-fit; for ReLU any fixed small step — correctness is enforced by line 10, not the step size)

1: $\widetilde{A}, \widetilde{C} \leftarrow$ ridge heads (3) on $\{(u_{\theta_t}(x_s), v_s)\}_{s\le t}$ via warm-started GD (Section 7); clip $\left\|\widetilde{A}\right\|_{\mathrm{op}}, \left\|\widetilde{C}\right\|_{\mathrm{op}} \le H_T$

2: $\theta^{(0)} \leftarrow \theta_t$

3: **for** $j = 0, \dots, J_{\mathrm{rep}} - 1$ **do**

4: sample a minibatch $S_j \subseteq \{1, \dots, t\}$, $|S_j| = B$ (full batch: $S_j = \{1, \dots, t\}$)

5: $g_j \leftarrow \nabla_\theta \Big[\frac{1}{|S_j|} \sum_{s\in S_j} \big(\frac{1}{2} \|\widetilde{A} u_\theta(x_s) - v_s\|^2 + \frac{\alpha_{\mathrm{inv}}}{2} \|\widetilde{C} v_s - u_\theta(x_s)\|^2 \big) \Big]_{\theta=\theta^{(j)}}$

6: $\theta^{(j+1)} \leftarrow \mathcal{O}\text{-step}\big(\theta^{(j)}, g_j\big)$, with the anchor applied as decay toward θ_0: $\theta^{(j+1)} \leftarrow \theta^{(j+1)} - \eta_{\mathrm{rep}} \nu \, (\theta^{(j)} - \theta_0)$

7: **end for**

8: $\theta^+ \leftarrow \theta^{(J_{\mathrm{rep}})}$; $\beta \leftarrow 1$

9: **while** $\sup_{x\in P_t} \|u_{\theta_t + \beta(\theta^+ - \theta_t)}(x) - u_{\theta_t}(x)\| > m_t$ **do**

10: $\beta \leftarrow \beta/2$

11: **end while**

12: **return** $\theta_{t+1} = \theta_t + \beta(\theta^+ - \theta_t)$

Auxiliary heads. *Before the θ-steps of round t, set $\widetilde{A}, \widetilde{C}$ to the ridge solutions (3) on the stored pairs under the* current *features (warm-started GD as in Section 7), clip $\left\|\widetilde{A}\right\|_{\mathrm{op}}, \left\|\widetilde{C}\right\|_{\mathrm{op}} \le H_T$, and* hold them fixed *during the θ-steps (alternating minimization). The auxiliary heads are then discarded: the UCB heads A_t, C_t of Algorithm 1, line 5, are always refit on the stored pairs and are never these.*

Subroutine. *The full update is Algorithm 2. Correctness is decoupled from optimization:* any *proposal θ^+ is projected onto the trust region by the geometric backtracking of line 10, which terminates because the pool-feature movement is continuous in β and vanishes at $\beta = 0$. This is what licenses aggressive optimizers.*

Which optimizer is more efficient? *The question splits along the two cost dimensions, and the trust region changes the usual calculus.* Per-step cost: *full-batch methods pay $O(t)$ Jacobian–vector products per step (the history grows), so over the horizon a dense full-batch schedule costs $O(T^2)$ feature passes; minibatch methods pay $O(B)$ per step regardless of t — the dominant efficiency lever, ahead of any choice among first-order rules.* Steps needed: *because line 10 truncates the move to the budget m_t anyway, optimization speed buys little — what matters is the* direction *quality of a small number of steps. There, an AdamW-style diagonal preconditioner is the better default: the loss (19) is badly scaled through $\widetilde{A}$ (operator norm up to $H_T = TL_UL_V/\lambda$, so the data-fit Hessian $\approx J_\theta^\top(\widetilde{A}^\top\widetilde{A} + \alpha_{\mathrm{inv}} I)J_\theta + \nu I$ is severely ill-conditioned), and diagonal preconditioning is approximately invariant to that scaling where plain (S)GD follows the dominant curvature directions.* Memory: *Adam state costs two extra parameter copies — at NTK widths $p = O(Lm^2)$ this is the one regime where plain SGD is preferable.* Provable rates: *only full-batch GD/Nesterov admit them here, and only toward a stationary point of the ν-strongly-convex-anchored objective (condition number $(L_{\mathrm{rep}} + \nu)/\nu$; $\sqrt{\cdot}$ with Nesterov) — but the regret theory never requires loss decrease, so these guarantees buy nothing the budget does not already enforce. Recommended default:* ***minibatch AdamW-style with decay-to-init (line 6), $J_{\mathrm{rep}} = O(1)$, on the sparse doubling schedule*** *— update only at rounds $t_k = 2^k$ with per-update budgets $m_{(k)} \le c\, 2^{-k(1/2+\delta)}$, so the tail condition of Lemma P still holds and the total enforcement cost of (17) over the horizon is $O(TK)$ feature evaluations, amortized $O(K)$ per round; fall back to SGD when optimizer state is memory-bound, and to full-batch GD only at small t or for parity with the analysis. Note the contrast with the* heads *(Section 7): the head problems are quadratic, so there acceleration has a provable payoff*

($\sqrt{\kappa_T}$); the representation problem is nonconvex and budget-truncated, so optimizer choice there is an empirical-efficiency question, not a guarantee question.

