Shuffle and Joint Differential Privacy for Generalized Linear Contextual Bandits

Sahasrajit Sarmasarkar
Stanford University
sahasras@stanford.edu

Abstract

We present the first algorithms for generalized linear contextual bandits under shuffle differential privacy and joint differential privacy. While prior work on private contextual bandits has been restricted to linear reward models—which admit closed-form estimators—generalized linear models (GLMs) pose fundamental new challenges: no closed-form estimator exists, requiring private convex optimization; privacy must be tracked across multiple evolving design matrices; distributional optimal design needs to be privatized and optimization error must be explicitly incorporated into regret analysis.

We address these challenges under two privacy models and context settings. For stochastic contexts, we design a shuffle-DP algorithm using a privately learned distributional design. Its leading regret term is $\widetilde{O}\big(d\sqrt{T\log T/\kappa^*} + \sqrt{T/(\kappa^*\varepsilon)}\,\big(d\sqrt{\log T} + d^{3/4}(\log T)^{3/4}\big)\big)$, with a lower-order $T^{1/3}$ term. For adversarial contexts, we provide a joint-DP algorithm with regret $\tilde{O}\big(d\sqrt{T}\log T + d^{3/4}\sqrt{T/\varepsilon}\,(\log T)\,(d+\log T)^{1/4}\big)$. Both of these bounds match their corresponding non-private rate in the leading term. Unlike prior work on locally private GLM bandits, our methods require no spectral assumptions on the context distribution beyond ℓ_2 boundedness.

1 Introduction

In the contextual bandit problem [Abbasi-yadkori et al., 2011, Chu et al., 2011], at each time step t, the learner observes a set of K arms $\mathcal{X}_t$, each represented by a d-dimensional feature vector. The learner selects an arm $x_t \in \mathcal{X}_t$ and receives a reward r_t. The goal is to minimize cumulative regret over T rounds. This problem has been extensively studied for linear models [Auer, 2003, Abbasi-yadkori et al., 2011], logistic models [Faury et al., 2020, 2022], and generalized linear models [Filippi et al., 2010, Li et al., 2017, Sawarni et al., 2024].

Given the wide applicability of contextual bandits in settings involving sensitive personal data—such as news recommendation [Li et al., 2010], healthcare [Varatharajah and Berry, 2022], and education [Lin et al., 2023]—protecting both context and reward information from adversaries is paramount. This motivates the study of *differentially private* contextual bandits.

In this paper, we study GLMs with DP constraints under two setups. In **M1**, the context set $\mathcal{X}_t$ is drawn i.i.d. from an unknown distribution $\mathcal{D}$, and we study (ε, δ)-shuffle DP. In **M2**, the context sets may be chosen adversarially, and we study (ε, δ)-joint DP. Under shuffle DP, each user transmits noisy messages to a shuffler that randomly permutes all messages before the learner receives them. Joint DP requires that the actions at all time steps other than t be differentially private with respect to the context-reward pair at time step t. Our main contributions are as follows.

- **First GLM bandits under shuffle/joint DP.** Prior work on private GLM bandits [Han et al., 2021] considers only the local-DP setting and relies on stronger spectral assumptions, which

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

also appear in its regret bounds. In contrast, we provide the first shuffle-DP and joint-DP results for GLM bandits under only ℓ_2 boundedness.

- **Shuffle DP for stochastic contexts (M1).** Using a privately learned distributional design, our shuffle-DP algorithm achieves a leading regret term of $\widetilde{O}\big(d\sqrt{T\log T/\kappa^*} + \sqrt{T/(\kappa^*\varepsilon)}\,\big(d\sqrt{\log T} + d^{3/4}(\log T)^{3/4}\big)\big)$; the lower-order term is stated in Theorem 5.1.
- **Joint DP for adversarial contexts (M2).** We give a joint-DP algorithm with regret $\tilde{O}\big(d\sqrt{T}\log T + d^{3/4}\sqrt{T/\varepsilon}\,(\log T)\,(d+\log T)^{1/4}\big)$ (Theorem 6.1), matching the non-private rate $\tilde{O}(d\sqrt{T}\log T)$ in the leading term; privacy contributes only an additive correction. Such additive privacy-induced regret penalties have also been observed in linear contextual bandits by Azize and Basu [2024b].
- **κ-free leading term.** Our regret bounds inherit the κ-free leading $\sqrt{T}$ term from the non-private framework of Sawarni et al. [2024]; preserving this under privacy is not automatic and requires care in how noise enters the design matrix. The parameter κ can grow as e^S for logistic rewards, where S upper-bounds $\|\theta^*\|$.

Remark 1.1 (Anytime extension). While our algorithms require the horizon T to be known, they can be made anytime via the standard doubling trick [Besson and Kaufmann, 2018], incurring only a constant-factor increase in regret.

1.1 Technical Overview

Extending private contextual bandits from linear to generalized linear models introduces four fundamental obstacles, each requiring a new technique rather than off-the-shelf composition. We sketch them here; full treatment is in Sections 5 and 6.

Challenge 1: Privatizing Distributional optimal design Distributional optimal design (DOD), introduced by Ruan et al. [2021] for linear bandits and extended to generalized linear bandits by Sawarni et al. [2024], attains $\tilde{O}(d\sqrt{T}\log T)$ regret, improving by a factor of $\sqrt{d}$ over what G-optimal design [Kiefer and Wolfowitz, 1960] can guarantee. Privatizing DOD is difficult, however. Its construction [Ruan et al., 2021, Algorithm 4] sub-selects a data-dependent core of the context sets, which requires repeated, adaptive access to the raw user contexts and makes privacy accounting under shuffle-DP hard. We instead propose a simplified construction that splits the users into predetermined phases and computes the DOD policy from the design matrix released at the end of each phase (see Appendix E). Because each user belongs to exactly one phase, privacy needs no composition across phases.

Challenge 2: Lack of closed form estimator Unlike linear bandits, generalized linear bandits have no closed-form ridge estimator, so privacy cannot be obtained by noising sufficient statistics. Instead, we privatize the iterative MLE computation using the shuffle-private convex optimizer of Cheu et al. [2022]. We show that the resulting optimization error ν enters the confidence radius only as $O(\sqrt{\nu})$ (Lemma I.4), so the number of iterations can be traded off against utility.

Challenge 3: Non-monotone design matrices under continual release. In the adversarial setting (M2), the binary-tree mechanism [Chan et al., 2011, Dwork et al., 2010] is needed to release design matrices at every round, but its noise makes the released $\{\mathbf{H}_t\}$ non-monotone, breaking the standard determinant-doubling rarely-switching argument [Abbasi-yadkori et al., 2011]. We replace it with a *directional growth criterion* (line 6 of Algorithm 2) which on noiseless monotone matrices implies determinant doubling; Lemma S.5 shows it preserves the $O(d\log T)$ switch bound. In Appendix S.2, we additionally analyze a variant that uses the *determinant-doubling switch* ($\det(\mathbf{H}_t) \geq 2\det(\mathbf{H}_\tau)$) in place of the directional growth criterion; this variant incurs an extra $d^{1/4}$ factor in the private regret term (Theorem S.8).

Challenge 4: Privacy leakage through data-dependent switching. In Algorithm 2, whether round t explores depends on $\mathcal{X}_t$ and on $\mathbf{V}_t$, so the index set $\mathcal{T}_o$ is data-dependent; a user's own data also decides which of the two binary trees and optimizer sequences receives it, and if estimator recomputations were triggered by exploration decisions, their timing would reveal those decisions. We therefore (i) trigger every estimator recomputation only from released matrices (directional growth of $\mathbf{V}_t$ and $\mathbf{H}_t$), and (ii) charge each tree and optimizer sequence for a change of any single user's data,

composing them adaptively in zCDP. The switching set $\mathcal{T}_o \setminus \{t\}$ is then a post-processing of the tree releases (Section 6.2).

2 Related Work

Non-private GLM bandits. Generalized linear contextual bandits have been studied extensively in the non-private setting [Filippi et al., 2010, Li et al., 2017, Sawarni et al., 2024]. The key challenge is that, unlike linear models, GLMs lack closed-form estimators for $\widehat{\theta}$, requiring iterative optimization. Sawarni et al. [2024] recently achieved order-optimal regret with limited adaptivity, which inspires our algorithmic framework.

Private linear bandits. Differential privacy in linear contextual bandits has been studied under various trust models. Under joint DP, Shariff and Sheffet [2018] first achieved sublinear regret using tree-based mechanisms; Azize and Basu [2024a] improved bounds under relaxed assumptions. Under shuffle DP, Chowdhury and Zhou [2022], Garcelon et al. [2022], Tenenbaum et al. [2023] obtained amplified privacy guarantees. All these works exploit closed-form ridge estimators for linear models.

Private GLM bandits. The only prior work on private GLM bandits, Han et al. [2021], considers the local DP setting under restrictive spectral assumptions: their regret bound scales inversely with a constant κ_l ([Han et al., 2021, Assumption 3]) which, as shown in [Han et al., 2021, Appendix A], is proportional to the minimum eigenvalue of the context covariance for Gaussian contexts—so the regret depends inversely on this minimum eigenvalue. In contrast, our results remove any such spectral dependence and achieve $\widetilde{O}(\sqrt{T/\epsilon})$ regret under both shuffle and joint DP assuming only ℓ_2-bounded contexts.

Comparison with prior techniques. We highlight the key technical differences from prior private bandit algorithms that necessitate our new techniques:

1. Sawarni et al. [2024] provide the non-private baseline we build upon. Their key insight—using policy switches based on design matrix determinants to remove κ from regret—does not directly transfer to the private setting because (a) the switch times leak information, and (b) noisy matrices are non-monotone. Additionally, we also need to privatize the construction of distributional optimal design (DOD) itself which requires new techniques.
2. Shariff and Sheffet [2018], Azize and Basu [2024a] study joint-DP linear bandits using tree-based mechanisms for the ridge estimator. Their analysis exploits the closed-form solution $\widehat{\theta} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{r}$, where adding noise to $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{X}^\top \mathbf{r}$ suffices. For GLMs, we must instead privatize the optimization process itself.

See Appendix D for extended discussion of related work on privacy amplification by shuffling and limited adaptivity in bandits.

3 Privacy Models

We consider two privacy models for contextual bandits, depending on the context generation process with formal definitions in Appendix C.

Shuffle DP (stochastic contexts). In the shuffle model [Erlingsson et al., 2019, Cheu et al., 2019], each user applies a local randomizer to their data and sends noisy messages to a trusted shuffler, which randomly permutes all messages before forwarding them to the learner. This amplifies privacy: even weak local randomization yields strong central-DP-level guarantees after shuffling. We adopt the multi-round shuffle protocol of Cheu et al. [2022], Chowdhury and Zhou [2022], where each user submits multiple randomized messages to the shuffler and the protocol proceeds over multiple rounds. In our setup, each time step t corresponds to a distinct user, who sends a noisy function of their (context, reward) pair to the shuffler.

Joint DP (adversarial contexts). Joint DP [Kearns et al., 2014, Shariff and Sheffet, 2018] is a relaxation appropriate when contexts are adversarial. It requires that, for every t, the actions $(a_s)_{s \neq t}$

Paper	Model	DP Type	Contexts	Regret
Shariff and Sheffet [2018]	Linear	Joint	Adv.	$\widetilde{O}(\sqrt{T/\varepsilon})$
Chowdhury and Zhou [2022]	Linear	Shuffle	Adv.	$\widetilde{O}(T^{3/5}/\sqrt{\varepsilon})$
Tenenbaum et al. [2023]	Linear	Shuffle*	Adv.	$\widetilde{O}(T^{1/2}/\sqrt{\varepsilon})$
Han et al. [2021]†	GLM	Local	Stoch.	$\widetilde{O}(\sqrt{T}/\varepsilon)^{\dagger}$
Ours (Algorithm 1)	**GLM**	**Shuffle**	**Stoch.**	$\widetilde{O}(\sqrt{T/\varepsilon})$
Ours (Algorithm 2)	**GLM**	**Joint**	**Adv.**	$\widetilde{O}(\sqrt{T/\varepsilon})$

Table 1: Comparison with prior work on private contextual bandits. All prior shuffle/joint DP results rely on closed-form estimators available only for linear models.

†Requires spectral assumptions on context distribution. * Uses multiple concurrent shufflers

recommended to all other users be differentially private with respect to the context-reward pair $(\mathcal{X}_t, r_t)$ of user t. This avoids the impossibility result of [Shariff and Sheffet, 2018], which shows that standard central DP incurs linear regret in contextual bandits.

Comparison with prior work. Table 1 summarizes how our results relate to prior work on private contextual bandits. All prior shuffle/joint DP results assume linear models with closed-form ridge estimators. Prior GLM work [Han et al., 2021] uses only local DP and requires spectral assumptions.

4 Notation and Preliminaries

A policy π is a function that maps any given arm set $\mathcal{X}$ to a probability distribution over the same set, i.e., $\pi(\mathcal{X}) \in \Delta(\mathcal{X})$, where $\Delta(\mathcal{X})$ is the probability simplex supported on $\mathcal{X}$. We denote matrices in bold upper case (e.g., $\mathbf{M}$). $\|x\|$ denotes the ℓ_2 norm of vector x. We write $\|x\|_{\mathbf{M}}$ to denote $\sqrt{x^\top \mathbf{M} x}$ for a positive semi-definite matrix $\mathbf{M}$ and vector x. For any two real numbers a and b, we denote by $a \wedge b$ the minimum of a and b. Throughout, $\widetilde{O}(\cdot)$ denotes big-O notation but suppresses factors in $\mathrm{poly}(R, S, \log d, \log(1/\delta), \log\log T)$. [1] For $n \in \mathbb{N}$, we denote the set $\{1, \ldots, n\}$ by $[n]$, and for $m < n$ we denote $\{m, \ldots, n\}$ by $[m, n]$. We use $A \succeq B$ to denote that $A - B$ is positive semi-definite.

Definition 4.1. A generalized linear model (GLM) with parameter vector $\theta^* \in \mathbb{R}^d$ is a real valued random variable r that belongs to the exponential family of distributions with density function

$$P(r \mid x) = \exp(r \cdot \langle x, \theta^* \rangle - b\left(\langle x, \theta^* \rangle\right) + c(r)).$$

Function $b(\cdot)$ (called the log-partition function) is assumed to be twice differentiable and $\dot{b}$ is assumed to be monotone. Further, we assume $r \in [0, R]$ and $\|\theta^*\| \leq S$.[2] Standard properties of GLMs include $\mathbb{E}[r] = \dot{b}(\langle x, \theta^* \rangle)$ and $\mathrm{Var}(r) = \ddot{b}(\langle x, \theta^* \rangle)$. We define the link function as $\mu(z) := \dot{b}(z)$. This formulation is standard in the contextual bandit literature [Filippi et al., 2010, Sawarni et al., 2024]. Next we describe the two contextual bandit problems with GLM rewards that we address in this paper. Let $T \in \mathbb{N}$ be the total number of rounds. At each round $t \in [T]$, we receive an arm set $\mathcal{X}_t \subset \mathbb{R}^d$ with $K = |\mathcal{X}_t|$ arms, and must select an arm $x_t \in \mathcal{X}_t$. Following this selection, we receive a reward r_t sampled from the GLM distribution $P(r|x)$ with unknown parameter θ^*.

