IMPROVED DIMENSION DEPENDENCE FOR LOGISTIC BANDITS VIA SECANT ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study logistic bandits with K arms and d-dimensional features, where rewards are binary and their means follow an unknown logistic model. Existing algorithms attain a leading $\sqrt{dT}$ regret term either with a large additive $d^{5/2}$ term or under the assumption that the feature covariance has a strictly positive minimum eigenvalue. We propose two algorithms that reduce the dependence of the additive term on d without a feature covariance assumption. First, SupSplitLog+ reduces the additive term to $d^{3/2}$ using one-step estimation with splitting samples according to their roles. One subset provides an initial parameter estimate, whose prediction is used to approximate the logistic function with a tangent line. The other subset is used to compute a one-step estimate based on this tangent line. Our improvement comes from a sharper analysis of the error introduced by the tangent approximation. Second, SupSecantLog replaces the tangent line with a secant line through two independent predictions. The leading approximation error then changes from the square of one prediction error to the product of two independent prediction errors. Reducing the bias of both estimates lets us exploit cancellation between positive and negative products, further reducing the additive term to $d^{4/3}$. Finally, we provide an algorithm based on a doubling rule that achieves both regret bounds with d replaced by a data-dependent complexity measure of the revealed features, yielding smaller bounds when these features lie in a low-dimensional subspace.

1 INTRODUCTION

Logistic bandits model sequential decisions with binary feedback, where the probability of a positive outcome is given by the logistic function applied to the inner product between the selected arm's feature vector and an unknown parameter vector (Filippi et al., 2010; Faury et al., 2020). Motivating applications include personalized recommendation and online advertising, where the learner observes whether a user clicks on the selected item or advertisement (Jun et al., 2021). Another example is treatment selection based on patient features, where binary feedback indicates whether the patient responds to the chosen treatment (Li et al., 2017).

Previous work on logistic bandits has focused on improving curvature dependence (Faury et al., 2020; Abeille et al., 2021), computational efficiency (Faury et al., 2022), and confidence bounds (Jun et al., 2021). Reducing the dependence on the feature dimension d, particularly in lower-order regret terms, has received relatively less attention (Mason et al., 2022). Yet a high dimensional dependence can be a major challenge in real-world applications involving high-dimensional feature representations, as illustrated by neural logistic bandits (Bae & Lee, 2025).

We consider the finite-armed setting, where the learner chooses among K arms with potentially time-varying features over T rounds. The algorithms of Li et al. (2017) and Jun et al. (2021) attain a leading term with $\widetilde{\mathcal{O}}(\sqrt{dT})$ dependence on d and T under the assumption that the feature covariance matrix has a strictly positive minimum eigenvalue. This assumption, however, excludes contexts concentrated in a lower-dimensional subspace. Moreover, both algorithms require an initial sampling phase, and their regret upper bounds contain additive terms that grow at least as fast as d^5.

Recently, Bae & Lee (2026) made progress towards reducing the additive dimension dependence without requiring a feature covariance assumption. They proposed SupSplitLog that achieves a regret upper bound of order $\widetilde{\mathcal{O}}(\kappa\sqrt{dT} + \kappa^{5/2}d^{5/2})$ without the assumption, where κ is a curva-

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.

ture parameter of the logistic function. This bound improves upon the previous results in terms of dimension dependence. This result raises the natural question of how far the additive dimension dependence can be reduced. For the fixed-arm setting, Mason et al. (2022) conjectured that bias reduction could lower this dependence to $d^{4/3}$.

This paper answers this question affirmatively by proposing two algorithms that reduce the additive dimension dependence without requiring a feature covariance assumption. The first algorithm improves upon SupSplitLog by reducing the additive term from $d^{5/2}$ to $d^{3/2}$. The second algorithm further reduces the additive term to $d^{4/3}$. Both achieve the leading $\sqrt{dT}$ term in their regret bounds. More detailed descriptions of the algorithms and our contributions are as follows, suppressing logarithmic and curvature factors.

- Building on the sample splitting and one-step estimation framework of SupSplitLog (Bae & Lee, 2026), we propose SupSplitLog+ (Algorithm 1), which reduces the additive dimension dependence from $d^{5/2}$ to $d^{3/2}$ while retaining the leading $\sqrt{dT}$ term. Our sharper analysis controls the accumulated tangent approximation error rather than applying the largest initial prediction error to every observation. Together with a revised sample splitting rule and a modified, appropriately regularized initial estimator, this analysis permits a larger uncertainty threshold for initial estimation while preserving the confidence guarantee for arm elimination, yielding the smaller additive regret term.
- We propose SupSecantLog (Algorithm 2), which further reduces the additive dependence to $d^{4/3}$ with the same leading term. The algorithm replaces the tangent line with a secant line through two independently computed predictions. The leading approximation error changes from the square of one prediction error to the product of two independent prediction errors. Whereas squared errors are nonnegative, these products can have either sign. Reducing the bias of both estimates allows us to exploit cancellation between positive and negative products and obtain a smaller approximation-error bound.
- We provide an algorithm based on a doubling rule (Algorithm 3) that achieves both regret bounds with d replaced by a data-dependent complexity measure of the revealed features. This yields smaller bounds when these features lie in a low-dimensional subspace, without requiring prior knowledge of the subspace.

2 Related Work

Logistic bandits. Work on generalized linear and logistic bandits (Filippi et al., 2010) has improved curvature dependence (Faury et al., 2020; Abeille et al., 2021), computation (Jun et al., 2017; Faury et al., 2022; Zhang & Sugiyama, 2023), and confidence sets (Lee et al., 2024a;b). Randomized methods include Thompson sampling and perturbed estimates (Dong et al., 2019; Kveton et al., 2020; Janz et al., 2024; Kim et al., 2023). Jun et al. (2021) derived fixed-design confidence bounds for the logistic maximum likelihood estimate. Their contextual algorithm and that of Li et al. (2017) use the sample separation of Auer (2002); Chu et al. (2011) to obtain a leading $\sqrt{dT}$ dependence under context diversity. For a fixed finite arm set, HOMER (Mason et al., 2022) uses experimental design to obtain variance-dependent regret with an additive $\widetilde{\mathcal{O}}\left(\kappa d^2\right)$ term. Our setting instead allows the available features to change between rounds. SupSplitLog (Bae & Lee, 2026) removes context diversity using sample splitting and one-step correction.

One-step estimation and bias reduction. One-step estimation refines an initial estimate using a correction computed from the estimating equation (Van der Vaart, 2000). Finite-sample likelihood expansions separate linear noise from nonlinear errors (Spokoiny, 2017). Bias reduction and adjusted scores for generalized linear models have been studied by Cordeiro & McCullagh (1991); Firth (1993); Kosmidis & Firth (2009). In logistic bandits, Mason et al. (2022, Section 4 and Appendix F) identified estimation bias as an obstacle to reducing the dimension dependence of initial exploration. They conjectured that bias reduction could lower the additive dependence to $d^{4/3}$, leaving the required finite-sample concentration bound open. We combine two independent, bias-adjusted estimates with secant estimation to establish this exponent for time-varying context sets.

3 Problem Setting

We consider the K-armed logistic bandit setting of Bae & Lee (2026). At round $t \in [T]$, the learner observes $\mathcal{X}_t = \{x_{t,a} : a \in [K]\} \subset \mathbb{R}^d$, selects $a_t \in [K]$, and receives $r_t \in \{0, 1\}$ with conditional mean $\mu(\langle x_t, \theta_\star \rangle)$, where $x_t = x_{t,a_t}$, $\mu(z) = (1 + e^{-z})^{-1}$, and $\theta_\star$ is unknown. The goal is to minimize cumulative regret $R_T = \sum_{t=1}^{T} [\max_{a \in [K]} \mu(\langle x_{t,a}, \theta_\star \rangle) - \mu(\langle x_t, \theta_\star \rangle)]$.

Assumption 1. For every t, a, $\|x_{t,a}\|_2 \le 1$, and $\|\theta_\star\|_2 \le B$ for a known $B > 0$.

Assumption 2. There are independent $U_t \sim \mathrm{Unif}[0, 1]$, independent of the entire context sequence, such that $r_t = \mathbb{1}\{U_t \le \mu(\langle x_t, \theta_\star \rangle)\}$.

Assumption 1 is standard (Faury et al., 2020). Assumption 2 formalizes the independence condition underlying prior finite-armed bandit analyses (Li et al., 2017; Jun et al., 2021; Bae & Lee, 2026), separating the context process from the randomness generating the rewards. The context sets may vary over time, be dependent across rounds, and lie in a lower-dimensional subspace. As in Bae & Lee (2026), we do not impose a minimum eigenvalue assumption on the feature covariance matrix.

Notation. Write $\dot{\mu} = \mu(1 - \mu)$ and $\ddot{\mu} = \dot{\mu}(1 - 2\mu)$, and define $\kappa = 1/\dot{\mu}(B) = 2 + e^B + e^{-B}$. For $V \succ 0$, let $\|x\|_V = (x^\top V x)^{1/2}$. The notation $A \preceq B$ means that $B - A$ is positive semidefinite, $\|A\|_{\mathrm{F}} = (\mathrm{tr}(A^\top A))^{1/2}$ is the Frobenius norm, and $[z]_c = \max\{-c, \min\{c, z\}\}$ clips a scalar to $[-c, c]$. We use $T \ge 2$ and failure probability $\delta \in (0, 1)$. The notation $\widetilde{\mathcal{O}}(\cdot)$ suppresses logarithmic factors in $T, K, d, 1/\delta$. Comparisons that suppress curvature factors are stated explicitly.

4 SupSplitLog+: Improved One-Step Estimation

In this section, we introduce our first algorithm, SupSplitLog+. We begin with the prediction error needed to obtain a leading $\sqrt{dT}$ regret term and explain how the existing SupSplitLog algorithm obtains such predictions. We then identify why its analysis requires many observations, and show how to reduce this requirement. The resulting additive dependence on d improves from $d^{5/2}$ to $d^{3/2}$.

4.1 Preliminaries

The required prediction error. An arm's expected reward is determined by its logit $\langle x, \theta_\star \rangle$. Thus we can compare arms using an estimate $\langle x, \hat{\theta} \rangle$ and a bound on its error. If E is the set of observations used for this estimate, the desired bound has the form

$$|\langle x, \hat{\theta} - \theta_\star \rangle| \le \alpha \|x\|_{V_E^{-1}}, \qquad V_E = \lambda_E I + \sum_{i \in E} x_i x_i^\top. \tag{1}$$

Here $\lambda_E > 0$ is a regularization parameter, and $\|x\|_{V_E^{-1}}$ measures uncertainty in the direction x. The important requirement is that α have no polynomial dependence on d. On rounds assigned to E, summing these widths using the elliptical potential lemma then gives a leading $\widetilde{\mathcal{O}}(\sqrt{dT})$ term. In comparison, applying Cauchy-Schwarz to a parameter-error bound $\|\hat{\theta} - \theta_\star\|_{V_E} = \widetilde{\mathcal{O}}(\sqrt{d})$ introduces another $\sqrt{d}$ into the prediction bound.

The Auer framework. The first difficulty is that a bandit algorithm chooses future observations using past rewards. The level construction of Auer (2002); Chu et al. (2011) separates observations so that their rewards can still be analyzed as independent. Each level s maintains its own estimate and aims for prediction accuracy $\epsilon_s = 2^{-s}$. If an arm's width exceeds ϵ_s, the algorithm observes an arm with the largest width. Otherwise, it removes arms whose estimated logits are more than $2\epsilon_s$ below the largest estimate and passes the remaining arms to level $s + 1$.

Crucially, the widths used to collect observations at level s depend only on contexts. Rewards at this level determine which arms reach later levels, but do not determine the observations stored at this level. Given the context sequence and observations at lower levels, the contexts at level s are fixed and its rewards are independent. Lemma 15 states this property precisely. It allows scalar concentration for each offered arm, instead of first bounding the error of the entire parameter vector.

SupSplitLog and the one-step estimate. Logistic regression presents a second difficulty. Its estimate is a nonlinear function of the rewards, so fixed contexts alone do not make its prediction error a sum with fixed coefficients. SupSplitLog (Bae & Lee, 2026) addresses this by splitting each level's observations into two sets. The set P supplies an initial-point estimate $\bar{\theta}$. The set E supplies the rewards used to correct this estimate. These sets are assigned using contexts, so their rewards are independent under the level conditioning above. We omit level and time indices when discussing one such pair of sets.

The correction uses a tangent to the logistic function at the initial prediction. Write $u_i = \langle x_i, \theta_\star \rangle$, $z_i = [\langle x_i, \bar{\theta} \rangle]_B$, and $\varepsilon_i = r_i - \mu(u_i)$. Taylor expansion gives

$$r_i = \mu(z_i) + \dot{\mu}(z_i)(u_i - z_i) + \varepsilon_i + R_i, \qquad |R_i| \le |\langle x_i, \bar{\theta} - \theta_\star \rangle|^2 / 20. \tag{2}$$

The first two terms describe a line with known slope $\dot{\mu}(z_i)$. Fitting this linear relation to the rewards in E gives

$$H = \lambda_E I + \sum_{i \in E} \dot{\mu}(z_i) x_i x_i^\top, \qquad \hat{\theta} = H^{-1} \sum_{i \in E} \big(r_i - \mu(z_i) + \dot{\mu}(z_i) z_i\big) x_i. \tag{3}$$

This is the one-step estimate used here. Its prediction error is

$$\langle x, \hat{\theta} - \theta_\star \rangle = \underbrace{x^\top H^{-1} \sum_{i \in E} \varepsilon_i x_i}_{\text{reward noise}} \underbrace{- \lambda_E x^\top H^{-1} \theta_\star}_{\text{regularization}} + \underbrace{x^\top H^{-1} \sum_{i \in E} R_i x_i}_{\text{approximation error}}. \tag{4}$$

Conditioning also on the rewards in P fixes $\bar{\theta}$, and hence H, while the noises ε_i in E remain independent and centered. Hoeffding's inequality therefore controls the first term without a $\sqrt{d}$ factor. The regularization term is bounded directly. To obtain (1), we must also control the last term, which is the error introduced by replacing the logistic curve with a tangent line.

4.2 Reducing the Approximation Error

The number of observations assigned to P depends on how accurately $\bar{\theta}$ must predict the contexts in E. We can reduce this number if we bound the approximation error in (4) more accurately. By (2) and Cauchy-Schwarz,

$$\left| x^\top H^{-1} \sum_{i \in E} R_i x_i \right| \le \frac{\kappa}{20} \|x\|_{V_E^{-1}} \left(\sum_{i \in E} \langle x_i, \bar{\theta} - \theta_\star \rangle^4 \right)^{1/2}. \tag{5}$$

Each tangent error R_i is quadratic in the initial prediction error, and Cauchy-Schwarz introduces the sum $\sum_i R_i^2$. The existing analysis uses the largest initial prediction error to control every estimation observation. This gives

$$\left(\sum_{i \in E} \langle x_i, \bar{\theta} - \theta_\star \rangle^4 \right)^{1/2} \le \sqrt{|E|} \max_{i \in E} |\langle x_i, \bar{\theta} - \theta_\star \rangle|^2.$$

A large error at one context is charged to every observation on the right-hand side. We instead bound the sum on the left. Two changes are needed, one in how we assign observations and one in how we construct and analyze $\bar{\theta}$.

Controlling the sum through the collected contexts. Let $V = \lambda_p I + \sum_{i \in P} x_i x_i^\top$ and $X_E = \sum_{i \in E} x_i x_i^\top$. We maintain

$$\|x_i\|_{V^{-1}}^2 \le \tau_s \quad (i \in E), \qquad X_E \preceq \gamma_s V. \tag{6}$$

The threshold τ_s limits the uncertainty of $\bar{\theta}$ at each estimation context. The factor γ_s limits how much the estimation contexts repeat directions that are uncertain under P. Indeed, for any estimation error v,

$$\sum_{i \in E} \langle x_i, v \rangle^2 = v^\top X_E v \le \gamma_s \|v\|_V^2, \qquad \sum_{i \in E} \langle x_i, v \rangle^4 \le \tau_s \gamma_s \|v\|_V^4.$$

Thus the second condition controls the accumulated error of the same vector $\bar{\theta} - \theta_\star$ across E. For repeated copies of one unit context, it reduces to $|E| \le \gamma_s(\lambda_p + |P|)$. Once E becomes too large relative to P, new observations must improve $\bar{\theta}$ before they can be used for correction. This is a rule on collected observations and requires no covariance assumption.

Analyzing the initial estimation error. The two conditions alone do not give our desired improvement. Using only $\|\bar{\theta} - \theta_\star\|_V^2 = \widetilde{\mathcal{O}}(d)$ would bound the fourth-power sum by $\widetilde{\mathcal{O}}(\tau_s \gamma_s d^2)$. We obtain a smaller bound by separating the part of $\bar{\theta} - \theta_\star$ that is linear in the reward noise.

Our initial estimate solves

$$\lambda_p \bar{\theta} + \sum_{i \in P} \big(\mu_B(\langle x_i, \bar{\theta}\rangle) - r_i\big) x_i = 0, \qquad \mu_B(z) = \mu(0) + \int_0^z \dot{\mu}([u]_B)\, du. \tag{7}$$

The function μ_B equals μ on $[-B, B]$ and extends it linearly outside, keeping its derivative at least $1/\kappa$. To isolate reward noise, replace each reward in this equation by its mean and denote the resulting solution by θ°. Define

$$\Psi(\theta) = \lambda_p \theta + \sum_{i \in P} \big(\mu_B(\langle x_i, \theta\rangle) - \mu(u_i)\big) x_i, \qquad \Psi(\theta^\circ) = 0, \qquad H^\circ = \nabla \Psi(\theta^\circ).$$

Unlike $\bar{\theta}$, the vector θ° depends only on contexts and $\theta_\star$. Since $\Psi(\bar{\theta}) = \sum_{i \in P} \varepsilon_i x_i$, expanding around θ° and writing $\mathcal{R}(v) = \Psi(\theta^\circ + v) - \Psi(\theta^\circ) - H^\circ v$ gives

$$\bar{\theta} - \theta_\star = \underbrace{(H^\circ)^{-1} \sum_{i \in P} \varepsilon_i x_i}_{\xi \text{ (linear reward noise)}} + \underbrace{\theta^\circ - \theta_\star - (H^\circ)^{-1} \mathcal{R}(\bar{\theta} - \theta^\circ)}_{b \text{ (regularization bias and nonlinear error)}}. \tag{8}$$

The coefficients in ξ are fixed when contexts are fixed. We can therefore bound its prediction at each context by scalar concentration. The residual b contains regularization bias and nonlinear error, which we bound separately.

Increasing λ_p reduces the nonlinear remainder but increases regularization bias. Their squared V-norm bounds have orders d/λ_p and λ_p, suppressing curvature and logarithmic factors. Choosing λ_p of order $\sqrt{d}$ balances them and gives

$$\max_{i \in E} \langle x_i, \xi\rangle^2 = \widetilde{\mathcal{O}}(\tau_s), \qquad \sum_{i \in E} \langle x_i, \xi\rangle^2 = \widetilde{\mathcal{O}}(\gamma_s d), \qquad \|b\|_V^2 = \widetilde{\mathcal{O}}\Big(\sqrt{d}\Big).$$

The first two bounds imply $\sum_{i \in E} \langle x_i, \xi\rangle^4 = \widetilde{\mathcal{O}}(\tau_s \gamma_s d)$. For b, (6) gives $\sum_{i \in E} \langle x_i, b\rangle^4 \le \tau_s \gamma_s \|b\|_V^4 = \widetilde{\mathcal{O}}(\tau_s \gamma_s d)$. Consequently, $\sum_{i \in E} \langle x_i, \bar{\theta} - \theta_\star\rangle^4 = \widetilde{\mathcal{O}}(\tau_s \gamma_s d)$, improving the d^2 bound above. Section C gives the complete calculation.

4.3 Algorithm and Confidence Bound

SupSplitLog+ in Algorithm 1 uses fixed tie-breaking and maintains separate sample sets $P^{(s)}$ and $E^{(s)}$ at each level. Each round starts with all K arms at level 0. At level s, it computes $\bar{\theta}$ from $P^{(s)}$ by (7) and $\hat{\theta}$ from $E^{(s)}$ by (3). If all confidence widths are at most $T^{-1/2}$, it plays the arm with the largest estimated logit. Otherwise, it explores the widest arm if its width exceeds $\epsilon_s = 2^{-s}$. In the remaining case, it eliminates arms whose estimated logits are more than $2\epsilon_s$ below the largest estimate and passes the remaining arms to level $s+1$. These sample sets accumulate across rounds. An exploration observation enters $E^{(s)}$ only if

$$\|x\|_{V^{-1}}^2 \le \tau_s, \qquad \|x\|_{(\gamma_s V - X_E)^{-1}}^2 \le 1/2. \tag{9}$$

Otherwise it enters $P^{(s)}$. The first test supplies the first condition in (6). The second ensures $\gamma_s V - X_E - xx^\top \succ 0$, so the matrix comparison remains valid after the observation enters E.

Set $M_d = \max\{1, d \log(1 + T/d)\}$ and $L = \log(32STK/\delta)$, where $S = \lceil \log_2 T \rceil$. For design budget $M = M_d$, the thresholds are $\tau_s = \rho \epsilon_s / (\alpha_1 L \sqrt{M})$ and $\gamma_s = \alpha_1^2 \tau_s / \epsilon_s^2$. The regularization parameters, confidence multiplier α_1, and factor ρ are specified in Section G.

Lemma 1. *Under Assumptions 1 and 2, SupSplitLog+(M_d, δ) satisfies, with probability at least $1 - \delta$, at every visited level s and round $t \le T$,*

$$\sum_{i \in E} \langle x_i, \bar{\theta} - \theta_\star\rangle^4 \le c_m L^2 \tau_s \gamma_s M_d, \qquad |\langle x_{t,a}, \hat{\theta} - \theta_\star\rangle| \le \alpha_1 \|x_{t,a}\|_{V_E^{-1}} \quad (a \in [K]),$$

where $c_m = 120\kappa^4(1 + B\kappa^2)^4$ and all sets and estimates use observations before round t.

The first bound controls the approximation error in (5). Adding the noise and regularization bounds gives the second, which is the prediction guarantee sought in (1).