Hyperparameters. *α_{inv} trades realizability against margin (Section 5.3's τ-tradeoff is the analogous output-side dial); ν and the budgets (m_s) jointly control laziness, with (m_s) the binding constraint; $J_{\mathrm{rep}}, \eta_{\mathrm{rep}}$ affect only how much of each round's budget is actually spent. Under Tier C$^+$ the budgets are so small that (Equation* (19)*) functions as a tie-breaker among nearly-frozen features; under Tier L it performs genuine (monitored, conditional) representation learning.*

5.6 Finite-width NTK regret

Theorem 4 (Finite-width NTK regret)**.** *Assume the candidate pool $\mathcal{X}_{T,K}$ is **fixed (oblivious) before initialization** - required in* both *Lemma 7 settings, because the Section 5.2 concentration event and Lemma 5 live on that pool; the parameter surrogate controls feature* drift *at unseen points but not their kernel* concentration*, so adaptively generated candidate sets are outside the certified scope. Assume further Sections 5.1-5.4 in one of Lemma 7's two settings - so Lemmas 6 and 7 supply A2 (budget form), A4, A5; A1 holds via Lemma 5; **A3 with** $s_A > 0$ **is assumed separately**; distinct selected contexts only - and $\lambda \geq \max\{1, L_U^2\}$ (A5), together with the A4 smallness condition (6) at the Lemma 7 constants:*

$$\frac{(s_A + S_A)E_U + E_\infty}{\kappa_0 - E_U} \leq \frac{s_A}{2}.$$

Under Tier C this requires c_b small relative to the structural constants — explicitly, $c_b[(s_A + S_A) + (\tau + S_A)] \leq \frac{s_A}{2}(\kappa_0 - c_b)$ — and under Tier C$^+$ the left side carries an extra $T^{-1/4}$ and the constraint is asymptotically vacuous. Define the stability-corrected *reference information gain*

$$\widetilde{\gamma}_T^0 := \gamma_T^0 + 2E_U\sqrt{\gamma_T^0/\lambda} + \frac{E_U^2}{\lambda},$$

where γ_T^0 is the empirical NTK information gain of the reference *features on the selected actions. W.p. $\geq 1 - O(\delta)$ over initialization (initialization, feature-bound, and any output-side stability events union-bounded), in the noiseless setting:*

$$\begin{aligned}\mathrm{Reg}(T) \leq{}& \frac{4}{\sqrt{3}}\|w\|B_{\mathrm{lift}}(\tau)\Big(1 + \frac{4}{s_A^2}\Big)\sqrt{2\lambda T\widetilde{\gamma}_T^0} \\ &+ 2^{5/4}\|w\|E_\infty\sqrt{\tfrac{L_U}{\lambda}}\bar{M}^{1/2}\lambda^{1/4}(\widetilde{\gamma}_T^0)^{1/4}T^{3/4} + 2\|w\|\Sigma_\Delta + \mathcal{E}_{\mathrm{head}}.\end{aligned} \tag{20}$$

*The middle term is inherited from Theorem 3 and is **not** dropped: under Tier C, $E_\infty \leq (\tau + S_A)c_b = O(1)$, so (20) is $\widetilde{O}(T^{3/4})$ — trainability at the $T^{-1/2}$ total budget costs the rate. Two regimes restore $\sqrt{T}$: **frozen** ($m_s \equiv 0$: $E_\infty = \Sigma_\Delta = 0$, recovering Theorem 2 with $S_A \leq \frac{2}{\sqrt{3}}B_{\mathrm{lift}}(\tau)$, $\bar{M} = 1 + s_A^{-2}$), and **Tier C$^+$** ($E_\infty = O(T^{-1/4})$ makes the middle term $\widetilde{O}(\sqrt{T})$ and $\Sigma_\Delta = O(T^{1/4})$ is lower order). Under Tier L the bound holds conditionally on the monitored assumptions, with $E_\infty = O(1)$, $\Sigma_\Delta = O(T^{1/2-\delta})$ — again $\widetilde{O}(T^{3/4})$ unless the tail constants are horizon-calibrated as in Corollary 1. In the noisy setting, Theorem 5 below applies on the same event with $\sigma_v = \sigma\sqrt{1 + \tau^2 L_\phi^2}$.*

Proof. Work on the intersection of: the kernel concentration event of Section 5.2 (probability $\geq 1 - \delta$ at width (13) by the cited concentration), the feature-bound event $L_u^2 = O(L)$, and (if ψ trains) the output-side analogues; a union bound gives total failure probability $O(\delta)$.

Step 1 (assumptions discharged). On this event: A4 holds with $\kappa_0 = \sqrt{3\lambda_0/4}$ (Section 5.2, via Fact A.5); A5 holds by the feature-bound event and the choice of λ; A2 holds in budget form with the Lemma 7 constants (in whichever of the two settings is assumed — both require the oblivious pool, the first directly and the second because Lemma 5's event lives on it); A1 holds with $S_A \leq \frac{2}{\sqrt{3}}B_{\mathrm{lift}}(\tau)$ by Lemma 5; A3 is assumed. The displayed smallness condition is (6) at the Lemma 7 constants.

Step 2 (apply Theorem 3). All hypotheses of Theorem 3 hold, giving (12) in terms of the *dynamic* information gain γ_T of the stored design, with S_A replaced by its bound $\frac{2}{\sqrt{3}}B_{\mathrm{lift}}(\tau)$ in the leading coefficient ($2 \cdot \frac{2}{\sqrt{3}} = \frac{4}{\sqrt{3}}$).

Step 3 (information-gain stability). By Lemma 4(b) with the initialization reference, $\gamma_T \le \gamma_T^0 + 2E_U\sqrt{\gamma_T^0/\lambda} + E_U^2/\lambda = \widetilde{\gamma}_T^0$. Since $x \mapsto \sqrt{x}$ and $x \mapsto x^{1/4}$ are increasing, every occurrence of γ_T in (12) may be replaced by $\widetilde{\gamma}_T^0$, giving Equation (20). Under Tier C / Tier C$^+$, $E_U \le c_b$ (resp. $c_b T^{-1/4}$) and $\lambda \ge 1$ by A5, so the correction is additive lower order: $\widetilde{\gamma}_T^0 = \gamma_T^0(1 + o(1)) + O(1)$.