Problem 1 (M1). In this setting, we assume that at each round t, the context set $\mathcal{X}_t \subseteq \mathbb{R}^d$ is drawn from an unknown distribution $\mathcal{D}$. The algorithm decides the lengths of each batch a priori, where the policy remains fixed within each batch and updates only at the start of each batch. We study (ϵ, δ)-shuffle differential privacy in this setup. The goal is to minimize the expected cumulative regret:

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \mu(\langle x, \theta^* \rangle) - \sum_{t=1}^{T} \mu(\langle x_t, \theta^* \rangle)\right], \tag{1}$$

where the expectation is over the randomness of the algorithm, the distribution of rewards r_t, and the distribution of arm sets $\mathcal{D}$.

[1] While we show multiplicative factors of $\log T$ in the dominant term, we drop additive terms that are $\mathrm{poly}(\log T)$ in T, even when they carry other multiplicative factors such as κ, d, or $1/\varepsilon$.

[2] We assume that R and S are lower bounded by a constant.

Problem 2 (M2). In this setting, we make no assumptions on the context sets $\mathcal{X}_t$—they may be chosen adversarially. The goal is to minimize, with high probability, the cumulative regret while satisfying (ϵ, δ)-joint differential privacy:

$$R_T = \sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \mu(\langle x, \theta^* \rangle) - \sum_{t=1}^{T} \mu(\langle x_t, \theta^* \rangle). \tag{2}$$

For both problems, the Maximum Likelihood Estimator (MLE) of θ^* is obtained by minimizing the cumulative log-loss. For an arm x, reward r, and parameter $\theta \in \mathbb{R}^d$, the log-loss is $\ell(\theta, x, r) := -r \cdot \langle x, \theta \rangle + \int_0^{\langle x, \theta \rangle} \mu(z)\, dz$. This loss is convex and $O(R + RS)$-Lipschitz over $\{\theta : \|\theta\| \leq S\}$. After t rounds, we define the regularized MLE objective as $\mathcal{L}_t(\theta; \lambda) := \sum_{s=1}^{t} \ell(\theta, x_s, r_s) + \frac{\lambda}{2}\|\theta\|_2^2$.

4.1 Instance-dependent non-linear parameters

Following Faury et al. [2020, 2022], Sawarni et al. [2024], we define instance-dependent parameters that capture the non-linearity of the link function. The parameter κ can be exponentially large in S—for example, $\kappa = \Theta(e^S)$ in logistic regression. For any arm set $\mathcal{X}$, let $x^* = \arg\max_{x \in \mathcal{X}} \mu(\langle x, \theta^* \rangle)$ denote the optimal arm. In setup **M1**, define:

$$\kappa := \max_{\mathcal{X} \in \mathrm{supp}(\mathcal{D})} \max_{x \in \mathcal{X}} \frac{1}{\dot{\mu}(\langle x, \theta^* \rangle)}, \qquad \frac{1}{\kappa^*} := \max_{\mathcal{X} \in \mathrm{supp}(\mathcal{D})} \max_{x \in \mathcal{X}} \dot{\mu}(\langle x, \theta^* \rangle). \tag{3}$$

In setup **M2**, we redefine these parameters as:

$$\kappa := \max_{x \in \cup_{t \in [T]} \mathcal{X}_t} \frac{1}{\dot{\mu}(\langle x, \theta^* \rangle)}, \qquad \frac{1}{\kappa^*} := \max_{x \in \cup_{t \in [T]} \mathcal{X}_t} \dot{\mu}(\langle x, \theta^* \rangle). \tag{4}$$

Unlike κ, which can grow exponentially in S, the parameter κ^* is benign: in general $\kappa^* \geq 1/R^2$ (since $\dot{\mu}(\langle x, \theta^* \rangle) = \mathrm{Var}(r \mid x) \leq R^2$ for rewards bounded by R), independent of d and S. For specific link functions one often has a sharper bound: e.g., $\kappa^* \geq 4$ for logistic rewards (since $\sup_z \dot{\mu}(z) = 1/4$). Similar to the non-private bounds in Sawarni et al. [2024], our regret bounds have no κ dependence in the dominant $\widetilde{O}(\sqrt{T})$ term. In practice, it suffices to have access to an upper bound on κ and a lower bound on κ^* as the true values may not be known a priori. The table of symbols used throughout the paper is given at the start of the appendix (Appendix A).

Assumption 4.2 (Boundedness)**.** The reward r_t at each time step t is bounded by a known constant R. Every context vector has ℓ_2 norm bounded by 1, i.e., $\|x\| \leq 1$ for all $x \in \mathcal{X}_t$ and $t \in [T]$. The unknown parameter satisfies $\|\theta^*\| \leq S$ for a known constant S. We further assume $R, S \geq 1$.[3]

5 Shuffle-DP GLM Bandits with Stochastic Contexts

We present our algorithm for the stochastic-context setting (**M1**). The first batch samples from the G-optimal design $\pi_G(\mathcal{X})$ of each context set $\mathcal{X}$ (Lemma E.1). In each later batch, the algorithm privately learns a distributional optimal-design (DOD) policy π_k from design users (Algorithm 3) and applies it to subsequent estimation users. It privately aggregates the resulting statistics via the shuffle model. Our design adapts the batched framework of Sawarni et al. [2024], with key modifications for privacy:

- **Private distributional optimal design**: We partition the design contexts D_k into $L + 1 = 4d + 1$ phases with boundaries fixed in advance. At the end of each phase ℓ, we compute a design matrix W_ℓ and use it to obtain the arg-max policy π_{DOD}. Unlike Ruan et al. [2021], which switches phases adaptively, this keeps privacy accounting tractable.
- **Private covariance estimation.** Design matrices $\mathbf{V}$ and $\mathbf{H}_k$ are computed via shuffled vector summation [Cheu et al., 2019], which aggregates noisy outer products across users.
- **Private parameter estimation.** The parameter estimate $\hat{\theta}_k$ at each batch is computed using a shuffle-private convex optimizer [Cheu et al., 2022] on the regularized negative log-likelihood.

[3] Any absolute constant lower bound would suffice.

Batch structure. For $M \geq 2$, before observing any users, partition the horizon into consecutive batches $\mathcal{B}_1, \ldots, \mathcal{B}_M$, with $n_k := |\mathcal{B}_k|$. The warm-up batch $\mathcal{B}_1$ has length n_1 and uses π_G throughout. For $k \geq 2$, each batch is ordered as $\mathcal{B}_k = D_k \cup E_k$, where the design group D_k precedes the estimation group E_k and $|D_k| = |E_k| = n_k/2$, and $\sum_{k=1}^{M} n_k \leq T$. Following Sawarni et al. [2024], set $\alpha := T^{1/(2(1-2^{-M+1}))}$ if $M \leq \log_2 \log_2 T$ and $\alpha := 2\sqrt{T}$ otherwise, and set

$$n_1 = \left\lceil \left(\frac{\sqrt{\kappa} e^{3S}}{S} \alpha R d^2 \gamma^2 \right)^{2/3} \right\rceil, n_2 = \alpha, n_k = \alpha \sqrt{n_{k-1}} \; (k \geq 3), \tag{5}$$

with γ the confidence radius in (6). We take $M = \lceil \log_2 \log_2 T \rceil + 1$, so $\alpha = 2\sqrt{T}$. The batch in which the horizon is reached is truncated so that $\sum_k n_k \leq T$, and any later batches are empty; this may happen before batch M. The algorithm uses $O(\log \log T)$ batches (Lemma I.8).

Privacy-aware parameters. The DOD step uses $L = 4d$ design phases and a regularizer Γ that absorbs the shuffle noise accumulated over them; the estimator regularizer λ adds the covariance-noise margin to Γ. Under the paper's convention that $\widetilde{O}$ suppresses factors in $\mathrm{poly}(R, S, \log d, \log(1/\delta), \log \log T)$, but keeps multiplicative $\log T$ factors in the dominant term visible and ignores only additive $\mathrm{poly}(\log T)$ terms (exact forms are in Appendix G.1),

$$\Gamma = \widetilde{O}((d + \log T)(1 + 1/\varepsilon)) \, , \lambda = \widetilde{O}(d \log T + (d + \log T)/\varepsilon) \, , \gamma = \widetilde{O}(\sqrt{\lambda} + \sqrt{\nu}). \tag{6}$$

Here $\sigma = \widetilde{O}(1/\varepsilon)$ and $\nu = \widetilde{O}(\sqrt{d}(\log T)^{3/2}/\varepsilon)$. The parameter σ is the sub-Gaussian noise of the shuffled vector-summation protocol and ν the optimization error of the shuffle-private convex optimizer (Appendix G.2). Further, we define upper/lower bounds as $\lambda_{\pm} = \lambda \pm \tilde{\Theta}\sigma(\sqrt{d} + \sqrt{\log T})$ [4]

Warm-up and estimation. In the warm-up batch $\mathcal{B}_1$, every user samples an arm from π_G to build the initial estimate. These users send shuffled covariance and GLM messages; in particular, $\mathbf{V} \approx \sum_{t \in \mathcal{B}_1} x_t x_t^\top + \lambda \mathbf{I}$. The parameter estimate $\hat{\theta}_1$ is computed by the shuffle-private convex optimizer on their regularized negative log-likelihood. The protocols are detailed in Appendices W and X.

Arm scaling and elimination. For batches $k \geq 2$, arms are scaled before selection (see Remark 5.2). For optimal design guarantees on scaled arms, see Appendix F:

$$\tilde{\mathcal{X}}_t := \left\{ \sqrt{\frac{\dot{\mu}(\langle x, \hat{\theta}_1 \rangle)}{\beta(x)}} \, x : x \in \mathcal{X}_t \right\}, \beta(x) := \exp\Big(R \min\big(2S, \gamma\sqrt{\kappa} \, \|x\|_{\mathbf{V}^{-1}}\big) \Big), \tag{7}$$

Arms with $\mathrm{UCB}_j(x) < \max_y \mathrm{LCB}_j(y)$ are eliminated for every $j \in [k-1]$. The elimination rule and scaling map are fixed before D_k begins. The design users construct π_k from their scaled, surviving arm sets, and the estimation users apply that policy to fresh sets obtained by the same transformation. The weighted design matrix is then computed as $\mathbf{H}_k \approx \sum_{t \in E_k} \frac{\dot{\mu}(\langle x_t, \hat{\theta}_1 \rangle)}{\beta(x_t)} x_t x_t^\top + \lambda \mathbf{I}$ using the shuffled vector summation protocol. A detailed description is given in Appendix W.

Confidence bounds. Upper and lower confidence bounds are defined as

$$\mathrm{UCB}_k(x) := \langle x, \hat{\theta}_k \rangle + \gamma_k \|x\|_{\mathbf{V}_k^{-1}}, \qquad \mathrm{LCB}_k(x) := \langle x, \hat{\theta}_k \rangle - \gamma_k \|x\|_{\mathbf{V}_k^{-1}}, \tag{8}$$

where $\mathbf{V}_1 = \mathbf{V}$, $\gamma_1 = \gamma\sqrt{\kappa}$, and $\mathbf{V}_k = \mathbf{H}_k$, $\gamma_k = \gamma$ for $k \geq 2$. The factor $\sqrt{\kappa}$ appears only in the warm-up bound: for $k \geq 2$ the curvature is already built into $\mathbf{H}_k$ (Remark 5.2). These bounds follow Sawarni et al. [2024] with γ adjusted for privacy noise under our noisy private estimates.

Private DOD. The design users D_k construct π_k with a shuffle-private version of DOD (Algorithm 3). Before observing any contexts, D_k is split into $L + 1$ disjoint phases $D_{k,0}, \ldots, D_{k,L}$, with $|D_{k,0}| = n_k/4$ and $|D_{k,\ell}| = n_k/(4L)$ for $\ell \geq 1$. In phase 0, user t samples $z_{k,0} \sim \pi_G(\tilde{\mathcal{X}}_t)$

[4] We choose λ_- and λ_+ such that the perturbations Z_k satisfy $\lambda_- \boldsymbol{I} \preceq Z_k + \lambda I \preceq \lambda_+ \boldsymbol{I}$ (see Assumption G.1 for a precise definition).

and reports $z_{k,0}z_{k,0}^\top$. For $\ell \geq 1$, a user in phase ℓ receives the published matrix $W_{k,\ell-1}$, selects $z_{k,\ell} = \arg\max_{x \in \tilde{\mathcal{X}}_t} x^\top W_{k,\ell-1}^{-1} x$, and reports $z_{k,\ell}z_{k,\ell}^\top$. In either case, the user plays the original arm corresponding to the selected scaled arm and discards its reward. Each phase is aggregated by the shuffled vector-summation protocol (Appendix W) so that

$$W_{k,\ell} \approx \Gamma I + \sum_{j=0}^{\ell} \sum_{t \in D_{k,j}} z_{k,j}(\tilde{\mathcal{X}}_t) z_{k,j}(\tilde{\mathcal{X}}_t)^\top$$

The published policy is $\pi_k(\mathcal{X}) = \pi_{\mathrm{DOD}}(\mathcal{X}) := \frac{1}{2}\pi_G(\mathcal{X}) + \frac{1}{2L}\sum_{\ell=1}^{L} \delta_{z_{k,\ell}(\mathcal{X})}$ (δ_z: point mass at z), which attains distributional G-variation (defined in Ruan et al., 2021, Definition 1) $\mathcal{V}_{\mathcal{D}}^{(\varrho_k)}(\pi_{\mathrm{DOD}}) = O(d \log d)$ for regularizer $\varrho_k := 2\Gamma/n_k$[5]; see Definition E.3 and Appendix E. Because phase boundaries are fixed and each phase uses fresh users, the privacy guarantee follows naturally.

The public transcript consists solely of the shuffle outputs $\mathbf{V}$ and $\{\mathbf{H}_k\}_{k=2}^M$, the parameter estimates $\{\hat{\theta}_k\}_{k=1}^M$, and the shuffle-private DOD policies $\{\pi_k\}_{k=2}^M$. It contains no raw contexts, actions, or rewards. In each batch, the algorithm uses only the public transcript from preceding batches.

Protocol instantiation. We use the shuffle convex optimizer $\mathcal{P}_{\mathrm{GD}}$ and vector summation protocol (R_2, S_2, A_2) of Cheu et al. [2022, 2019]. In batch k, the optimizer is a multi-round shuffle protocol running for $T_{\mathrm{PGD}} = \widetilde{\Theta}(\varepsilon^2 N_k^2/d)$ iterations, where $N_1 = n_1$ and $N_k = n_k/2$ for $k \geq 2$, with step size $\eta = D/(L\sqrt{T_{\mathrm{PGD}}}) = 1/(R\sqrt{T_{\mathrm{PGD}}})$ for $D = 2S$ and gradients clipped to $L = 2RS$ (Appendix I). The summation protocol privately aggregates outer products from $\mathcal{B}_1$ and, in later batches, E_k via shuffled bit summation. The resulting algorithm is (ε, δ)-shuffle DP; since shuffle DP implies joint DP via the Billboard lemma [Kearns et al., 2014], the same guarantee holds in the joint-DP sense as well. Full details are in Appendix I.