Algorithm 1 SupSplitLog+(M, δ)

input: T, K, B, budget $M \geq 1$, failure probability δ, parameters in Section G
initialize: $S \leftarrow \lceil \log_2 T \rceil$, $\epsilon_s \leftarrow 2^{-s}$, $P^{(s)}, E^{(s)} \leftarrow \varnothing$ for $s = 0, \ldots, S$
for $t = 1, \ldots, T$ **do**
  Observe $\mathcal{X}_t$, set $\mathcal{A} \leftarrow [K]$
  for $s = 0, \ldots, S$ **do**
    Compute V, V_E, X_E from $P^{(s)}, E^{(s)}$
    Compute $\bar{\theta}$ by (7) // one initial-point estimate
    Compute $\hat{\theta}$ by (3) // tangent estimate
    $m(a) \leftarrow \langle x_{t,a}, \hat{\theta} \rangle$, $w(a) \leftarrow \alpha_1 \|x_{t,a}\|_{V_E^{-1}}$ for $a \in \mathcal{A}$
    if $\max_{a \in \mathcal{A}} w(a) \leq T^{-1/2}$ **then**
      Play $a_t \in \arg\max_{a \in \mathcal{A}} m(a)$, discard its reward, and break // exploit
    else if $\max_{a \in \mathcal{A}} w(a) > \epsilon_s$ **then**
      $a_t \in \arg\max_{a \in \mathcal{A}} w(a)$, $x \leftarrow x_{t,a_t}$
      Play a_t and observe r_t
      if $\|x\|^2_{V^{-1}} \leq \tau_s$ and $\|x\|^2_{(\gamma_s V - X_E)^{-1}} \leq 1/2$ **then**
        $E^{(s)} \leftarrow E^{(s)} \cup \{t\}$ // use for the one-step correction
      else
        $P^{(s)} \leftarrow P^{(s)} \cup \{t\}$ // improve the initial-point estimate
      end if
      break
    else
      $\mathcal{A} \leftarrow \{a \in \mathcal{A} : m(a) \geq \max_{b \in \mathcal{A}} m(b) - 2\epsilon_s\}$ // eliminate
    end if
  end for
end for

4.4 Regret Upper Bound

Theorem 1. *Under Assumptions 1 and 2, SupSplitLog+(M_d, δ) satisfies, with probability at least $1 - \delta$,*

$$R_T = \widetilde{\mathcal{O}}\Big(\kappa\sqrt{dT} + \kappa^3(1 + B\kappa^2)^2 d^{3/2}\Big).$$

Proof sketch. By Lemma 1, an optimal arm survives every elimination step. An arm reaching level $s \geq 1$ has survived elimination at level $s-1$, so its true logit is within $4\epsilon_{s-1} = 8\epsilon_s$ of the optimum. Since μ is $1/4$-Lipschitz, each exploration round at level s incurs regret at most $2\epsilon_s$. The same bound holds at level 0 because rewards lie in $[0, 1]$. We therefore count observations assigned to $E^{(s)}$ and $P^{(s)}$. Exploitation rounds contribute at most $T/(2\sqrt{T}) = \sqrt{T}/2$.

Summing the exploration widths over the estimation sets using the elliptical potential lemma bounds regret by $C\alpha_1\sqrt{(S+1)M_d T}$. For $P^{(s)}$, the first test in (9) fails at most $2M_d/\tau_s$ times, again by the elliptical potential lemma. To count failures of the second test, consider $Z = \gamma_s V - X_E$. An observation assigned to $P^{(s)}$ on such a failure multiplies $\det Z$ by more than $1 + \gamma_s/2$, whereas an observation assigned to $E^{(s)}$ reduces it by at most a factor of 2. Comparing these changes bounds the second-test count by $(M_d + |E^{(s)}| \log 2)/\log(1+\gamma_s/2)$. Combining both counts with $\gamma_s = \alpha_1^2 \tau_s/\epsilon_s^2$,

$$R_T \leq C\alpha_1\sqrt{(S+1)M_d T} + C\sum_{s=0}^{S} M_d \epsilon_s/\tau_s + \sqrt{T}/2.$$

The improvement to $d^{3/2}$ comes from the larger threshold τ_s allowed by our approximation-error bound. By (5) and Lemma 1, its multiplier has order $\sqrt{\tau_s \gamma_s M_d}$. Substituting $\gamma_s = \alpha_1^2 \tau_s/\epsilon_s^2$ permits τ_s of order $\epsilon_s/\sqrt{M_d}$, suppressing curvature and logarithmic factors. Hence $M_d \epsilon_s/\tau_s$ has order $M_d^{3/2}$ at each level. Summing over $S+1$ levels and using $M_d = \widetilde{\mathcal{O}}(d)$ proves the result. See Section B for the full proof. □

5 SUPSECANTLOG: FROM TANGENT TO SECANT ESTIMATION

We now introduce SupSecantLog, which further reduces the additive dimension dependence from $d^{3/2}$ to $d^{4/3}$. The remaining limitation of SupSplitLog+ is the squared initial-point prediction error in (2). Even if this prediction error were centered, its square would generally have positive expectation. We replace the tangent line by a secant line through two independent initial-point predictions. Its approximation error is a product of the two prediction errors, allowing us to use their independence. We explain the required changes to the regression and the initial-point estimates below.

5.1 REPLACING SQUARED ERRORS BY INDEPENDENT PRODUCTS

Secant approximation. Split the observations at a level into two initial-point sets P_1, P_2 and an estimation set E. Compute $\bar{\theta}_1, \bar{\theta}_2$ from P_1, P_2, respectively. For an estimation context x_i, set $z_{k,i} = [\langle x_i, \bar{\theta}_k \rangle]_{B+1}$. The secant line connects the two predicted probabilities $(z_{1,i}, \mu(z_{1,i}))$ and $(z_{2,i}, \mu(z_{2,i}))$. Its slope and the resulting observation model are

$$\omega_i = \frac{\mu(z_{1,i}) - \mu(z_{2,i})}{z_{1,i} - z_{2,i}}, \qquad r_i = \mu(z_{1,i}) + \omega_i(u_i - z_{1,i}) + \varepsilon_i + \rho_i,$$

where $\omega_i = \dot{\mu}(z_{1,i})$ when $z_{1,i} = z_{2,i}$. Here $u_i = \langle x_i, \theta_\star \rangle$ and $\varepsilon_i = r_i - \mu(u_i)$, as before. Secant interpolation gives $|\rho_i| \le |z_{1,i} - u_i||z_{2,i} - u_i|/20$. Thus the approximation becomes exact if either initial-point prediction equals u_i. More importantly, its leading error contains the product of two errors, instead of the square of one error.

Normalizing the observations. A direct replacement of the tangent slope by ω_i in (3) would give the regression matrix $\lambda_E I + \sum_{i \in E} \omega_i x_i x_i^\top$. Its inverse depends on both initial-point estimates. Consequently, the coefficients multiplying their errors also depend on those errors, and independence cannot be used directly. We instead divide the observation equation by ω_i and regress the response $z_{1,i} + (r_i - \mu(z_{1,i}))/\omega_i$ on x_i. This gives

$$\hat{\theta} = V_E^{-1} \sum_{i \in E} \left(z_{1,i} + \frac{r_i - \mu(z_{1,i})}{\omega_i} \right) x_i, \qquad V_E = \lambda_E I + \sum_{i \in E} x_i x_i^\top. \tag{10}$$

The inverse matrix now depends only on contexts. Conditionally on the initial-point estimates, the response has mean $u_i + \rho_i/\omega_i$. It remains to understand the normalized approximation error ρ_i/ω_i.

Write $e_{k,i} = \langle x_i, \bar{\theta}_k - \theta_\star \rangle$. When $|e_{k,i}| \le 1/2$, clipping at $B+1$ does not change the predictions. The observation-assignment tests below ensure this accuracy with high probability. Expanding the normalized approximation error gives

$$\frac{\rho_i}{\omega_i} = \underbrace{\frac{\ddot{\mu}(u_i)}{2\dot{\mu}(u_i)} e_{1,i} e_{2,i}}_{\text{quadratic term}} + \mathcal{R}_{3,i}, \qquad |\mathcal{R}_{3,i}| \le \frac{5e^2}{12} |e_{1,i} e_{2,i}| (|e_{1,i}| + |e_{2,i}|). \tag{11}$$

The quadratic coefficient depends on u_i, not on either estimate. Both this expansion and the context-only inverse in (10) are needed to use independence. The remaining term is cubic in the initial-point prediction errors and is smaller when these errors are small.

5.2 REDUCING THE BIAS OF THE INITIAL-POINT ESTIMATES

Recall that for the regret bounds in terms of the ambient dimension, we set the design budget to $M = M_d = \widetilde{\mathcal{O}}(d)$.

Two independent estimates do not make the entire product $e_{1,i} e_{2,i}$ centered. In (8), only ξ is centered, whereas b contains the regularization bias and the nonlinear remainder. The bound $\|b\|_V^2 = \widetilde{\mathcal{O}}(M^{1/2})$ is sufficient for SupSplitLog+, but products involving these residuals would retain its $M^{3/2}$ additive regret. We therefore reduce the residuals before using secant estimation.

The leading nonlinear term in the initial-point score expansion is quadratic in ξ. Its mean is generally nonzero. We subtract an estimate of this mean from the score equation. Let g be the smooth

extension of μ in Lemma 11, which agrees with μ on $[-B, B]$. For an initial-point set P, define

$$H_P(\theta) = \lambda_p I + \sum_{i \in P} g'(\langle x_i, \theta \rangle) x_i x_i^\top,$$

$$A_P(\theta) = \frac{1}{2} \sum_{i \in P} g''(\langle x_i, \theta \rangle) \big[x_i^\top H_P(\theta)^{-1} (H_P(\theta) - \lambda_p I) H_P(\theta)^{-1} x_i \big] x_i.$$

The bracket estimates the variance of the linear prediction error at x_i. Multiplying it by $g''(\langle x_i, \theta \rangle)/2$ estimates the mean quadratic term. We compute $\bar{\theta}_1, \bar{\theta}_2$ from

$$\lambda_p \bar{\theta}_k + \sum_{i \in P_k} (g(\langle x_i, \bar{\theta}_k \rangle) - r_i) x_i - A_{P_k}(\bar{\theta}_k) = 0, \qquad k = 1, 2. \tag{12}$$

The parameter choice in Section G ensures a unique solution.

For each set P_k, let θ° solve the mean-reward equation before adjustment

$$\lambda_p \theta^\circ + \sum_{i \in P_k} \big(g(\langle x_i, \theta^\circ \rangle) - \mu(u_i) \big) x_i = 0.$$

Write $H^\circ = H_{P_k}(\theta^\circ)$ and define ξ as in (8). The quadratic term in the unadjusted expansion is $-\frac{1}{2}(H^\circ)^{-1} \mathcal{T}[\xi, \xi]$, where $\mathcal{T}[v, w] = \sum_{i \in P_k} g''(\langle x_i, \theta^\circ \rangle) \langle x_i, v \rangle \langle x_i, w \rangle \, x_i$. The adjustment in (12) compensates for this mean, up to an error included in b_k. We obtain

$$\bar{\theta}_k - \theta_\star = \underbrace{\xi - \tfrac{1}{2}(H^\circ)^{-1} \big(\mathcal{T}[\xi, \xi] - \mathbb{E}[\mathcal{T}[\xi, \xi]] \big)}_{\zeta_k} + b_k. \tag{13}$$

Conditionally on the context sequence and observations below the level, $\mathbb{E}[\zeta_k] = 0$ and $\mathrm{Cov}(\zeta_k) \preceq C_v V_k^{-1}$ for $V_k = \lambda_p I + \sum_{i \in P_k} x_i x_i^\top$. The residual satisfies $\|b_k\|_{V_k}^2 = \widetilde{\mathcal{O}}\big(M^{1/3}\big)$. This follows by balancing its regularization and cubic-remainder bounds, of orders λ_p and $L^3 M / \lambda_p^2$. We choose $\lambda_p = C_p L M^{1/3}$. Here and in the dimension comparisons below, the suppressed factors depend on κ, B and logarithms. Lemma 7 gives the complete expansion and bounds.

5.3 Algorithm and Regret Upper Bound

SupSecantLog in Algorithm 2 uses the same levels, confidence-width selection, and elimination rule as SupSplitLog+. At each level it maintains the three sets P_1, P_2, E. An exploration observation enters E only when both tests in (9) hold for each initial-point set. Otherwise, it enters an initial-point set whose test fails. Thus (6) holds with both V_1 and V_2. The tests use only contexts, preserving the conditional independence of the two initial-point estimates and the estimation rewards.

We use $\tau_s = \epsilon_s / (4 C_2 L^2 M^{1/3})$ and $\gamma_s = \alpha_2^2 \tau_s / \epsilon_s^2$, where $\alpha_2 = O(\sqrt{(1+B)\kappa L})$. All parameters and constants, including the exact value of α_2, are given in Section G. The larger threshold τ_s, compared with its $M^{-1/2}$ dependence in SupSplitLog+, reduces the number of observations assigned to the initial-point sets. The following confidence bound justifies using this threshold.

Lemma 2. *Under Assumptions 1 and 2, SupSecantLog(M_d, δ) satisfies, with probability at least $1 - \delta$, at every visited level and round $t \le T$, $|\langle x_{t,a}, \hat{\theta} - \theta_\star \rangle| \le \alpha_2 \|x_{t,a}\|_{V_E^{-1}}$ for all $a \in [K]$. The estimate and design use observations before t.*

The reason this bound permits a larger τ_s is the product of independent centered errors in (13). For a fixed test context x, define

$$B_x = \sum_{i \in E} \frac{\ddot{\mu}(u_i)}{2 \dot{\mu}(u_i)} (x^\top V_E^{-1} x_i) x_i x_i^\top.$$

The quadratic term in the prediction error is $(\bar{\theta}_1 - \theta_\star)^\top B_x (\bar{\theta}_2 - \theta_\star)$. Condition on the context sequence and observations below the level, so that B_x, V_1, V_2 are fixed and ζ_1, ζ_2 are independent. Their covariance bounds and (6) give

$$\mathbb{E}[(\zeta_1^\top B_x \zeta_2)^2] \le C_v^2 \|V_1^{-1/2} B_x V_2^{-1/2}\|_{\mathrm{F}}^2 \le \frac{C_v^2 \tau_s \gamma_s}{4} \|x\|_{V_E^{-1}}^2. \tag{14}$$

Algorithm 2 SupSecantLog(M, δ)

input: T, K, B, budget $M \geq 1$, failure probability δ, parameters in Section G
initialize: $S \leftarrow \lceil \log_2 T \rceil$, $\epsilon_s \leftarrow 2^{-s}$, $P_1^{(s)}, P_2^{(s)}, E^{(s)} \leftarrow \varnothing$ for $s = 0, \ldots, S$
for $t = 1, \ldots, T$ **do**
  Observe $\mathcal{X}_t$, set $\mathcal{A} \leftarrow [K]$
  for $s = 0, \ldots, S$ **do**
    Compute V_1, V_2, V_E, X_E from $P_1^{(s)}, P_2^{(s)}, E^{(s)}$
    Compute $\bar{\theta}_1, \bar{\theta}_2$ by (12) // two initial-point estimates
    Compute $\hat{\theta}$ by (10) // normalized secant estimate
    $m(a) \leftarrow \langle x_{t,a}, \hat{\theta} \rangle$, $w(a) \leftarrow \alpha_2 \|x_{t,a}\|_{V_E^{-1}}$ for $a \in \mathcal{A}$
    if $\max_{a \in \mathcal{A}} w(a) \leq T^{-1/2}$ **then**
      Play $a_t \in \arg\max_{a \in \mathcal{A}} m(a)$, discard its reward, and break // exploit
    else if $\max_{a \in \mathcal{A}} w(a) > \epsilon_s$ **then**
      $a_t \in \arg\max_{a \in \mathcal{A}} w(a)$, $x \leftarrow x_{t,a_t}$ // explore
      Play a_t and observe r_t
      if both tests in (9) hold for $k = 1, 2$ **then** // check both initial-point sets
        $E^{(s)} \leftarrow E^{(s)} \cup \{t\}$
      else
        Choose the first $k \in \{1, 2\}$ whose test fails
        $P_k^{(s)} \leftarrow P_k^{(s)} \cup \{t\}$
      end if
      break
    else
      $\mathcal{A} \leftarrow \{a \in \mathcal{A} : m(a) \geq \max_{b \in \mathcal{A}} m(b) - 2\epsilon_s\}$ // eliminate
    end if
  end for
end for

Unlike the tangent fourth-moment bound, this second moment contains no factor M. Each ζ_k is a polynomial of degree at most two in its reward noise. A polynomial concentration inequality therefore converts (14) into a high-probability bound with only additional logarithmic factors (Lemma 8). The terms involving b_1 or b_2 and the cubic remainder also fit within the confidence width, as summarized in the proof sketch.

Theorem 2. *Under Assumptions 1 and 2, SupSecantLog(M_d, δ) satisfies, with probability at least* $1 - \delta$,

$$R_T = \widetilde{\mathcal{O}}\Big(\sqrt{(1+B)\kappa d T} + (\kappa^{9/2} + B^2\kappa^{15/2})d^{4/3}\Big).$$

Proof sketch. We first account for the terms not included in (14). Expanding the quadratic prediction error gives $(\bar{\theta}_1 - \theta_\star)^\top B_x(\bar{\theta}_2 - \theta_\star) = \zeta_1^\top B_x \zeta_2 + \zeta_1^\top B_x b_2 + b_1^\top B_x \zeta_2 + b_1^\top B_x b_2$. For the second term, condition further on (ζ_2, b_2) and apply the polynomial concentration inequality to ζ_1. The third term is treated in the same way, and the last term uses $|b_1^\top B_x b_2| \leq \|b_1\|_{V_1} \|b_2\|_{V_2} \|V_1^{-1/2} B_x V_2^{-1/2}\|_{\mathrm{op}}$. Since $\|b_k\|_{V_k}^2 = \widetilde{\mathcal{O}}(M^{1/3})$, the entire quadratic term is at most $\widetilde{\mathcal{O}}\big(M^{1/3}\sqrt{\tau_s \gamma_s}\big)\|x\|_{V_E^{-1}}$. The cubic remainder in (11) is at most $\widetilde{\mathcal{O}}\big(M^{1/2}\tau_s\sqrt{\gamma_s}\big)\|x\|_{V_E^{-1}}$. Substituting $\gamma_s = \alpha_2^2 \tau_s / \epsilon_s^2$ and τ_s proportional to $\epsilon_s / M^{1/3}$ makes these contributions of order $\alpha_2 \|x\|_{V_E^{-1}}$. Adding the reward-noise and regularization bounds proves Lemma 2. See Section E for the full calculation.

The regret bound now follows from the same observation counts as Theorem 1, applied to both initial-point sets. Their contribution is at most $C \sum_s M_d \epsilon_s / \tau_s = 4C(S+1)C_2 L^2 M_d^{4/3}$. Together with the leading term $C\alpha_2\sqrt{(S+1)M_d T}$, this gives the stated bound. □

Remark 1. Both regret bounds can replace d, up to logarithmic factors, by a log-determinant of the revealed contexts. A doubling rule selects the required design budget without knowing this quantity in advance. The construction and bounds appear in Section F.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 3691–3699. PMLR, 2021.

Arya Akhavan, Amitis Shidani, Alex Ayoub, and David Janz. Bernstein-type dimension-free concentration for self-normalised martingales. *arXiv preprint arXiv:2507.20982*, 2025.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of machine learning research*, 3(Nov):397–422, 2002.

Seoungbin Bae and Dabeen Lee. Neural logistic bandits. *arXiv preprint arXiv:2505.02069*, 2025.

Seoungbin Bae and Dabeen Lee. Logistic bandits with $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret without context diversity assumptions. *arXiv preprint arXiv:2604.22161*, 2026.

Pierre Boudart, Pierre Gaillard, and Alessandro Rudi. Enjoying non-linearity in multinomial logistic bandits: A minimax-optimal algorithm. *arXiv preprint arXiv:2507.05306*, 2025.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pp. 844–853. PMLR, 2017.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Eugenio Clerico, Hamish Flynn, Wojciech Kotłowski, and Gergely Neu. Confidence sequences for generalized linear models via regret analysis. *arXiv preprint arXiv:2504.16555*, 2025.

Gauss M Cordeiro and Peter McCullagh. Bias correction in generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 53(3):629–643, 1991.

Qin Ding, Cho-Jui Hsieh, and James Sharpnack. An efficient algorithm for generalized linear bandit: Online stochastic gradient descent and thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pp. 1585–1593. PMLR, 2021.

Shi Dong, Tengyu Ma, and Benjamin Van Roy. On the performance of thompson sampling on logistic bandits. In *Conference on Learning Theory*, pp. 1158–1160. PMLR, 2019.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pp. 3052–3060. PMLR, 2020.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 546–580. PMLR, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. *Advances in neural information processing systems*, 23, 2010.

David Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, pp. 27–38, 1993.

Amaury Gouverneur, Tobias Oechtering, and Mikael Skoglund. An information-theoretic analysis of thompson sampling for logistic bandit problems. *Transactions on Machine Learning Research*, 2026.

Daniel Hsu, Sham Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(52):1–6, 2012. doi: 10.1214/ECP.v17-2079.

David Janz, Shuai Liu, Alex Ayoub, and Csaba Szepesvári. Exploration via linearly perturbed loss minimisation. In *International Conference on Artificial Intelligence and Statistics*, pp. 721–729. PMLR, 2024.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. *Advances in Neural Information Processing Systems*, 30, 2017.

Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved confidence bounds for the linear logistic model and applications to bandits. In *International Conference on Machine Learning*, pp. 5148–5157. PMLR, 2021.

Parnian Kassraie and Andreas Krause. Neural contextual bandits without regret. In *International Conference on Artificial Intelligence and Statistics*, pp. 240–278. PMLR, 2022.

Wonyoung Kim, Kyungbok Lee, and Myunghee Cho Paik. Double doubly robust thompson sampling for generalized linear contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8300–8307, 2023.

Johannes Kirschner, Andreas Krause, Michele Meziu, and Mojmir Mutny. Confidence estimation via sequential likelihood mixing. *arXiv preprint arXiv:2502.14689*, 2025.

Ioannis Kosmidis and David Firth. Bias reduction in exponential family nonlinear models. *Biometrika*, 96(4):793–804, 2009.

Ioannis Kosmidis and Nicola Lunardon. Empirical bias-reducing adjustments to estimating functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):62–89, 2024.

Branislav Kveton, Manzil Zaheer, Csaba Szepesvari, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 2066–2076. PMLR, 2020.

Joongkyu Lee and Min-hwan Oh. Improved online confidence bounds for multinomial logistic bandits. *arXiv preprint arXiv:2502.10020*, 2025.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved regret bounds of (multinomial) logistic bandits via regret-to-confidence-set conversion. In *International Conference on Artificial Intelligence and Statistics*, pp. 4474–4482. PMLR, 2024a.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. *arXiv preprint arXiv:2407.13977*, 2024b.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pp. 2071–2080. PMLR, 2017.

Yingkai Li, Yining Wang, and Yuan Zhou. Nearly minimax-optimal regret for linearly parameterized bandits. *IEEE Transactions on Information Theory*, 70(1):372–388, 2023.

Licong Lin, Fangzhou Su, Wenlong Mou, Peng Ding, and Martin Wainwright. When is it worthwhile to jackknife? breaking the quadratic barrier for z-estimators. *arXiv preprint arXiv:2411.02909*, 2024.

Shuai Liu, Alex Ayoub, Flore Sentenac, Xiaoqi Tan, and Csaba Szepesvári. Almost free: Self-concordance in natural exponential families and an application to bandits. *Advances in Neural Information Processing Systems*, 37:70828–70874, 2024.