Step 4 (tier computations). Substitute the Lemma C aggregates: Tier C gives $E_\infty \le (\tau + S_A)c_b$ and $\Sigma_\Delta \le (\tau + S_A)c_b\sqrt{T}$, so the middle term of (20) is $\Theta(T^{3/4})$ up to logs and the last is $O(\sqrt{T})$; Tier C$^+$ gives $E_\infty \le (\tau + S_A)c_b T^{-1/4}$ (middle term $\widetilde{O}(\sqrt{T})$) and $\Sigma_\Delta \le (\tau + S_A)c_b T^{1/4}$ (lower order); frozen sets both to zero and Theorem 2 applies directly. Under Tier L, Lemma 6 supplies A2 with the limit reference and the same arithmetic applies, conditionally on A1/A3/A4 at $\bar{u}$. □

6 The Noisy Setting

Now $y_t = F^\star(x_t) + e_t$. The stored column becomes $v_s = v_s(F^\star(x_s)) + n_s$ with

$$n_s = \begin{bmatrix} e_s \\ \tau\big(\phi_s(y_s) - \phi_s(F^\star(x_s))\big) \end{bmatrix}, \qquad \|n_s\| \le \sqrt{\|e_s\|^2 + \tau^2 L_\phi^2 \|e_s\|^2} = \sqrt{1 + \tau^2 L_\phi^2}\,\|e_s\|,$$

using the lift's L_ϕ-Lipschitz property (A5). Hence $V_{t-1} = A^\star U_{t-1} + E_{t-1} + N_{t-1}$. Decompose $n_s = b_s + m_s$: a predictable Jensen-type lift bias $b_s := \mathbb{E}[n_s \mid \mathcal{F}_{s-1}, x_s]$ (supported on the ϕ-block, non-decaying) plus a martingale difference m_s.

Fact 6.1 (Reward projection kills lift noise)**.** *$w^\top n_s = \mu^\top e_s =: \eta_s$ and $w^\top b_s = 0$, exactly; η_s is a scalar martingale difference, conditionally $\|\mu\|\sigma$-sub-Gaussian.*

Proof. $w = [\mu; 0]$ annihilates the ϕ-block of n_s, and the first block of n_s is $y_s - F^\star(x_s) = e_s$. Thus $w^\top n_s = \mu^\top e_s$. Taking conditional expectations, $w^\top b_s = \mathbb{E}[\mu^\top e_s \mid \mathcal{F}_{s-1}, x_s] = 0$ by Equation (2).

Sub-Gaussianity: for unit a, $a^\top e_s$ is σ-sub-Gaussian, and $\mu^\top e_s = \|\mu\| \cdot (\mu/\|\mu\|)^\top e_s$. (The nonlinear lift noise and its bias never enter the reward-error path; the inverse/margin path is another matter - there, realizability should be re-anchored to the mean lift $\mathbb{E}[\bar{v}(F^\star(x) + e)]$ to absorb b_s.) □

Fact 6.2 (Store-once preserves the martingale structure)**.** *$u_s = u_s(x_s)$ is $\mathcal{F}_{s-1} \vee \sigma(x_s)$-measurable while $\mathbb{E}[\eta_s \mid \mathcal{F}_{s-1}, x_s] = 0$: (u_s, η_s) is a predictable-design martingale pair.*

Proof. The map $u_s(\cdot)$ is determined by θ_s, which is trained on data from rounds $< s$ (line 12 of Algorithm 1 runs at the end of round $s-1$), and x_s is selected before y_s is observed; so $u_s(x_s)$ is measurable w.r.t. $\mathcal{F}_{s-1} \vee \sigma(x_s)$, while e_s is conditionally centered given exactly that σ-algebra by Equation (2). Re-embedding would destroy this: features re-trained at round $t > s$ depend on e_s, so re-embedded columns $u_t(x_s)$ correlate the design with the noise it is summed against. Store-once is what makes the noisy analysis possible. □

The one imported concentration result, stated in our indexing:

Proposition 2 (Abbasi-Yadkori et al., 2011, Theorem 1)**.** *Let $\{\eta_s\}$ be a scalar martingale-difference sequence, conditionally ρ-sub-Gaussian given $\mathcal{F}_{s-1} \vee \sigma(u_s)$, with u_s predictable as in Fact 6.2. Then w.p. $\ge 1 - \delta$, simultaneously for all $t \ge 1$,*

$$\Big\|\sum_{s<t} \eta_s u_s\Big\|^2_{G_t^{-1}} \le 2\rho^2 \log\Big(\frac{\det(G_t)^{1/2}\det(\lambda I)^{-1/2}}{\delta}\Big) = \rho^2\Big(\log\det\big(I + \lambda^{-1}U_{t-1}^\top U_{t-1}\big) + 2\log\tfrac{1}{\delta}\Big).$$

Lemma 8 (Noisy reward error)**.** *W.p. $\ge 1 - \delta$, for all t and all candidates,*

$$|\widehat{r}_t(x) - r^\star(x)| \le \|w\| S_A R_t(x) + \big(\|w\| E_\infty + \beta_t(\delta)\big)\omega_t(x) + \|w\|\Delta_t, \qquad \beta_t(\delta) := \sigma\|\mu\|\sqrt{\gamma_t + 2\log(1/\delta)}, \quad \gamma_t := \log\det\big(I + \lambda^{-1}U_{t-1}^\top U_{t-1}\big). \tag{21}$$

Proof. As in Lemma 2 Step 1, now with $V = A^\star U + E + N$:

$$A_t - A^\star = \big((E + N)U^\top - \lambda A^\star\big)G_t^{-1}.$$

Project by w and split into three pieces at a candidate $u = u_t(x)$:

$$w^\top(A_t - A^\star)u = \underbrace{w^\top E U^\top G_t^{-1}u}_{\text{(i)}} + \underbrace{w^\top N U^\top G_t^{-1}u}_{\text{(ii)}} - \underbrace{\lambda\, w^\top A^\star G_t^{-1}u}_{\text{(iii)}}.$$

$|(\text{iii})| \le \|w\| S_A \lambda \|G_t^{-1}u\| \le \|w\| S_A R_t(x)$ (Theorem 1); $|(\text{i})| \le \|w\| E_\infty \omega_t(x)$ exactly as in Lemma 2 Step 3. For (ii), by Fact 6.1, $w^\top N U^\top = \sum_{s<t}(w^\top n_s)u_s^\top = \sum_{s<t}\eta_s u_s^\top$, so

$$\begin{aligned}|(\text{ii})| &= \Big|\Big(\sum_{s<t}\eta_s u_s\Big)^\top G_t^{-1}u\Big| = \Big|\Big\langle G_t^{-1/2}\sum_{s<t}\eta_s u_s,\; G_t^{-1/2}u\Big\rangle\Big| \\ &\le \Big\|\sum_{s<t}\eta_s u_s\Big\|_{G_t^{-1}}\omega_t(x) \le \beta_t(\delta)\,\omega_t(x),\end{aligned}$$

by Cauchy-Schwarz and Proposition 2 with $\rho = \|\mu\|\sigma$ (valid by Facts 6.1–6.2), on its probability-$(1-\delta)$ event, uniformly in t.