For each warm-up user and each later estimation user, the privacy budget splits as $(\varepsilon/2, \delta/2)$ for covariance estimation and $(\varepsilon/2, \delta/2)$ for parameter estimation. Each design user in a later batch sends one context-only DOD report with budget (ε, δ). These groups are disjoint, so parallel composition gives (ε, δ) shuffle DP. Elimination, scaling, and publication of π_k use postprocessing of earlier private outputs. Here R_2 and A_2 are the randomizer and analyzer of Appendix W; $\mathcal{P}_{\mathrm{GD}}$ is the optimizer in Appendix X.

Theorem 5.1 (Regret bound for Algorithm 1; full proof in Appendix G)**.** *For $T \geq d$, under the conditions of Theorem J.1, Algorithm 1 satisfies (ε, δ)-shuffle DP. With $L = \Theta(d)$, its regret is*

$$\widetilde{O}\Bigg(\frac{d\sqrt{T \log T}}{\sqrt{\kappa^*}} + \frac{\sqrt{T}\big(d\sqrt{\log T} + d^{3/4}(\log T)^{3/4}\big)}{\sqrt{\kappa^* \varepsilon}} + e^{2S}\kappa^{1/3} d^2 T^{1/3} (\log T)^{2/3} + \frac{e^{2S}\kappa^{1/3} T^{1/3}\big(d^2 (\log T)^{2/3} + d^{5/3} \log T\big)}{\varepsilon^{2/3}} \Bigg).$$

We note two points about Theorem 5.1.

1. The non-private leading $\sqrt{T}$ term, $d\sqrt{T \log T}/\sqrt{\kappa^*}$, matches the rate of Sawarni et al. [2024] up to logarithmic factors, and privacy adds $\sqrt{T}\big(d\sqrt{\log T} + d^{3/4}(\log T)^{3/4}\big)/\sqrt{\kappa^* \varepsilon}$. The parameters κ and e^{2S} appear only in the $T^{1/3}$ terms (Remark 5.2).
2. Using policy π_G in every batch would give design width d^2/n_k and an extra $\sqrt{d}$ in the leading regret term; the shuffle-private DOD achieves $d \log d/n_k$ and avoids it (Appendix E and Lemma F.1).

Remark 5.2 (Removing κ from the leading term). The estimator $\hat{\theta}_k$ itself is computed by an ordinary (shuffle-private) convex optimizer on the standard GLM negative log-likelihood—no κ-aware reweighting of the loss. Instead, κ is absorbed into the *confidence-set geometry*: from batch $k \geq 2$ onward we build confidence ellipsoids against the weighted design matrix $\mathbf{H}_k$ (see Lemma I.4), rather than the naive design matrix $\sum_s x_s x_s^\top$. In Lemma I.7, the design matrix $\mathbf{H}_k$ is bounded by the Hessian of the regularized MLE loss $\mathcal{L}(\theta^*)$, confining the κ dependence to a lower-order term.

[5] ΓI is added to a sum of $n_k/2$ contributions, whereas Definition E.3 averages them.

[6] The bound $\Delta = 4R^2$ holds unconditionally since $\beta(x_t) = \exp\big(R \min(2S, \gamma\sqrt{\kappa}\|x_t\|_{\mathbf{V}^{-1}})\big) \geq 1$ and $\dot{\mu}(\langle x_t, \hat{\theta}_1 \rangle) = \mathrm{Var}(r \mid x_t; \hat{\theta}_1) \leq 4R^2$ for bounded rewards; hence $\frac{\dot{\mu}(\langle x_t, \hat{\theta}_1 \rangle)}{\beta(x_t)} \leq 4R^2$.

Algorithm 1 SHUFFLEGLM: Shuffle-DP GLM Bandits

Input: Batches M, horizon T, privacy (ε, δ), estimator regularizer λ, DOD regularizer Γ, and confidence radius γ from (6).

1: Initialize batch lengths $n_1, \ldots, n_M$ via (5).
2: **for** $t \in \mathcal{B}_1$ **do**
3: Observe arm set $\mathcal{X}_t$, sample $x_t \sim \pi_G(\mathcal{X}_t)$, and observe reward r_t.
4: LOCALRANDOMIZER: send $M_t \leftarrow R_2(x_t x_t^\top,\ \Delta{=}1,\ \varepsilon/2,\ \delta/2)$.
5: **end for**
6: CENTRAL ANALYSER: $\mathbf{V} \leftarrow A_2(\text{Shuffle}(\{M_t\}_{t\in\mathcal{B}_1}),\ \varepsilon/2,\ \delta/2) + \lambda\mathbf{I}$.
7: PRIVATEOPTIMIZER: $\hat{\theta}_1 \leftarrow \mathcal{P}_{\text{GD}}\big(\{(r_t, x_t)\}_{t\in\mathcal{B}_1};\ \varepsilon/2,\ \delta/2,\ \ell(\cdot) + \frac{\lambda_+}{2n_1}\|\theta\|^2\big)$.
8: **for** $k = 2$ to M **do**
9: **for** $t \in D_k$ **do**
10: Observe arm set $\mathcal{X}_t$; set $\widetilde{\mathcal{X}}_t \leftarrow$ ELIMINATEANDSCALE$(\mathcal{X}_t, k)$.
11: User t participates in shuffle private DOD construction of Algorithm 3 using $\widetilde{\mathcal{X}}_t$ with regularizer Γ.
12: **end for**
13: At the end of D_k, publish the resulting policy π_k.
14: **for** $t \in E_k$ **do**
15: Observe arm set $\mathcal{X}_t$; set $\widetilde{\mathcal{X}}_t \leftarrow$ ELIMINATEANDSCALE$(\mathcal{X}_t, k)$.
16: Sample $\widetilde{x}_t \sim \pi_k(\widetilde{\mathcal{X}}_t)$, play the corresponding x_t, and observe reward r_t.
17: LOCALRANDOMIZER: $M_t \leftarrow R_2\Big(\frac{\dot{\mu}(\langle x_t, \hat{\theta}_1\rangle)}{\beta(x_t)} x_t x_t^\top,\ \Delta{=}4R^2,\ \varepsilon/2,\ \delta/2\Big)$.[6]
18: **end for**
19: CENTRAL ANALYSER: $\mathbf{H}_k \leftarrow A_2(\text{Shuffle}(\{M_t\}_{t\in E_k}),\ \varepsilon/2,\ \delta/2) + \lambda\mathbf{I}$.
20: PRIVATEOPTIMIZER: $\hat{\theta}_k \leftarrow \mathcal{P}_{\text{GD}}\big(\{(r_t, x_t)\}_{t\in E_k};\ \varepsilon/2,\ \delta/2,\ \ell(\cdot) + \frac{\lambda_+}{n_k}\|\theta\|^2\big)$.
21: **end for**

Subroutine ELIMINATEANDSCALE$(\mathcal{X}, k)$

22: **for** $j = 1, \ldots, k-1$ **do** $\mathcal{X} \leftarrow \{x \in \mathcal{X} : \text{UCB}_j(x) \geq \max_{y\in\mathcal{X}} \text{LCB}_j(y)\}$.
23: **return** $\widetilde{\mathcal{X}} \leftarrow \{\widetilde{x} : x \in \mathcal{X}\}$, where $\widetilde{x} := \sqrt{\dot{\mu}(\langle x, \hat{\theta}_1\rangle)/\beta(x)}\, x$.

6 Joint-DP GLM Bandits with Adversarial Contexts

We present Algorithm 2 for setting **M2**: GLM bandits with adversarial contexts under (ε, δ)-joint differential privacy. Our design adapts the rarely-switching framework of Sawarni et al. [2024], but faces key technical challenges under privacy, each requiring a new ingredient:

- **Private design matrix release.** The design matrices $\mathbf{V}$ (for Criterion I rounds, line 9) and $\mathbf{H}_t$ (for Criterion II rounds, line 6) must be released at every round. We use the binary-tree mechanism [Chan et al., 2011, Dwork et al., 2010] to achieve this under continual observation.
- **Switching analysis under noise.** Binary-tree noise makes $\{\mathbf{H}_t\}$ non-monotone, breaking the standard determinant-doubling argument; we replace it with a directional growth criterion at line 6 (PSD comparison against the last-update matrix). The data-dependent switching set $\mathcal{T}_o$ could also leak information; we show it is a post-processing of the tree releases (Section 6.2).

The algorithm has two switching criteria. *Criterion I* triggers when some arm has high uncertainty under $\mathbf{V}$: we select the most uncertain arm and update $\mathbf{V}$; the coarse estimate $\hat{\theta}_o$ is recomputed only when the released $\mathbf{V}_t \not\preceq 2\mathbf{V}_{\tau_o}$, a public trigger. Otherwise, we use a scaled design matrix $\mathbf{H}_t$ incorporating the link derivative to remove κ from the dominant regret term. *Criterion II* triggers a policy update when $\mathbf{H}_t \not\preceq 2\mathbf{H}_\tau$ (directional growth criterion). In Appendix S.2, we additionally analyze a variant that replaces the directional growth criterion in Criterion II with a standard determinant-doubling switch ($\det(\mathbf{H}_t) \geq 2\det(\mathbf{H}_\tau)$); this variant incurs an extra $d^{1/4}$ factor in the private regret term (Theorem S.8).

Optimizer and loss details. Lemma T.2 bounds (with high probability) the number of exploration rounds by $N_o := 16dR^2\kappa\gamma^2 \log T$ (used only in the regret analysis), the number of updates of $\hat{\theta}_o$ by $\text{count1}_{\text{cutoff}} := 4d\log_2(1 + T/d)$, and the number of policy switches by $\text{count2}_{\text{cutoff}} := 4d\log_2(1 +$

Algorithm 2 Joint-DP Generalized Linear Bandits with Adversarial Contexts

Input: Privacy (ε, δ), horizon T, error probability ζ, parameters λ, γ, β as in (10)
Initialize: $\mathbf{V}, \mathbf{H}_1 \leftarrow \lambda \boldsymbol{I}$; $\mathcal{T}_o \leftarrow \emptyset$; $\tau, \tau_o \leftarrow 0$; binary tree mechanisms $\mathcal{B}_V, \mathcal{B}_H$

for $t \leftarrow 1$ **to** T **do**
  Observe $\mathcal{X}_t$; query $\mathbf{V}_t, \mathbf{H}_t$ from $\mathcal{B}_V, \mathcal{B}_H$
  if $\mathbf{V}_t \not\preceq 2\mathbf{V}_{\tau_o}$ and fewer than $\text{count1}_{\text{cutoff}}$ updates of $\hat{\theta}_o$ so far **then**
    $\tau_o \leftarrow t$; compute $\hat{\theta}_o$ via DP-PGD optimizer on $\mathcal{T}_o$ data
  end if
  if $\mathbf{H}_t \not\preceq 2\mathbf{H}_\tau$ and fewer than $\text{count2}_{\text{cutoff}}$ updates of $\hat{\theta}_\tau$ so far **then** {**Crit. II**}
    $\tau \leftarrow t$; compute $\hat{\theta}_\tau$ via DP-PGD optimizer on $[t-1] \setminus \mathcal{T}_o$ data
  end if
  if $\max_{x \in \mathcal{X}_t} \|x\|^2_{\mathbf{V}^{-1}} \geq \frac{1}{2\gamma^2 \kappa R^2}$ **then** {**Crit. I**}
    Play $x_t = \arg\max_{x \in \mathcal{X}_t} \|x\|_{\mathbf{V}^{-1}}$, observe r_t, $\mathcal{T}_o \leftarrow \mathcal{T}_o \cup \{t\}$
    PRIVATEUPDATEEXPLORE($\mathbf{V}, \mathbf{H}_t, x_t$)
  else
    $\mathcal{X}_t \leftarrow \{x \in \mathcal{X}_t \mid \text{UCB}_o(x) \geq \max_z \text{LCB}_o(z)\}$ {Arm elimination via (9)}
    Play $x_t = \arg\max_{x \in \mathcal{X}_t} \text{UCB}(x, \mathbf{H}_\tau, \hat{\theta}_\tau)$, observe r_t
    PRIVATEUPDATEEXPLOIT($\mathbf{V}, \mathbf{H}_t, x_t, \hat{\theta}_o$)
  end if
end for

Privacy Mechanisms

PRIVATEUPDATEEXPLORE($\mathbf{V}, \mathbf{H}, x$): $\mathbf{V} \leftarrow \mathbf{V} + xx^\top + \mathcal{R}^V$ via $\mathcal{B}_V$; $\mathbf{H} \leftarrow \mathbf{H} + \mathcal{R}^H$ via $\mathcal{B}_H$

PRIVATEUPDATEEXPLOIT($\mathbf{V}, \mathbf{H}, x, \hat{\theta}$): $\mathbf{H} \leftarrow \mathbf{H} + \frac{\dot{\mu}(\langle x, \hat{\theta} \rangle)}{e} xx^\top + \mathcal{R}^H$ via $\mathcal{B}_H$; $\mathbf{V} \leftarrow \mathbf{V} + \mathcal{R}^V$ via $\mathcal{B}_V$

$TR^3/d)$. The DP-PGD optimizer (Lemma X.5) minimizes the *summed* (not averaged) regularized loss $\mathcal{L}(\theta; \lambda_+)$ over $\Theta_I := \{\theta : \|\theta\|_2 \leq S\}$ (Criterion I) and $\Theta_{II} := \{\theta : \|\theta - \hat{\theta}_o\|_{\mathbf{V}_{\tau_o}} \leq \gamma\sqrt{\kappa}$ and $||\theta||_2 \leq S\}$ (Criterion II), with zero-Concentrated DP [Bun and Steinke, 2016] budget $\rho := \rho_{\text{tot}}/(4c)$ per call for $c \in \{\text{count1}_{\text{cutoff}}, \text{count2}_{\text{cutoff}}\}$, where $\rho_{\text{tot}} := \left(\sqrt{\log(1/\delta) + \varepsilon} - \sqrt{\log(1/\delta)}\right)^2$ is the zCDP budget that converts to exactly (ε, δ)-DP.[7] Since which users explored is private, the size of each optimizer's dataset must not influence the mechanism. We therefore minimize the summed loss, whose per-step gradient sensitivity is $2L_0$ (for L_0-Lipschitz per-sample losses) regardless of the dataset size, and set the noise level, step size and number of iterations from the public bound $t - 1$ on that size (Lemma X.5).

Binary-tree mechanism. To release $\mathbf{V}$ and $\mathbf{H}_t$ at every round, we use the binary-tree mechanism [Chan et al., 2011, Dwork et al., 2010]. Each node of a binary tree over $[T]$ stores a partial sum plus independent Gaussian noise; each data point enters $\log T$ nodes and each prefix sum is a sum of at most $\log T$ nodes, with noise calibrated so that each tree is $\rho_{\text{tot}}/4$-zCDP with respect to any single user's leaf. $\mathcal{R}^V, \mathcal{R}^H$ denote the implied per-round noise (Appendix L.3).