Blake Mason, Kwang-Sung Jun, and Lalit Jain. An experimental design approach for regret minimization in logistic bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7736–7743, 2022.

Alberto Maria Metelli, Simone Drago, and Marco Mussi. Generalized kernelized bandits: A novel self-normalized bernstein-like dimension-free inequality and regret bounds. *arXiv preprint arXiv:2508.01681*, 2025.

Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

Robert L Schaefer. Bias correction in maximum likelihood logistic regression. *Statistics in Medicine*, 2(1):71–78, 1983.

Vladimir Spokoiny. Penalized maximum likelihood estimation and effective dimension. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):389–429, 2017. doi: 10.1214/15-AIHP720.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Arun Verma, Zhongxiang Dai, Xiaoqiang Lin, Patrick Jaillet, and Bryan Kian Hsiang Low. Neural dueling bandits: Preference-based optimization with human feedback. In *International Conference on Learning Representations*, volume 2025, pp. 97145–97182, 2025.

Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. *Advances in Neural Information Processing Systems*, 36:29741–29782, 2023.

Yu-Jie Zhang, Sheng-An Xu, Peng Zhao, and Masashi Sugiyama. Generalized linear bandits: Almost optimal regret with one-pass update. *Advances in Neural Information Processing Systems*, 38:69244–69277, 2025.

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with ucb-based exploration. In *International conference on machine learning*, pp. 11492–11502. PMLR, 2020.

A Additional Related Work

Logistic bandits. Logistic bandits have been studied as a special case of generalized linear bandits (Filippi et al., 2010). Subsequent work improves the dependence on logistic curvature through self-concordant analysis and obtains instance-dependent regret bounds (Faury et al., 2020; Abeille et al., 2021; Faury et al., 2022; Liu et al., 2024). Online estimation reduces the computational cost of these algorithms (Jun et al., 2017; Zhang & Sugiyama, 2023; Zhang et al., 2025). Recent confidence bounds use likelihood ratios, sequential prediction, and online regret (Lee et al., 2024a;b; Clerico et al., 2025; Kirschner et al., 2025). Thompson sampling and randomized estimators provide alternative exploration rules (Dong et al., 2019; Kveton et al., 2020; Ding et al., 2021; Janz et al., 2024; Gouverneur et al., 2026). In particular, Kim et al. (2023) combine doubly robust estimation with Thompson sampling under i.i.d. contexts and a positive covariance condition. Recent multinomial logistic bandit algorithms also improve confidence bounds and exploit reward variance and the curvature at optimal actions (Lee & Oh, 2025; Boudart et al., 2025).

Dimension dependence in contextual, kernel and neural bandits. For finite contextual action sets, the observation splitting of Auer (2002) and Chu et al. (2011) gives leading $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret for linear rewards, with logarithmic dependence on the number of actions (Li et al., 2023). SupCB-GLM and SupLogistic extend this approach to nonlinear rewards using fixed-design confidence bounds, but require context diversity for initial exploration (Li et al., 2017; Jun et al., 2021). SupSplitLog removes this requirement through sample splitting and one-step estimation for context sequences independent of past actions and rewards (Bae & Lee, 2026). We reduce its additive dimension dependence from $d^{5/2}$ to $d^{3/2}$ and $d^{4/3}$. For a fixed finite arm set, HOMER uses experimental design to obtain variance-dependent regret with an additive $\widetilde{\mathcal{O}}(\kappa d^2)$ term (Mason et al., 2022). Our algorithms instead allow the available features to change across rounds.

Kernel bandits express regret through spectral effective dimension or information gain (Srinivas et al., 2009; Valko et al., 2013; Chowdhury & Gopalan, 2017). Neural bandit analyses similarly express their leading regret bounds in terms of the neural tangent kernel's effective dimension or information gain (Zhou et al., 2020; Kassraie & Krause, 2022). Removing context diversity and retaining $\sqrt{dT}$ dependence are important for extending no-regret guarantees to kernel and neural models. A positive covariance lower bound need not hold for infinite-dimensional or linearly dependent features. Moreover, when d is replaced by an effective dimension growing with T, the $\sqrt{dT}$ leading term remains sublinear under weaker growth conditions than $d\sqrt{T}$ (Valko et al., 2013; Kassraie & Krause, 2022). Neural methods for binary feedback (Verma et al., 2025; Bae & Lee, 2025) and Bernstein-type inequalities in Hilbert spaces (Akhavan et al., 2025; Metelli et al., 2025) extend these ideas to nonlinear reward models. Our doubling rule adapts the finite-dimensional logistic regret bounds to the log-determinant of the revealed features.

One-step estimation and bias reduction. One-step estimators refine an initial estimate through its estimating equation (Van der Vaart, 2000). Classical corrections and adjusted scores reduce the leading bias of maximum likelihood estimates (Schaefer, 1983; Cordeiro & McCullagh, 1991; Firth, 1993; Kosmidis & Firth, 2009), with more general estimating-function adjustments studied by Kosmidis & Lunardon (2024). Lin et al. (2024) show that jackknife correction relaxes dimension-dependent sample requirements for inference with Z-estimators. For logistic bandits, Mason et al. (2022, Section 4 and Appendix F) identify estimation bias as an obstacle to reducing initial exploration and conjecture an additive $d^{4/3}$ dependence through bias reduction. Like the finite-sample likelihood expansions of Spokoiny (2017), our analysis separates the initial estimation error into a linear score term and a remainder. We sharpen the accumulated tangent error bound in SupSplitLog+ and exploit products of independent, bias-adjusted errors in SupSecantLog.

B Proofs of Theorems 1 and 2

We prove both regret bounds using the same sampling analysis. Write A_1 for SupSplitLog+ and A_2 for SupSecantLog. Algorithm A_v uses $p = v$ initial-point sets. In this section only, we write $P_1 = P$ when $v = 1$. Thus every sum over the initial-point sets has at most two terms.

For a prefix of n rounds, let $V_k^{(s)}(n)$ and $V_E^{(s)}(n)$ denote the designs after the observations in that prefix have been added. We say that the budget is valid through n if

$$\max_{0\le s\le S} \max\left\{ \max_{1\le k\le p} \log\det(V_k^{(s)}(n)/\lambda_p), \log\det(V_E^{(s)}(n)/\lambda_E) \right\} \le M. \tag{15}$$

Define $N(M) = \max\{n \in \{0,\dots,T\} : (15) \text{ holds}\}$. The designs increase with n, so the budget is valid at every prefix up to $N(M)$. The parameters of both algorithms satisfy

$$\lambda_p > 0, \quad \lambda_E \ge 1, \quad \alpha \ge 1, \quad 0 < \tau_s \le \epsilon_s, \quad \gamma_s = \alpha^2\tau_s/\epsilon_s^2, \quad F = \epsilon_s/\tau_s,$$

where F does not depend on s. In particular, $F = \alpha_1 L\sqrt{M}/\rho$ for A_1 and $F = 4C_2L^2M^{1/3}$ for A_2.

The following lemma separates the regret due to estimating rewards from the regret due to collecting initial-point observations. We prove the lemma below and first use it to obtain the two theorems.

Lemma 3 (Regret up to the design budget)**.** *Under Assumptions 1 and 2, fix $M \ge 1$ and run A_v with the corresponding parameters in Section G. With probability at least $1-\delta$, simultaneously for all $n \le N(M)$,*

$$R_n \le 13\alpha\sqrt{(S+1)Mn} + n/(2\sqrt{T}) + 64(S+1)MF. \tag{16}$$

More generally, this inequality holds on every realization on which the confidence bounds of A_v hold at all visited levels through round $N(M)$.

Proof of Theorems 1 and 2. Set $M_d = \max\{1, d\log(1+T/d)\}$. Every design has regularization parameter at least one and at most T observations. By Lemma 14,

$$\log\det(V/\lambda) \le d\log(1+T/(d\lambda)) \le d\log(1+T/d) \le M_d.$$

Hence $N(M_d) = T$ and Lemma 3 gives, for either algorithm,

$$R_T \le 13\alpha\sqrt{(S+1)M_dT} + \sqrt{T}/2 + 64(S+1)M_dF.$$

For A_1, (91) gives

$$\alpha_1 \le 4\kappa\sqrt{L}, \qquad \alpha_1/\rho = \max\{2a_0, 2c_R\} \le 4\kappa\sqrt{L} + 2\kappa^3(1+B\kappa^2)^2.$$

Substituting $F = \alpha_1 L\sqrt{M_d}/\rho$ therefore yields

$$R_T = \widetilde{\mathcal{O}}\left(\kappa\sqrt{dT} + \kappa^3(1+B\kappa^2)^2 d^{3/2}\right).$$

For A_2, (94) gives $\alpha_2 \le 10\sqrt{(1+B)\kappa L}$ and $C_2 = \mathcal{O}\big(\kappa^{9/2} + B^2\kappa^{15/2}\big)$. Substituting $F = 4C_2L^2M_d^{1/3}$ yields

$$R_T = \widetilde{\mathcal{O}}\left(\sqrt{(1+B)\kappa dT} + (\kappa^{9/2} + B^2\kappa^{15/2})d^{4/3}\right).$$

Here we used $S = \mathcal{O}(\log T)$, $L = \mathcal{O}(\log(TK/\delta))$, and $M_d = \mathcal{O}(d\log(1+T/d))$. This proves both claims. □

B.1 Proof of Lemma 3

We first bound the number of observations assigned by each sampling test. Partition each initial-point set into $P_{k,\mathrm{loc}}^{(s)}(n)$ and $P_{k,\mathrm{comp}}^{(s)}(n)$. The first contains observations assigned to initial-point set k whose pre-update leverage exceeds τ_s. The second contains its remaining observations, for which the first test passes and the second test fails. Thus, for t in these two sets respectively,

$$x_t^\top (V_k^{(s)}(t-1))^{-1} x_t > \tau_s, \qquad x_t^\top (Z_k^{(s)}(t-1))^{-1} x_t > 1/2,$$

where $Z_k^{(s)} = \gamma_s V_k^{(s)} - X_E^{(s)}$.

Lemma 4 (Numbers of estimation and initial-point observations). *For every level s and $k \in [p]$, the sampling tests ensure*

$$Z_k^{(s)} \succ 0, \qquad X_E^{(s)} \preceq \gamma_s V_k^{(s)}, \qquad x_i^\top (V_k^{(s)})^{-1} x_i \le \tau_s \quad (i \in E^{(s)}).$$

For every $n \le N(M)$, they also give

$$\sum_{t \in E^{(s)}(n)} \|x_t\|^2_{(V_E^{(s)}(t-1))^{-1}} \le 3M/2, \qquad |E^{(s)}(n)| \le 3\alpha^2 M/(2\epsilon_s^2),$$

$$|P_{k,\mathrm{loc}}^{(s)}(n)| \le 2M/\tau_s, \qquad |P_{k,\mathrm{comp}}^{(s)}(n)| \le \frac{M + |E^{(s)}(n)| \log 2}{\log(1+\gamma_s/2)}.$$

Consequently,

$$\sum_{s=0}^{S} \sum_{t \in E^{(s)}(n)} \|x_t\|_{(V_E^{(s)}(t-1))^{-1}} \le \sqrt{3(S+1)Mn/2}.$$

Proof. Initially $Z_k^{(s)} = \gamma_s \lambda_p I \succ 0$. An observation assigned to $P_k^{(s)}$ adds $\gamma_s x x^\top$ to this matrix. An estimation observation is accepted only if $x^\top (Z_k^{(s)})^{-1} x \le 1/2$ for every $k \in [p]$. Its update satisfies

$$Z_k^{(s)} - x x^\top = (Z_k^{(s)})^{1/2} \left[I - (Z_k^{(s)})^{-1/2} x x^\top (Z_k^{(s)})^{-1/2} \right] (Z_k^{(s)})^{1/2} \succ 0,$$

because the rank-one matrix in brackets has largest eigenvalue at most $1/2$ before being subtracted from I. Induction proves $Z_k^{(s)} \succ 0$ and $X_E^{(s)} \preceq \gamma_s V_k^{(s)}$. The first sampling test gives the leverage bound when a context enters $E^{(s)}$. It remains valid because $V_k^{(s)}$ only increases.

For the estimation count, enumerate the contexts in $E^{(s)}(n)$ as $x_1, \dots, x_m$, and write $V_i = \lambda_E I + \sum_{j=1}^{i} x_j x_j^\top$ and $h_i = x_i^\top V_{i-1}^{-1} x_i$. By (104),

$$\sum_{i=1}^{m} h_i \le \frac{\lambda_E^{-1}}{\log(1+\lambda_E^{-1})} \log\det(V_m/\lambda_E) \le 3M/2,$$

$$|E^{(s)}(n)|\, \epsilon_s^2 \le \alpha^2 \sum_{i=1}^{m} h_i \le 3\alpha^2 M/2.$$

The first line uses $\lambda_E \ge 1$ and budget validity. The second uses $\alpha\sqrt{h_i} > \epsilon_s$ on every exploration round. Cauchy–Schwarz now gives

$$\sum_{s=0}^{S} \sum_{t \in E^{(s)}(n)} \|x_t\|_{(V_E^{(s)}(t-1))^{-1}} \le \sqrt{\left(\sum_{s=0}^{S} |E^{(s)}(n)| \right) \left(\sum_{s=0}^{S} \sum_{t \in E^{(s)}(n)} \|x_t\|^2_{(V_E^{(s)}(t-1))^{-1}} \right)}$$
$$\le \sqrt{3(S+1)Mn/2},$$

where $\sum_s |E^{(s)}(n)| \le n$ since the estimation sets are disjoint.

For observations assigned by the first test in (9), $\tau_s \le 1$ implies

$$\tau_s |P_{k,\mathrm{loc}}^{(s)}(n)| \le \sum_{t \in P_k^{(s)}(n)} \min\{1, x_t^\top (V_k^{(s)}(t-1))^{-1} x_t\}$$
$$\le 2 \log\det(V_k^{(s)}(n)/\lambda_p) \le 2M.$$

The second inequality is (103) applied to all observations assigned to that initial-point set.

For observations assigned by the second test in (9), consider $\det Z_k^{(s)}$. An estimation update multiplies it by $1 - x^\top (Z_k^{(s)})^{-1} x \ge 1/2$. An update in $P_{k,\mathrm{comp}}^{(s)}(n)$ multiplies it by $1 + \gamma_s x^\top (Z_k^{(s)})^{-1} x >$

$1+\gamma_s/2$. Other updates to initial-point set k increase it, and updates to another initial-point set leave it unchanged. The determinant identity in Lemma 14 therefore gives

$$|P^{(s)}_{k,\mathrm{comp}}(n)|\log(1+\gamma_s/2) - |E^{(s)}(n)|\log 2 \le \log\frac{\det Z^{(s)}_k(n)}{\det(\gamma_s\lambda_p I)} \le \log\det(V^{(s)}_k(n)/\lambda_p) \le M.$$

The second inequality follows from $Z^{(s)}_k(n) \preceq \gamma_s V^{(s)}_k(n)$. Rearranging proves the claimed count. □

Proof of Lemma 3. The proofs in Sections C and E establish the confidence bounds of Lemmas 1 and 2 for every deterministic $M \ge 1$ through $N(M)$, with probability at least $1-\delta$. We prove (16) on this event. Fix $n \le N(M)$. The standard elimination argument of Bae & Lee (2026, Lemmas 14 and 30) applies because it uses only the confidence bounds and elimination threshold. An optimal arm remains active. Since $\dot\mu \le 1/4$, the regret is at most $2\epsilon_s$ on exploration rounds at level $s \ge 1$, and at most $1/(2\sqrt{T})$ on exploitation rounds. At level zero, the regret is at most $1 \le 2\epsilon_0$. Every exploration observation belongs to exactly one estimation or initial-point set. Therefore,

$$R_n \le \underbrace{\sum_{s=0}^{S} 2\epsilon_s|E^{(s)}(n)|}_{R_E} + \underbrace{\sum_{s=0}^{S}\sum_{k=1}^{p} 2\epsilon_s|P^{(s)}_{k,\mathrm{loc}}(n)|}_{R_{P,1}} + \underbrace{\sum_{s=0}^{S}\sum_{k=1}^{p} 2\epsilon_s|P^{(s)}_{k,\mathrm{comp}}(n)|}_{R_{P,2}} + n/(2\sqrt{T}). \tag{17}$$

We bound the three terms separately. For R_E, the exploration condition and Lemma 4 give

$$R_E = 2\sum_{s=0}^{S}\sum_{t\in E^{(s)}(n)} \epsilon_s \le 2\alpha\sum_{s=0}^{S}\sum_{t\in E^{(s)}(n)} \|x_t\|_{(V^{(s)}_E(t-1))^{-1}} \le 2\alpha\sqrt{3(S+1)Mn/2}. \tag{18}$$

For $R_{P,1}$, Lemma 4 and $p \le 2$ give

$$R_{P,1} \le \sum_{s=0}^{S} 2\epsilon_s \cdot 2 \cdot (2M/\tau_s) = 8(S+1)MF. \tag{19}$$

For $R_{P,2}$, we first use $1/\log(1+u/2) \le 3(1+1/u)$ for $u > 0$. Indeed, for $u \le 1$, $\log(1+u/2) \ge (u/2)/(1+u/2) \ge u/3$. For $u > 1$, $\log(1+u/2) \ge \log(3/2) > 1/3$. Applying this inequality to the second count in Lemma 4 gives

$$R_{P,2} \le 12\sum_{s=0}^{S}\epsilon_s(M + |E^{(s)}(n)|\log 2)(1+1/\gamma_s) = \underbrace{12M\sum_{s=0}^{S}\epsilon_s(1+1/\gamma_s)}_{I_1} + \underbrace{12\log 2\sum_{s=0}^{S}\epsilon_s|E^{(s)}(n)|(1+1/\gamma_s)}_{I_2}. \tag{20}$$

Since $\gamma_s = \alpha^2/(F\epsilon_s)$ and $\alpha \ge 1$,

$$I_1 = 12M\sum_{s=0}^{S}\epsilon_s + 12MF\alpha^{-2}\sum_{s=0}^{S}\epsilon_s^2 \le 24M + 16MF.$$

Here $\sum_s \epsilon_s \le 2$ and $\sum_s \epsilon_s^2 \le 4/3$. For I_2, (18) and Lemma 4 give

$$I_2 = 12\log 2 \sum_{s=0}^{S} \epsilon_s |E^{(s)}(n)| + 12F\alpha^{-2}\log 2 \sum_{s=0}^{S} \epsilon_s^2 |E^{(s)}(n)|$$
$$\le 12\log 2\, \alpha\sqrt{3(S+1)Mn/2} + 18\log 2\,(S+1)MF.$$

The first term uses the exploration-width sum, and the second uses $\epsilon_s^2|E^{(s)}(n)| \le 3\alpha^2 M/2$ for each level. Substituting these bounds into (20) and then (17), together with (18) and (19), yields

$$R_n \le (2+12\log 2)\alpha\sqrt{3(S+1)Mn/2} + n/(2\sqrt{T})$$
$$+ (8+18\log 2)(S+1)MF + 24M + 16MF$$
$$\le 13\alpha\sqrt{(S+1)Mn} + n/(2\sqrt{T}) + 64(S+1)MF.$$

The last inequality uses $F \ge 1$, $(2+12\log 2)\sqrt{3/2} < 13$, and $8+18\log 2+24+16 < 64$. The argument holds for every $n \le N(M)$ on the same confidence event, which proves the lemma. □

C Proof of Lemma 1

We prove a stronger form of Lemma 1 for every deterministic budget $M \ge 1$. With probability at least $1-\delta$, both bounds in that lemma hold with M in place of M_d at every visited level and every $t \le N(M)$, where $N(M)$ is defined in (15). This also supplies the confidence bound used in Section F.

We bound the reward noise, regularization bias, and Taylor remainder in the tangent estimate. The first two terms follow from scalar concentration and Cauchy–Schwarz. For the Taylor remainder, the following lemma separates the initial-point estimation error into a centered linear term and a residual. We state the lemma first and prove it after the confidence bound.

Lemma 5 (Expansion of the initial-point estimate)**.** *Let $x_1,\dots,x_n$ be fixed with $\|x_i\|_2 \le 1$, let $\|\theta_\star\|_2 \le B$, and let r_i be independent with $r_i \sim \mathrm{Bern}(\mu(\langle x_i,\theta_\star\rangle))$. Define $\varepsilon_i = r_i - \mu(\langle x_i,\theta_\star\rangle)$, $\lambda_p = \kappa^4 L\sqrt{M}$, and $V = \lambda_p I + \sum_i x_i x_i^\top$, where $L, M \ge 1$. Suppose $\sum_i \|x_i\|_{V^{-1}}^2 \le M$. Let $\bar\theta$ and θ° be the unique solutions of*

$$\lambda_p\bar\theta + \sum_i \big(\mu_B(\langle x_i,\bar\theta\rangle) - r_i\big)x_i = 0, \qquad \lambda_p\theta^\circ + \sum_i \big(\mu_B(\langle x_i,\theta^\circ\rangle) - \mu(\langle x_i,\theta_\star\rangle)\big)x_i = 0,$$

and define

$$H^\circ = \lambda_p I + \sum_i \mu_B'(\langle x_i,\theta^\circ\rangle)x_i x_i^\top, \qquad \xi = (H^\circ)^{-1}\sum_i \varepsilon_i x_i, \qquad b = \bar\theta - \theta_\star - \xi.$$

Then $\mathbb{E}[\xi] = 0$. For every fixed x, the scalar $\langle x,\xi\rangle = \sum_i a_i\varepsilon_i$ has coefficients satisfying $\sum_i a_i^2 \le \kappa^2\|x\|_{V^{-1}}^2$. On the event

$$|\langle x_i,\xi\rangle|^2 \le \kappa^2 L\|x_i\|_{V^{-1}}^2 \quad \text{for every } i \in [n],$$

we have $\|b\|_V^2 \le c_r L\sqrt{M}$, where $c_r = 3\kappa^2(1+B\kappa^2)^2$.

C.1 Proof of the Confidence Bound

Proof of Lemma 1. Fix a round t and a visited level s, and omit these indices. Let P, E be the initial-point estimate and estimation samples collected before t. Write $V = \lambda_p I + \sum_{i\in P} x_i x_i^\top$, $X_E = \sum_{i\in E} x_i x_i^\top$, and $V_E = I + X_E$. For each $i \in E$, define

$$z_i = [\langle x_i,\bar\theta\rangle]_B, \qquad \omega_i = \dot\mu(z_i), \qquad \varepsilon_i = r_i - \mu(\langle x_i,\theta_\star\rangle),$$
$$R_i = \mu(\langle x_i,\theta_\star\rangle) - \mu(z_i) - \omega_i(\langle x_i,\theta_\star\rangle - z_i).$$

The tangent estimate and its design matrix are

$$H = I + \sum_{i\in E}\omega_i x_i x_i^\top, \qquad \hat\theta = H^{-1}\sum_{i\in E}\big(r_i - \mu(z_i) + \omega_i z_i\big)x_i.$$

Since $z_i \in [-B, B]$, we have $1/\kappa \le \omega_i \le 1/4$, and hence

$$\kappa^{-1} V_E \preceq H \preceq V_E, \qquad H^{-1} \preceq \kappa V_E^{-1}. \tag{21}$$

Subtracting $H\theta_\star$ from the normal equation for $\hat{\theta}$ yields

$$\langle x, \hat{\theta} - \theta_\star \rangle = \underbrace{x^\top H^{-1} \sum_{i \in E} \varepsilon_i x_i}_{I_1(x)} - \underbrace{x^\top H^{-1} \theta_\star}_{I_2(x)} + \underbrace{x^\top H^{-1} \sum_{i \in E} R_i x_i}_{I_3(x)}. \tag{22}$$

We first construct an event on which the linear term ξ and $I_1(x)$ concentrate. We then bound $I_3(x)$ by the sum of fourth powers of the initial-point prediction errors.