Adding the score-time term $\|w\|\Delta_t$ as in Lemma 2 Step 4 (Lemma 1's score-time part is deterministic and unaffected by noise) completes Equation (21). □

The noisy error is width-shaped with coefficient $\beta_T = \widetilde{O}(\sigma\sqrt{\gamma_T})$, which is non-decaying and, unlike the drift budget, not under the algorithm's control.

Theorem 5 (Noisy regret; conditional)**.** *Run (4) with $b = \|w\|S_A$, $\tilde{b} = (\|w\|E_\infty + \beta_T(\delta))\sqrt{L_U/\lambda}$. Assume A1–A5 and the **noisy residual bound (NRB)**: $\|M_{t-1}\|_{\mathrm{op}} \le \bar{M}$ for all t under noise. NRB is* not *implied by A3–A4; it is certified in the small-noise/finite-horizon regime of Proposition 4 below, and is otherwise conditional (an unproven lift-noise nondegeneracy assumption is the other known route). Then w.p. $1 - O(\delta)$,*

$$\mathrm{Reg}(T) \le 2\|w\|S_A\bar{M}\sqrt{2\lambda T\gamma_T} + 2^{5/4}\tilde{b}\,\bar{M}^{1/2}\lambda^{1/4}\gamma_T^{1/4}T^{3/4} + 2\|w\|\Sigma_\Delta + \mathcal{E}_{\mathrm{head}}, \tag{22}$$

with leading noise contribution $\widetilde{O}(\sigma\|\mu\|\sqrt{L_U}\,\bar{M}^{1/2}\lambda^{-1/4}\gamma_T^{3/4}T^{3/4})$.

Proof. Work on the event of Lemma 8.

Optimism: since $\beta_t(\delta) \le \beta_T(\delta)$ (the map $t \mapsto \gamma_t$ is nondecreasing because $\det(I + \lambda^{-1}K_t)$ is nondecreasing in t by Fact A.6) and, by Theorem 3 Step 1, $(\|w\|E_\infty + \beta_T)\,\omega_t(x) \le \tilde{b}\sqrt{R_t(x)}$, inequality (21) gives, for all $x \in \mathcal{X}_t$,

$$r^\star(x) \le \widehat{r}_t(x) + bR_t(x) + \tilde{b}\sqrt{R_t(x)} + \|w\|\Delta_t, \qquad \widehat{r}_t(x_t) \le r^\star(x_t) + bR_t(x_t) + \tilde{b}\sqrt{R_t(x_t)} + \|w\|\Delta_t.$$

The UCB step then yields per-round regret $\le 2bR_t(x_t) + 2\tilde{b}\sqrt{R_t(x_t)} + 2\|w\|\Delta_t$ exactly as in Theorem 3 Step 2.

Summation: the input-side design and the potential argument are untouched by output noise, so Lemma 4(a) applies verbatim; the residual upper bound $R_t(x_t) \le \bar{M}\sqrt{\lambda}\,\omega_t(x_t)$ holds by *assumption* NRB together with Theorem 1 ($R_t = \lambda\|M_{t-1}G^{-1}u\| \le \|M_{t-1}\|_{\mathrm{op}}\lambda\|G^{-1}u\| \le \bar{M}\sqrt{\lambda}\,\omega_t$; this is the only place the noisy V-spectrum enters). The two sums are then bounded exactly as in Theorem 3 Step 3, giving (22); the leading noise contribution substitutes $\tilde{b} \ge \beta_T\sqrt{L_U/\lambda}$ with $\beta_T = \widetilde{O}(\sigma\|\mu\|\sqrt{\gamma_T})$. Heads add $\mathcal{E}_{\mathrm{head}}$ (Lemma 9); union-bounding the Lemma 8 event with any NTK events gives $1 - O(\delta)$. □

Corollary 2 (λ-tuning and its limit)**.** *Take the projected choice*

$$\lambda^\star = \max\Big\{1,\; L_U^2,\; \lambda_{\mathrm{bal}}\Big\}, \qquad \lambda_{\mathrm{bal}} := \Theta\big((\sigma\sqrt{L_U}/S_A)^{4/3}\gamma_T^{1/3}T^{1/3}\big)$$

(the maximum enforcing A5). If λ_{bal} is the maximizer — the large-T/non-degenerate-noise regime — then

$$\mathrm{Reg}(T) = \widetilde{O}\Big(S_A^{1/3}(\sigma\sqrt{L_U})^{2/3}\gamma_T^{2/3}T^{2/3}\Big), \tag{23}$$

and no λ does better within this template. If the maximizer is $\max\{1, L_U^2\}$ (small noise, small horizon, or large S_A), the untuned Theorem 5 bound at $\lambda = \max\{1, L_U^2\}$ applies instead, with leading noise term $\widetilde{O}(\sigma\|\mu\|\bar{M}^{1/2}\gamma_T^{3/4}T^{3/4})$.