6.1 Confidence Bounds and Parameters

The confidence bounds used in arm elimination and selection are:

$$\text{UCB}_o/\text{LCB}_o(x) := \langle x, \hat{\theta}_o \rangle \pm \gamma\sqrt{2\kappa}\, \|x\|_{\mathbf{V}^{-1}}, \qquad \text{UCB}(x, \mathbf{H}, \theta) := \langle x, \theta \rangle + \beta\, \|x\|_{\mathbf{H}^{-1}}. \tag{9}$$

The algorithm uses an estimator regularizer λ and two confidence-width scalings β (Criterion II) and γ (Criterion I), scaling as Equation (10). Exact closed-form expressions are deferred to Appendix L.1; they account for the binary-tree noise, the privacy cost of switching, and the DP-PGD optimizer error.

$$\lambda = \widetilde{O}\big(d \log T + \tfrac{\sqrt{d} \log T}{\varepsilon} \sqrt{\max\{d, \log T\}}\big),\ \beta = \widetilde{O}\Big(\sqrt{d \log T} + \tfrac{d^{1/4} (\log T)^{1/2} (d + \log T)^{1/4}}{\sqrt{\varepsilon}}\Big)$$
$$\gamma = \widetilde{O}\Big(\sqrt{d \log T} + \tfrac{d^{1/4} (\log T)^{1/2} (d + \log T)^{1/4}}{\sqrt{\varepsilon}}\Big). \tag{10}$$

[7] Algorithm 2 enforces $\text{count1}_{\text{cutoff}}, \text{count2}_{\text{cutoff}}$ as hard caps on the (publicly triggered) estimator updates, so the number of optimizer calls is bounded for any data; on the high-probability event the caps never bind, so regret is unaffected.

6.2 Privacy of Switching Times

A user's own data decides whether it explores, and hence which tree and optimizer sequence receive it, so we charge each of them for a change of *any* single user's data (including whether that user's leaf is zero). The set $\mathcal{T}_o$ is never released; for $s \neq t$, whether s explores depends only on the released $\mathbf{V}_s$ and on $\mathcal{X}_s$, and the estimators' update times depend only on released matrices, so switching reveals nothing beyond the tree releases (Appendix T). Adaptive zCDP composition of the two trees and two optimizer sequences ($\rho_{\mathrm{tot}}/4$ each) gives (ε, δ)-DP for the transcript, and the billboard lemma gives JDP (Appendix L.3).

6.3 Main Result

Theorem 6.1 (Joint-DP regret bound; full proof in Appendix L)**.** *Algorithm 2 satisfies (ε, δ)-joint differential privacy. Moreover, if $\zeta < \delta \leq \min\left(1/e, \frac{d}{R^3}\right)$, $d \leq T$, and $\varepsilon \leq 8\log(1/\delta)$,*[8] *then with probability at least $1-\zeta$, its regret is bounded by*

$$R_T \leq \widetilde{O}\left(\sqrt{T}\left(\frac{d\log T}{\sqrt{\kappa^*}} + \frac{d^{3/4}\log T(\log T + d)^{1/4}}{\sqrt{\varepsilon\kappa^*}}\right)\right).$$

Comparison with non-private bound. Our bound matches the non-private rate $\widetilde{O}(d\sqrt{T/\kappa^*}\,\log T)$ of Sawarni et al. [2024] in the leading-in-T term (optimal up to logs for logistic bandits [Abeille et al., 2021, Thm. 2]), with privacy contributing only an additive correction. Neither $\sqrt{T}$ term carries κ or e^S; κ enters only via additive $\mathrm{poly}(\log T)$ terms from the exploration rounds $\mathcal{T}_o$, suppressed by $\widetilde{O}$ (Section 4; see Theorem L.8).

Remark 6.2 (Removing κ from the leading term). As in Section 5, κ is absorbed into the weighted design matrix $\mathbf{H}_t$ (Lemmas O.1 and O.4), so non-exploration rounds are κ-free; the $|\mathcal{T}_o| \leq 16dR^2\kappa\gamma^2\log T$ exploration rounds (Lemma T.2) give the sub-leading κ-dependent terms.

Remark 6.3 (Gap in privacy-dependent term). Our privacy term $\widetilde{O}(d^{3/4}\sqrt{T/\varepsilon}\,(\log T)\,(d+\log T)^{1/4})$ exceeds the $\Omega(d/(\varepsilon+\delta))$ lower bound for generalized linear contextual bandits (Appendix U).[9] The same gap in joint-DP *linear* bandits [Azize and Basu, 2024b] is resolved only for stochastic contexts [Chen et al., 2025].

7 Simulations

We evaluate Algorithm 2 on probit bandits with $K = 20$, $d = 3$, $T = 5{,}000$ (10-run average), varying $S \in \{2, 2.5, 3\}$ to sweep κ. θ^* is sampled uniformly on the radius-S sphere and arms from the unit ball. Table 2 shows the κ-dependent baselines (GLM-UCB, GLOC) degrade sharply as κ grows, whereas Private-GLM and the non-private RS-GLinUCB stay stable; the Private-GLM and RS-GLinUCB gap shrinks with ε. See Appendix B for additional experiments.

Algorithm	$\kappa \approx 18.39$	$\kappa \approx 56.44$	$\kappa \approx 222.15$
GLM-UCB [Filippi et al., 2010]	1400.27 ± 1.94	1943.23 ± 1.81	2323.06 ± 1.21
GLOC [Jun et al., 2017]	384.68 ± 3.12	631.60 ± 4.29	1346.59 ± 8.62
Private-GLM (ours, $\varepsilon = 4$)	461.77 ± 12.72	543.17 ± 12.96	712.86 ± 16.15
Private-GLM (ours, $\varepsilon = 6$)	348.84 ± 7.35	380.23 ± 7.72	522.69 ± 7.80
Private-GLM (ours, $\varepsilon = 8$)	300.81 ± 6.95	336.37 ± 12.27	417.73 ± 9.54
RS-GLinUCB [Sawarni et al., 2024]	294.36 ± 98.85	150.69 ± 14.27	186.97 ± 8.14

Table 2: Cumulative regret on probit bandits; $\pm$ values are standard errors over 10 runs.

Conclusion. We give the first shuffle-DP and joint-DP algorithms for GLM contextual bandits with stochastic and adversarial contexts, respectively, whose leading $\sqrt{T}$ regret term is κ-free and matches the non-private counterparts.

[8] Privacy holds for every ε; this condition is only for utility (the optimizer errors ν_1, ν_2 in Lemma L.6). For larger ε, the bound holds with ε replaced by $\varepsilon\sqrt{\log(1/\delta)/(\log(1/\delta)+\varepsilon)} \approx \sqrt{\varepsilon\log(1/\delta)}$.

[9] Extending He et al. [2022, Prop. 4.3]; only local curvature bounds on $\dot{\mu}$ are needed. The lower bound is proved for JDP with respect to users' potential rewards, a slightly stronger notion [Azize and Basu, 2024a] that Algorithm 2 also satisfies.

Acknowledgements. The author thanks Ayush Sawarni for helpful technical discussions on batched algorithms in generalized linear bandits, Justin Whitehouse for valuable insights regarding likelihood analysis for differential privacy, and Mohak Goyal and Siddharth Chandak for their thoughtful suggestions on the exposition.

A previous version of this draft utilized G-optimal design for arm selection under the shuffle-private setting (Algorithm 1). The current shuffle-private distributional optimal design, constructed with the assistance of GPT-5.6 Sol, improves upon the previous formulation by reducing the regret bound by a factor of $\sqrt{d}$. Claude Opus 5.5 identified a bug in an earlier version of the joint differential privacy argument for Algorithm 2, and proposed the fix adopted here: updating the estimators $\hat{\theta}_o$ and $\hat{\theta}_\tau$ outside the Criterion I exploration step, only at publicly triggered rounds, which also makes the privacy accounting easier to track.

References

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/e1d5be1c7f2f456670de3d53c7b54f4a-Paper.pdf.

Marc Abeille, Louis Faury, and Clement Calauzenes. Instance-wise minimax-optimal algorithms for logistic bandits. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 3691–3699. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/abeille21a.html.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3 (null):397–422, March 2003. ISSN 1532-4435.

Achraf Azize and Debabrota Basu. Concentrated Differential Privacy for Bandits . In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 78–109, Los Alamitos, CA, USA, April 2024a. IEEE Computer Society. doi: 10.1109/SaTML59370.2024.00013. URL https://doi.ieeecomputersociety.org/10.1109/SaTML59370.2024.00013.

Achraf Azize and Debabrota Basu. Open problem: What is the complexity of joint differential privacy in linear contextual bandits? In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 5306–5311. PMLR, 30 Jun–03 Jul 2024b. URL https://proceedings.mlr.press/v247/azize24a.html.

Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Private summation in the multi-message shuffle model. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 657–676. ACM, 2020. doi: 10.1145/3372297.3417242.

Lilian Besson and Emilie Kaufmann. What doubling tricks can and can't do for multi-armed bandits. *arXiv preprint arXiv:1803.06971*, 2018. URL https://arxiv.org/abs/1803.06971.

Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Justin Rudominer, Uttam Kode, Marc Tinnes, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 441–459. ACM, 2017.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3–4):231–357, November 2015. ISSN 1935-8237. doi: 10.1561/2200000050. URL https://doi.org/10.1561/2200000050.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985*, page 635–658, Berlin, Heidelberg, 2016. Springer-Verlag. ISBN 9783662536407. doi: 10.1007/978-3-662-53641-4_24. URL https://doi.org/10.1007/978-3-662-53641-4_24.

T.-H. Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Trans. Inf. Syst. Secur.*, 14(3), November 2011. ISSN 1094-9224. doi: 10.1145/2043621.2043626. URL https://doi.org/10.1145/2043621.2043626.

Fan Chen, Jiachun Li, Alexander Rakhlin, and David Simchi-Levi. Beyond covariance matrix: The statistical complexity of private linear regression, 2025. URL https://arxiv.org/abs/2502.13115.

Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. *Distributed Differential Privacy via Shuffling*, page 375–403. Springer International Publishing, 2019. ISBN 9783030176532. doi: 10.1007/978-3-030-17653-2_13. URL http://dx.doi.org/10.1007/978-3-030-17653-2_13.

Albert Cheu, Matthew Joseph, Jieming Mao, and Binghui Peng. Shuffle private stochastic convex optimization. In *International Conference on Learning Representations (ICLR)*, 2022. Published: 28 Jan 2022.

Sayak Ray Chowdhury and Xingyu Zhou. Shuffle private linear contextual bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 3984–4009. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/chowdhury22a.html.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 208–214, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL https://proceedings.mlr.press/v15/chu11a.html.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, aug 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 715–724, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300506. doi: 10.1145/1806689.1806787. URL https://doi.org/10.1145/1806689.1806787.

Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2468–2479. SIAM, 2019.

Louis Faury, Marc Abeille, Clement Calauzenes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3052–3060. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/faury20a.html.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clement Calauzenes. Jointly efficient and optimal algorithms for logistic bandits. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 546–580. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/faury22a.html.

Vitaly Feldman, Audra McMillan, and Kunal Talwar. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 954–964. IEEE, 2021. doi: 10.1109/FOCS52979.2021.00096.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/c2626d850c80ea07e7511bbae4c76f4b-Paper.pdf.

Evrard Garcelon, Kamalika Chaudhuri, Vianney Perchet, and Matteo Pirotta. Privacy amplification via shuffling for linear contextual bandits. In Sanjoy Dasgupta and Nika Haghtalab, editors, *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pages 381–407. PMLR, 29 Mar–01 Apr 2022. URL https://proceedings.mlr.press/v167/garcelon22a.html.

Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Pure differentially private summation from anonymous messages. In *11th Innovations in Theoretical Computer Science Conference (ITC 2020)*, volume 151 of *LIPIcs*, pages 15:1–15:23. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. doi: 10.4230/LIPIcs.ITC.2020.15.

Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Amer Sinha. Differentially private aggregation in the shuffle model: Almost central accuracy in almost a single message. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3692–3701. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/ghazi21a.html.

Yanjun Han, Zhengqing Zhou, Zhengyuan Zhou, Jose Blanchet, Peter W. Glynn, and Yinyu Ye. Sequential batch learning in finite-action linear contextual bandits. *arXiv preprint arXiv:2004.06321*, 2020. URL https://arxiv.org/abs/2004.06321.

Yuxuan Han, Zhipeng Liang, Yang Wang, and Jiheng Zhang. Generalized linear bandits with local differential privacy. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Osama Hanna, Antonious M. Girgis, Christina Fragouli, and Suhas Diggavi. Differentially private stochastic linear bandits: (almost) for free. *IEEE Journal on Selected Areas in Information Theory*, 5:135–147, 2024. doi: 10.1109/JSAIT.2024.3389954.

Osama A Hanna, Lin Yang, and Christina Fragouli. Contexts can be cheap: Solving stochastic contextual bandits with linear bandit algorithms. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1791–1821. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/hanna23a.html.

Jiahao He, Jiheng Zhang, and Rachel Q. Zhang. A reduction from linear contextual bandit lower bounds to estimation lower bounds. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8660–8677. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/he22e.html.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: online computation and hashing. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 98–108, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 44:1–44:9, 2018.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2008. URL https://api.semanticscholar.org/CorpusID:1935.

Michael Kearns, Mallesh M. Pai, Aaron Roth, and Jonathan Ullman. Mechanism design in large games: Incentives and privacy. *American Economic Review*, 104(5):431–35, May 2014. doi: 10.1257/aer.104.5.431. URL https://www.aeaweb.org/articles?id=10.1257/aer.104.5.431.

J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4.

Jiachun Li, David Simchi-Levi, and Yining Wang. On the optimal regret of locally private linear contextual bandit. *arXiv preprint arXiv:2404.09413*, 2024. URL https://arxiv.org/abs/2404.09413.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 661–670, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605587998. doi: 10.1145/1772690.1772758. URL https://doi.org/10.1145/1772690.1772758.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2071–2080. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/li17c.html.

Chien-Chang Lin, Anna Y. Q. Huang, and Owen H. T. Lu. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments*, 10: 1–22, 2023. URL https://api.semanticscholar.org/CorpusID:261215251.

Andrew Lowy and Meisam Razaviyayn. Private federated learning without a trusted server: Optimal algorithms for convex losses. In *International Conference on Learning Representations (ICLR)*, 2023. Published as a conference paper at ICLR 2023.

Ilya Mironov. Rényi Differential Privacy . In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, Los Alamitos, CA, USA, August 2017. IEEE Computer Society. doi: 10.1109/CSF.2017.11. URL https://doi.ieeecomputersociety.org/10.1109/CSF.2017.11.

Saptarshi Roy, Sunrit Chakraborty, and Debabrota Basu. Fliphat: Joint differential privacy for high dimensional linear bandits. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan, editors, *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 2359–2367. PMLR, 03–05 May 2025. URL https://proceedings.mlr.press/v258/roy25b.html.

Yufei Ruan, Jiaqi Yang, and Yuan Zhou. Linear bandits with limited adaptivity and learning distributional optimal design. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 74–87, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380539. doi: 10.1145/3406325.3451004. URL https://doi.org/10.1145/3406325.3451004.

Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized linear bandits with limited adaptivity. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 8329–8369. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/0faa0019b0a8fcab8e6476bc43078e2e-Paper-Conference.pdf.