Concentration of the linear terms. Use $\theta^\circ, H^\circ, \xi, b$ from Lemma 5 for the samples in P, and let $\mathcal{U} = \{x_i : i \in P \cup E\} \cup \mathcal{X}_t$. Let $\mathcal{E}_1$ be the event that, at every visited (s,t), the following bounds hold for all $u \in \mathcal{U}$ and $x \in \mathcal{X}_t$,

$$\begin{aligned} |\langle u, \xi \rangle|^2 &\le \kappa^2 L \|u\|_{V^{-1}}^2, \\ \xi^\top X_E \xi &\le \frac{3}{4} \kappa^2 \big(\operatorname{tr}(V^{-1} X_E) + L \|V^{-1/2} X_E V^{-1/2}\|_{\mathrm{op}} \big), \\ |I_1(x)| &\le \kappa \sqrt{L/2} \|x\|_{V_E^{-1}}. \end{aligned} \tag{23}$$

Condition on $\mathcal{F}_{s,t}$ from Lemma 15. The samples and their features are then fixed, and their rewards are independent with the original Bernoulli means.

For the first bound in (23), use the coefficient bound in Lemma 5 and apply Hoeffding's inequality with $z = L + \log(2|\mathcal{U}|)$. A union bound over $\mathcal{U}$ gives failure probability at most e^{-L}. Since $|\mathcal{U}| \le T + K$ and $\log(2(T+K)) \le L$, the squared threshold is at most $\kappa^2 L \|u\|_{V^{-1}}^2$.

For the second bound, let $X = [x_i]_{i \in P}$ and $N = X^\top (H^\circ)^{-1} X_E (H^\circ)^{-1} X$. Then $\xi^\top X_E \xi = \varepsilon^\top N \varepsilon$, where $\varepsilon = (\varepsilon_i)_{i \in P}$. Since $H^\circ \succeq \kappa^{-1} V$,

$$(H^\circ)^{-1} V (H^\circ)^{-1} = V^{-1/2} (V^{-1/2} H^\circ V^{-1/2})^{-2} V^{-1/2} \preceq \kappa^2 V^{-1}.$$

Together with $X X^\top \preceq V$, this gives

$$\begin{aligned} \operatorname{tr} N &\le \operatorname{tr}\big(X_E (H^\circ)^{-1} V (H^\circ)^{-1} \big) \le \kappa^2 \operatorname{tr}(V^{-1} X_E), \\ \|N\|_{\mathrm{op}} &\le \|X_E^{1/2} (H^\circ)^{-1} V (H^\circ)^{-1} X_E^{1/2}\|_{\mathrm{op}} \le \kappa^2 \|V^{-1/2} X_E V^{-1/2}\|_{\mathrm{op}}. \end{aligned}$$

Applying Lemma 17 with $z = L$ proves the second bound in (23), with failure probability at most e^{-L}. For the third bound, condition additionally on the rewards in P. The matrix H is then fixed, and the rewards in E remain independent. Moreover,

$$\begin{aligned} \sum_{i \in E} (x^\top H^{-1} x_i)^2 &= x^\top H^{-1} X_E H^{-1} x \\ &\le \kappa x^\top H^{-1} (H - I) H^{-1} x \le \kappa x^\top H^{-1} x \le \kappa^2 \|x\|_{V_E^{-1}}^2. \end{aligned}$$

The first inequality uses $X_E \preceq \kappa(H - I)$, and the last uses (21). Hoeffding's inequality therefore gives failure probability at most $2e^{-L}$ for each of the K arms. Taking expectations and a union bound over at most $(S+1)T \le 2ST$ pairs (s,t) gives

$$\mathbb{P}(\mathcal{E}_1^c) \le 2ST(2 + 2K) e^{-L} \le 8STK e^{-L} = \delta/4 \le \delta.$$

Fourth powers of the initial-point estimation errors. Work on $\mathcal{E}_1$ and fix $t \le N(M)$. The design budget and (105) imply $\sum_{i \in P} \|x_i\|_{V^{-1}}^2 \le \log\det(V/\lambda_p) \le M$. Thus Lemma 5 gives $\|b\|_V^2 \le c_r L \sqrt{M}$. By Lemma 4,

$$\max_{i \in E} \|x_i\|_{V^{-1}}^2 \le \tau_s, \qquad X_E \preceq \gamma_s V, \qquad \operatorname{tr}(V^{-1} X_E) \le \tau_s |E| \le \frac{3}{2} \gamma_s M.$$

Here and below a maximum over an empty estimation set is interpreted as zero. The last inequality uses $|E| \le 3\alpha_1^2 M/(2\epsilon_s^2)$ and $\gamma_s = \alpha_1^2\tau_s/\epsilon_s^2$. Applying (23) gives

$$\max_{i\in E} |\langle x_i, \xi\rangle|^2 \le \kappa^2 L\tau_s,$$
$$\sum_{i\in E}\langle x_i,\xi\rangle^2 \le \frac{3}{4}\kappa^2\gamma_s(3M/2+L) \le 3\kappa^2 L\gamma_s M.$$

For the residual b, Cauchy–Schwarz and $X_E \preceq \gamma_s V$ give

$$\max_{i\in E} |\langle x_i, b\rangle|^2 \le c_r L\tau_s\sqrt{M}, \qquad \sum_{i\in E}\langle x_i,b\rangle^2 = b^\top X_E b \le c_r L\gamma_s\sqrt{M}.$$

Since $\bar\theta - \theta_\star = \xi + b$ and $(a+b)^4 \le 8a^4 + 8b^4$,

$$\begin{aligned}\sum_{i\in E}\langle x_i, \bar\theta-\theta_\star\rangle^4 &\le 8\sum_{i\in E}\langle x_i,\xi\rangle^4 + 8\sum_{i\in E}\langle x_i,b\rangle^4\\ &\le 8\max_{i\in E}\langle x_i,\xi\rangle^2\sum_{i\in E}\langle x_i,\xi\rangle^2 + 8\max_{i\in E}\langle x_i,b\rangle^2\sum_{i\in E}\langle x_i,b\rangle^2\\ &\le 8(3\kappa^4+c_r^2)L^2\tau_s\gamma_s M \le c_m L^2\tau_s\gamma_s M. \end{aligned} \tag{24}$$

For the final inequality, with $P_B = (1+B\kappa^2)^2 \ge 1$,

$$8(3\kappa^4+c_r^2) = \kappa^4(24+72P_B^2) \le 96\kappa^4 P_B^2 \le 120\kappa^4 P_B^2 = c_m.$$

Prediction error. The reward noise $I_1(x)$ is bounded in (23). For the regularization bias, $H \succeq I$ and (21) give

$$|I_2(x)| \le \|x\|_{H^{-1}}\|\theta_\star\|_{H^{-1}} \le B\|x\|_{H^{-1}} \le \sqrt{\kappa}B\|x\|_{V_E^{-1}}.$$

For the Taylor remainder, Lemma 9 with $z_1 = z_2 = z_i$, followed by its clipping bound, gives $|R_i| \le |\langle x_i,\theta_\star\rangle - z_i|^2/20 \le \langle x_i, \bar\theta-\theta_\star\rangle^2/20$. Weighted Cauchy–Schwarz therefore yields

$$\begin{aligned}|I_3(x)| &\le \Big(\sum_{i\in E}\omega_i(x^\top H^{-1}x_i)^2\Big)^{1/2}\Big(\sum_{i\in E}R_i^2/\omega_i\Big)^{1/2}\\ &\le \sqrt{\kappa}\|x\|_{V_E^{-1}}\Big(\frac{\kappa}{400}\sum_{i\in E}\langle x_i,\bar\theta-\theta_\star\rangle^4\Big)^{1/2}\\ &\le \frac{\kappa\sqrt{c_m}}{20}L\sqrt{\tau_s\gamma_s M}\|x\|_{V_E^{-1}}.\end{aligned}$$

The second inequality uses $\omega_i \ge 1/\kappa$ and

$$\sum_{i\in E}\omega_i(x^\top H^{-1}x_i)^2 = x^\top H^{-1}(H-I)H^{-1}x \le x^\top H^{-1}x \le \kappa\|x\|^2_{V_E^{-1}}.$$

The third follows from (24). The parameter choices in (91) satisfy

$$L^2\tau_s\gamma_s M = \alpha_1^2 L^2\tau_s^2 M/\epsilon_s^2 = \rho^2, \qquad \kappa\sqrt{c_m}/20 \le c_R, \qquad c_R\rho \le a_0.$$

Combining these bounds in (22), we obtain

$$|\langle x, \hat\theta - \theta_\star\rangle| \le \big(\kappa\sqrt{L/2} + \sqrt{\kappa}B + a_0\big)\|x\|_{V_E^{-1}} = \alpha_1\|x\|_{V_E^{-1}}.$$

The bound holds for all the stated rounds, levels, and arms on $\mathcal{E}_1$. □

C.2 Proof of the Initial-Point Estimate Expansion

Proof of Lemma 5. We expand the score at θ°, whose definition uses the mean rewards. The Jacobian at θ° is deterministic, so the first term of the expansion is linear in the independent reward noise. We then bound the displacement $\theta^\circ - \theta_\star$ due to regularization and the nonlinear remainder.

Define

$$\Psi(\theta) = \lambda_p\theta + \sum_i\big(\mu_B(\langle x_i,\theta\rangle) - \mu(\langle x_i,\theta_\star\rangle)\big)x_i, \qquad Z = \sum_i \varepsilon_i x_i.$$

By definition, $\mu_B = \mu$ on $[-B, B]$ and $1/\kappa \le \mu'_B \le 1/4$. Since clipping is 1-Lipschitz, (95) gives $|\mu'_B(a) - \mu'_B(b)| \le |a-b|/(6\sqrt{3})$. Therefore,

$$\begin{aligned} |\mu_B(a) - \mu_B(b) - \mu'_B(b)(a-b)| &\le \int_{\min\{a,b\}}^{\max\{a,b\}} |\mu'_B(u) - \mu'_B(b)|\, du \\ &\le (a-b)^2/(12\sqrt{3}). \end{aligned} \tag{25}$$

For every θ, ϑ, the lower bound on μ'_B gives

$$\begin{aligned} \langle \Psi(\theta) - \Psi(\vartheta), \theta - \vartheta \rangle &\ge \lambda_p \|\theta - \vartheta\|_2^2 + \kappa^{-1} \sum_i \langle x_i, \theta - \vartheta \rangle^2 \\ &\ge \kappa^{-1} \|\theta - \vartheta\|_V^2. \end{aligned}$$

Thus Lemma 12 gives unique solutions $\Psi(\theta^\circ) = 0$ and $\Psi(\bar\theta) = Z$, as well as a continuous path $\theta(a) = \Psi^{-1}(aZ)$ for $a \in [0,1]$. The endpoints of this path are $\theta(0) = \theta^\circ$ and $\theta(1) = \bar\theta$.

The derivative bounds $1/\kappa \le \mu'_B \le 1/4$ imply

$$\kappa^{-1} V \preceq H^\circ \preceq V, \qquad (H^\circ)^{-1} V (H^\circ)^{-1} \preceq \kappa (H^\circ)^{-1} \preceq \kappa^2 V^{-1}. \tag{26}$$

The matrix H° is fixed because θ° uses the mean rewards. Therefore $\mathbb{E}[\xi] = 0$. With $a_i = x^\top (H^\circ)^{-1} x_i$,

$$\sum_i a_i^2 = x^\top (H^\circ)^{-1} (V - \lambda_p I)(H^\circ)^{-1} x \le \kappa^2 \|x\|_{V^{-1}}^2.$$

This proves the assertion about the linear term. If $n = 0$, then $\bar\theta = \theta^\circ = \xi = 0$ and $\|b\|_V^2 = \lambda_p \|\theta_\star\|_2^2 \le c_r L \sqrt{M}$. We henceforth take $n \ge 1$.

To obtain the error decomposition, define the score remainder

$$\mathcal{R}(v) = \sum_i \big(\mu_B(\langle x_i, \theta^\circ + v \rangle) - \mu_B(\langle x_i, \theta^\circ \rangle) - \mu'_B(\langle x_i, \theta^\circ \rangle) \langle x_i, v \rangle \big) x_i.$$

The identity $\Psi(\theta^\circ + v) = H^\circ v + \mathcal{R}(v)$, evaluated at $v = \bar\theta - \theta^\circ$, gives

$$\bar\theta - \theta_\star = \underbrace{(H^\circ)^{-1} Z}_{\xi} + \underbrace{\theta^\circ - \theta_\star - (H^\circ)^{-1} \mathcal{R}(\bar\theta - \theta^\circ)}_{b}. \tag{27}$$

The same identity along $\theta(a)$ gives

$$\theta(a) - \theta^\circ - a\xi = -(H^\circ)^{-1} \mathcal{R}(\theta(a) - \theta^\circ). \tag{28}$$

This expansion separates the linear noise from the regularization and nonlinear errors, as in Fisher expansions for penalized estimators (Spokoiny, 2017, Section 2.4). The remaining calculations bound b for the present initial-point estimate.

Displacement due to regularization. Since $\|\theta_\star\|_2 \le B$ and $\|x_i\|_2 \le 1$, $\mu_B(\langle x_i, \theta_\star \rangle) = \mu(\langle x_i, \theta_\star \rangle)$ and $\Psi(\theta_\star) = \lambda_p \theta_\star$. Strong monotonicity therefore yields

$$\begin{aligned} \kappa^{-1} \|\theta^\circ - \theta_\star\|_V^2 &\le \langle \Psi(\theta^\circ) - \Psi(\theta_\star), \theta^\circ - \theta_\star \rangle \\ &= -\lambda_p \langle \theta_\star, \theta^\circ - \theta_\star \rangle \\ &\le \lambda_p \|\theta_\star\|_{V^{-1}} \|\theta^\circ - \theta_\star\|_V \le B\sqrt{\lambda_p} \|\theta^\circ - \theta_\star\|_V. \end{aligned}$$

Consequently,

$$\|\theta^\circ - \theta_\star\|_V \le \kappa B \sqrt{\lambda_p}. \tag{29}$$

Nonlinear remainder. Let $X = [x_1, \dots, x_n]$. The inequality $XX^\top \preceq V$ implies $X^\top V^{-1} X \preceq I$. Applying (25) to the coefficients of $\mathcal{R}(v)$ gives

$$\|\mathcal{R}(v)\|_{V^{-1}}^2 \le \frac{1}{432} \sum_i \langle x_i, v \rangle^4.$$

Set $y(a) = \|\theta(a) - \theta^\circ - a\xi\|_V$. By (26) and (28),

$$y(a) \le \frac{\kappa}{20}\Big(\sum_i \langle x_i, \theta(a) - \theta^\circ\rangle^4\Big)^{1/2}. \tag{30}$$

On the event in the statement,

$$\begin{aligned}\sum_i \langle x_i, \xi\rangle^4 &\le \kappa^4 L^2 \sum_i \|x_i\|_{V^{-1}}^4 \\ &\le \frac{\kappa^4 L^2}{\lambda_p}\sum_i \|x_i\|_{V^{-1}}^2 \le \frac{\kappa^4 L^2 M}{\lambda_p}.\end{aligned}$$

The second inequality uses $\|x_i\|_{V^{-1}}^2 \le 1/\lambda_p$. For any v, Cauchy–Schwarz also gives

$$\begin{aligned}\sum_i \langle x_i, v\rangle^4 &\le \max_i \langle x_i, v\rangle^2 \sum_i \langle x_i, v\rangle^2 \\ &\le \lambda_p^{-1}\|v\|_V^2 v^\top\Big(\sum_i x_i x_i^\top\Big)v \le \lambda_p^{-1}\|v\|_V^4.\end{aligned}$$

Writing $\theta(a) - \theta^\circ = a\xi + (\theta(a) - \theta^\circ - a\xi)$ and using $(u+v)^4 \le 8u^4 + 8v^4$, we obtain

$$\sum_i \langle x_i, \theta(a) - \theta^\circ\rangle^4 \le 8\kappa^4 L^2 M/\lambda_p + 8y(a)^4/\lambda_p.$$

Substitution into (30), with $\sqrt{8}/20 \le 1/7$ and $\lambda_p = \kappa^4 L\sqrt{M}$, gives

$$y(a) \le \frac{\kappa^3 L}{7}\sqrt{M/\lambda_p} + \frac{\kappa}{7\sqrt{\lambda_p}}y(a)^2 = \frac{\kappa\sqrt{L}\,M^{1/4}}{7} + \frac{\kappa}{7\sqrt{\lambda_p}}y(a)^2. \tag{31}$$

We use continuity to select the smaller solution of this quadratic inequality. Since $y(0) = 0$, if y reached $2\kappa\sqrt{L}\,M^{1/4}/7$, then at the first such point (31) would imply

$$\frac{2\kappa\sqrt{L}\,M^{1/4}}{7} \le \frac{\kappa\sqrt{L}\,M^{1/4}}{7}(1 + 4/49) < \frac{2\kappa\sqrt{L}\,M^{1/4}}{7}.$$

This is a contradiction. Hence $\|\bar\theta - \theta^\circ - \xi\|_V^2 = y(1)^2 \le 4\kappa^2 L\sqrt{M}/49$. Finally, (27) and (29) give

$$\begin{aligned}\|b\|_V^2 &\le 2\|\theta^\circ - \theta_\star\|_V^2 + 2\|\bar\theta - \theta^\circ - \xi\|_V^2 \\ &\le 2\kappa^2 B^2\lambda_p + 8\kappa^2 L\sqrt{M}/49 \\ &= \big(2\kappa^6 B^2 + 8\kappa^2/49\big)L\sqrt{M} \le 3\kappa^2(1 + B\kappa^2)^2 L\sqrt{M}.\end{aligned}$$

This proves the residual bound. □

D Proof of Lemma 7

We prove that the adjusted initial-point estimate has an expansion $\bar\theta - \theta_\star = \zeta + b$, where ζ is centered and $\|b\|_V^2 \le C_b L M^{1/3}$ with high probability. The linear and centered quadratic terms of the initial-point estimation error form ζ. We first establish existence of the adjusted initial-point estimate and then bound the residual b.

Let $x_1, \dots, x_n$ be fixed vectors with $\|x_i\|_2 \le 1$, $\|\theta_\star\|_2 \le B$, and independent rewards $r_i \sim \mathrm{Bern}(\mu(\langle x_i, \theta_\star\rangle))$. Set $\varepsilon_i = r_i - \mu(\langle x_i, \theta_\star\rangle)$ and $Z = \sum_i \varepsilon_i x_i$. Let $M \ge 1$, $L \ge \max\{1, \log(2(T+K))\}$, $\lambda_p = C_p L M^{1/3}$, and $V = \lambda_p I + \sum_i x_i x_i^\top$, and assume

$$\sum_i x_i^\top V^{-1} x_i \le M. \tag{32}$$

The constants are defined in Section G. For the function g in Lemma 11, define

$$H(\theta) = \lambda_p I + \sum_i g'(\langle x_i, \theta\rangle) x_i x_i^\top,$$
$$v_i(\theta) = x_i^\top H(\theta)^{-1}(H(\theta) - \lambda_p I)H(\theta)^{-1} x_i,$$
$$A(\theta) = \frac{1}{2}\sum_i g''(\langle x_i, \theta\rangle) v_i(\theta) x_i,$$
$$F_u(\theta) = \sum_i (g(\langle x_i, \theta\rangle) - r_i)x_i + \lambda_p \theta - u^2 A(\theta), \qquad u \in [0,1].$$

The adjusted initial-point estimate is the solution of $F_1(\bar\theta) = 0$.

Lemma 6 (Existence and uniqueness of the adjusted initial-point estimate)**.** *For every $\theta, a, b \in \mathbb{R}^d$,*

$$|a^\top DA(\theta)[b]| \le \frac{c_A}{\lambda_p}\|a\|_{H(\theta)}\|b\|_{H(\theta)} \le \frac{c_A}{\lambda_p}\|a\|_V\|b\|_V, \qquad c_A = 2\bar\kappa + 8\bar\kappa^2/3. \tag{33}$$

For every $u \in [0,1]$ and $\theta, \vartheta \in \mathbb{R}^d$,

$$\langle F_u(\theta) - F_u(\vartheta), \theta - \vartheta\rangle \ge \frac{1}{4\bar\kappa}\|\theta - \vartheta\|_V^2. \tag{34}$$

Thus each F_u is a bijection. In particular, $\bar\theta$ exists uniquely.