Proof. Ignoring constants, $\bar{M}$, and logarithmic factors, the two leading terms of (22) scale as

$$\mathcal{T}_1(\lambda) = S_A\sqrt{\lambda T\gamma_T}, \qquad \mathcal{T}_2(\lambda) = \sigma\sqrt{L_U}\,\lambda^{-1/4}\gamma_T^{3/4}T^{3/4}.$$

$\mathcal{T}_1$ increases and $\mathcal{T}_2$ decreases in λ, so the order-optimal choice equates them: $\lambda^{3/4} = \Theta\big((\sigma\sqrt{L_U}/S_A)\,\gamma_T^{1/4}T^{1/4}\big)$, i.e., λ_{bal} as displayed. Substituting back,

$$\mathcal{T}_1(\lambda_{\mathrm{bal}}) = S_A\sqrt{T\gamma_T}\cdot\Theta\big((\sigma\sqrt{L_U}/S_A)^{2/3}\gamma_T^{1/6}T^{1/6}\big) = \Theta\big(S_A^{1/3}(\sigma\sqrt{L_U})^{2/3}\gamma_T^{2/3}T^{2/3}\big),$$

which is (23). Any other λ makes $\max\{\mathcal{T}_1,\mathcal{T}_2\}$ larger, by monotonicity. If the projection binds, evaluate (22) at $\lambda = \max\{1, L_U^2\} = O(1)$ to get the stated $T^{3/4}$ form. (Caveat: $\gamma_T = \gamma_T(\lambda)$ also depends on λ; since it is nonincreasing in λ — each eigenvalue term $\log(1+\mu_i/\lambda)$ decreases — and $\lambda^\star \geq L_U^2$ by construction, the statement is rigorous with γ_T read as the upper bound $\gamma_T(L_U^2)$, treated as slowly varying in the balancing.) □

Proposition 3 (Degradation intrinsic to the template)**.** *By Proposition 1 with $c = \beta_T = \Theta(\sigma\sqrt{\gamma_T})$, any R-only bonus covering the noise error pays $\Omega(\sigma\sqrt{\gamma_T}\,T^{3/4})$-type sums on benign trajectories. Nor does R_t self-inflate to compensate: noise enters M_{t-1} only through $V^\top V$, and for any fixed vector z, enlarging $V^\top V$ in the positive-semidefinite order shrinks $(V^\top V+\lambda I)^{-1}$ (Fact A.2), hence shrinks M_{t-1} toward I — the residual becomes a* cleaner*, not larger, shrinkage proxy under noise. Width-based bonuses achieve $\widetilde{O}(\sigma\gamma_T\sqrt{T} + S_A\sqrt{\lambda T\gamma_T})$ under the same noise (run the proof of Theorem 5 with the bonus $(\,\|w\|S_A\sqrt{\lambda} + \|w\|E_\infty + \beta_T)\,\omega_t(x)$, whose validity is (21) plus $\lambda\|G^{-1}u\| \leq \sqrt{\lambda}\,\omega_t$, and sum with Lemma 4(a)): noise separates the designs by a polynomial factor.*

6.1 The margin under noise: a certified regime and its limit

Proposition 4 (Small-noise certification of NRB)**.** *Let $\bar{\sigma}_v := \sqrt{1+\tau^2L_\phi^2}\,B_e$ (so $\|n_s\| \leq \bar{\sigma}_v$ deterministically). Suppose the* noiseless part *of the stored design satisfies the Lemma 3 margin: under A1–A4,*

$$\big\|(A^\star U_{t-1} + E_{t-1})\,a\big\| \;\geq\; \frac{s_A}{2}\,\|U_{t-1}a\| \qquad \forall a$$

(this is exactly Lemma 3's conclusion, which never involves N). If moreover

$$\sqrt{T}\,\bar{\sigma}_v \;\leq\; \frac{s_A}{4}\,(\kappa_0 - E_U),$$

then the noisy *stored design $V_{t-1} = A^\star U_{t-1} + E_{t-1} + N_{t-1}$ satisfies $\|V_{t-1}a\| \geq \frac{s_A}{4}\|U_{t-1}a\|$, i.e., NRB holds with $\bar{M} = 1 + 16/s_A^2$.*

Proof. $\|N_{t-1}\|_{\mathrm{op}} \leq \|N_{t-1}\|_F = (\sum_{s<t}\|n_s\|^2)^{1/2} \leq \sqrt{t-1}\,\bar{\sigma}_v \leq \sqrt{T}\bar{\sigma}_v$, and $\sigma_{\min}(U_{t-1}) \geq \kappa_0 - E_U$ as in Lemma 3's proof (Fact A.4). Hence for every a, using $\|a\| \leq \|U_{t-1}a\|/(\kappa_0 - E_U)$,

$$\|V_{t-1}a\| \geq \big\|(A^\star U_{t-1}+E_{t-1})a\big\| - \|N_{t-1}a\| \geq \frac{s_A}{2}\|U_{t-1}a\| - \frac{\sqrt{T}\bar{\sigma}_v}{\kappa_0 - E_U}\|U_{t-1}a\| \geq \frac{s_A}{4}\|U_{t-1}a\|,$$

by the displayed smallness condition. Theorem 1's sandwich with $s = s_A/4$ gives $\|M_{t-1}\|_{\mathrm{op}} \leq 1 + 16/s_A^2$. □

Proposition 4 makes Theorem 5 unconditional in a small-noise/finite-horizon regime — but note what it says: the certified noise amplitude shrinks as $B_e \lesssim s_A\kappa_0/(\bar{\sigma}_v$-scale $\sqrt{T})$. For constant-amplitude noise the certified margin route collapses *at $T \gtrsim \kappa_0^2 s_A^2/\bar{\sigma}_v^2$, and no improvement is available from concentration alone ($\|N\|_{\mathrm{op}} \gtrsim \sigma_v\sqrt{T}$ is generic). The opposing effect — noise adds energy to $V^\top V$, and under a (strong, unproven) lift-noise nondegeneracy assumption $\big\|U(V^\top V+\lambda I)^{-1}U^\top\big\|_{\mathrm{op}} \leq \|U\|^2/(\lambda+\sigma_{\min}(V)^2) = O(L_U^2/\sigma_v^2)$, bounding $\bar{M}$ with no margin on $A^\star$ at all — remains a conjecture. Either way, NRB stabilizes only the $\sqrt{T}$ term of (22); the $T^{3/4}$ term is untouched.*