Roshan Shariff and Or Sheffet. Differentially private contextual linear bandits. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 4301–4311, Red Hook, NY, USA, 2018. Curran Associates Inc.

Jay Tenenbaum, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Differentially private multi-armed bandits in the shuffle model. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34,

pages 24956–24967. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d14388bb836687ff2b16b7bee6bab182-Paper.pdf.

Jay Tenenbaum, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Concurrent shuffle differential privacy under continual observation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 33961–33982. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/tenenbaum23a.html.

Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011. doi: 10.1214/ECP.v16-1624. URL https://doi.org/10.1214/ECP.v16-1624.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. ISSN 1935-8237. doi: 10.1561/2200000048. URL http://dx.doi.org/10.1561/2200000048.

Yogatheesan Varatharajah and Brent Berry. A contextual-bandit-based approach for informed decision-making in clinical trials. *Life*, 12(8), 2022. ISSN 2075-1729. doi: 10.3390/life12081277. URL https://www.mdpi.com/2075-1729/12/8/1277.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. ISBN 978-1-108-41519-4. doi: 10.1017/9781108231596. URL https://www.cambridge.org/core/books/highdimensional-probability/797C466DA29743D2C8213493BD2D2102.

Kai Zheng, Tianle Cai, Weiran Huang, Zhenguo Li, and Liwei Wang. Locally differentially private (contextual) bandits learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12300–12310. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/908c9a564a86426585b29f5335b619bc-Paper.pdf.

Appendix roadmap

The appendix is organized around the two privacy models—shuffle differential privacy and joint differential privacy (JDP)—that underlie our results:

- **Notation.** Appendix A collects the symbols used throughout the paper and proofs.
- **Synthetic experiments.** Appendix B presents synthetic experiments for Algorithm 2.
- **Privacy definitions.** Appendix C collects the formal definitions of local, shuffle, and joint differential privacy used throughout the paper, together with the convention that a user's data is their context set and observed reward.
- **Extended related work.** Appendix D discusses additional related work on privacy amplification by shuffling and limited-adaptivity bandits.
- **Optimal design.** Appendix E provides details on optimal experimental design for the generalized linear model.
- **Shuffle-private population design.** Appendix R proves the fresh-phase population-DOD guarantee, and Appendix F instantiates it after elimination and scaling in the stochastic-context GLM algorithm.
- **Shuffled privacy, stochastic contexts.** Appendices G to J present the full privacy and regret analysis of Algorithm 1.
- **Background material.** Standard facts on generalized linear models are collected in Appendix K.
- **Joint differential privacy, adversarial contexts.** A detailed JDP treatment appears in Appendices L to Q. Supporting lemmas and privacy proofs for policy-switching arguments are given in Appendices S and T.
- **Privacy lower bound.** Appendix U proves that every JDP algorithm for GLM contextual bandits incurs regret $\Omega\big(\frac{m_0}{\sqrt{M_0}} \cdot \frac{d}{\varepsilon+\delta}\big)$, extending He et al. [2022, Prop. 4.3], under a slightly stronger privacy notion that Algorithm 2 also satisfies.

- **Approximate optimization.** Appendix V bounds the error in the estimated parameter $\hat{\theta}$ arising from approximate convex minimization due to privacy constraints.
- **Shuffle-private primitives.** Appendices W and X summarize the shuffle-private vector-summation protocol and shuffle-private convex optimizer of Cheu et al. [2022].

A Notation summary

For convenience, we collect the symbols used throughout the paper and proofs. Generic notation introduced in passing in Section 4 ($\|x\|$, $\|x\|_M$, $\Delta(\mathcal{X})$, $\widetilde{O}(\cdot)$, $a \wedge b$, $[n]$, $[m, n]$) is not repeated here.

Table 3: Sets, GLM constants, and privacy parameters.

Symbol	Meaning	Defined in
Group 1: Sets, indices, horizon		
T, K, d	horizon, number of arms per round, feature dimension	Section 4
$\mathcal{X}_t$, $x_t \in \mathcal{X}_t$	context set / pulled arm at round t	Section 4
$\widetilde{\mathcal{X}}_t$	scaled arm set under arm reweighting	Equation (7)
$\mathcal{B}_k$, n_k, M	k-th batch, its size, total number of batches	Equation (5)
$\mathcal{T}_o$	set of exploration rounds (Criterion I)	Algorithm 2
$\mathcal{D}$	context distribution (setting M1)	Section 4
$\mathcal{K}_{\mathcal{X}}$	G-optimal design distribution on $\mathcal{X}$	Appendix E
π_G	local G-optimal sampling policy	Appendix E
Group 2: GLM and instance constants		
θ^*, $\|\theta^*\| \le S$	unknown true parameter; norm bound	Definition 4.1
$r_t \in [0, R]$	reward, with known bound R	Assumption 4.2
$b(\cdot)$, $\mu(\cdot) = \dot{b}(\cdot)$	log-partition function, link (mean) function	Definition 4.1
κ	$1/\min_{(\mathcal{X},x)} \dot{\mu}(\langle x, \theta^* \rangle)$; can be $\Theta(e^S)$	Equations (3) and (4)
$\hat{\kappa}$	$1/\mathbb{E}[\dot{\mu}(\langle x^*, \theta^* \rangle)]$ at the optimal arm	Equation (3)
κ^*	$1/\max_{(\mathcal{X},x)} \dot{\mu}(\langle x, \theta^* \rangle)$; benign, $\ge 1/R^2$	Equations (3) and (4)
Group 3: Privacy parameters		
ε, δ	target (ε, δ)-DP parameters	Section 3
$U_t = (\mathcal{X}_t, r_t)$	data of user t: context set and observed reward	Appendix C
ζ	failure probability used in JDP analysis	Theorem 6.1
σ	sub-Gaussian parameter of the matrix-summation noise	Assumption G.1
ν, ν_1, ν_2	cumulative-loss optimizer error	Assumptions G.2 and L.2
ρ	zCDP budget per private optimizer call	Section 6

B Synthetic Experiments

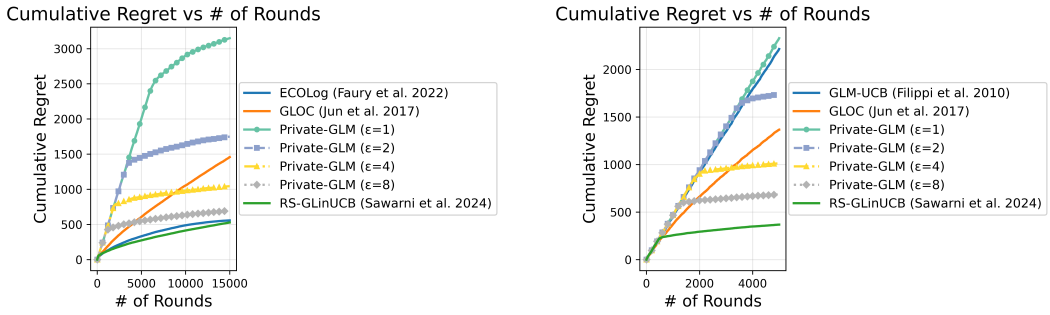
We evaluate the practicality of Algorithm 2 by comparing it against several non-private baselines for generalized linear bandits: **ECOLog** [Faury et al., 2022], **GLOC** [Jun et al., 2017], **GLM-UCB** [Filippi et al., 2010], and **RS-GLinUCB** [Sawarni et al., 2024].

ECOLog achieves computational efficiency with overall time complexity $\tilde{O}(T)$ and avoids κ-dependence in regret. However, it is designed specifically for logistic bandits and does not extend directly to general GLMs. GLOC is also computationally efficient, but its regret does depend on κ. GLM-UCB was the first algorithm proposed for generalized linear contextual bandits. Finally, RS-GLinUCB is both computationally efficient and achieves regret bounds without κ dependence for GLMs.

Table 4: Algorithm parameters, design matrices, and estimators.

Symbol	Meaning	Defined in
Group 4: Algorithm parameters (regularizer + confidence widths)		
λ	estimator regularizer	Equations (37) and (89)
λ_+, λ_-	regularizer bounds: high-prob. upper / lower bounds on the eigenvalues of the noisy regularizer	Equation (38)
γ	confidence-width scaling (Criterion I / shuffle batch)	Equations (39) and (91)
β	confidence-width scaling (Criterion II, JDP)	Equation (90)
$\beta(\lambda_+, \lambda_-)$	parameterized form used in proofs	Appendix O
$\beta(x)$	per-arm exponential reweighting factor	Equation (7)
η	step size of $\mathcal{P}_{\text{GD}}$	Appendix X
Group 5: Design matrices		
$\mathbf{V}$, V	noisy warm-up (and Criterion-I) design matrix	Algorithm 1,Algorithm 2
$\mathbf{H}_k$, H_k	noisy Hessian-weighted batch design matrix	Algorithm 1
H_t	noisy Hessian-weighted continual-release matrix (JDP)	Algorithm 2
H_k^*, H_o^*	true Hessian-weighted design matrix (used in proofs)	Appendices I and O
$\mathsf{H}_t^{(u),*}$, $\mathsf{H}_t^{(l),*}$	high-prob upper / lower envelope on H_t	Appendix O
$\mathcal{B}_V$, $\mathcal{B}_H$	binary-tree mechanisms releasing V, H_t	Algorithm 2
$\mathcal{R}_t^V$, $\mathcal{R}_t^H$	binary-tree noise increments at round t	Algorithm 2
Group 6: Estimators and confidence bounds		
$\hat{\theta}_k$	private MLE at the end of batch k (shuffle)	Algorithm 1
$\hat{\theta}_o$, $\hat{\theta}_\tau$	private MLE on $\mathcal{T}_o$ resp. on $[t-1] \setminus \mathcal{T}_o$ (JDP)	Algorithm 2
$\widetilde{\theta}$, $\widetilde{\theta}_o$, $\widetilde{\theta}_\tau$	non-private (unconstrained) MLE counterparts (proofs only)	Appendices I and O
Θ_I, Θ_{II}	constraint sets for the two DP-PGD calls	Section 6
$\text{UCB}_k(x)$, $\text{LCB}_k(x)$	confidence bounds at batch k (shuffle)	Equation (8)
UCB_o, LCB_o	exploration-time confidence bounds (JDP)	Equation (9)
$\text{UCB}(x, \mathbf{H}, \theta)$	generic Hessian-weighted UCB	Equation (9)
$\mathcal{P}_{\text{GD}}$	shuffle-private projected gradient descent optimizer	Algorithm 7
R_2, S_2, A_2	shuffle vector-summation triplet (randomizer / shuffler / analyzer)	Appendix W
Group 7: Privacy lower bound (local to Appendix U)		
r_0; m_0, M_0	local predictor range; $\inf$ / $\sup$ of $\dot{\mu}$ on $[-r_0, r_0]$	Appendix U
Θ_r^*, ν_r	product-of-circles parameter set of radius r; uniform prior on it	Appendix U
$U_t = (\mathcal{X}_t, (y_{t,a})_a)$	data of user t with potential rewards of all arms (Table JDP)	Definition U.1
$\overline{\mathcal{H}}_t$, $\mathcal{C}_t$	full-information history; past contexts	Appendix U
N_t^i	number of past rounds in which block i is active	Appendix U
φ	rotation angle in the two-point argument	Proposition U.2

For our simulations, we fix the privacy parameter $\delta = 2 \times 10^{-2}$ and vary $\varepsilon \in \{1, 2, 4, 8\}$. We consider both logistic and probit models with $\zeta = 2 \times 10^{-2}$.



(a) Logistic model

(b) Probit model

Figure 1: Cumulative regret under different generalized linear models. Private-GLM variants are plotted with distinct colors/markers.

Logistic model. We compare Private-GLM (Algorithm 2) against RS-GLinUCB, ECOLog, and GLOC, which are the only algorithms with time complexity $\tilde{O}(T)$ in Figure 1a. The setting is $d = 5$ dimensions and $K = 20$ arms. The parameter θ^* is sampled from a d-dimensional sphere of radius $S = 3.5$, and arms are sampled uniformly from the unit ball. We run $T = 15{,}000$ rounds, repeated 10 times. At $\varepsilon = 1$, the privacy budget is too tight and Private-GLM degrades substantially, as expected. While $\varepsilon = 2$ results in some performance degradation, the regret of Private-GLM at $\varepsilon = 8$ is nearly identical to RS-GLinUCB.

Probit model. We compare Private-GLM against RS-GLinUCB, GLM-UCB, and GLOC in Figure 1b. The setting is again $d = 5$ dimensions and $K = 20$ arms, with θ^* sampled from a sphere of radius $S = 3$ and arms drawn from the unit ball. We run $T = 5{,}000$ rounds, repeated 10 times. Similar to the logistic case, $\varepsilon = 1$ is too tight a privacy budget and Private-GLM degrades substantially. Regret at $\varepsilon = 2$ is somewhat higher, but at $\varepsilon = 8$ Private-GLM performs almost on par with RS-GLinUCB.

Remark B.1. For the simulations, we calibrate the per-step noise using Rényi-DP accounting [Mironov, 2017], a generalization of zero-CDP [Bun and Steinke, 2016] that gives tighter privacy accounting for iterative Gaussian mechanisms such as DP-SGD.

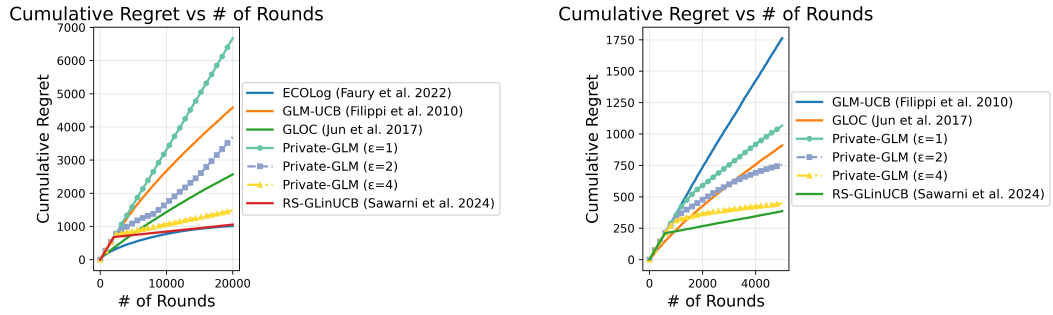
The constants λ, γ, and β are treated as hyperparameters rather than fixed as in the theoretical construction. The caps $\text{count1}_{\text{cutoff}}$ and $\text{count2}_{\text{cutoff}}$ on the number of updates of $\hat{\theta}_o$ and $\hat{\theta}_\tau$ are scaled by constants in practice. However, the algorithm stops updating $\hat{\theta}_o$ when $\text{count1}_{\text{cutoff}}$ is crossed and $\hat{\theta}_\tau$ when $\text{count2}_{\text{cutoff}}$ is crossed. This ensures that the overall privacy budget is never exceeded.