Proof. We bound the two terms obtained by differentiating A. The derivative bounds on g imply

$$(2\bar\kappa)^{-1} V \preceq H(\theta) \preceq V. \tag{35}$$

Fix θ and write $H = H(\theta)$, $w_i = H^{-1/2}x_i$, $W = [w_1, \dots, w_n]$, $G = W^\top W$, and $\Lambda = \lambda_p H^{-1}$. For $\alpha = H^{1/2}a$ and $\beta = H^{1/2}b$, we have

$$\|W\|_{\mathrm{op}}^2 \le 2\bar\kappa, \qquad \|w_i\|_2^2 \le \lambda_p^{-1}, \qquad 0 \preceq \Lambda \preceq I, \qquad v_i(\theta) = w_i^\top(I - \Lambda)w_i \le \lambda_p^{-1}. \tag{36}$$

Indeed, $\sum_i x_i x_i^\top \preceq V \preceq 2\bar\kappa H$ and $H \succeq \lambda_p I$. Differentiating A gives

$$a^\top DA(\theta)[b] = \underbrace{\frac{1}{2}\sum_i g'''(\langle x_i, \theta\rangle)(x_i^\top b)v_i(\theta)(x_i^\top a)}_{I_1} + \underbrace{\frac{1}{2}\sum_i g''(\langle x_i, \theta\rangle)(x_i^\top a)Dv_i(\theta)[b]}_{I_2}. \tag{37}$$

For I_1, $|g'''| \le 2$, (36), and Cauchy-Schwarz yield

$$|I_1| \le \lambda_p^{-1}\sum_i |w_i^\top\alpha||w_i^\top\beta| \le \lambda_p^{-1}\|W\|_{\mathrm{op}}^2\|\alpha\|_2\|\beta\|_2 \le \frac{2\bar\kappa}{\lambda_p}\|\alpha\|_2\|\beta\|_2. \tag{38}$$

To bound I_2, put $D_\beta = H^{-1/2}DH(\theta)[b]H^{-1/2} = \sum_j g''(\langle x_j, \theta\rangle)(w_j^\top\beta)w_j w_j^\top$. Differentiating $H^{-1}(H - \lambda_p I)H^{-1} = H^{-1} - \lambda_p H^{-2}$ gives

$$D(H^{-1} - \lambda_p H^{-2})[b] = -H^{-1}DH[b]H^{-1} + \lambda_p H^{-1}DH[b]H^{-2} + \lambda_p H^{-2}DH[b]H^{-1},$$
$$Dv_i(\theta)[b] = w_i^\top(-D_\beta + D_\beta\Lambda + \Lambda D_\beta)w_i.$$

With $D_g = \mathrm{diag}(g''(\langle x_i, \theta\rangle))_i$ and $G_\Lambda = W^\top\Lambda W$, substitution into I_2 gives

$$I_2 = -\frac{1}{2}\alpha^\top W D_g(G\circ G)D_g W^\top\beta + \frac{1}{2}\alpha^\top W D_g(G\circ G_\Lambda)D_g W^\top\beta$$
$$+ \frac{1}{2}\alpha^\top W D_g(G_\Lambda\circ G)D_g W^\top\beta. \tag{39}$$

Here $\circ$ denotes the entrywise matrix product. By (100) and (36),

$$\|G\circ G\|_{\mathrm{op}} \le \frac{2\bar\kappa}{\lambda_p}, \qquad \|G\circ G_\Lambda\|_{\mathrm{op}} = \|G_\Lambda\circ G\|_{\mathrm{op}} \le \frac{2\bar\kappa}{\lambda_p}.$$

Since $\|D_g\|_{\mathrm{op}} \le 2/3$, each term in (39) has absolute value at most

$$\tfrac{1}{2}(2\bar\kappa)(2/3)^2(2\bar\kappa/\lambda_p)\|\alpha\|_2\|\beta\|_2 = 8\bar\kappa^2\|\alpha\|_2\|\beta\|_2/(9\lambda_p).$$

Together with (37) and (38), this proves (33).

For $h = \theta - \vartheta$, integration along $\vartheta + vh$, $v \in [0,1]$, now gives

$$\begin{aligned}\langle F_u(\theta) - F_u(\vartheta), h\rangle &= \int_0^1 h^\top H(\vartheta + vh)h\,dv - u^2\int_0^1 h^\top DA(\vartheta+vh)[h]\,dv \\ &\ge \left(\frac{1}{2\bar\kappa} - \frac{c_A}{\lambda_p}\right)\|h\|_V^2 \ge \frac{1}{4\bar\kappa}\|h\|_V^2.\end{aligned}$$

The first inequality uses (33) and (35), and the second uses $\lambda_p \ge C_p \ge 4\bar\kappa c_A$. Lemma 12 proves bijectivity. □

Define θ° as the solution of

$$\Phi(\theta^\circ) = 0, \qquad \Phi(\theta) = \sum_i \big(g(\langle x_i,\theta\rangle) - \mu(\langle x_i,\theta_\star\rangle)\big)x_i + \lambda_p\theta. \tag{40}$$

By (35) and Lemma 12, this solution exists uniquely. It is deterministic since $x_1,\dots,x_n$ are fixed. Let $H_\circ = H(\theta^\circ)$ and define

$$\xi = H_\circ^{-1}Z, \qquad T_\circ[v,w] = \sum_i g''(\langle x_i,\theta^\circ\rangle)(x_i^\top v)(x_i^\top w)x_i.$$

The centered quadratic term and residual are

$$Q = -\frac{1}{2}H_\circ^{-1}\big(T_\circ[\xi,\xi] - \mathbb{E}[T_\circ[\xi,\xi]]\big), \qquad \zeta = \xi + Q, \qquad b = \bar\theta - \theta_\star - \zeta. \tag{41}$$

Lemma 7 (Expansion of the adjusted initial-point estimate)**.** *Let $\mathcal{U} \subset \mathbb{R}^d$ be fixed and finite, with $|\mathcal{U}| + n \le T + K$. Then $\mathbb{E}[\zeta] = 0$, ζ is a polynomial of degree at most two in $(\varepsilon_i)_i$, and $\mathrm{Cov}(\zeta) \preceq C_vV^{-1}$. With probability at least $1 - 5e^{-L}$,*

$$\|b\|_V^2 \le C_bLM^{1/3}, \qquad \|\zeta\|_V^2 \le C_ELM, \qquad |v^\top\zeta|^2 \le C_xL\,v^\top V^{-1}v \quad (v\in\mathcal{U}).$$

Proof. We first write the equation for the residual b. For any $v \in \mathbb{R}^d$, Taylor expansion of (40) at θ° gives

$$\Phi(\theta^\circ + v) = H_\circ v + \frac{1}{2}T_\circ[v,v] + R_3(v), \qquad R_3(v) = \sum_i \rho_i(v)(x_i^\top v)^3x_i, \tag{42}$$

where $\rho_i(v) = \frac{1}{2}\int_0^1(1-a)^2g'''(\langle x_i,\theta^\circ\rangle + ax_i^\top v)\,da$ and $|\rho_i(v)| \le 1/3$. Let $A_\star = \mathbb{E}[T_\circ[\xi,\xi]]/2$. The definition of Q gives $H_\circ\xi = Z$ and $H_\circ Q = -T_\circ[\xi,\xi]/2 + A_\star$. Thus subtracting $\xi + Q$ from the adjusted initial-point estimate cancels the linear term and the centered quadratic term in (42).

To bound the resulting error, for $u \in [0,1]$ define δ_u by

$$\Phi(\theta^\circ + \delta_u) = uZ + u^2A(\theta^\circ + \delta_u). \tag{43}$$

Lemma 6 implies that this equation has a unique solution, with $\delta_0 = 0$ and $\delta_1 = \bar\theta - \theta^\circ$. Write $e_u = \delta_u - u\xi - u^2Q$ and $z_u = \|e_u\|_{H_\circ}$. Substituting (42) into (43) yields

$$H_\circ e_u = u^2\big(A(\theta^\circ+\delta_u) - A_\star\big) - \frac{1}{2}\big(T_\circ[\delta_u,\delta_u] - u^2T_\circ[\xi,\xi]\big) - R_3(\delta_u), \tag{44}$$

$$b = (\theta^\circ - \theta_\star) + e_1. \tag{45}$$

We bound ξ and Q, then use (44) to show $z_1 \le 2C_1(L^{3/2}\sqrt{M}/\lambda_p + 1/\sqrt{\lambda_p})$. Together with a bound on $\|\theta^\circ - \theta_\star\|_V$, this will give the claimed bound on b.

Bounds on the centered terms. Let $X = [x_1, \dots, x_n]$, $w_i = H_\circ^{-1/2} x_i$, $W = [w_1, \dots, w_n]$, and $h_i = x_i^\top H_\circ^{-1} x_i$. (32) and (35) imply

$$h_i \le \lambda_p^{-1}, \qquad \sum_i h_i \le 2\bar\kappa M, \qquad X^\top H_\circ^{-1} X \preceq 2\bar\kappa I, \qquad \sum_i h_i^j \le \frac{2\bar\kappa M}{\lambda_p^{j-1}} \quad (j = 2, 3). \tag{46}$$

Since $\Sigma_\star = \mathrm{Cov}(Z) = \sum_i \dot\mu(\langle x_i, \theta_\star\rangle) x_i x_i^\top \preceq XX^\top/4$,

$$\mathrm{Cov}(\xi) = H_\circ^{-1}\Sigma_\star H_\circ^{-1} \preceq \tfrac14 H_\circ^{-1} XX^\top H_\circ^{-1} \preceq \tfrac{\bar\kappa}{2} H_\circ^{-1}. \tag{47}$$

For a fixed $x \in \mathbb{R}^d$, let $w = H_\circ^{-1/2} x$ and $c_i = g''(\langle x_i, \theta^\circ\rangle)$. Expanding (41) gives

$$x^\top Q = -\frac12 \sum_i c_i (w^\top w_i)\big((w_i^\top W\varepsilon)^2 - \mathbb{E}[(w_i^\top W\varepsilon)^2]\big) = \varepsilon^\top N_x \varepsilon - \mathbb{E}[\varepsilon^\top N_x \varepsilon],$$

$$N_x = -\frac12 W^\top \left(\sum_i c_i (w^\top w_i) w_i w_i^\top\right) W. \tag{48}$$

Let $\sigma_i^2 = \mathbb{E}[\varepsilon_i^2]$. By independence and centering,

$$\begin{aligned}\mathbb{E}[(x^\top Q)^2] &= \sum_i (N_x)_{ii}^2 \,\mathrm{Var}(\varepsilon_i^2) + 4\sum_{i<j} (N_x)_{ij}^2 \sigma_i^2 \sigma_j^2 \\ &\le \sum_i (N_x)_{ii}^2 + 4\sum_{i<j} (N_x)_{ij}^2 \le 2\|N_x\|_{\mathrm F}^2.\end{aligned}$$

The equality follows from (48), and the first inequality uses $|\varepsilon_i| \le 1$. Using (102) with $\tau = 1/\lambda_p$, $\Gamma = 2\bar\kappa$, and $|c_i| \le 2/3$, we therefore obtain

$$\begin{aligned}\|N_x\|_{\mathrm F} &\le \tfrac12 \|W\|_{\mathrm{op}}^2 \left\|\sum_i c_i (w^\top w_i) w_i w_i^\top\right\|_{\mathrm F} \le \frac{4\bar\kappa^2}{3}\sqrt{\frac{x^\top H_\circ^{-1} x}{\lambda_p}}, \\ \mathbb{E}[(x^\top Q)^2] &\le 2\|N_x\|_{\mathrm F}^2 \le \frac{32\bar\kappa^4}{9\lambda_p}\, x^\top H_\circ^{-1} x.\end{aligned} \tag{49}$$

Both ξ and Q are centered. Applying $(a+b)^2 \le 2a^2 + 2b^2$ and (47) and (49) gives

$$\mathbb{E}[(x^\top \zeta)^2] \le 2\left(\frac{\bar\kappa}{2} + \frac{32\bar\kappa^4}{9C_p}\right) 2\bar\kappa\, x^\top V^{-1} x = C_v x^\top V^{-1} x.$$

This proves the covariance bound. The degree of ζ is at most two by (41) and (48).

We next bound $x^\top\xi$ and $x^\top Q$ for $x \in \mathcal{U} \cup \{x_1, \dots, x_n\}$. The coefficients of $x^\top\xi$ satisfy

$$x^\top \xi = \sum_i (x^\top H_\circ^{-1} x_i)\varepsilon_i, \qquad \sum_i (x^\top H_\circ^{-1} x_i)^2 \le 2\bar\kappa\, x^\top H_\circ^{-1} x.$$

Apply Lemma 16 with $z = L + \log(2(T+K)) \le 2L$. For $x^\top Q$, apply Lemma 18 with degree two, $z = L + \log(T+K) \le 2L$, and (49). The Bernoulli probabilities lie in $[p_0, 1-p_0]$ because $|\langle x_i, \theta_\star\rangle| \le B$. A union bound over at most $T+K$ vectors gives, with probability at least $1 - 2e^{-L}$,

$$|x^\top\xi|^2 \le c_\xi L\, x^\top H_\circ^{-1} x, \qquad |x^\top Q|^2 \le \frac{c_Q^2 L^2}{\lambda_p}\, x^\top H_\circ^{-1} x \quad \big(x \in \mathcal{U} \cup \{x_i\}_i\big). \tag{50}$$

For the second bound, the coefficient given by Lemma 18 is $64\sqrt{2}e\,\bar\kappa^2/(3p_0) \le c_Q$. Since $H_\circ^{-1} \preceq 2\bar\kappa V^{-1}$ and $\lambda_p \ge C_p L$, (50) implies

$$|x^\top\zeta| \le \sqrt{2\bar\kappa L}\left(\sqrt{c_\xi} + \frac{c_Q}{\sqrt{C_p}}\right)\sqrt{x^\top V^{-1} x} = \sqrt{C_x L\, x^\top V^{-1} x}. \tag{51}$$

For $\|\zeta\|_V$, first note that

$$\mathbb{E}[\|\xi\|_V^2] = \mathrm{tr}(V H_\circ^{-1}\Sigma_\star H_\circ^{-1}) \le 2\bar\kappa\ \mathrm{tr}(\Sigma_\star H_\circ^{-1}) \le \frac{\bar\kappa}{2}\sum_i h_i \le \bar\kappa^2 M. \tag{52}$$

The first inequality uses (35), and the last uses (46). Writing $y_i = c_i((x_i^\top \xi)^2 - \mathbb{E}[(x_i^\top \xi)^2])$ gives $Q = -H_\circ^{-1} X y/2$ and

$$\|Q\|_V^2 = \tfrac{1}{4} y^\top X^\top H_\circ^{-1} V H_\circ^{-1} X y \le \tfrac{\bar\kappa}{2} y^\top X^\top H_\circ^{-1} X y \le \bar\kappa^2 \|y\|_2^2. \tag{53}$$

For $Y = \sum_j a_j \varepsilon_j$, independence and centering give

$$\begin{aligned}\mathbb{E}[Y^4] &= \sum_j a_j^4 \mathbb{E}[\varepsilon_j^4] + 6 \sum_{j<k} a_j^2 a_k^2 \mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_k^2] \\ &\le \tfrac{1}{4} \sum_j a_j^4 + \tfrac{3}{8} \sum_{j<k} a_j^2 a_k^2 \le \tfrac{1}{4} \left(\sum_j a_j^2 \right)^2 .\end{aligned}$$

We used $\mathbb{E}[\varepsilon_j^4] \le \mathbb{E}[\varepsilon_j^2] \le 1/4$. For $Y = x_i^\top \xi$, the squared coefficients sum to $\sum_j (w_i^\top w_j)^2 \le 2\bar\kappa h_i$. Consequently, (53) yields

$$\mathbb{E}[\|Q\|_V^2] \le \bar\kappa^2 \sum_i \mathbb{E}[y_i^2] \le \frac{4\bar\kappa^4}{9} \sum_i h_i^2 \le \frac{8\bar\kappa^5 M}{9\lambda_p}. \tag{54}$$

Apply the Hilbert-space version of Lemma 18 with $z = L$ to (52) and (54). With probability at least $1 - 2e^{-L}$,

$$\|\xi\|_V^2 \le \frac{8e^2\bar\kappa^2}{p_0} LM, \qquad \|Q\|_V^2 \le \frac{512 e^2 \bar\kappa^5}{9 p_0^2 C_p} LM, \qquad \|\zeta\|_V^2 \le C_E LM. \tag{55}$$

The last inequality uses $\|\xi + Q\|_V^2 \le 2\|\xi\|_V^2 + 2\|Q\|_V^2$.

Regularization and the mean quadratic term. We bound $\theta^\circ - \theta_\star$ in (45) and $A(\theta^\circ) - A_\star$ in (44). Since $g = \mu$ on $[-B, B]$, we have $\Phi(\theta_\star) = \lambda_p \theta_\star$ and $\Phi(\theta^\circ) = 0$. By (35) and Cauchy-Schwarz,

$$\begin{aligned}\frac{1}{2\bar\kappa}\|\theta^\circ - \theta_\star\|_V^2 &\le \langle \Phi(\theta_\star) - \Phi(\theta^\circ), \theta_\star - \theta^\circ \rangle = \lambda_p \langle \theta_\star, \theta_\star - \theta^\circ \rangle \\ &\le \lambda_p \|\theta_\star\|_{V^{-1}} \|\theta_\star - \theta^\circ\|_V \le B\sqrt{\lambda_p}\|\theta_\star - \theta^\circ\|_V .\end{aligned}$$

Therefore

$$\|\theta^\circ - \theta_\star\|_V \le 2\bar\kappa B \sqrt{\lambda_p}. \tag{56}$$

The correction $A(\theta^\circ)$ uses $H_\circ - \lambda_p I$, whereas $A_\star$ uses the reward covariance $\Sigma_\star$. In particular,

$$\begin{aligned}A(\theta^\circ) - A_\star &= \frac{1}{2} \sum_i c_i x_i^\top H_\circ^{-1} (H_\circ - \lambda_p I - \Sigma_\star) H_\circ^{-1} x_i \, x_i, \\ H_\circ - \lambda_p I - \Sigma_\star &= \sum_j \left(\int_0^1 g''(\langle x_j, \theta_\star + v(\theta^\circ - \theta_\star)\rangle)\, dv \right) (x_j^\top (\theta^\circ - \theta_\star)) x_j x_j^\top .\end{aligned} \tag{57}$$

The second equality follows from $g'(\langle x_j, \theta_\star \rangle) = \dot\mu(\langle x_j, \theta_\star \rangle)$ and the fundamental theorem of calculus. Let $D_\circ$ be the diagonal matrix of the integrals in (57), and $D_g = \mathrm{diag}(c_i)_i$. For $\|a\|_{H_\circ} = 1$, set $\alpha = H_\circ^{1/2} a$, $\beta = H_\circ^{1/2}(\theta^\circ - \theta_\star)$, and $G = W^\top W$. By (46), (56), and (100),

$$\begin{aligned}a^\top (A(\theta^\circ) - A_\star) &= \tfrac{1}{2} \alpha^\top W D_g (G \circ G) D_\circ W^\top \beta, \\ |a^\top (A(\theta^\circ) - A_\star)| &\le \tfrac{1}{2}(2\bar\kappa)(2/3)^2 (2\bar\kappa/\lambda_p)\|\beta\|_2 \le \frac{16\bar\kappa^3 B}{9\sqrt{\lambda_p}}.\end{aligned}$$

Taking the supremum over $\|a\|_{H_\circ} = 1$ gives

$$\|A(\theta^\circ) - A_\star\|_{H_\circ^{-1}} \le \frac{16\bar\kappa^3 B}{9\sqrt{\lambda_p}}. \tag{58}$$

Bounds on δ_u. We first bound δ_u uniformly over $u \in [0,1]$. Applying (34) to $\theta \mapsto \Phi(\theta) - u^2 A(\theta)$, whose derivative is that of F_u, gives

$$\|\delta_u\|_V \le 4\bar{\kappa}\big(\|Z\|_{V^{-1}} + \|A(\theta^\circ)\|_{V^{-1}}\big).$$

Lemma 17 and (32) give, with probability at least $1 - e^{-L}$,

$$\|Z\|_{V^{-1}}^2 = \varepsilon^\top X^\top V^{-1} X \varepsilon \le \tfrac{3}{4}(M + L) \le \tfrac{3}{2} LM.$$

Here $\mathrm{tr}(X^\top V^{-1} X) \le M$ and $\|X^\top V^{-1} X\|_{\mathrm{op}} \le 1$. Writing $A(\theta^\circ) = Xa$ gives $|a_i| \le h_i/3$, so

$$\|A(\theta^\circ)\|_{V^{-1}}^2 \le \|a\|_2^2 \le \tfrac{1}{9} \sum_i h_i^2 \le \frac{2\bar{\kappa} M}{9\lambda_p}.$$

Hence, on the same event,

$$\sup_{u \in [0,1]} \|\delta_u\|_V \le 4\bar{\kappa} \left(\sqrt{3/2} + \frac{\sqrt{2\bar{\kappa}}}{3} \right) \sqrt{LM} \le \sqrt{C_\delta LM}. \tag{59}$$

By (33), $\|A(\theta) - A(\vartheta)\|_{V^{-1}} \le (c_A/\lambda_p)\|\theta - \vartheta\|_V$. Subtracting (43) at u and v and applying (34) thus gives

$$\begin{aligned} \|\delta_u - \delta_v\|_V &\le 4\bar{\kappa}|u - v|\big(\|Z\|_{V^{-1}} + 2\|A(\theta^\circ + \delta_v)\|_{V^{-1}}\big) \\ &\le 4\bar{\kappa}|u - v| \left(\|Z\|_{V^{-1}} + 2\|A(\theta^\circ)\|_{V^{-1}} + \frac{2c_A}{\lambda_p}\sqrt{C_\delta LM} \right). \end{aligned}$$

It follows that δ_u and z_u are continuous in u, with $z_0 = 0$.