7 IMPLEMENTATION

GD heads. *With $G_U = UU^\top + \lambda I$, full-batch GD with weight decay on the forward ridge objective is*

$$A^{k+1} = A^k - \eta_A \big[A^k G_U - V U^\top\big], \qquad 0 < \eta_A \le \|G_U\|_{\mathrm{op}}^{-1}.$$

Since the ridge minimizer satisfies $A_t G_U = V U^\top$, subtracting gives $A^{k+1} - A_t = (A^k - A_t)(I - \eta_A G_U)$, and $\|I - \eta_A G_U\|_{\mathrm{op}} = 1 - \eta_A \lambda_{\min}(G_U) \le 1 - \eta_A \lambda$ (all eigenvalues of G_U lie in $[\lambda, \|G_U\|_{\mathrm{op}}]$), so

$$\|A^k - A_t\|_F \le (1 - \eta_A \lambda)^k \|A^0 - A_t\|_F;$$

symmetrically for C^k with $G_V = V V^\top + \lambda I$. Head norms obey $\|A_t\|_{\mathrm{op}} \le \|V U^\top\|_F \big\|G_U^{-1}\big\|_{\mathrm{op}} \le T L_U L_V / \lambda =: H_T$, and likewise $\|C_t\|_{\mathrm{op}} \le H_T$.

Lemma 9 (Approximate-head score perturbation)**.** *If $\|\bar{A} - A_t\|_F \le \varepsilon \le 1$ and $\|\bar{C} - C_t\|_F \le \varepsilon$, then uniformly over candidates: $|w^\top \bar{A} u - w^\top A_t u| \le \|w\| L_U \varepsilon$ and $|\bar{R}_t(x) - R_t(x)| \le D_R \varepsilon$ with $D_R := L_U(2H_T + 1)$. Consequently, the UCB score computed with $(\bar{A}, \bar{C})$ differs from the exact score by at most $\eta := \|w\| L_U \varepsilon + b D_R \varepsilon + \tilde{b}\sqrt{D_R \varepsilon}$ at every candidate, and choosing*

$$\varepsilon_{\mathrm{head}} := \min\Big\{ \frac{1}{T^2 \max\{1, \|w\| L_U + b D_R\}},\ \frac{1}{T^4 \max\{1, \tilde{b}^2 D_R\}} \Big\}$$

gives $\eta \le 2/T^2$, hence $\mathcal{E}_{\mathrm{head}} = O(T^{-1})$; since $\log(1/\varepsilon_{\mathrm{head}}) = O(\log T)$, the GD step counts are unaffected in order.

Proof. Mean. $|w^\top(\bar{A} - A_t)u| \le \|w\| \big\|\bar{A} - A_t\big\|_{\mathrm{op}} \|u\| \le \|w\| L_U \varepsilon$.

Residual. Decompose

$$(\bar{C}\bar{A} - C_t A_t)u = \big[(\bar{C} - C_t)A_t + C_t(\bar{A} - A_t) + (\bar{C} - C_t)(\bar{A} - A_t)\big]u,$$

so by the triangle inequality, submultiplicativity, $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_F$, and $\varepsilon \le 1$:

$$\big|\bar{R}_t(x) - R_t(x)\big| \le \big\|(\bar{C}\bar{A} - C_t A_t)u\big\| \le \big(\varepsilon H_T + H_T \varepsilon + \varepsilon^2\big) L_U \le L_U(2H_T + 1)\,\varepsilon = D_R \varepsilon.$$

Square-root bonus. By Fact A.7(iv), $|\sqrt{\bar{R}} - \sqrt{R}| \le \sqrt{|\bar{R} - R|} \le \sqrt{D_R \varepsilon}$. Collecting the three score components gives the stated η.

Tolerance. With $\varepsilon = \varepsilon_{\mathrm{head}}$: the first branch makes $\|w\| L_U \varepsilon + b D_R \varepsilon \le 1/T^2$; the second makes $\tilde{b}\sqrt{D_R \varepsilon} \le 1/T^2$; so $\eta \le 2/T^2$.

Regret accounting. In the optimism chain (Theorems 2–5), replacing the exact score by the computed score perturbs both the upper-confidence inequality at $x_t^\star$ and the chosen-action inequality at x_t by at most η each, adding $\le 2\eta \le 4/T^2$ per round to the regret; summing, $\mathcal{E}_{\mathrm{head}} \le 4/T = O(T^{-1})$. The required iteration count to reach $\varepsilon_{\mathrm{head}}$ from a warm start is $\le \frac{\log(\|A^0 - A_t\|_F / \varepsilon_{\mathrm{head}})}{\eta_A \lambda} \le \kappa_T \cdot O(\log T)$ with $\kappa_T := \|G_U\|_{\mathrm{op}} / \lambda \le 1 + T L_U^2 / \lambda$, by the contraction above. □

Honest cost: the step count carries $\kappa_T = 1 + T L_U^2 / \lambda$ — $O(\kappa_T \log T)$ plain-GD steps per round, $O(\sqrt{\kappa_T} \log T)$ accelerated; warm starts (one new column per round) shrink the log factor but provably not κ_T; an amortized polylog schedule is open.

Kernelized reference. *With $K_X = U^\top U$, $K_V = V^\top V$, $k_x = U^\top u$: $\alpha_t(x) = (K_X + \lambda I)^{-1} k_x$, $\widehat{y}_t(x) = Y_{t-1}\alpha_t(x)$, $\beta_t(x) = (K_V + \lambda I)^{-1} K_V \alpha_t(x)$, and*

$$R_t(x)^2 = \beta_t^\top K_X \beta_t - 2\beta_t^\top k_x + \|u\|^2. \tag{24}$$

(Derivation: by Fact A.1, $U^\top G_t^{-1} u = (K_X + \lambda I)^{-1} k_x = \alpha_t(x)$ and $V^\top (V V^\top + \lambda I)^{-1} V = (K_V + \lambda I)^{-1} K_V$, so $C_t A_t u = U \beta_t(x)$ and $R_t(x)^2 = \|U\beta_t - u\|^2$ expands to (24).) Store-once means the Grams only append: rank-one Cholesky updates apply. At $t = 1$: $\widehat{y}_1 \equiv 0$, $R_1(x)^2 = \|u\|^2$. NTK kernels are computable by Jacobian–vector products at stored parameters.