For our experiments, we set

$$\lambda = \frac{0.4\,\log(4/\delta)\,\log^2(4T/\zeta)}{\varepsilon\,\kappa^*} + \frac{0.75\,d\,\log(4T/\zeta)}{R^2},$$
$$\gamma = R\,S\,\sqrt{d\,\log(T/\zeta)},$$
$$\beta = 4\sqrt{d\,\log(T/\zeta)}.$$

B.1 Experiments on higher dimension $d \in \{20, 40\}$

Note that while the directional switching condition $\mathbf{H}_t \not\preceq 2\mathbf{H}_\tau$ yields a tighter regret upper bound, as discussed in Appendix S.2, it often performs poorly in practice, especially when the dimension d is large. This is because switching occurs more rarely, particularly for noiseless design matrices. The reason is that, assuming $\mathbf{H}_i \succeq \mathbf{H}_j$, the directional trigger $\mathbf{H}_i \not\preceq 2\mathbf{H}_j$ implies the determinant trigger $\det(\mathbf{H}_i) > 2\det(\mathbf{H}_j)$, but the converse does not hold in general: the determinant can double through moderate growth in many directions without any single direction doubling. As a result, the directional criterion triggers less often than the determinant doubling criterion, leading to fewer switches and higher regret. The issue worsens as the privacy parameter ε increases, since the design matrices grow more monotonically. To alleviate this, our simulations in higher dimensions d use the determinant doubling criterion.



(a) Logistic model

(b) Probit model

Figure 2: Cumulative regret in higher dimensions ($d = 40$ logistic, $d = 20$ probit). Private-GLM variants are plotted with distinct colors/markers.

We now perform experiments on both the logistic and probit models. For the logistic model, we use $d = K = 40$, $R = 1$, and $S = 5$ with horizon $T = 20{,}000$, giving $\kappa = 27.93$. For the probit model, we set $d = K = 20$ and $S = 3$ with horizon $T = 5{,}000$, giving $\kappa = 42.38$. The sampling of arms and θ^* is kept identical to previous experiments for low d. The plots have been shown in Figure 2.

The observations are similar to the previous subsection: our private algorithm Private-GLM performs better than algorithms with κ-dependent regret.

Compute. All experiments are low-compute and run on an Apple M1 chip within a few hours per run.

C Privacy Definitions

We first present the definition of central differential privacy (DP), followed by local, shuffle, and joint DP. Let $\mathcal{D}$ denote the data universe and $n \in \mathbb{N}$ the number of users. Let $D_i \in \mathcal{D}$ denote the data of user i, and $D_{-i} \in \mathcal{D}^{n-1}$ denote the data of all users except i. Let $\varepsilon > 0$ and $\delta \in (0, 1]$ be privacy parameters.

Definition C.1 (Differential Privacy [Dwork and Roth, 2014])**.** A mechanism $\mathcal{M}$ satisfies (ε, δ)-DP if, for each user $i \in [n]$, each pair of datasets $D, D' \in \mathcal{D}^n$ differing only in user i's data, and each event E in the range of $\mathcal{M}$,

$$\mathbb{P}\left[\mathcal{M}(D_i, D_{-i}) \in E\right] \leq e^{\varepsilon}\mathbb{P}\left[\mathcal{M}(D_i', D_{-i}) \in E\right] + \delta.$$

Definition C.2 (Local Differential Privacy [Kasiviswanathan et al., 2008])**.** A mechanism $\mathcal{M}$ satisfies (ε, δ)-LDP if for each user $i \in [n]$, each pair of data points $D_i, D_i' \in \mathcal{D}$, and each event E,

$$\mathbb{P}\left[\mathcal{M}(D_i) \in E\right] \leq e^{\varepsilon}\mathbb{P}\left[\mathcal{M}(D_i') \in E\right] + \delta.$$

A shuffle protocol $\mathcal{P} = (\mathcal{R}, \mathcal{S}, \mathcal{A})$ consists of: (i) a local randomizer $\mathcal{R}$, (ii) a shuffler $\mathcal{S}$, and (iii) an analyzer $\mathcal{A}$. Each user i applies $\mathcal{R}$ to their data D_i and sends the resulting messages to the shuffler, which uniformly permutes all messages before forwarding them to the analyzer. We consider a multi-round protocol [Cheu et al., 2022, Lowy and Razaviyayn, 2023] where users participate across I rounds.

Definition C.3 (Shuffle Differential Privacy [Erlingsson et al., 2019])**.** A multi-round protocol $\mathcal{P} = (\mathcal{R}, \mathcal{S}, \mathcal{A})$ for n users satisfies (ε, δ)-SDP if the mechanism $\mathcal{S} \circ \mathcal{R}^n = \left[\mathcal{S}(\mathcal{R}^{(i)}(D_1), \ldots, \mathcal{R}^{(i)}(D_n))\right]_{i=1}^{I}$ satisfies (ε, δ)-DP.

Recall from Section 4 that at each round t a user arrives with a context set $\mathcal{X}_t$ and, once an arm is selected, a reward r_t is observed. Throughout, the data of user t is therefore $U_t = (\mathcal{X}_t, r_t)$, its context set and observed reward.

Definition C.4 (Joint Differential Privacy [Kearns et al., 2014, Shariff and Sheffet, 2018])**.** A learning algorithm $\mathcal{M} : \mathcal{U}^T \to \mathcal{A}^T$ for contextual bandits is (ε, δ)-JDP if for all $t \in [T]$, all user sequences U_T, U_T' differing only in user t, and all action sets $A_{-t} \subseteq \mathcal{A}^{T-1}$,

$$\mathbb{P}\left[\mathcal{M}_{-t}(U_T) \in A_{-t}\right] \leq e^{\varepsilon}\mathbb{P}\left[\mathcal{M}_{-t}(U_T') \in A_{-t}\right] + \delta,$$

where $\mathcal{M}_{-t}$ denotes the actions given to all users except user t.

Joint DP requires that each user's action be private with respect to all other users' data. This is weaker than central DP but sufficient to avoid the linear regret lower bound of Shariff and Sheffet [2018]. By the Billboard Lemma [Kearns et al., 2014], shuffle DP implies joint DP.

D Extended Related Work

Privacy amplification by shuffling. The shuffle model bridges local and central DP, showing that randomly permuting locally privatized data achieves central-DP-level protection [Bittau et al., 2017, Erlingsson et al., 2019, Cheu et al., 2019, Feldman et al., 2021]. Multi-message protocols for binary summation and histograms demonstrate that sending a few short messages per user significantly tightens shuffling guarantees [Ghazi et al., 2020, Balle et al., 2020, Ghazi et al., 2021].

Limited adaptivity in bandits. Several works explore limited adaptivity in linear contextual bandits [Ruan et al., 2021, Abbasi-yadkori et al., 2011, Han et al., 2020, Hanna et al., 2023], motivated by settings where policies cannot change frequently, such as clinical trials and online advertising. Extensions to logistic bandits [Faury et al., 2022] and GLMs [Li et al., 2017, Sawarni et al., 2024] show that polylogarithmic policy switches suffice for optimal regret. Our algorithms build on the batching framework of Sawarni et al. [2024].

Shuffle DP convex optimization. Cheu et al. [2022] developed shuffle-DP algorithms for stochastic convex optimization, achieving near-central-DP accuracy. We adapt their techniques for the GLM estimation subroutine in our stochastic-context algorithm.

Detailed comparison with private linear bandits. Shuffle DP for linear contextual bandits has been studied in [Garcelon et al., 2022, Chowdhury and Zhou, 2022, Tenenbaum et al., 2023], and for K-armed bandits in [Tenenbaum et al., 2021]. By implementing $\log T$ concurrent shufflers in the tree-based model, Tenenbaum et al. [2023] tightened the regret bound to $\tilde{O}(\sqrt{T/\varepsilon})$ from the $\tilde{O}(T^{3/5}/\sqrt{\varepsilon})$ bound established by Chowdhury and Zhou [2022]. Joint DP for linear contextual bandits was first introduced in Shariff and Sheffet [2018], using binary tree mechanisms [Dwork et al., 2010] to update design matrix estimators at each step. Azize and Basu [2024a] improved regret bounds under relaxed assumptions where context sets are public and only rewards require privacy protection. Roy et al. [2025] further tightened regret bounds for sparse linear contextual bandits.

Local DP in contextual bandits has been studied in [Zheng et al., 2020], but their regret scales as $\tilde{O}(T^{3/4}/\varepsilon)$. More recently, Li et al. [2024] proposed an alternative approach that achieves $\tilde{O}(\sqrt{T}/\varepsilon)$ regret, though their method incurs exponential dependence on the dimension d.

Near-free privacy in linear bandits. In linear bandits (without contexts), Hanna et al. [2024] show that joint DP can be achieved almost for free, with only an additive logarithmic term involving ε. Azize and Basu [2024a] extend this to linear contextual bandits under relaxed assumptions—specifically, with public contexts and private rewards—and demonstrate that privacy comes at negligible cost. Roy et al. [2025] further show that in sparse linear contextual bandits, one can obtain nearly optimal joint DP guarantees when the sparsity parameter satisfies $s \ll d$. Azize and Basu [2024b] identifies the characterization of optimal regret for (ε, δ)-joint DP linear contextual bandits in the general case as an open problem.

E Optimal design

In the stochastic-context regime, batches are determined ahead of time. Previous work has leveraged G-optimal design to construct policies for both linear [Ruan et al., 2021] and generalized linear [Sawarni et al., 2024] reward settings. We restate the fundamental lemma, which follows from the General Equivalence Theorem [Kiefer and Wolfowitz, 1960].

Lemma E.1 (Ruan et al., 2021)**.** *For any subset $\mathcal{X} \subseteq \mathbb{R}^d$, there exists a distribution $\mathcal{K}(\mathcal{X})$ supported on $\mathcal{X}$ such that for any $\varsigma > 0$,*

$$\max_{x \in \mathcal{X}} x^\top \left(\varsigma \boldsymbol{I} + \mathop{\mathbb{E}}_{y \sim \mathcal{K}(\mathcal{X})} y y^\top \right)^{-1} x \leq d. \tag{11}$$

Furthermore, if $\mathcal{X}$ is finite, one can find a distribution achieving the relaxed bound $2d$ in time $\mathrm{poly}(|\mathcal{X}|)$.

Throughout the paper, $\pi_G(\mathcal{X})$ denotes the polynomial-time design from the second part of the lemma, and hence satisfies, for every $\varsigma > 0$,

$$\max_{x\in\mathcal{X}} x^\top \left(\varsigma I + \mathop{\mathbb{E}}_{y\sim\pi_G(\mathcal{X})}[yy^\top]\right)^{-1} x \le 2d. \tag{12}$$

Thus the factor-2 approximation is built into the notation π_G and will not be stated repeatedly.

For a context set $\mathcal{X}$, write

$$G_{\mathcal{X}} := \mathop{\mathbb{E}}_{y\sim\pi_G(\mathcal{X})}[yy^\top]. \tag{13}$$

We now state a corollary of G-optimal design (Lemma E.1) that shows the domination of G-optimal design matrix $G_{\mathcal{X}}$ on every arm $x \in \mathcal{X}$. This was shown in Ruan et al. [2021] and we reprove it here.

Corollary E.2 (Rank-one domination by the local G-design)**.** *For every $x \in \mathcal{X}$,*

$$xx^\top \preceq 2d\, G_{\mathcal{X}}. \tag{14}$$

Proof. Fix $a \in \mathbb{R}^d$ and $\eta > 0$. Cauchy–Schwarz in the inner product induced by $G_{\mathcal{X}} + \eta I$, followed by (12), gives

$$(a^\top x)^2 \le \big(a^\top (G_{\mathcal{X}} + \eta I)a\big)\big(x^\top (G_{\mathcal{X}} + \eta I)^{-1}x\big) \tag{15}$$

$$\le 2d\big(a^\top G_{\mathcal{X}} a + \eta\|a\|_2^2\big). \tag{16}$$

Letting $\eta \downarrow 0$ yields $(a^\top x)^2 \le 2d\, a^\top G_{\mathcal{X}} a$ for every $a \in \mathbb{R}^d$ proving the desired corollary. □

For stochastically generated context sets, we define λ-distributional G-variation [Ruan et al., 2021] to evaluate policies that sample an action from a context set. Optimizing this leads to distributional G-optimal design which removes a factor of $\sqrt{d}$ from the dominant term of the cumulative regret for linear and generalized linear bandits under stochastic contexts [Ruan et al., 2021, Sawarni et al., 2024].

Definition E.3 (λ-distributional G-variation Ruan et al. [2021])**.** Let $\mathcal{D}$ be a distribution over context sets in $\mathbb{R}^d$. For a policy π, define

$$\mathcal{I}_{\mathcal{D}}(\pi) := \mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}} \mathop{\mathbb{E}}_{x\sim\pi(\mathcal{X})}[xx^\top]. \tag{17}$$

For $\lambda > 0$, its λ-distributional G-variation is

$$\mathcal{V}_{\mathcal{D}}^{(\lambda)}(\pi) := \mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}}\left[\max_{x\in\mathcal{X}} x^\top \big(\lambda I + \mathcal{I}_{\mathcal{D}}(\pi)\big)^{-1} x\right], \tag{18}$$

and $\mathcal{V}_{\mathcal{D}}^{(0)}(\pi) := \lim_{\lambda\downarrow 0} \mathcal{V}_{\mathcal{D}}^{(\lambda)}(\pi)$ whenever the limit is finite.

In Ruan et al. [2021], the authors show that the local G-optimal policy π_G attains $\mathcal{V}_{\mathcal{D}}^{(0)}(\pi_G) = O(d^2)$ and construct a policy π_{DOD} for which

$$\mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}}\left[\max_{x\in\mathcal{X}} \sqrt{x^\top \big(\lambda I + \mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}})\big)^{-1} x}\right] = O(\sqrt{d\log d}) \tag{19}$$

for the relevant range of λ. This removes the additional $\sqrt{d}$ factor from the dominant regret term. We obtain a shuffle-private policy π_{DOD} with the stronger guarantee $\mathcal{V}_{\mathcal{D}}^{(\lambda)}(\pi_{\mathrm{DOD}}) = O(d\log d)$ at a suitable regularizer. Its construction and guarantees appear in Algorithm 3 and Theorems E.6 and E.7.

The non-private learner of Ruan et al. [2021, Algorithm 3] selects a data-dependent core of the context sets $\mathcal{X}_t$ and builds its design from that core (see Algorithm 4 therein). Computing the core requires repeated access to the raw context sets, which makes privacy accounting difficult. We avoid this by splitting the context sets into predetermined phases. At the end of each phase, the server publishes a

noisy design matrix (noise added for privacy) built from arms selected in the previous phases; each such arm is chosen by the argmax policy described below.

In Algorithm 3, the m independent context sets are split into $L+1$ predetermined phases. Phase 0 uses $m/2$ of them; each user draws an arm from its local G-design policy and reports the corresponding rank-one matrix. Each later phase r uses $m/(2L)$ fresh context sets. Given the published matrix W_{r-1}, a user in phase r selects the arm

$$z_r(\mathcal{X}_i) \in \arg\max_{x \in \mathcal{X}_i} x^\top W_{r-1}^{-1} x \tag{20}$$

with deterministic tie-breaking and sends a randomized report of $z_r(\mathcal{X}_i) z_r(\mathcal{X}_i)^\top$. After normalization, the shuffled aggregate estimates the conditional population covariance

$$M_r := \mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}} \left[z_r(\mathcal{X}) z_r(\mathcal{X})^\top \mid W_{r-1} \right]. \tag{21}$$

The server adds each aggregate to the regularized cumulative matrix W_r. Thus W_r reflects all phases through r, rather than the latest phase alone. The matrix-summation protocol is described in Appendix W.