Combining (33), (58), and (59), we bound the adjustment term of (44) by

$$\begin{aligned} \|A(\theta^\circ + \delta_u) - A_\star\|_{H_\circ^{-1}} &\le \sqrt{2\bar{\kappa}} \frac{c_A}{\lambda_p} \|\delta_u\|_V + \|A(\theta^\circ) - A_\star\|_{H_\circ^{-1}} \\ &\le \sqrt{2\bar{\kappa}}\, c_A \sqrt{C_\delta} \frac{\sqrt{LM}}{\lambda_p} + \frac{16\bar{\kappa}^3 B}{9\sqrt{\lambda_p}}. \end{aligned} \tag{60}$$

Quadratic and cubic remainders. We next bound the other terms in (44) on (50). Since $z_u = \|e_u\|_{H_\circ}$, (46) gives

$$\sum_i (x_i^\top e_u)^2 \le 2\bar{\kappa} z_u^2, \qquad \max_i (x_i^\top e_u)^2 \le z_u^2/\lambda_p. \tag{61}$$

Expanding $\delta_u = u\xi + u^2 Q + e_u$ gives

$$\begin{aligned} (x_i^\top \delta_u)^2 - u^2 (x_i^\top \xi)^2 = 2u^3 (x_i^\top \xi)(x_i^\top Q) + 2u (x_i^\top \xi)(x_i^\top e_u) + u^4 (x_i^\top Q)^2 \\ + 2u^2 (x_i^\top Q)(x_i^\top e_u) + (x_i^\top e_u)^2. \end{aligned} \tag{62}$$

For any coefficients $(a_i)_i$, (46) and $|c_i| \le 2/3$ imply

$$\left\| \sum_i c_i a_i x_i \right\|_{H_\circ^{-1}} \le \sqrt{8\bar{\kappa}/9} \left(\sum_i a_i^2 \right)^{1/2}. \tag{63}$$

The coefficient products in (62) satisfy

$$\begin{aligned} \sum_i (x_i^\top \xi)^2 (x_i^\top Q)^2 &\le \frac{c_\xi c_Q^2 L^3}{\lambda_p} \sum_i h_i^2 \le \frac{2\bar{\kappa} c_\xi c_Q^2 L^3 M}{\lambda_p^2}, \\ \sum_i (x_i^\top \xi)^2 (x_i^\top e_u)^2 &\le \frac{c_\xi L}{\lambda_p} \sum_i (x_i^\top e_u)^2 \le \frac{2\bar{\kappa} c_\xi L}{\lambda_p} z_u^2, \\ \sum_i (x_i^\top Q)^4 &\le \frac{c_Q^4 L^4}{\lambda_p^2} \sum_i h_i^2 \le \frac{2\bar{\kappa} c_Q^4 L^4 M}{\lambda_p^3}, \\ \sum_i (x_i^\top Q)^2 (x_i^\top e_u)^2 &\le \frac{c_Q^2 L^2}{\lambda_p^2} \sum_i (x_i^\top e_u)^2 \le \frac{2\bar{\kappa} c_Q^2 L^2}{\lambda_p^2} z_u^2, \\ \sum_i (x_i^\top e_u)^4 &\le \frac{z_u^2}{\lambda_p} \sum_i (x_i^\top e_u)^2 \le \frac{2\bar{\kappa}}{\lambda_p} z_u^4. \end{aligned} \tag{64}$$

These inequalities use (46), (50), and (61). Substitute (64) into (62) and (63), and use $u \le 1$ and $\sqrt{L/\lambda_p} \le C_p^{-1/2}$ to obtain

$$\|T_\circ[\delta_u,\delta_u] - u^2 T_\circ[\xi,\xi]\|_{H_\circ^{-1}} \le \left(\frac{8\bar\kappa}{3} c_Q\sqrt{c_\xi} + \frac{4\bar\kappa c_Q^2}{3\sqrt{C_p}}\right)\frac{L^{3/2}\sqrt{M}}{\lambda_p} + \frac{8\bar\kappa}{3}\left(\sqrt{\frac{c_\xi L}{\lambda_p}} + \frac{c_Q L}{\lambda_p}\right) z_u + \frac{4\bar\kappa}{3\sqrt{\lambda_p}} z_u^2. \tag{65}$$

For R_3 in (42), the bound $|\rho_i(v)| \le 1/3$ and (46) imply $\|R_3(v)\|_{H_\circ^{-1}}^2 \le (2\bar\kappa/9)\sum_i |x_i^\top v|^6$. Using $\delta_u = u\xi + u^2 Q + e_u$, $u \le 1$, and $|a+b+c|^6 \le 3^5(|a|^6+|b|^6+|c|^6)$ gives

$$\|R_3(\delta_u)\|_{H_\circ^{-1}} \le \sqrt{54\bar\kappa}\left[\left(\sum_i |x_i^\top \xi|^6\right)^{1/2} + \left(\sum_i |x_i^\top Q|^6\right)^{1/2} + \left(\sum_i |x_i^\top e_u|^6\right)^{1/2}\right]. \tag{66}$$

By (46), (50), and (61),

$$\sum_i |x_i^\top \xi|^6 \le c_\xi^3 L^3 \sum_i h_i^3 \le \frac{2\bar\kappa c_\xi^3 L^3 M}{\lambda_p^2},$$
$$\sum_i |x_i^\top Q|^6 \le \frac{c_Q^6 L^6}{\lambda_p^3}\sum_i h_i^3 \le \frac{2\bar\kappa c_Q^6 L^6 M}{\lambda_p^5},$$
$$\sum_i |x_i^\top e_u|^6 \le \frac{z_u^4}{\lambda_p^2}\sum_i (x_i^\top e_u)^2 \le \frac{2\bar\kappa z_u^6}{\lambda_p^2}.$$

Substitution into (66) gives

$$\|R_3(\delta_u)\|_{H_\circ^{-1}} \le 11\bar\kappa\left(c_\xi^{3/2} + \frac{c_Q^3}{C_p^{3/2}}\right)\frac{L^{3/2}\sqrt{M}}{\lambda_p} + \frac{11\bar\kappa}{\lambda_p} z_u^3, \tag{67}$$

where we used $\sqrt{108} < 11$ and $L/\lambda_p \le 1/C_p$.

Completing the residual bound. Taking the $H_\circ^{-1}$-norm of (44) and substituting (60), (65), and (67) gives

$$z_u \le \mathsf{A} + \frac{4\bar\kappa}{3}\left(\sqrt{\frac{c_\xi L}{\lambda_p}} + \frac{c_Q L}{\lambda_p}\right) z_u + \frac{2\bar\kappa}{3\sqrt{\lambda_p}} z_u^2 + \frac{11\bar\kappa}{\lambda_p} z_u^3, \tag{68}$$

where $\mathsf{A} = C_1(L^{3/2}\sqrt{M}/\lambda_p + 1/\sqrt{\lambda_p})$ and $C_1 = C(C_p)$ in (92). The definition of C_1 bounds the constant terms of (60), (65), and (67), using $\sqrt{LM} \le L^{3/2}\sqrt{M}$. Furthermore, $\lambda_p = C_p L M^{1/3}$ implies

$$\frac{\mathsf{A}}{\sqrt{\lambda_p}} = C_1\left(C_p^{-3/2} + (C_p L M^{1/3})^{-1}\right) \le \frac{2C_1}{C_p}, \qquad \frac{\mathsf{A}^2}{\lambda_p} \le \frac{4C_1^2}{C_p^2}.$$

Since $z_0 = 0$ and z_u is continuous, $z_u \ge 2\mathsf{A}$ at some $u \in [0,1]$ would imply $z_v = 2\mathsf{A}$ at some $v \in [0,1]$. At this value, (68) would give

$$\begin{aligned} 2\mathsf{A} &\le \mathsf{A} + \frac{8\bar\kappa}{3}\left(\sqrt{\frac{c_\xi L}{\lambda_p}} + \frac{c_Q L}{\lambda_p}\right)\mathsf{A} + \frac{8\bar\kappa}{3\sqrt{\lambda_p}}\mathsf{A}^2 + \frac{88\bar\kappa}{\lambda_p}\mathsf{A}^3 \\ &\le \mathsf{A}\left[1 + \frac{8\bar\kappa}{3}\left(\sqrt{\frac{c_\xi}{C_p}} + \frac{c_Q}{C_p}\right) + \frac{16\bar\kappa C_1}{3C_p} + \frac{352\bar\kappa C_1^2}{C_p^2}\right] \\ &= \mathsf{A}[1 + \Upsilon(C_p)] \le 31\mathsf{A}/16 < 2\mathsf{A}, \end{aligned}$$

which is a contradiction. The last inequality follows from (93). Thus $z_u < 2\mathsf{A}$ for every $u \in [0,1]$. In particular,

$$\|e_1\|_V^2 \le 2\bar\kappa z_1^2 \le 16\bar\kappa C_1^2\left(\frac{L^3 M}{\lambda_p^2} + \frac{1}{\lambda_p}\right) \le 16\bar\kappa C_1^2(C_p^{-2} + C_p^{-1}) L M^{1/3}. \tag{69}$$

Finally, (45), (56), and (69) imply

$$\begin{aligned}\|b\|_V &\le \|\theta^\circ - \theta_\star\|_V + \|e_1\|_V \\ &\le \left(2\bar\kappa B\sqrt{C_p} + 4\sqrt{\bar\kappa}\, C_1\sqrt{C_p^{-2} + C_p^{-1}}\right)\sqrt{LM^{1/3}} = \sqrt{C_b L M^{1/3}}.\end{aligned}$$

The events in (50), (55), and (59) hold simultaneously with probability at least $1 - 2e^{-L} - 2e^{-L} - e^{-L} = 1 - 5e^{-L}$. On this event, (51) and (55) and the last display prove the three asserted bounds. □

E Proof of Lemma 2

We prove that the normalized secant estimate has a confidence multiplier independent of the design budget M. We first bound the product of the two independent initial-point estimation errors. We then decompose the prediction error into reward noise, regularization, this product, and a cubic remainder.

We prove the following stronger form of Lemma 2. For any deterministic $M \ge 1$, with probability at least $1 - \delta$,

$$|\langle x_{t,a}, \hat\theta - \theta_\star\rangle| \le \alpha_2 \|x_{t,a}\|_{V_E^{-1}}$$

for every $t \le N(M)$, every visited level, and every $a \in [K]$. The definition of $N(M)$ is in (15). This prefix statement is also used in Section F.

E.1 Product of Independent Initial-Point Estimation Errors

Lemma 8. *Let $\mathcal{F}$ fix x, $(x_i)_{i\in E}$, positive definite matrices V_1, V_2, and $V_E = \lambda_E I + \sum_{i\in E} x_i x_i^\top$, where $\lambda_E > 0$. Suppose that, for $k = 1, 2$,*

$$x_i^\top V_k^{-1} x_i \le \tau \quad (i \in E), \qquad X_E = \sum_{i\in E} x_i x_i^\top \preceq \gamma V_k.$$

Conditionally on $\mathcal{F}$, suppose that (ζ_1, b_1) and (ζ_2, b_2) are independent. Each ζ_k is a centered polynomial of degree at most two in residuals $R_i - p_i$ of independent $R_i \sim \mathrm{Bern}(p_i)$ with success probabilities in $[p_0, 1 - p_0]$, and $\mathrm{Cov}(\zeta_k \mid \mathcal{F}) \preceq C_v V_k^{-1}$. Define

$$B_x = \sum_{i\in E} \frac{\ddot\mu(\langle x_i, \theta_\star\rangle)}{2\dot\mu(\langle x_i, \theta_\star\rangle)} (x^\top V_E^{-1} x_i) x_i x_i^\top, \qquad \bar\theta_k - \theta_\star = \zeta_k + b_k.$$

For $L \ge 1$ and $R \ge 0$, with conditional probability at least $1 - 3e^{-L}$, whenever $\max_{k=1,2}\|b_k\|_{V_k} \le R$, we have

$$|(\bar\theta_1 - \theta_\star)^\top B_x (\bar\theta_2 - \theta_\star)| \le \frac{\sqrt{\tau\gamma}}{2}\|x\|_{V_E^{-1}} \left(\frac{64 e C_v L^2}{p_0^2} + \frac{16 e\sqrt{C_v} L R}{p_0} + R^2\right).$$

Proof. We bound the four terms in

$$(\bar\theta_1 - \theta_\star)^\top B_x (\bar\theta_2 - \theta_\star) = \underbrace{\zeta_1^\top B_x \zeta_2}_{J_1} + \underbrace{\zeta_1^\top B_x b_2}_{J_2} + \underbrace{b_1^\top B_x \zeta_2}_{J_3} + \underbrace{b_1^\top B_x b_2}_{J_4}. \tag{70}$$

The matrix in each term is fixed conditionally on $\mathcal{F}$. Write $B_x = \sum_i a_i x_i x_i^\top$ and $\widetilde B = V_1^{-1/2} B_x V_2^{-1/2}$. By (95), $|\ddot\mu(u)/(2\dot\mu(u))| \le 1/2$, so

$$\|a\|_2^2 \le \frac{1}{4}\sum_{i\in E}(x^\top V_E^{-1} x_i)^2 = \frac{1}{4} x^\top V_E^{-1}(V_E - \lambda_E I) V_E^{-1} x \le \frac{1}{4}\|x\|_{V_E^{-1}}^2,$$

$$\|\widetilde B\|_{\mathrm{op}} \le \|\widetilde B\|_{\mathrm{F}} \le \sqrt{\tau\gamma}\|a\|_2 \le \frac{\sqrt{\tau\gamma}}{2}\|x\|_{V_E^{-1}}. \tag{71}$$

The middle inequality in (71) is (101) with the two design conditions in the statement.

For J_1, conditional independence and centering give $\mathbb{E}[J_1 \mid \mathcal{F}] = 0$ and

$$\begin{aligned}\mathbb{E}[J_1^2 \mid \mathcal{F}] &= \mathbb{E}\big[\zeta_1^\top B_x \operatorname{Cov}(\zeta_2 \mid \mathcal{F}) B_x^\top \zeta_1 \,\big|\, \mathcal{F}\big] \\ &= \operatorname{tr}\big(B_x \operatorname{Cov}(\zeta_2 \mid \mathcal{F}) B_x^\top \operatorname{Cov}(\zeta_1 \mid \mathcal{F})\big) \\ &\le C_v^2 \operatorname{tr}(B_x V_2^{-1} B_x^\top V_1^{-1}) = C_v^2 \|\widetilde{B}\|_{\mathrm{F}}^2. \end{aligned} \tag{72}$$

The scalar J_1 is a polynomial of degree at most four. Applying Lemma 18 with $m = 4$ and $z = L$ to (72) yields

$$\mathbb{P}\left(|J_1| > \frac{64 e C_v L^2}{p_0^2} \|\widetilde{B}\|_{\mathrm{F}} \,\middle|\, \mathcal{F} \right) \le e^{-L}.$$

For J_2, condition further on (ζ_2, b_2). The conditional law of (ζ_1, b_1) is unchanged, and b_2 is fixed. Thus

$$\begin{aligned}\mathbb{E}[J_2^2 \mid \mathcal{F}, \zeta_2, b_2] &\le C_v b_2^\top B_x^\top V_1^{-1} B_x b_2 \\ &= C_v \|\widetilde{B} V_2^{1/2} b_2\|_2^2 \le C_v \|\widetilde{B}\|_{\mathrm{op}}^2 \|b_2\|_{V_2}^2. \end{aligned}$$

Applying Lemma 18 with $m = 2$ and $z = L$, and then integrating over (ζ_2, b_2), gives

$$\mathbb{P}\left(\|b_2\|_{V_2} \le R,\ |J_2| > \frac{8 e L \sqrt{C_v} R}{p_0} \|\widetilde{B}\|_{\mathrm{op}} \,\middle|\, \mathcal{F} \right) \le e^{-L}. \tag{73}$$

Exchanging indices 1 and 2 in (73) gives the same bound for J_3. Finally,

$$|J_4| = |(V_1^{1/2} b_1)^\top \widetilde{B} (V_2^{1/2} b_2)| \le \|b_1\|_{V_1} \|\widetilde{B}\|_{\mathrm{op}} \|b_2\|_{V_2} \le R^2 \|\widetilde{B}\|_{\mathrm{op}}$$

when both residual norms are at most R. A union bound for J_1, J_2, J_3, followed by substitution into (70) and (71), proves the result. □

E.2 Proof of Lemma 2

Proof. Fix a round t and a level s, and condition on $\mathcal{F}_{s,t}$ from Lemma 15. All sets, contexts, and designs in this proof are those of level s before round t. A maximum over an empty estimation set is interpreted as zero. Write $\tau = \tau_s$ and $\gamma = \gamma_s$. We first establish a conditional bound on the $\mathcal{F}_{s,t}$-measurable event

$$\mathcal{E}_{s,t}^0 = \left\{ \text{round } t \text{ visits level } s,\ \max_{k=1,2} \log\det(V_k/\lambda_p) \le M \right\}.$$

On this event, (105) gives

$$\sum_{i \in P_k} x_i^\top V_k^{-1} x_i \le \log\det(V_k/\lambda_p) \le M.$$

We can therefore apply Lemma 7 to both initial-point estimates. Use the test set $\mathcal{U} = \{x_i : i \in E\} \cup \mathcal{X}_t$, which satisfies $|P_k| + |\mathcal{U}| \le T + K$. The lemma gives independent decompositions $\bar{\theta}_k - \theta_\star = \zeta_k + b_k$ with the covariance and polynomial properties in Lemma 8. Let $\mathcal{E}_{s,t}^1$ denote the event that, for $k = 1, 2$,

$$\|\zeta_k\|_{V_k}^2 \le C_E L M, \qquad |\langle x_i, \zeta_k \rangle|^2 \le C_x L\, x_i^\top V_k^{-1} x_i \ (i \in E), \qquad \|b_k\|_{V_k}^2 \le C_b L M^{1/3}. \tag{74}$$

By Lemma 7, $\mathbb{P}((\mathcal{E}_{s,t}^1)^c \mid \mathcal{F}_{s,t}) \le 10 e^{-L}$ on $\mathcal{E}_{s,t}^0$.

Decomposition of the prediction error. We first show that Lemma 10 applies to the initial-point estimate predictions. On $\mathcal{E}_{s,t}^1$, Lemma 4 and Cauchy-Schwarz give

$$\begin{aligned}|\langle x_i, \bar{\theta}_k - \theta_\star \rangle| &\le |\langle x_i, \zeta_k \rangle| + \|x_i\|_{V_k^{-1}} \|b_k\|_{V_k} \\ &\le \sqrt{C_x L \tau} + \sqrt{C_b L M^{1/3} \tau} \\ &\le (\sqrt{C_x} + \sqrt{C_b}) \sqrt{L\tau}\, M^{1/6} = \frac{\sqrt{C_x} + \sqrt{C_b}}{\sqrt{C_2}} \sqrt{\frac{\epsilon_s}{4L}} \le \frac{1}{2}. \end{aligned} \tag{75}$$

The equality uses $\tau_s = \epsilon_s/(4C_2L^2M^{1/3})$. The last inequality uses $C_2 \geq (\sqrt{C_x} + \sqrt{C_b})^2$, $\epsilon_s \leq 1$, and $L \geq 1$. Since $|u_i| = |\langle x_i, \theta_\star\rangle| \leq B$, clipping to $[-B-1, B+1]$ does not change $z_{k,i} = \langle x_i, \bar{\theta}_k\rangle$.

Put $e_{k,i} = z_{k,i} - u_i$ and $\varepsilon_i = r_i - \mu(u_i)$. By (75) and Lemma 10, the secant slope and normalized error satisfy

$$\begin{aligned} e^{-1}\dot{\mu}(u_i) \leq \omega_i \leq e\dot{\mu}(u_i), \qquad \omega_i \geq 1/\bar{\kappa}, \\ z_{1,i} + \frac{r_i - \mu(z_{1,i})}{\omega_i} - u_i = \frac{\varepsilon_i}{\omega_i} + \frac{\ddot{\mu}(u_i)}{2\dot{\mu}(u_i)} e_{1,i}e_{2,i} + \mathcal{R}_{3,i}, \\ |\mathcal{R}_{3,i}| \leq \frac{5e^2}{12}(|e_{1,i}|^2|e_{2,i}| + |e_{1,i}||e_{2,i}|^2). \end{aligned} \tag{76}$$

Substituting (76) into the normal equation (10), we obtain

$$\begin{aligned} \langle x, \hat{\theta} - \theta_\star\rangle = \underbrace{\sum_{i\in E}(x^\top V_E^{-1}x_i)\varepsilon_i/\omega_i}_{I_1} - \underbrace{\lambda_E x^\top V_E^{-1}\theta_\star}_{I_2} \\ + \underbrace{(\bar{\theta}_1 - \theta_\star)^\top B_x(\bar{\theta}_2 - \theta_\star)}_{I_3} + \underbrace{\sum_{i\in E}(x^\top V_E^{-1}x_i)\mathcal{R}_{3,i}}_{I_4}. \end{aligned} \tag{77}$$

Here B_x is defined in Lemma 8, and $\sum_{i\in E} x_i x_i^\top = V_E - \lambda_E I$ gives the regularization term I_2. We bound I_1, I_2, I_3, I_4 in this order.

Reward noise and regularization. Condition further on the reward vectors in P_1, P_2. The coefficients of I_1 are now fixed and, by Lemma 15, the estimation rewards remain independent. The following variances use this conditioning. On $\mathcal{E}^1_{s,t}$,

$$\begin{aligned} \sum_{i\in E}\mathrm{Var}\big((x^\top V_E^{-1}x_i)\varepsilon_i/\omega_i\big) &= \sum_{i\in E}(x^\top V_E^{-1}x_i)^2\dot{\mu}(u_i)/\omega_i^2 \\ &\leq e\bar{\kappa}\sum_{i\in E}(x^\top V_E^{-1}x_i)^2 \\ &= e\bar{\kappa}\, x^\top V_E^{-1}(V_E - \lambda_E I)V_E^{-1}x \leq e\bar{\kappa}\|x\|^2_{V_E^{-1}}. \end{aligned}$$

The first inequality follows from $\dot{\mu}(u_i)/\omega_i^2 \leq e/\omega_i \leq e\bar{\kappa}$ in (76). Each summand is bounded by

$$|(x^\top V_E^{-1}x_i)\varepsilon_i/\omega_i| \leq \bar{\kappa}\|x\|_{V_E^{-1}}\|x_i\|_{V_E^{-1}} \leq \frac{\bar{\kappa}}{\sqrt{\lambda_E}}\|x\|_{V_E^{-1}},$$

where we used $|\varepsilon_i| \leq 1$, $\|x_i\|_2 \leq 1$, and $V_E \succeq \lambda_E I$. Applying Lemma 16 with $z = L$ and taking a union bound over the K arms gives

$$|I_1| \leq \left(\sqrt{2e\bar{\kappa}L} + \frac{\bar{\kappa}L}{3\sqrt{\lambda_E}}\right)\|x\|_{V_E^{-1}}$$

except with conditional probability $2Ke^{-L}$. The event $\mathcal{E}^1_{s,t}$ is determined by the rewards in $P_1 \cup P_2$. Taking the conditional expectation over their rewards thus bounds the probability that $\mathcal{E}^1_{s,t}$ holds but this inequality fails by $2Ke^{-L}$. For I_2, Cauchy-Schwarz gives

$$|I_2| \leq \lambda_E\|x\|_{V_E^{-1}}\|\theta_\star\|_{V_E^{-1}} \leq B\sqrt{\lambda_E}\|x\|_{V_E^{-1}}.$$

Using $\lambda_E = \bar{\kappa}L/(1+B)$ and the definition of α_2,

$$\begin{aligned} |I_1| + |I_2| &\leq \left(\sqrt{2e\bar{\kappa}L} + \frac{\bar{\kappa}L}{3\sqrt{\lambda_E}} + B\sqrt{\lambda_E}\right)\|x\|_{V_E^{-1}} \\ &= \left(\sqrt{2e\bar{\kappa}L} + \frac{1+4B}{3\sqrt{1+B}}\sqrt{\bar{\kappa}L}\right)\|x\|_{V_E^{-1}} = \frac{2\alpha_2}{3}\|x\|_{V_E^{-1}}. \end{aligned} \tag{78}$$

Product of the two initial-point estimation errors. Apply Lemma 8 with $R = \sqrt{C_b L M^{1/3}}$. Its two design conditions follow from Lemma 4. Conditionally on $\mathcal{F}_{s,t}$, the probability that $\mathcal{E}^1_{s,t}$ holds but the resulting bound fails for some current arm is at most $3Ke^{-L}$. For all other outcomes in $\mathcal{E}^1_{s,t}$,

$$\begin{aligned}|I_3| &\le \frac{\sqrt{\tau_s\gamma_s}}{2}\|x\|_{V_E^{-1}}\left(\frac{64eC_vL^2}{p_0^2}+\frac{16e\sqrt{C_vC_b}}{p_0}L^{3/2}M^{1/6}+C_bLM^{1/3}\right)\\ &\le \frac{C_qL^2M^{1/3}}{2}\sqrt{\tau_s\gamma_s}\|x\|_{V_E^{-1}}\\ &\le \frac{C_2L^2M^{1/3}}{2}\frac{\alpha_2\tau_s}{\epsilon_s}\|x\|_{V_E^{-1}}=\frac{\alpha_2}{8}\|x\|_{V_E^{-1}}.\end{aligned} \tag{79}$$

The second inequality uses $L, M \ge 1$ and the definition of C_q. The third uses $C_q \le C_2$ and $\gamma_s = \alpha_2^2\tau_s/\epsilon_s^2$. The equality substitutes τ_s.