8 DISCUSSION

What the round trip is. *Theorem 1 settles it: ridge shrinkage modulated by the output-embedding conditioning — a geometric quantity, blind to fit quality, noise, and misspecification. It is the right tool exactly when the error budget is shrinkage-shaped: noiseless (or effectively noiseless, structured) vector feedback with frozen or drift-budgeted features, where it certifies the forward error exactly, inverse-free, with no confidence-set construction (Theorems 2–3). A byproduct stands on its own: $R_t/(\lambda\|G_t^{-1}u\|)$-type ratios and generalized singular values of (V, U) monitor the margin s_A — the identifiability of inputs from outcomes — online.*

What it is not. *When the error has a width-shaped component, the residual's direction-wise compression (Proposition 1) is fundamentally lossy; observation noise makes that component mandatory (Lemma 8), and the order degrades to $\gamma_T^{2/3}T^{2/3}$ at best (Corollary 2, Proposition 3). Width-based bonuses — including inverse-free ones computed by a trained input-side head recovering ω_t^2 exactly as a reconstruction correlation — keep $\widetilde{O}(\gamma_T\sqrt{T})$ with fewer assumptions (no margin, no lift). We state it plainly: as a* bonus*, the forward–inverse round trip through observation space is dominated outside the noiseless regime. Its durable contributions are the exact identity and the assumption-light noiseless theory; the store-once design principle (monotone mixed-feature potentials; martingale preservation), usable by any uncertainty method with trainable features; and the margin diagnostic. As a* representation-learning objective *the round trip plausibly retains a distinct role — Section 5.5 adopts it as the (guarantee-free) default training signal, but proving that it improves the constants it heuristically targets is open.*

Open problems. *(1) Upgrade Proposition 1 to a formal lower-bound instance for R-only scores. (2) With S_A grounded by Lemma K, exhibit a model class with s_A bounded below; the τ-tradeoff of Section 5.3 is the design space. (3) Prove or refute NRB under lift-noise nondegeneracy; whether the noise-regularized $M_{t-1} \to I$ regime is exploitable for some purpose other than the bonus. (4) A span-relative replacement for A4 tolerating repeated contexts. (5) Uniform-domain kernel concentration to lift the oblivious-pool restriction of Theorem 3′. (6) The nonlinear, non-lazy regime, where the linearized theory is silent; the present results predict self-reconstruction signals dominate cycle-consistency signals there as well.*

REPRODUCIBILITY STATEMENT

All results are theoretical; complete proofs appear in the main text, with auxiliary linear-algebra facts proved in Appendix A and a single imported concentration inequality stated precisely in Section 6. The constants-and-conditions table in Appendix B indexes every assumption and where it is consumed.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In Advances in Neural Information Processing Systems*, volume 24, 2011.*

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In Advances in Neural Information Processing Systems*, volume 31, 2018.*

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with UCB-based exploration. In International Conference on Machine Learning*, pp. 11492–11502, 2020.*

A AUXILIARY FACTS

Fact A.1 (Push-through identity)**.** *For any matrix $B \in \mathbb{R}^{n\times m}$ and $\lambda > 0$: $(B^\top B + \lambda I)^{-1}B^\top = B^\top(BB^\top + \lambda I)^{-1}$, and consequently*

$$B^\top(BB^\top + \lambda I)^{-1}B = I - \lambda(B^\top B + \lambda I)^{-1}.$$

Proof. From $B^\top(BB^\top + \lambda I) = (B^\top B + \lambda I)B^\top$, multiply on the left by $(B^\top B + \lambda I)^{-1}$ and on the right by $(BB^\top + \lambda I)^{-1}$ (both invertible for $\lambda > 0$) to get the first identity. Then

$$\begin{aligned} B^\top(BB^\top + \lambda I)^{-1}B &= (B^\top B + \lambda I)^{-1}B^\top B = (B^\top B + \lambda I)^{-1}\big[(B^\top B + \lambda I) - \lambda I\big] \\ &= I - \lambda(B^\top B + \lambda I)^{-1}. \end{aligned}$$

□

Fact A.2 (Inverse monotonicity)**.** *If $A \succeq B \succ 0$ then $B^{-1} \succeq A^{-1}$.*

Proof. $A \succeq B$ implies $B^{-1/2}AB^{-1/2} \succeq I$, hence all eigenvalues of the symmetric matrix $B^{-1/2}AB^{-1/2}$ are ≥ 1, so its inverse $B^{1/2}A^{-1}B^{1/2}$ has eigenvalues ≤ 1, i.e., $B^{1/2}A^{-1}B^{1/2} \preceq I$. Conjugating by $B^{-1/2}$ gives $A^{-1} \preceq B^{-1}$. □

Fact A.3 (Lower bound for $M \succeq I$)**.** *If M is symmetric with $M \succeq I$, then $\|Mz\| \geq \|z\|$ for all z.*

Proof. Diagonalize $M = Q\Lambda Q^\top$ with eigenvalues $\lambda_i \geq 1$. Then $\|Mz\|^2 = \sum_i \lambda_i^2 (Q^\top z)_i^2 \geq \sum_i (Q^\top z)_i^2 = \|z\|^2$. □

Fact A.4 (Singular-value perturbation)**.** *For matrices A, B of equal shape, $\sigma_{\min}(A + B) \geq \sigma_{\min}(A) - \|B\|_{\mathrm{op}}$.*