After the final phase, the returned policy mixes the local G-design with the L max-variance rules:

$$\pi_{\mathrm{DOD}}(\mathcal{X}) = \frac{1}{2}\pi_G(\mathcal{X}) + \frac{1}{2L} \sum_{r=1}^{L} \delta_{z_r(\mathcal{X})}, \tag{22}$$

where δ_z denotes the point mass at z. The formal accuracy and privacy proofs appear in Appendix R.

Remark E.4 (Distinction from Ruan et al. [2021]). We use the arg-max policy of Ruan et al. [2021], but differ in two respects, in addition to avoiding core identification entirely. First, we never reuse context sets: each phase draws fresh ones. Second, our design matrices W_r are updated at predetermined phase boundaries, whereas Ruan et al. [2021] update only when the determinant of the covariance matrix reaches twice that of the current design matrix, a data-dependent rule.

Together, these choices make the context sets in phase r conditionally i.i.d. given the matrices published in earlier phases. This conditional independence is central to our analysis: it lets us bound the population design matrix by the empirical one in Lemma R.2. As a result, we can build the design directly from all sampled context sets, without identifying a core.

The calls to R_2, S_2, A_2 are precisely the matrix-summation randomizer, shuffler, and analyzer from Algorithms 4 to 6. If any of W_r, $\mathbf{V}$, or $\mathbf{H}_k$ is not positive definite, every rule that requires its inverse returns a feasible arm selected by a fixed deterministic convention.[10] We assume below that $2L$ divides m only to avoid integer rounding.

Remark E.5 (Choice of L). We take $L = 4d$ (any $L \geq 4d$ suffices). This ensures that the population increment H_r estimated by $\widetilde{H}_r = W_r - W_{r-1}$ is dominated, up to a constant factor, by the population matrix B_0 estimated by W_0: $H_r \preceq \frac{1}{2} B_0 \preceq \frac{1}{2} B_{r-1}$ (Lemma R.7). This domination is what the elliptical-potential lemma (Corollary S.2) requires in the proof of Theorem E.6.

Theorem E.6 (Accuracy of shuffle-private DOD)**.** *Suppose the context sets $\{\mathcal{X}_i\}_{i=1}^m$ are independently drawn from $\mathcal{D}$ with $\|x\|_2^2 \leq \Delta$ for every arm x in $\mathcal{X}_i$. Assume $0 < \varepsilon \leq 15$ and $0 < \delta < 1/2$ (the range of the matrix-summation protocol, Lemma W.2), and let $L \geq 4d$ with $2L$ dividing m. Let $\beta \in (0,1)$. Suppose the DOD matrix-summation noises satisfy the corresponding conditions in Assumption G.1 with sub-Gaussian noise parameter σ. It suffices to take*

$$\Gamma \geq 2\Delta \log \frac{4d(L+1)}{\beta} + 4\sigma\sqrt{L+1} \left(4\sqrt{d} + 2\sqrt{\log \frac{4(L+1)}{\beta}} \right). \tag{23}$$

Then, with probability at least $1-\beta$, the output of Algorithm 3 satisfies

$$\mathcal{V}_{\mathcal{D}}^{(\Gamma/m)}(\pi_{\mathrm{DOD}}) = O(d \log d). \tag{24}$$

In particular,

$$\mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}} \sqrt{\max_{x \in \mathcal{X}} x^\top \big((\Gamma/m) I + \mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}}) \big)^{-1} x} = O(\sqrt{d \log d}). \tag{25}$$

[10] On the high-probability events used in the analysis, all three matrices are positive definite, so this convention is never invoked. On the complementary event, its contribution to expected regret is covered by the uniform per-round regret bound.

Algorithm 3 Shuffle-private distributional optimal design

Input: Users with contexts $\{\mathcal{X}_i\}_{i=1}^m$, number of design phases $L = 4d$, privacy parameters ε, δ, sum-scale regularizer $\Gamma > 0$, context bound Δ, and the local design π_G.

Set $b \leftarrow m/(2L)$ and, before observing any contexts, partition the users into $D_0, D_1, \ldots, D_L$, where $|D_0| = Lb = m/2$ and $|D_r| = b$ for $r \geq 1$.
for each local user $i \in D_0$ **do**
    Observe $\mathcal{X}_i$ and sample $z_{i,0} \sim \pi_G(\mathcal{X}_i)$.
    Send $M_{i,0} \leftarrow R_2(z_{i,0}z_{i,0}^\top, \Delta; \varepsilon, \delta, |D_0|)$ to the shuffler.
end for
$Y_0 \leftarrow S_2(\{M_{i,0}\}_{i \in D_0})$ and $\widetilde{S}_0 \leftarrow A_2(Y_0)$.
$W_0 \leftarrow \Gamma I + \widetilde{S}_0$.
for $r = 1, \ldots, L$ **do**
    Publish W_{r-1} to the users in D_r.
    for each local user $i \in D_r$ **do**
        Observe $\mathcal{X}_i$ and compute, using deterministic tie-breaking,

$$z_{i,r} \leftarrow \arg\max_{x \in \mathcal{X}_i} x^\top W_{r-1}^{-1} x.$$

        Send $M_{i,r} \leftarrow R_2(z_{i,r}z_{i,r}^\top, \Delta; \varepsilon, \delta, b)$ to the shuffler.
    end for
    $Y_r \leftarrow S_2(\{M_{i,r}\}_{i \in D_r})$ and $\widetilde{H}_r \leftarrow A_2(Y_r)$.
    $W_r \leftarrow W_{r-1} + \widetilde{H}_r$.
end for
For a new context set $\mathcal{X}$, define $z_r(\mathcal{X}) \in \arg\max_{x \in \mathcal{X}} x^\top W_{r-1}^{-1} x$ for every $r \in [L]$.
return $\pi_{\mathrm{DOD}}(\mathcal{X}) = \dfrac{1}{2}\pi_G(\mathcal{X}) + \dfrac{1}{2L}\displaystyle\sum_{r=1}^{L} \delta_{z_r(\mathcal{X})}$.

The appearance of Γ/m is only a normalization: the algorithm adds ΓI to a sum of m design contributions, whereas (18) is defined using their population average. The proof is given in Appendix R.

Theorem E.7 (Privacy of shuffle-private DOD)**.** *Assume each invocation of the matrix-summation protocol is (ε, δ) shuffle differentially private for the users in that phase. Then the complete adaptive transcript of Algorithm 3, including all published matrices and the returned policy, is (ε, δ) shuffle differentially private with respect to the users.*

Although later phases are chosen adaptively from earlier public matrices, the phases contain disjoint users and their membership is fixed in advance. A changed user can therefore affect only one private shuffle release; every subsequent release and the returned policy are adaptive postprocessing of that release using fresh users. Hence parallel composition, rather than an $(L+1)$-fold sequential composition, gives the stated privacy parameters.

F Private DOD Guarantees after Elimination and Scaling

We now instantiate the general shuffle-private DOD construction from Appendix E inside the stochastic-context GLM algorithm. The only point requiring care is that the DOD algorithm is applied not to the original context sets, but to the surviving arms after elimination and the deterministic scaling used by the GLM design matrix.

F.1 The induced distribution of scaled survivor sets

Fix a batch $k \geq 2$ and condition on the public transcript available before its design users arrive. This transcript fixes all previous parameter estimates, the elimination rule, and the scaling. Let $\mathcal{X}^{(k)}$ denote the surviving subset of a fresh context set and let $\mathcal{D}_k$ be its conditional distribution. Let $\mathcal{D}_1$ be the original context-set distribution. In the warm-up batch all n_1 users sample from the local policy

π_G. For $k \geq 2$, recall the scaled arm

$$\widetilde{x} := \sqrt{\frac{\dot{\mu}(\langle x, \hat{\theta}_1 \rangle)}{\beta(x)}}\, x, \qquad x \in \mathcal{X}^{(k)}, k \geq 2, \tag{26}$$

where $\beta(x)$ and the arm scaling are defined in (7). The transformed survivor set is

$$\widetilde{\mathcal{X}}^{(k)} := \{\widetilde{x} : x \in \mathcal{X}^{(k)}\}, \tag{27}$$

and $\widetilde{\mathcal{D}}_k$ denotes its conditional distribution for $k \geq 2$. Thus $\widetilde{\mathcal{D}}_k$ is simply the pushforward of $\mathcal{D}_k$ under the fixed map in (27). Conditional on the public transcript, fresh users still provide independent draws from $\widetilde{\mathcal{D}}_k$, exactly as required by Theorem E.6.

Width. For a positive-semidefinite matrix A and $j \geq 1$, define the *width* of A under $\mathcal{D}_j$ by

$$\omega_j(A) := \mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}_j} \max_{x \in \mathcal{X}} \|x\|_A^2, \qquad \widetilde{\omega}_j(A) := \mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}_j} \max_{x \in \mathcal{X}} \|\widetilde{x}\|_A^2. \tag{28}$$

This is the expected worst-case squared confidence width $\|x\|_A^2$ of (8) under the batch-j survivor sets. The scaled width $\widetilde{\omega}_j$ is used only for $j \geq 2$.

By Assumption 4.2, $\|x\|_2^2 \leq 1$ for every arm. For $k \geq 2$, $\beta(x) \geq 1$, $\|x\|_2 \leq 1$, and bounded rewards give

$$\|\widetilde{x}\|_2^2 = \frac{\dot{\mu}(\langle x, \hat{\theta}_1 \rangle)}{\beta(x)} \|x\|_2^2 \leq 4R^2. \tag{29}$$

For $k \geq 2$, let n_k denote the full batch length. The $n_k/2$ users in D_k construct a DOD policy, and the disjoint $n_k/2$ users in E_k estimate the GLM parameter. As in Algorithm 3, assume $4L$ divides n_k only to avoid integer rounding. Run Algorithm 3 on $\widetilde{\mathcal{X}}^{(k)}$ for the D_k users. Use regularizer Γ, and denote the resulting policy by π_k. In a scaled batch, each design user plays the original arm corresponding to the scaled arm $z_{k,\ell}(\widetilde{\mathcal{X}}_t)$ that it reports in its phase (and discards the reward), and when an estimation user samples $\widetilde{x}$ from π_k, the bandit plays its corresponding original arm x. In both cases the played arm lies in the surviving set $\mathcal{X}^{(k)}$.

Define

$$\mathcal{I}_1 := \mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}_1} \mathop{\mathbb{E}}_{x \sim \pi_G(\mathcal{X})} [xx^\top] \qquad \mathcal{I}_k := \mathop{\mathbb{E}}_{\widetilde{\mathcal{X}} \sim \widetilde{\mathcal{D}}_k} \mathop{\mathbb{E}}_{\widetilde{x} \sim \pi_k(\widetilde{\mathcal{X}})} [\widetilde{x}\widetilde{x}^\top], \qquad k \geq 2. \tag{30}$$

The local bound (12) gives $xx^\top \preceq 2dG_{\mathcal{X}}$ for each $x \in \mathcal{X}$. Consequently, for every $\eta > 0$ (the ridge keeps the inverse well defined even when $\mathcal{I}_1$ is singular),

$$\mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}_1} \max_{x \in \mathcal{X}} \|x\|^2_{(\mathcal{I}_1 + \eta I)^{-1}} \leq 2d \operatorname{tr}\left(\mathop{\mathbb{E}}_{\mathcal{X} \sim \mathcal{D}_1} G_{\mathcal{X}} \left(\mathcal{I}_1 + \eta I\right)^{-1} \right) = 2d \operatorname{tr}\big(\mathcal{I}_1 (\mathcal{I}_1 + \eta I)^{-1}\big) \leq 2d^2. \tag{31}$$

Lemma F.1 (DOD guarantee for scaled survivor sets)**.** *Suppose $L \geq 4d$ and, for every $k \geq 2$, Γ satisfies (23) with $m = n_k/2$, failure probability T^{-3}/M, and $\Delta = 4R^2$. Then for each $k \geq 2$, conditional on the public transcript preceding D_k, the bound below holds with probability at least $1 - T^{-3}/M$; consequently, by a union bound, it holds simultaneously for every $k \geq 2$ with probability at least $1 - T^{-3}$:*

$$\mathop{\mathbb{E}}_{\mathcal{X}^{(k)} \sim \mathcal{D}_k} \max_{x \in \mathcal{X}^{(k)}} \|\widetilde{x}\|^2_{((2\Gamma/n_k) I + \mathcal{I}_k)^{-1}} = O(d \log d), \qquad k \geq 2. \tag{32}$$

The released policy π_k is also (ε, δ) shuffle differentially private with respect to the users in D_k.

Proof. Fix $k \geq 2$ and condition on the public transcript preceding the design users of batch k. The estimate $\hat{\theta}_1$, the function β, and the elimination rule are functions of this transcript, so the map $\mathcal{X} \mapsto \widetilde{\mathcal{X}}^{(k)}$ is deterministic under the conditioning. The users in D_k have not yet interacted with the protocol, so their context sets are independent of the transcript and of each other. Applying the same fixed map to each of them, their transformed survivor sets are i.i.d. draws from $\widetilde{\mathcal{D}}_k$.

Accuracy. We apply Theorem E.6 with $\mathcal{D} = \widetilde{\mathcal{D}}_k$ and $m = n_k/2$. Its hypotheses hold: every transformed arm satisfies $\|\widetilde{x}\|_2^2 \le 4R^2$ by (29), so $\Delta = 4R^2$; $L \ge 4d$ by assumption; $2L$ divides m because $4L$ divides n_k; and Γ satisfies (23). Since $\Gamma/m = 2\Gamma/n_k$ and $\mathcal{I}_{\widetilde{\mathcal{D}}_k}(\pi_k) = \mathcal{I}_k$ by (30), the theorem gives, with probability at least $1 - T^{-3}/M$ conditional on the transcript preceding D_k,

$$\mathop{\mathbb{E}}_{\widetilde{\mathcal{X}} \sim \widetilde{\mathcal{D}}_k} \max_{\widetilde{x} \in \widetilde{\mathcal{X}}} \|\widetilde{x}\|^2_{((2\Gamma/n_k)I + \mathcal{I}_k)^{-1}} = O(d \log d).$$

Because $\widetilde{\mathcal{D}}_k$ is the pushforward of $\mathcal{D}_k$, this expectation equals the expectation over $\mathcal{X}^{(k)} \sim \mathcal{D}_k$ of the same quantity evaluated at $\widetilde{\mathcal{X}}^{(k)}$. The maximum over $\widetilde{x} \in \widetilde{\mathcal{X}}^{(k)}$ is the maximum over $x \in \mathcal{X}^{(k)}$ of $\|\widetilde{x}\|^2$, which is (32). A union bound over the at most M batches gives the simultaneous statement.