Cubic remainder. It remains to bound I_4 on $\mathcal{E}^1_{s,t}$. Write $a_{k,i} = \langle x_i, \zeta_k\rangle$ and $b_{k,i} = \langle x_i, b_k\rangle$, so $e_{k,i} = a_{k,i} + b_{k,i}$. By (74) and Lemma 4,

$$\max_{i\in E}|a_{k,i}| \le \sqrt{C_xL\tau}, \qquad \max_{i\in E}|b_{k,i}| \le \sqrt{C_bLM^{1/3}\tau}. \tag{80}$$

The same results imply

$$\begin{aligned}\sum_{i\in E}a_{1,i}^2a_{2,i}^2 &\le C_xL\tau\sum_{i\in E}a_{2,i}^2 = C_xL\tau\,\zeta_2^\top X_E\zeta_2 \le C_xC_EL^2M\tau\gamma,\\ \sum_{i\in E}a_{1,i}^2b_{2,i}^2 &\le C_xL\tau\,b_2^\top X_Eb_2 \le C_xC_bL^2M^{1/3}\tau\gamma,\\ \sum_{i\in E}b_{1,i}^2a_{2,i}^2 &\le C_xL\tau\,b_1^\top X_Eb_1 \le C_xC_bL^2M^{1/3}\tau\gamma,\\ \sum_{i\in E}b_{1,i}^2b_{2,i}^2 &\le C_bLM^{1/3}\tau\,b_2^\top X_Eb_2 \le C_b^2L^2M^{2/3}\tau\gamma.\end{aligned} \tag{81}$$

Each last inequality uses $X_E \preceq \gamma V_k$ and the corresponding norm bound in (74).

To use these bounds, expand

$$e_{1,i}^2e_{2,i} = a_{1,i}^2a_{2,i} + a_{1,i}^2b_{2,i} + 2a_{1,i}b_{1,i}a_{2,i} + 2a_{1,i}b_{1,i}b_{2,i} + b_{1,i}^2a_{2,i} + b_{1,i}^2b_{2,i}.$$

For these six terms, (80) and (81) give

$$\begin{aligned}\|(a_{1,i}^2a_{2,i})_{i\in E}\|_2 &\le \max_i|a_{1,i}|\left(\sum_i a_{1,i}^2a_{2,i}^2\right)^{1/2} \le C_x\sqrt{C_E}L^{3/2}M^{1/2}\tau\sqrt{\gamma},\\ \|(a_{1,i}^2b_{2,i})_{i\in E}\|_2 &\le \max_i|a_{1,i}|\left(\sum_i a_{1,i}^2b_{2,i}^2\right)^{1/2} \le C_x\sqrt{C_b}L^{3/2}M^{1/6}\tau\sqrt{\gamma},\\ \|(a_{1,i}b_{1,i}a_{2,i})_{i\in E}\|_2 &\le \max_i|a_{1,i}|\left(\sum_i b_{1,i}^2a_{2,i}^2\right)^{1/2} \le C_x\sqrt{C_b}L^{3/2}M^{1/6}\tau\sqrt{\gamma},\\ \|(a_{1,i}b_{1,i}b_{2,i})_{i\in E}\|_2 &\le \max_i|b_{1,i}|\left(\sum_i a_{1,i}^2b_{2,i}^2\right)^{1/2} \le \sqrt{C_x}C_bL^{3/2}M^{1/3}\tau\sqrt{\gamma},\\ \|(b_{1,i}^2a_{2,i})_{i\in E}\|_2 &\le \max_i|b_{1,i}|\left(\sum_i b_{1,i}^2a_{2,i}^2\right)^{1/2} \le \sqrt{C_x}C_bL^{3/2}M^{1/3}\tau\sqrt{\gamma},\\ \|(b_{1,i}^2b_{2,i})_{i\in E}\|_2 &\le \max_i|b_{1,i}|\left(\sum_i b_{1,i}^2b_{2,i}^2\right)^{1/2} \le C_b^{3/2}L^{3/2}M^{1/2}\tau\sqrt{\gamma}.\end{aligned}$$

All sums and maxima are over E. By the triangle inequality and $M^{1/6}, M^{1/3} \le M^{1/2}$,

$$\begin{aligned}\|(e_{1,i}^2 e_{2,i})_{i\in E}\|_2 &\le \left(C_x\sqrt{C_E} + 3C_x\sqrt{C_b} + 3\sqrt{C_x}C_b + C_b^{3/2}\right) L^{3/2}M^{1/2}\tau\sqrt{\gamma} \\ &= C_3 L^{3/2}M^{1/2}\tau\sqrt{\gamma}. \end{aligned} \tag{82}$$

Exchanging indices 1 and 2 proves the same bound for $\|(e_{1,i}e_{2,i}^2)_{i\in E}\|_2$. Consequently,

$$\begin{aligned}|I_4| &\le \left(\sum_{i\in E}(x^\top V_E^{-1}x_i)^2\right)^{1/2} \|(\mathcal{R}_{3,i})_{i\in E}\|_2 \\ &\le \frac{5e^2}{12}\|x\|_{V_E^{-1}}\left(\|(e_{1,i}^2 e_{2,i})_{i\in E}\|_2 + \|(e_{1,i}e_{2,i}^2)_{i\in E}\|_2\right) \\ &\le \frac{5e^2 C_3}{6} L^{3/2}M^{1/2}\tau_s\sqrt{\gamma_s}\|x\|_{V_E^{-1}} \\ &= \frac{5e^2 C_3}{48C_2^{3/2}}\frac{\alpha_2\sqrt{\epsilon_s}}{L^{3/2}}\|x\|_{V_E^{-1}} \le \frac{\alpha_2}{6}\|x\|_{V_E^{-1}}. \end{aligned} \tag{83}$$

The first inequality is Cauchy-Schwarz. The second uses (76) and $\sum_i (x^\top V_E^{-1}x_i)^2 \le \|x\|_{V_E^{-1}}^2$. The third is (82). The equality substitutes τ_s, γ_s, and the last inequality uses $C_2^{3/2} \ge 5e^2C_3/8$, $\epsilon_s \le 1$, and $L \ge 1$.

Simultaneous confidence. On $\mathcal{E}_{s,t}^0$, the conditional probability that (74) fails, or it holds and either the noise or product bound fails for a current arm, is at most

$$10e^{-L} + 2Ke^{-L} + 3Ke^{-L} \le 15Ke^{-L}.$$

On all remaining outcomes, substituting (78), (79), and (83) into (77) gives

$$|\langle x, \hat{\theta} - \theta_\star\rangle| \le (|I_1| + |I_2|) + |I_3| + |I_4| \le \left(\frac{2}{3} + \frac{1}{8} + \frac{1}{6}\right)\alpha_2\|x\|_{V_E^{-1}} = \frac{23}{24}\alpha_2\|x\|_{V_E^{-1}}.$$

Since $\mathcal{E}_{s,t}^0$ is $\mathcal{F}_{s,t}$-measurable, integrating the conditional failure bound gives the same bound for failure intersected with $\mathcal{E}_{s,t}^0$. A union bound over $(S+1)T \le 2ST$ level-round pairs yields $30STKe^{-L} < \delta$. If $t \le N(M)$, every visited level satisfies $\mathcal{E}_{s,t}^0$. The resulting simultaneous event therefore implies the claimed confidence bound. □

F Adaptation to the Revealed Features

The log-determinants used in the proofs can be smaller than the ambient dimension bound M_d. We give a restart rule that adapts to this quantity without knowing a suitable budget in advance. Define

$$d_{\text{eff}} = \max\left\{1, \log\det\left(I + \sum_{t=1}^{T}\sum_{a=1}^{K} x_{t,a}x_{t,a}^\top\right)\right\}. \tag{84}$$

This quantity includes every offered feature. It is used in the regret bound, not as an input to the algorithm.

For $v \in \{1,2\}$, let $\mathsf{A}_v(M,\delta)$ denote the base algorithm with design budget M. Thus A_1 is Sup-SplitLog+ with one unadjusted initial-point estimate and A_2 is SupSecantLog with two adjusted initial-point estimates. Algorithm 3 runs fresh instances with $M_j = 2^j$ and $\delta_j = 6\delta/(\pi^2(j+1)^2)$. Before accepting an exploration proposal, it checks whether the updated design satisfies the current budget. If it does not, the algorithm restarts before playing the arm and makes another proposal on the same context set. Every instance uses the original horizon parameter T.

Theorem 3. *Under Assumptions 1 and 2, Algorithm 3 satisfies, with probability at least* $1-\delta$,

$$R_T = \begin{cases} \widetilde{\mathcal{O}}\left(\kappa\sqrt{d_{\text{eff}}T} + \kappa^3(1+B\kappa^2)^2 d_{\text{eff}}^{3/2}\right), & v = 1, \\ \widetilde{\mathcal{O}}\left(\sqrt{(1+B)\kappa d_{\text{eff}}T} + (\kappa^{9/2} + B^2\kappa^{15/2})d_{\text{eff}}^{4/3}\right), & v = 2. \end{cases}$$

If the offered features span a subspace of dimension $r \ge 1$, *then* $d_{\text{eff}} \le \max\{1, r\log(1+TK/r)\}$.

Algorithm 3 Log-determinant doubling for A_v

input: d, T, K, B, algorithm index $v \in \{1, 2\}$, failure probability δ
initialize: $j \leftarrow 0$, $M_j \leftarrow 1$, $\delta_j \leftarrow 6\delta/\pi^2$, and a new instance of $\mathsf{A}_v(M_j, \delta_j)$
for $t = 1, \ldots, T$ **do**
  Observe $\mathcal{X}_t$
  loop
    Obtain the instance's proposed arm a and receiving set
    if the proposal is exploitation **then**
      Play a, discard its reward, and **break**
    end if
    Let V, λ be the design matrix and regularization parameter of the receiving set, and $x \leftarrow x_{t,a}$
    if $\log\det((V + xx^\top)/\lambda) > M_j$ **then**
      $j \leftarrow j + 1$, $M_j \leftarrow 2^j$, $\delta_j \leftarrow 6\delta/(\pi^2(j+1)^2)$
      Start a new instance of $\mathsf{A}_v(M_j, \delta_j)$ // reconsider the same context set
    else
      Play a, observe r_t, add the observation to the receiving set, and **break**
    end if
  end loop
end for

F.1 Proof of Theorem 3

The proof applies Lemma 3 to each completed phase. The budget check ensures that its conclusion holds up to that phase's last accepted round. We then sum over the geometrically increasing budgets. To keep the constants explicit, define

$$\Phi_1(M, L) = \frac{\alpha_1}{\rho} L M^{3/2}, \qquad \Phi_2(M, L) = 4C_2 L^2 M^{4/3}, \tag{85}$$

where α_1, ρ are evaluated at L. Let J be the last phase, $L_J = \log(32STK/\delta_J)$, and let α_J be its confidence multiplier. We will prove

$$R_T \le 26\alpha_J \sqrt{(S+1)d_{\text{eff}}T} + \sqrt{T}/2 + 300(S+1)\Phi_v(d_{\text{eff}}, L_J). \tag{86}$$

Proof of Theorem 3. For phase j, write $L_j = \log(32STK/\delta_j)$ and let α_j be its confidence multiplier. Set $F_j = \epsilon_s/\tau_s$, which does not depend on s. By the parameter choices in Section G, $M_j F_j = \Phi_v(M_j, L_j)$.

Valid prefixes and the number of phases. A new instance has design λI in every set, with $\lambda \ge 1$. Its first exploration proposal satisfies

$$\log\det((\lambda I + xx^\top)/\lambda) = \log(1 + \|x\|_2^2/\lambda) \le \log 2 < 1 \le M_j.$$

Thus a fresh instance accepts its first proposal, and a round requires at most one restart. On each accepted exploration round, the updated design passes the budget check. Other designs do not change, and an exploitation round does not change any design. Induction over accepted rounds therefore shows that every accepted prefix of phase j satisfies (15) with $M = M_j$.

For a rejected proposal, its design contains only earlier features in that phase and the proposed feature. Since $\lambda \ge 1$,

$$\begin{aligned}\log\det((V + xx^\top)/\lambda) &= \log\det\left(I + \lambda^{-1}\left(\sum_{i \text{ in the receiving set}} x_i x_i^\top + xx^\top\right)\right) \\ &\le \log\det\left(I + \sum_{t=1}^{T}\sum_{a=1}^{K} x_{t,a}x_{t,a}^\top\right) \le d_{\text{eff}}.\end{aligned}$$

The first inequality uses positive semidefinite ordering and monotonicity of the log-determinant. A rejection in phase j consequently implies $M_j < d_{\text{eff}}$. Since $M_j = 2^j$, we obtain

$$M_J \le 2d_{\text{eff}}, \qquad \sum_{j=0}^{J} M_j \le 2M_J \le 4d_{\text{eff}}, \qquad J \le \log_2(2d_{\text{eff}}). \tag{87}$$

When $J = 0$, the same inequalities follow from $d_{\text{eff}} \ge 1$.

Regret in each phase. Let σ_j be the starting round of phase j, and put $\sigma_j = T + 1$ if that phase is not reached. Use the reward representation $r_t = \mathbb{1}\{U_t \le \mu(\langle x_t, \theta_\star\rangle)\}$ from Section I.1. For $\mathcal{H}_t = \sigma(W, U_1, \dots, U_{t-1})$, the event $\{\sigma_j = t\}$ belongs to $\mathcal{H}_t$, because the proposal and budget check precede the reward in round t. The restart property proved in Section I.1 therefore gives independent future uniforms conditionally on the history at σ_j.

Starting at σ_j, consider running that instance on the remaining context sequence without further restarts. Before its first rejected proposal, this run has exactly the decisions and observations of the actual phase. The rejected proposal is not played by that instance. The actual phase is a prefix of this run, and every accepted update satisfies (15) with $M = M_j$. The run retains horizon parameter T, so the prefix statement of Lemma 3 applies to all its accepted prefix lengths.

Let n_j be the number of played rounds assigned to phase j and $R^{(j)}$ their regret. Apply Lemma 3 conditionally on the history at σ_j, using Lemma 15 for the level-wise samples after the restart. Except on an event of probability at most δ_j intersected with $\{\sigma_j \le T\}$,

$$R^{(j)} \le 13\alpha_j\sqrt{(S+1)M_j n_j} + n_j/(2\sqrt{T}) + 64(S+1)\Phi_v(M_j, L_j). \tag{88}$$

Here the statement for every valid prefix permits the random length n_j. Since $\sum_{j\ge0}\delta_j = \delta$, a union bound implies that (88) holds in all reached phases with probability at least $1 - \delta$.

Summing the phase bounds. Work on this event. The phases partition the played rounds, so $\sum_{j=0}^{J} n_j = T$. Summing (88) gives

$$R_T \le \underbrace{13\sum_{j=0}^{J}\alpha_j\sqrt{(S+1)M_j n_j}}_{I_1} + \underbrace{\sum_{j=0}^{J} n_j/(2\sqrt{T})}_{I_2} + \underbrace{64(S+1)\sum_{j=0}^{J}\Phi_v(M_j, L_j)}_{I_3}. \tag{89}$$

The sequences L_j and α_j are nondecreasing. Thus

$$\begin{aligned} I_1 &\le 13\alpha_J\sqrt{S+1}\sum_{j=0}^{J}\sqrt{M_j n_j} \\ &\le 13\alpha_J\sqrt{S+1}\sqrt{\left(\sum_{j=0}^{J} M_j\right)\left(\sum_{j=0}^{J} n_j\right)} \le 26\alpha_J\sqrt{(S+1)d_{\text{eff}}T}. \end{aligned}$$

The second inequality is Cauchy-Schwarz and the last is (87). Also $I_2 = \sqrt{T}/2$.

To bound I_3, let $p = 3/2$ for $v = 1$ and $p = 4/3$ for $v = 2$. For fixed M, $\Phi_v(M, L)$ is nondecreasing in L. For $v = 1$, this follows from $\alpha_1/\rho = \max\{2a_0(L), 2c_R\}$ in (91). For $v = 2$, it follows from the factor L^2 in (85). Therefore

$$\begin{aligned} \sum_{j=0}^{J}\Phi_v(M_j, L_j) &\le \sum_{j=0}^{J}\Phi_v(M_j, L_J) = \Phi_v(M_J, L_J)\sum_{j=0}^{J} 2^{-p(J-j)} \\ &\le \frac{\Phi_v(M_J, L_J)}{1 - 2^{-p}} \le \frac{2^p}{1 - 2^{-p}}\Phi_v(d_{\text{eff}}, L_J) < 4.4\Phi_v(d_{\text{eff}}, L_J). \end{aligned}$$

We used the geometric series formula and $M_J \le 2d_{\text{eff}}$. It follows that $I_3 \le 300(S+1)\Phi_v(d_{\text{eff}}, L_J)$. Substituting the bounds for I_1, I_2, I_3 into (89) proves (86).

Finally, with $L = \log(32STK/\delta)$, (87) gives

$$\begin{aligned} L_J &= L + \log(\pi^2/6) + 2\log(J+1) \\ &\le L + 1 + 2\log(1 + \log_2(2d_{\text{eff}})). \end{aligned}$$

Substituting the orders in (91) and (94) into (86) yields the two regret bounds.

For the rank statement, suppose $r \ge 1$ and let $\nu_1, \dots, \nu_r$ be the eigenvalues of $\sum_{t,a} x_{t,a} x_{t,a}^\top$ on a subspace containing the offered features. Then $\sum_{i=1}^r \nu_i \le TK$, and concavity of $\log$ gives

$$\log\det\left(I + \sum_{t,a} x_{t,a} x_{t,a}^\top\right) = \sum_{i=1}^r \log(1+\nu_i) \le r \log\left(1 + \frac{\sum_i \nu_i}{r}\right) \le r\log(1 + TK/r).$$

Combining this with (84) proves the claim. If all offered features are zero, (84) gives $d_{\text{eff}} = 1$. □

G Parameter Choices

Set $S = \lceil \log_2 T \rceil$, $\epsilon_s = 2^{-s}$, and $L = \log(32STK/\delta)$. The algorithms take a budget $M \ge 1$. The choice $M = \max\{1, d\log(1 + T/d)\}$ bounds $\log\det(V/\lambda)$ for every design $V = \lambda I + \sum_i x_i x_i^\top$ used by either algorithm, since their regularization parameters are at least one. For SupSplitLog+, set

$$\lambda_E = 1, \qquad \lambda_p = \kappa^4 L\sqrt{M}, \qquad \tau_s = \frac{\rho\epsilon_s}{\alpha_1 L\sqrt{M}}, \qquad \gamma_s = \frac{\alpha_1^2 \tau_s}{\epsilon_s^2}.$$

For SupSecantLog, set

$$\lambda_E = \frac{\bar\kappa L}{1+B}, \qquad \lambda_p = C_p L M^{1/3}, \qquad \tau_s = \frac{\epsilon_s}{4C_2 L^2 M^{1/3}}, \qquad \gamma_s = \frac{\alpha_2^2 \tau_s}{\epsilon_s^2}.$$

The constants below make both algorithms fully specified. They are retained in the proofs when verifying the parameter inequalities.

G.1 Parameters of SupSplitLog+

Recall $\kappa = 1/\dot\mu(B)$ and $\bar\kappa = 1/\dot\mu(B+1)$, and define $p_0 = \mu(-B)$. Since $\kappa = 2 + e^B + e^{-B}$ and $\bar\kappa = 2 + e^{B+1} + e^{-B-1}$,

$$\kappa \ge 4 + B^2 \ge 2(1+B), \qquad \kappa \le \bar\kappa \le e\kappa, \qquad \frac{1}{p_0} = 1 + e^B \le \kappa, \qquad \sqrt{\kappa}\, B \le \kappa. \tag{90}$$

With $P_B \coloneqq (1 + B\kappa^2)^2$,

$$\begin{aligned} c_r &\coloneqq 3\kappa^2 P_B && \text{(residual } b\text{, Lemma 5)}, \\ c_m &\coloneqq 120\kappa^4 P_B^2 && \text{(fourth moment, Lemma 1)}, \\ c_R &\coloneqq \kappa^3 P_B \ \ge\ \tfrac{\kappa}{20}\sqrt{c_m} && \text{(Taylor remainder, Lemma 1)}, \\ a_0 &\coloneqq \kappa\sqrt{L/2} + \sqrt{\kappa}B, \quad \alpha_1 \coloneqq 2a_0, \quad \rho \coloneqq \min\{1, a_0/c_R\}. \end{aligned} \tag{91}$$

By (90), $a_0 \le \kappa(\sqrt{L/2}+1) \le 2\kappa\sqrt{L}$, and $\alpha_1/\rho = \max\{2a_0, 2c_R\} \le 4\kappa\sqrt{L} + 2\kappa^3 P_B$.