Proof. For any unit a: $\|(A + B)a\| \geq \|Aa\| - \|Ba\| \geq \sigma_{\min}(A) - \|B\|_{\mathrm{op}}$; take the minimum over a. □

Fact A.5 (Eigenvalue interlacing for principal submatrices)**.** *If $M \in \mathbb{R}^{N\times N}$ is symmetric and M_S is the principal submatrix indexed by $S \subseteq [N]$, then $\lambda_{\min}(M_S) \geq \lambda_{\min}(M)$.*

Proof. For any unit $z \in \mathbb{R}^{|S|}$, let $\tilde{z} \in \mathbb{R}^N$ be its zero-padded extension (still unit). Then $z^\top M_S z = \tilde{z}^\top M \tilde{z} \geq \lambda_{\min}(M)$; minimize over z. □

Fact A.6 (Determinant identities)**.** *(i) $\det(G + uu^\top) = \det(G)\,(1 + u^\top G^{-1}u)$ for invertible G. (ii) $\det(I_n + AB) = \det(I_m + BA)$ for $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times n}$ (Sylvester). (iii) For PSD $K \preceq K'$: $\det(I + \lambda^{-1}K) \leq \det(I + \lambda^{-1}K')$; in particular $t \mapsto \det(I + \lambda^{-1}K_t)$ is nondecreasing when K_t gains PSD increments.*

Proof. (i) $\det(G + uu^\top) = \det(G)\det(I + G^{-1}uu^\top)$, and $\det(I + ab^\top) = 1 + b^\top a$ (the matrix $I + ab^\top$ has eigenvalue $1 + b^\top a$ on a and 1 on the orthogonal complement of b); apply with $a = G^{-1}u$, $b = u$. (ii) Both sides equal $\det \left(\begin{smallmatrix} I_n & -A \\ B & I_m \end{smallmatrix}\right)$ by the two block-triangular factorizations of that matrix (Schur complements in the two orders). (iii) $I + \lambda^{-1}K' \succeq I + \lambda^{-1}K \succ 0$, and $\det$ is monotone on the PSD order for ordered positive-definite matrices: writing $I + \lambda^{-1}K' = (I + \lambda^{-1}K)^{1/2}(I + R)(I + \lambda^{-1}K)^{1/2}$ with $R \succeq 0$, $\det(I + R) \geq 1$. □

Fact A.7 (Scalar inequalities)**.** *(i) $s \leq 2\log(1 + s)$ for $s \in [0, 1]$. (ii) $\log(1 + x) \leq x$ for $x > -1$. (iii) $x/(1 + x) \leq \log(1 + x)$ for $x \geq 0$. (iv) $|\sqrt{a} - \sqrt{b}| \leq \sqrt{|a - b|}$ for $a, b \geq 0$.*

Proof. (i) $g(s) := 2\log(1 + s) - s$ has $g(0) = 0$ and $g'(s) = \frac{2}{1+s} - 1 \geq 0$ on $[0, 1]$. (ii) Concavity of $\log$ at 1: $\log(1 + x) \leq \log 1 + x$. (iii) $h(x) := \log(1 + x) - \frac{x}{1+x}$ has $h(0) = 0$ and $h'(x) = \frac{1}{1+x} - \frac{1}{(1+x)^2} \geq 0$. (iv) WLOG $a \geq b \geq 0$: $(\sqrt{a} - \sqrt{b})^2 = a + b - 2\sqrt{ab} \leq a + b - 2b = a - b$. □

Fact A.8 (Trace Cauchy–Schwarz)**.** *$|\operatorname{tr}(A^\top B)| \leq \|A\|_F \|B\|_F$ for conformal matrices A, B.*

Proof. $\operatorname{tr}(A^\top B) = \langle A, B\rangle_F$ is the Euclidean inner product of the vectorizations; apply the vector Cauchy–Schwarz inequality. □

Fact A.9 (Minimum-norm interpolation)**.** *Let $U \in \mathbb{R}^{p\times N}$ have full column rank, $K := U^\top U \succ 0$, and $g \in \mathbb{R}^N$. The minimum-Euclidean-norm solution of $U^\top a = g$ is $a^\star = UK^{-1}g$, with $\|a^\star\|^2 = g^\top K^{-1} g$.*

Proof. $U^\top a^\star = U^\top U K^{-1} g = g$, so $a^\star$ is feasible, and $a^\star \in \mathrm{range}(U)$. Any feasible a writes $a = a^\star + n$ with $U^\top n = 0$, i.e., $n \perp \mathrm{range}(U) \ni a^\star$, so $\|a\|^2 = \|a^\star\|^2 + \|n\|^2 \geq \|a^\star\|^2$. Finally $\|a^\star\|^2 = g^\top K^{-1} U^\top U K^{-1} g = g^\top K^{-1} g$. □

B Constants and Conditions at a Glance

object	role	where
S_A	realizability norm; $b = \|w\|_{S_A}$	A1, Lemma K
s_A	output-embedding margin; $\bar{M} = 1 + 4/s_A^2$	A3, Lemma 3
$E_U, E_\infty, \Sigma_\Delta$	drift budgets (ℓ^2, lifted ℓ^2, ℓ^1)	A2, Theorem 3
κ_0	reference conditioning (restrictive; see A4)	A4, §5.2
$\lambda \geq \max\{1, L_U^2\}$	potential boundedness; $\widetilde{\gamma}_T^0$ simplification	A5, Lemma 4
$B_{\mathrm{lift}}(\tau)$	RKHS grounding of S_A; τ-tradeoff	§5.3
budgets m_s; Tiers C/C$^+$/L	certification of A2	§5.4
$\alpha_{\mathrm{inv}}, \nu, J_{\mathrm{rep}}, B$	representation training (guarantee-free)	§5.5
$\beta_T(\delta)$	self-normalized noise width	(21)
NRB	noisy residual bound (conditional)	Theorem 5, §6.1
$\kappa_T,\ \mathcal{E}_{\mathrm{head}}$	GD cost and tolerance	§7, Lemma 9

Table 1: Index of constants and the assumptions/results that consume them.