G.2 Parameters of SupSecantLog

The constants are defined in dependency order. First,

$$c_\xi \coloneqq 2\bar\kappa, \qquad c_Q \coloneqq \frac{83\bar\kappa^2}{p_0}, \qquad C_\delta \coloneqq \bar\kappa^2(5 + 2\sqrt{\bar\kappa})^2, \qquad c_A \coloneqq 2\bar\kappa + \frac{8\bar\kappa^2}{3}.$$

For $c \ge 1$ let

$$C(c) \coloneqq \max\Big\{\sqrt{2\bar\kappa}\, c_A \sqrt{C_\delta} + \frac{4\bar\kappa}{3} c_Q \sqrt{c_\xi} + \frac{2\bar\kappa}{3}\frac{c_Q^2}{\sqrt{c}} + 11\bar\kappa\Big(c_\xi^{3/2} + \frac{c_Q^3}{c^{3/2}}\Big),\ \frac{16\bar\kappa^3 B}{9}\Big\}, \tag{92}$$

which is nonincreasing in c, and define the regularization constant C_p

$$C_p := \min\{c \ge \max\{1, 4\bar{\kappa}c_A\} : \ \Upsilon(c) \le \tfrac{15}{16}\},$$
$$\Upsilon(c) := \frac{8\bar{\kappa}}{3}\Big(\sqrt{\tfrac{c_\xi}{c}} + \frac{c_Q}{c}\Big) + \frac{16\bar{\kappa}\,C(c)}{3c} + \frac{352\bar{\kappa}\,C(c)^2}{c^2}, \tag{93}$$

and $C_1 := C(C_p)$. Here Υ is continuous and nonincreasing on $[1,\infty)$ and tends to zero, so C_p is well defined. Next,

$$C_x := 2\bar{\kappa}\Big(\sqrt{c_\xi} + \frac{c_Q}{\sqrt{C_p}}\Big)^2, \qquad C_v := 2\bar{\kappa}^2 + \frac{128\bar{\kappa}^5}{9C_p}, \qquad C_E := \frac{16e^2\bar{\kappa}^2}{p_0} + \frac{1024e^2\bar{\kappa}^5}{9p_0^2C_p},$$
$$C_b := \Big(2\bar{\kappa}B\sqrt{C_p} + 4\sqrt{\bar{\kappa}}\,C_1\sqrt{C_p^{-2} + C_p^{-1}}\Big)^2, \qquad C_q := \frac{64eC_v}{p_0^2} + \frac{16e\sqrt{C_vC_b}}{p_0} + C_b,$$
$$C_3 := C_x\sqrt{C_E} + 3C_x\sqrt{C_b} + 3\sqrt{C_x}\,C_b + C_b^{3/2},$$
$$C_2 := \max\{C_q,\ (\sqrt{C_x} + \sqrt{C_b})^2,\ (\tfrac{5}{8}e^2C_3)^{2/3}\}.$$

The confidence multiplier is

$$\alpha_2 = \frac{3}{2}\sqrt{2e\bar{\kappa}L} + \frac{1+4B}{2\sqrt{1+B}}\sqrt{\bar{\kappa}L}.$$

Dependence on κ and B. The regret bounds use

$$\alpha_2 \le 10\sqrt{(1+B)\kappa L}, \qquad C_2 = \mathcal{O}\Big(\kappa^{9/2} + B^2\kappa^{15/2}\Big). \tag{94}$$

We verify these bounds from the parameter definitions. By (90), $c_\xi = \mathcal{O}(\kappa)$, $c_Q = \mathcal{O}(\kappa^3)$, $C_\delta = \mathcal{O}(\kappa^3)$, and $c_A = \mathcal{O}(\kappa^2)$. Substitute $c = R\kappa^{11/2}$ with $R \ge 1$ into (92). The six terms have orders κ^4, $\kappa^{9/2}$, $R^{-1/2}\kappa^{17/4}$, $\kappa^{5/2}$, $R^{-3/2}\kappa^{7/4}$, and $\kappa^{7/2}$, using $B \le \sqrt{\kappa}$. Hence $C(R\kappa^{11/2}) = \mathcal{O}(\kappa^{9/2})$ uniformly in $R \ge 1$. The three terms of $\Upsilon(c)$ in (93) then have orders $R^{-1/2}\kappa^{-5/4} + R^{-1}\kappa^{-3/2}$, R^{-1}, and $R^{-2}\kappa^{-1}$. Since $4\bar{\kappa}c_A = \mathcal{O}(\kappa^3)$, a sufficiently large constant R makes $c = R\kappa^{11/2}$ admissible. Thus $C_p = \mathcal{O}(\kappa^{11/2})$. Conversely, the middle term of $\Upsilon(C_p) \le 15/16$ gives $C_p \ge 256\bar{\kappa}C_1/45$, and $C_1 \ge (4\bar{\kappa}/3)c_Q\sqrt{c_\xi} = \Omega(\kappa^{9/2})$. Here $c_Q \ge 83\kappa^3/2$, since $1/p_0 \ge \kappa/2$. It follows that

$$C_p = \Theta(\kappa^{11/2}), \qquad C_1 \le 45C_p/(256\bar{\kappa}) = \mathcal{O}\Big(\kappa^{9/2}\Big).$$

The definitions then give $C_x = \mathcal{O}(\kappa^2)$, $C_v = \mathcal{O}(\kappa^2)$ and $C_E = \mathcal{O}(\kappa^3)$. Moreover,

$$C_b \le 8\bar{\kappa}^2B^2C_p + 64\bar{\kappa}C_1^2/C_p = \mathcal{O}\Big(B^2\kappa^{15/2} + \kappa^{9/2}\Big),$$

where we used $(a+b)^2 \le 2a^2 + 2b^2$ and $C_p^{-2} + C_p^{-1} \le 2C_p^{-1}$. For C_q, use $C_v/p_0^2 = \mathcal{O}(\kappa^4)$ and $2\sqrt{C_vC_b}/p_0 \le C_v/p_0^2 + C_b$. The four terms of $C_3^{2/3}$ are bounded by constants times $\kappa^{7/3}$, $C_x^{2/3}C_b^{1/3}$, $C_x^{1/3}C_b^{2/3}$, and C_b. Since $C_x = \mathcal{O}(\kappa^2)$ and $C_b = \Omega(\kappa^{9/2})$, all four are $\mathcal{O}(C_b)$. These inequalities and the definition of C_2 prove its bound in (94). Finally,

$$\alpha_2 \le \big(\tfrac{3}{2}\sqrt{2e} + 2\big)\sqrt{(1+B)\bar{\kappa}L} \le 10\sqrt{(1+B)\kappa L},$$

using $(1+4B)/(2\sqrt{1+B}) \le 2\sqrt{1+B}$, $\bar{\kappa} \le e\kappa$, and $(\tfrac{3}{2}\sqrt{2e} + 2)\sqrt{e} < 10$.

H Auxiliary Lemmas

We write $[z]_c = \max\{-c, \min\{c, z\}\}$, $\|A\|_{\mathrm{F}}^2 = \mathrm{tr}(A^\top A)$, and $\|A\|_{\mathrm{op}} = \sup_{\|x\|_2=1}\|Ax\|_2$. Recall $\kappa = 1/\dot{\mu}(B)$, $\bar{\kappa} = 1/\dot{\mu}(B+1)$, and $p_0 = \mu(-B)$.

H.1 Properties of the Logistic Function

Differentiating $\mu(u) = (1+e^{-u})^{-1}$ gives $\ddot{\mu}(u) = \dot{\mu}(u)(1-2\mu(u))$ and $\dddot{\mu}(u) = \dot{\mu}(u)(1-6\dot{\mu}(u))$. Since $0 < \dot{\mu} \le 1/4$,

$$\begin{aligned} |\ddot{\mu}(u)| &\le \dot{\mu}(u), & |\dddot{\mu}(u)| &\le \dot{\mu}(u), \\ \sup_u |\ddot{\mu}(u)| &= 1/(6\sqrt{3}), & \sup_u |\dddot{\mu}(u)| &= 1/8. \end{aligned} \tag{95}$$

The two suprema follow by maximizing $|a(1-a)(1-2a)|$ on $[0,1]$ and $|a(1-6a)|$ on $[0,1/4]$, respectively. Also, $|(\log\dot{\mu})'| = |1-2\mu| \le 1$, so

$$e^{-|u-v|}\dot{\mu}(v) \le \dot{\mu}(u) \le e^{|u-v|}\dot{\mu}(v). \tag{96}$$

Lemma 9. *For $f \in C^2(\mathbb{R})$, define $\omega_f(z_1,z_2) = (f(z_1)-f(z_2))/(z_1-z_2)$ when $z_1 \ne z_2$ and $\omega_f(z,z) = f'(z)$. Then*

$$f(u)-f(z_1)-\omega_f(z_1,z_2)(u-z_1) = (z_1-u)(z_2-u)\int_{\substack{a,b\ge 0\\ a+b\le 1}} f''(u+a(z_1-u)+b(z_2-u))\,da\,db. \tag{97}$$

For $f=\mu$, the absolute value of the left-hand side is at most $|u-z_1||u-z_2|/20$. If $|u| \le B \le c$, then $|u-[z]_c| \le |u-z|$ and $\dot{\mu}(c) \le \omega_\mu([z_1]_c,[z_2]_c) \le 1/4$.

Proof. Let $h_1 = z_1-u$ and $h_2 = z_2-u$. For $h_1h_2(h_1-h_2) \ne 0$,

$$\begin{aligned} h_1h_2\int_0^1\int_0^{1-a} f''(u+ah_1+bh_2)\,db\,da &= h_1\int_0^1 \{f'(u+h_2+a(h_1-h_2)) - f'(u+ah_1)\}\,da \\ &= h_1\frac{f(z_1)-f(z_2)}{h_1-h_2} - f(z_1) + f(u). \end{aligned}$$

Continuity proves (97) in the remaining cases. The integration region has area $1/2$. Thus (95) gives the coefficient $1/(12\sqrt{3}) < 1/20$. Clipping is projection onto an interval containing u, so it cannot increase $|u-z|$. The clipped secant slope is an average of $\dot{\mu}$ over a subinterval of $[-c,c]$, which gives the last bound. □

Lemma 10. *Let $|z_k-u| \le 1$ for $k=1,2$, $\omega = \omega_\mu(z_1,z_2)$, and $\rho = \mu(u)-\mu(z_1)-\omega(u-z_1)$. Then $e^{-1}\dot{\mu}(u) \le \omega \le e\dot{\mu}(u)$ and*

$$\frac{\rho}{\omega} = \frac{\ddot{\mu}(u)}{2\dot{\mu}(u)}(z_1-u)(z_2-u) + \mathcal{R}_3, \qquad |\mathcal{R}_3| \le \frac{5e^2}{12}|z_1-u||z_2-u|(|z_1-u|+|z_2-u|). \tag{98}$$

The coefficient $|\ddot{\mu}(u)/(2\dot{\mu}(u))|$ is at most $1/2$.

Proof. Every point between u, z_1, z_2 is within distance one of u. By (96), $e^{-1}\dot{\mu}(u) \le \omega \le e\dot{\mu}(u)$. Let ν denote the integral in (97), and set $h = |z_1-u|+|z_2-u|$. We have

$$\begin{aligned} \left|\nu - \frac{\ddot{\mu}(u)}{2}\right| &\le e\dot{\mu}(u)\int_{\substack{a,b\ge 0\\ a+b\le 1}} (a|z_1-u|+b|z_2-u|)\,da\,db = e\dot{\mu}(u)h/6, \\ |\omega-\dot{\mu}(u)| &\le e\dot{\mu}(u)\int_0^1 ((1-a)|z_1-u|+a|z_2-u|)\,da = e\dot{\mu}(u)h/2. \end{aligned}$$

The first inequality uses $|\dddot{\mu}(v)| \le e\dot{\mu}(u)$, and the second uses $|\ddot{\mu}(v)| \le e\dot{\mu}(u)$. Consequently,

$$\left|\frac{\nu}{\omega} - \frac{\ddot{\mu}(u)}{2\dot{\mu}(u)}\right| \le \frac{|\nu-\ddot{\mu}(u)/2|}{\omega} + \frac{|\ddot{\mu}(u)||\dot{\mu}(u)-\omega|}{2\omega\dot{\mu}(u)} \le e^2h/6 + e^2h/4 = 5e^2h/12.$$

Multiplying by $|(z_1-u)(z_2-u)|$ proves (98). The last claim follows from $\ddot{\mu}/\dot{\mu} = 1-2\mu$. □

H.2 Extension of the Logistic Function

Lemma 11. *There exists $g \in C^3(\mathbb{R})$ with $g = \mu$ on $[-B, B]$, $1/(2\bar{\kappa}) \le g' \le 1/4$, $|g''| \le 2/3$, and $|g'''| \le 2$.*

Proof. Let $\sigma(a) = 10a^3 - 15a^4 + 6a^5$ and define

$$\chi(z) = \begin{cases} 1, & |z| \le B, \\ 1 - \sigma(|z| - B), & B < |z| < B + 1, \\ 0, & |z| \ge B + 1. \end{cases} \qquad h(z) = \chi(z)\dot{\mu}(z) + (1 - \chi(z))/(2\bar{\kappa}).$$

Since σ' and σ'' vanish at both endpoints, $\chi \in C^2$, $\|\chi'\|_\infty \le 15/8$, and $\|\chi''\|_\infty \le 6$. The definition gives $1/(2\bar{\kappa}) \le h \le 1/4$ and $h = \dot{\mu}$ on $[-B, B]$. On the transition intervals,

$$|h'| \le \|\chi'\|_\infty/4 + \sup|\ddot{\mu}| \le 15/32 + 1/(6\sqrt{3}) < 2/3,$$
$$|h''| \le \|\chi''\|_\infty/4 + 2\|\chi'\|_\infty \sup|\ddot{\mu}| + \sup|\dddot{\mu}| \le 3/2 + 15/(24\sqrt{3}) + 1/8 < 2.$$

Outside these intervals, h equals $\dot{\mu}$ or $1/(2\bar{\kappa})$. Therefore $g(z) = \mu(0) + \int_0^z h(u)\,du$ has the stated properties. □

Lemma 12. *Suppose $V \succ 0$ and a continuous map $F : \mathbb{R}^d \to \mathbb{R}^d$ satisfies $\langle F(x) - F(y), x - y\rangle \ge m\|x - y\|_V^2$ for some $m > 0$. Then F is bijective and*

$$\|F^{-1}(u) - F^{-1}(v)\|_V \le m^{-1}\|u - v\|_{V^{-1}}. \tag{99}$$

Proof. For $F(x) = u$ and $F(y) = v$, Cauchy-Schwarz gives $m\|x - y\|_V^2 \le \langle u - v, x - y\rangle \le \|u - v\|_{V^{-1}}\|x - y\|_V$. This proves injectivity and (99). To prove surjectivity, fix u and choose $R > m^{-1}\|F(0) - u\|_{V^{-1}}$. On $\|x\|_V = R$,

$$\langle F(x) - u, x\rangle \ge mR^2 - \|F(0) - u\|_{V^{-1}}R > 0.$$

If $F(x) - u$ had no zero in the closed V-ball of radius R, the continuous map $x \mapsto -R(F(x) - u)/\|F(x) - u\|_V$ would have a fixed point in that ball by Brouwer's theorem. At this point $\|x\|_V = R$ and $\langle F(x) - u, x\rangle = -R\|F(x) - u\|_2^2/\|F(x) - u\|_V < 0$, a contradiction. □

H.3 Matrix Inequalities

Lemma 13. *If $A, C \succeq 0$, then*

$$0 \preceq A \circ C, \qquad \|A \circ C\|_{\mathrm{op}} \le (\max_i A_{ii})\|C\|_{\mathrm{op}}. \tag{100}$$

If $x_i^\top V_k^{-1} x_i \le \tau$ and $\sum_i x_i x_i^\top \preceq \gamma V_k$ for $V_k \succ 0$, $k = 1, 2$, then

$$\left\| \sum_i a_i V_1^{-1/2} x_i x_i^\top V_2^{-1/2} \right\|_{\mathrm{F}} \le \sqrt{\tau\gamma}\left(\sum_i a_i^2\right)^{1/2}. \tag{101}$$

If $\|w_i\|_2^2 \le \tau$, $\sum_i w_i w_i^\top \preceq \Gamma I$, and $|c_i| \le C$, then

$$\left\| \sum_i c_i \langle w, w_i\rangle w_i w_i^\top \right\|_{\mathrm{F}} \le C\sqrt{\tau}\,\Gamma\|w\|_2. \tag{102}$$

Proof. Write $A = \sum_j a_j a_j^\top$ and $D_j = \mathrm{diag}(a_j)$. Then $0 \preceq A \circ C = \sum_j D_j C D_j \preceq \|C\|_{\mathrm{op}} \sum_j D_j^2 = \|C\|_{\mathrm{op}}\,\mathrm{diag}(A)$, which proves (100). For (101), let $(G_k)_{ij} = x_i^\top V_k^{-1} x_j$. We have

$$\left\| \sum_i a_i V_1^{-1/2} x_i x_i^\top V_2^{-1/2} \right\|_{\mathrm{F}}^2 = \sum_{i,j} a_i a_j (G_1)_{ij}(G_2)_{ij}$$
$$\le (\max_i (G_1)_{ii})\|G_2\|_{\mathrm{op}} \sum_i a_i^2 \le \tau\gamma \sum_i a_i^2,$$

The first inequality follows from (100). The second uses $\max_i (G_1)_{ii} \le \tau$ and $\|G_2\|_{\mathrm{op}} \le \gamma$, since $\sum_i x_i x_i^\top \preceq \gamma V_2$. For (102), apply (101) with $V_1 = V_2 = I$, $x_i = w_i$, and $a_i = c_i \langle w, w_i \rangle$. This gives

$$\left\| \sum_i c_i \langle w, w_i \rangle w_i w_i^\top \right\|_{\mathrm{F}} \le \sqrt{\tau \Gamma} \left(\sum_i c_i^2 \langle w, w_i \rangle^2 \right)^{1/2} \le C \sqrt{\tau}\, \Gamma \|w\|_2,$$

where the second inequality uses $\sum_i w_i w_i^\top \preceq \Gamma I$. □

Lemma 14 (Elliptical potential lemma (Abbasi-Yadkori et al., 2011)). *Let $V_0 = \lambda I$, $V_i = V_{i-1} + x_i x_i^\top$, $\lambda > 0$, $\|x_i\|_2 \le 1$, and $h_i = x_i^\top V_{i-1}^{-1} x_i$. Then*

$$\det V_i = \det V_{i-1}(1 + h_i), \qquad \sum_{i=1}^n \min\{1, h_i\} \le 2 \log \det(V_n/\lambda), \tag{103}$$

$$\sum_{i=1}^n h_i \le \frac{\lambda^{-1}}{\log(1 + \lambda^{-1})} \log \det(V_n/\lambda), \qquad \log \det(V_n/\lambda) \le d \log(1 + n/(d\lambda)), \tag{104}$$

$$\sum_{i=1}^n x_i^\top V_n^{-1} x_i \le \log \det(V_n/\lambda). \tag{105}$$

The coefficient in (104) is at most $3/2$ when $\lambda \ge 1$.

The first two lines are standard consequences of the matrix determinant lemma. For (105), let $a_1, \dots, a_d$ be the eigenvalues of $\lambda^{-1} \sum_i x_i x_i^\top$. Then

$$\sum_i x_i^\top V_n^{-1} x_i = \mathrm{tr}(V_n^{-1}(V_n - \lambda I)) = \sum_j \frac{a_j}{1 + a_j} \le \sum_j \log(1 + a_j).$$

I Conditional Independence and Concentration Inequalities

I.1 Conditional Independence

Both algorithms use the level construction of Auer (2002). We verify that assigning observations to different sets within a level preserves the conditional independence required by the confidence proofs. Write $W = (\mathcal{X}_i)_{i \le T}$ and represent the rewards as $r_i = \mathbb{1}\{U_i \le \mu(\langle x_i, \theta_\star \rangle)\}$, where the U_i are independent uniform random variables independent of W. For level s and time t, let $\mathcal{F}_{s,t}$ be generated by W and the record of decisions made below level s before time t. This record includes the visits, active sets, and branch decisions below s. It includes a reward and the selected arm only if that observation is stored below s. The choices made after a round reaches s are not included. All ties use a fixed order independent of the rewards.

Lemma 15. *Under Assumption 2, conditionally on $\mathcal{F}_{s,t}$, the sets $P^{(s)}, E^{(s)}$ of SupSplitLog+, or $P_1^{(s)}, P_2^{(s)}, E^{(s)}$ of SupSecantLog, and their contexts are fixed. Their rewards are independent Bernoulli variables with means $\mu(\langle x_i, \theta_\star \rangle)$. Whether round t reaches level s and its active set upon arrival at s are also $\mathcal{F}_{s,t}$-measurable.*

Proof. We apply the level-wise independence argument of Auer (2002, Lemma 14), as used for logistic rewards and the initial-point/estimation split in Bae & Lee (2026, Appendix D.2). Whether a round reaches level s and its active set are determined by the contexts and observations below s. At level s, the widths, the selected exploration arm, and both tests in (9) depend only on contexts. Thus the additional assignment to P, E or P_1, P_2, E is also $\mathcal{F}_{s,t}$-measurable and uses no rewards stored at level s. Under Assumption 2, the rewards in these disjoint sets therefore retain the stated conditional product Bernoulli law. □

The same argument applies after a restart. If $\mathcal{H}_t = \sigma(W, U_1, \dots, U_{t-1})$ and $\{\tau = t\} \in \mathcal{H}_t$, the uniforms from τ onward are independent conditionally on the history at τ. This follows by conditioning on each event $\{\tau = t\}$ and using independence of $U_t, \dots, U_T$. We use this fact in Section F.

I.2 Concentration Inequalities

Throughout this subsection, ε_i are independent centered random variables supported in intervals of length one. The statements also hold conditionally on a sigma-field that fixes their coefficients and under which these assumptions hold.

Lemma 16 (Hoeffding and Bernstein inequalities)**.** *For fixed real a_i and $z > 0$,*

$$\mathbb{P}\left(\left|\sum_i a_i\varepsilon_i\right| > \sqrt{z\sum_i a_i^2/2}\right) \le 2e^{-z}.$$

If independent centered variables Y_i satisfy $|Y_i| \le c$ and $\sum_i \mathrm{Var}(Y_i) \le v$, then $\mathbb{P}(|\sum_i Y_i| > \sqrt{2vz} + cz/3) \le 2e^{-z}$.

Lemma 17 (Hsu et al. (2012, Theorem 2.1))**.** *For $\varepsilon = (\varepsilon_i)_i$, $N \succeq 0$, and $z \ge 0$, with probability at least $1 - e^{-z}$,*

$$\varepsilon^\top N\varepsilon \le \tfrac{1}{4}(\mathrm{tr}\, N + 2\sqrt{z\,\mathrm{tr}(N^2)} + 2z\|N\|_{\mathrm{op}}) \le \tfrac{3}{4}(\mathrm{tr}\, N + z\|N\|_{\mathrm{op}}).$$

The sub-Gaussian assumption in the cited theorem follows from $\mathbb{E}[\exp(\langle a, \varepsilon\rangle)] \le \exp(\|a\|_2^2/8)$. The second inequality uses $\mathrm{tr}(N^2) \le \|N\|_{\mathrm{op}}\,\mathrm{tr}\, N$ and $2\sqrt{z\|N\|_{\mathrm{op}}\,\mathrm{tr}\, N} \le \mathrm{tr}\, N + z\|N\|_{\mathrm{op}}$.

I.3 Polynomials of Bernoulli Variables

Lemma 18 (O'Donnell (2014, Theorem 10.21))**.** *Let $R_i \sim \mathrm{Bern}(p_i)$ be independent, with $p_i \in [p_0, 1 - p_0]$. For a polynomial f of degree at most m in $(R_i - p_i)_i$ and integers $q \ge 2$,*

$$\|f\|_{L^q} \le (4q/p_0)^{m/2}\|f\|_{L^2}.$$

Consequently, for $z \ge 1$, $\mathbb{P}(|f| > e(8z/p_0)^{m/2}\|f\|_{L^2}) \le e^{-z}$. The same statements hold for a polynomial with values in a finite-dimensional Hilbert space, with $\|f\|_{L^q} = (\mathbb{E}[\|f\|^q])^{1/q}$.

The cited hypercontractive inequality tensorizes over independent coordinates and allows different p_i. It gives $\|f\|_{L^q} \le ((q-1)^{1/2}p_0^{1/q-1/2})^m\|f\|_{L^2}$, which is smaller than the stated bound. Every polynomial can be made multilinear using $(R_i - p_i)^2 = (1 - 2p_i)(R_i - p_i) + p_i(1 - p_i)$. For the tail statement, choose $q = \max\{2, \lceil z\rceil\} \le 2z$. Markov's inequality gives

$$\mathbb{P}\left(|f| > e(8z/p_0)^{m/2}\|f\|_{L^2}\right) \le e^{-q}\left(\frac{q}{2z}\right)^{mq/2} \le e^{-z}.$$

For Hilbert-valued $f = (f_j)_j$, Minkowski's inequality gives $\|\sum_j f_j^2\|_{L^{q/2}} \le \sum_j \|f_j\|_{L^q}^2$. Applying the scalar inequality to each coordinate proves the Hilbert-valued statement.

Use Of Generative-AI Tools

GPT-5.6 Sol assisted the authors in identifying a secant-based approximation as an improvement over the tangent approximation used in one-step Newton estimation. It also assisted with manuscript polishing and reference exploration. The authors take responsibility for the final arguments and presentation.