Privacy. Conditional on the public transcript, each user in D_k applies the same deterministic map to their own context set and then runs the local randomizers of Algorithm 3. Thus neighboring datasets of original context sets induce neighboring datasets of transformed sets, and Theorem E.7 gives (ε, δ)-shuffle differential privacy for π_k with respect to the transformed sets, hence also with respect to the original context sets. Playing the corresponding original arm x is postprocessing of π_k and the acting user's own context, so it does not change the guarantee. □

Claim 1 (Survivor-width monotonicity)**.** *For every fixed positive-semidefinite matrix A and $j < k$, the widths of (28) satisfy*

$$\omega_k(A) \le \omega_j(A). \tag{33}$$

The same inequality holds for $\widetilde{\omega}$ whenever $2 \le j < k$, because the scaling rule is then common to both survivor sets.

Proof. The survivor sets are nested, so there is a coupling under which $\mathcal{X}^{(k)} \subseteq \mathcal{X}^{(j)}$. The maximum over the smaller set is no larger pointwise; taking expectations under this coupling proves the claim. □

This is the survivor-set monotonicity used when a matrix learned in an earlier batch controls widths in a later batch.

G Regret and Privacy Analysis of the Shuffle-DP GLM Algorithm

G.1 Parameter settings

We collect the explicit parameter settings for Algorithm 1. Throughout, $c_\sigma, c_\nu > 0$ denote the universal constants hidden in Lemmas W.2 and X.3.

Noise level and optimization error. Set

$$\sigma := \frac{c_\sigma R^2 \log\bigl(2d^2/\delta\bigr)}{\varepsilon}, \qquad \nu := \frac{c_\nu R S^2}{\varepsilon} \sqrt{d \log^3(Td/\delta) \log(2/\delta)}. \tag{34}$$

In Appendix G.5 we show that σ is a valid sub-Gaussian noise parameter for every matrix-summation release and that ν bounds the error of every private optimization call.

Logarithmic shorthands. Define

$$\iota := \log \frac{2dMT^3}{\delta}, \qquad d_T := d + \log(2dMT). \tag{35}$$

Since $\delta < 1$ and $M \ge 2$, we have $\iota \ge 1$, $\log T \le \iota$, $\log\bigl(2dMT^3\bigr) \le \iota$, and $d_T \le d + \iota \le 2d\iota$.

Regularizers and confidence radius. The DOD regularizer Γ, the estimator regularizer λ, the estimator-regularizer bounds $\lambda_\pm$, and the confidence radius γ are

$$\Gamma := (24R^2 + 72\sigma)\, d_T, \tag{36}$$

$$\lambda := 128R^2 d\iota + 13\sigma(d + \iota), \tag{37}$$

$$\lambda_\pm := \lambda \pm \sigma\big(4\sqrt{d} + 2\sqrt{\iota}\big), \tag{38}$$

$$\gamma := 24RS^2\sqrt{\lambda} + \sqrt{4RS\nu}. \tag{39}$$

With high probability, λ_- and λ_+ bound the eigenvalues of every noisy regularizer $\lambda I + Z_k$ (see (48)). The inequalities required of these choices are verified in Appendix G.3.

Batch schedule. The batch lengths are those of (5):

$$\alpha := \begin{cases} T^{\frac{1}{2(1-2^{-M+1})}}, & M \le \log_2\log_2 T, \\ 2\sqrt{T}, & M > \log_2\log_2 T, \end{cases} \qquad n_1 = \left\lceil \left(\tfrac{\sqrt{\kappa}e^{3S}}{S}\alpha R d^2\gamma^2\right)^{2/3} \right\rceil, \tag{40}$$

$$n_2 := \alpha, \qquad n_k := \alpha\sqrt{n_{k-1}}, \quad k \in [3, M].$$

These are the nominal batch lengths; n_1 is an integer by construction. We suppress rounding of the later batch lengths. Let

$$\bar{M} := \min\Big\{m \in [M] : \textstyle\sum_{k=1}^{m} n_k \ge T\Big\} \tag{41}$$

be the batch in which the horizon is reached; $\bar{M}$ is deterministic and well defined, since the nominal n_M alone is at least T (proof of Lemma I.8). Batches $k < \bar{M}$ have their nominal lengths; batch $\bar{M}$ is truncated to length $T - \sum_{k<\bar{M}} n_k \le \alpha\sqrt{n_{\bar{M}-1}}$; batches $k > \bar{M}$ are empty. The truncated batch need not be batch M (e.g., $\bar{M} \le M-1$ for $T = 2^{16}$). Its estimate $\hat{\theta}_{\bar{M}}$ is never used, because no later batch has users. We take $M = \lceil \log_2\log_2 T \rceil + 1$, so that $\alpha = 2\sqrt{T}$; by Lemma I.8, the algorithm then uses $O(\log\log T)$ batches.

We assume $\kappa \ge 1$ throughout. This is without loss of generality: in general only $\kappa \ge 4/R^2$ holds, and otherwise we replace κ by $\max\{\kappa, 1\}$ in β, n_1, and $\mathrm{UCB}_1, \mathrm{LCB}_1$. Every bound that uses an upper bound on $1/\kappa$ remains valid, and the change affects only the lower-order $T^{1/3}$ terms.

Throughout the analysis, for an original arm x we write $\widetilde{x} := \sqrt{\frac{\dot{\mu}(\langle x, \hat{\theta}_1\rangle)}{\beta(x)}}\, x$ as in (7).

G.2 Assumptions

We isolate the properties used by the regret proof. Assumption G.1 controls the shuffle perturbations of the design matrices, Assumption G.2 controls the error of the private optimizer, and Assumption G.3 records the only property required from the exploration policy. This separation lets us first prove a general regret theorem and then instantiate it with the shuffle-private DOD construction of Appendix F.

The covariance releases in Algorithm 1 define perturbations Z_k through

$$\mathbf{V} = \lambda I + \sum_{t \in \mathcal{B}_1} x_t x_t^\top + Z_1,$$

$$\mathbf{H}_k = \lambda I + \sum_{t \in E_k} \frac{\dot{\mu}(\langle x_t, \hat{\theta}_1\rangle)}{\beta(x_t)} x_t x_t^\top + Z_k, \qquad k \ge 2. \tag{42}$$

For the invocation of Algorithm 3 in batch $k \ge 2$, let $z_{k,0}(\widetilde{\mathcal{X}}_t)$ be the arm drawn from $\pi_G(\widetilde{\mathcal{X}}_t)$ in phase 0, and, for $\ell \ge 1$, let $z_{k,\ell}(\widetilde{\mathcal{X}}_t)$ be the arm selected by the max-variance rule from $W_{k,\ell-1}$. Write the phasewise shuffled releases as

$$\widetilde{H}_{k,\ell} = \sum_{t \in D_{k,\ell}} z_{k,\ell}(\widetilde{\mathcal{X}}_t) z_{k,\ell}(\widetilde{\mathcal{X}}_t)^\top + Z_{k,\ell}^{\mathrm{DOD}}, \qquad \ell = 0, \dots, L, \tag{43}$$

and define

$$Z_{k,0:r}^{\mathrm{DOD}} := \sum_{\ell=0}^{r} Z_{k,\ell}^{\mathrm{DOD}},$$

$$W_{k,r} := \Gamma I + \sum_{\ell=0}^{r} \widetilde{H}_{k,\ell}$$

$$= \Gamma I + \sum_{\ell=0}^{r} \sum_{t \in D_{k,\ell}} z_{k,\ell}(\widetilde{\mathcal{X}}_t) z_{k,\ell}(\widetilde{\mathcal{X}}_t)^\top + Z_{k,0:r}^{\mathrm{DOD}}. \tag{44}$$

Assumption G.1 (Matrix-summation noise)**.** There is a sub-Gaussian noise parameter $\sigma > 0$ such that, conditional on the public transcript preceding the corresponding matrix-summation call and on the inputs submitted to that call, every estimator error Z_k, $k \in [M]$, and every phasewise DOD error $Z_{k,\ell}^{\mathrm{DOD}}$, $k \geq 2$, $\ell = 0, \ldots, L$, is symmetric with independent, mean-zero sub-Gaussian upper-triangular entries having variance proxy at most σ^2.

The assumption is conditional: within one call the entries are independent, but noise matrices from different DOD phases need not be independent, since the inputs of a phase are selected from earlier noisy releases and the randomized rounding of the matrix-summation protocol depends on its inputs. For every $t \geq 0$, the sub-Gaussian random-matrix bound [Vershynin, 2018, Corollary 4.4.8] for a single release and Lemma R.3 for the cumulative DOD noise give[11]

$$\mathbb{P}\left[\left\{\|Z_k\|_{\mathrm{op}} > \sigma(4\sqrt{d} + 2\sqrt{t}) \;\middle|\; \text{the preceding public transcript}\right\}\right] \leq 2e^{-t}, \tag{45}$$

$$\mathbb{P}\left[\left\{\|Z_{k,0:r}^{\mathrm{DOD}}\|_{\mathrm{op}} > 2\sigma\sqrt{r+1}\,(4\sqrt{d} + 2\sqrt{t}) \;\middle|\; \text{the transcript preceding this DOD invocation}\right\}\right] \leq 2e^{-t}. \tag{46}$$

Taking $t = \iota$ in the first inequality and union bounding over the M estimator releases (total failure probability $2Me^{-\iota} = \delta/(dT^3) \leq T^{-3}$) shows that, with probability at least $1 - T^{-3}$, simultaneously for all batches,

$$\|Z_k\|_{\mathrm{op}} \leq \sigma\big(4\sqrt{d} + 2\sqrt{\iota}\big), \qquad k \in [M]. \tag{47}$$

On this event, the definition (38) of $\lambda_\pm$ gives

$$\lambda_- I \preceq \lambda I + Z_k \preceq \lambda_+ I, \qquad k \in [M]. \tag{48}$$

The concrete matrix-summation mechanism is verified to satisfy Assumption G.1 in Appendix G.5.

Assumption G.2. $\exists \nu > 0$ such that for every batch $k \in [\bar{M} - 1]$, $\|\hat{\theta}_k\|_2 \leq S$ and

$$\left|\mathcal{L}_{\mathcal{S}}(\hat{\theta}_k; \lambda_+) - \min_{\theta : ||\theta||_2 \leq S} \mathcal{L}_{\mathcal{S}}(\theta; \lambda_+)\right| < \nu,$$

where $\mathcal{S} = \{r_t, x_t\}_{t \in \mathcal{B}_1}$ for $k = 1$ and $\mathcal{S} = \{r_t, x_t\}_{t \in E_k}$ for $k \geq 2$, and

$$\mathcal{L}_{\mathcal{S}}(\theta; \lambda) := \sum_{(r,x) \in \mathcal{S}} \ell(\theta, x, r) + \frac{\lambda}{2}\|\theta\|_2^2.$$

This holds simultaneously for all batches $k \in [\bar{M} - 1]$ with probability at least $1 - T^{-3}$, where $\bar{M}$ is the last nonempty batch (41). (The estimates $\hat{\theta}_k$, $k \geq \bar{M}$, are never used by Algorithm 1, so no guarantee is required for them.)

Design width. Recall the widths $\omega_j, \widetilde{\omega}_j$ of (28). Writing $\mathbf{V}_1 := \mathbf{V}$ and $\mathbf{V}_j := \mathbf{H}_j$ for $j \geq 2$, the *design width* faced by batch $k \geq 2$ is the width of the previous batch's matrix,

$$w_k := \widetilde{\omega}_k\big(\mathbf{V}_{k-1}^{-1}\big), \qquad k \geq 2. \tag{49}$$

Assumption G.3 (Distributional design after elimination and scaling)**.** The warm-up batch $\mathcal{B}_1$ has n_1 local-design users. Every batch $k \geq 2$ uses a design group D_k and a disjoint estimation group E_k, with $|D_k| = |E_k| = n_k/2$. Elimination and scaling are frozen before the users in D_k arrive. Simultaneously for every batch, with probability at least $1 - 3T^{-3}$,

$$\omega_1\big(\mathbf{V}^{-1}\big) = O\left(\frac{d^2}{n_1}\right), \tag{50}$$

$$\widetilde{\omega}_k\big(\mathbf{H}_k^{-1}\big) = O\left(\frac{d \log d}{n_k}\right), \qquad k \geq 2. \tag{51}$$

The same bounds hold under every later survivor distribution by Claim 1.

[11] The single-release bound holds up to a universal constant factor in front of σ, which we normalize into the constant c_σ of (34); the cumulative bound is Lemma R.3 as stated.

For each $k \geq 2$, let π_k denote the policy produced from the transformed context sets of D_k by Algorithm 3 with the parameters L and Γ above. The warm-up batch uses the fixed local design policy π_G.

We first verify the parameter inequalities and establish shuffle privacy, and then verify Assumptions G.1 to G.3 for the concrete protocol. The regret proof under these abstract assumptions appears in Appendix I.

G.3 Verification of the parameter inequalities

DOD regularizer. Apply Theorem E.6 with $L = 4d$ phases, summand bound $\Delta = 4R^2$ (by (29)), and per-batch failure probability $\beta = T^{-3}/M$. Then $L + 1 \leq 5d$ and, since $3d_T = 3d + \log\bigl(8d^3M^3T^3\bigr)$,

$$\log \frac{4(L+1)}{\beta} \leq \log \frac{4d(L+1)}{\beta} \leq \log\bigl(20d^2MT^3\bigr) \leq 3d_T.$$

Hence the first term in (23) is at most $8R^2 \cdot 3d_T = 24R^2d_T$. For the second term, use $\sqrt{L+1} \leq \sqrt{5d_T}$ and $\sqrt{d} \leq \sqrt{d_T}$:

$$4\sigma\sqrt{L+1}\left(4\sqrt{d} + 2\sqrt{\log \frac{4(L+1)}{\beta}}\right) \leq 4\sqrt{5}\,\sigma\bigl(4+2\sqrt{3}\bigr)d_T = (16\sqrt{5}+8\sqrt{15})\,\sigma d_T \leq 72\sigma d_T.$$

Thus Γ in (36) satisfies (23) in every post-warm-up batch.

Regularizer bounds. Since $d, \iota \geq 1$, we have $4\sqrt{d} + 2\sqrt{\iota} \leq 4(d+\iota)$, so (37) and (38) give

$$\lambda_- \geq 128R^2d\iota + 9\sigma(d+\iota), \qquad \lambda_+ \leq 128R^2d\iota + 17\sigma(d+\iota) \leq 2\lambda. \tag{52}$$

Using $R \geq 1$, $\log T \leq \iota$, $\log\bigl(2dMT^3\bigr) \leq \iota$, and $d_T \leq d + \iota \leq 2d\iota$,

$$\begin{aligned} 20dR\log T + 96R^2\log\bigl(2dMT^3\bigr) + \frac{1}{8}\Gamma &\leq 20R^2d\iota + 96R^2d\iota + 3R^2d_T + 9\sigma d_T \\ &\leq 122R^2d\iota + 9\sigma(d+\iota) \leq \lambda_-. \end{aligned}$$

Consequently,

$$\lambda_- \geq 20dR\log T, \qquad \lambda_- \geq 96R^2\log\bigl(2dMT^3\bigr) + \frac{1}{8}\Gamma. \tag{53}$$

These are the regularizer conditions used by Lemmas H.2 and I.4.

Confidence radius. For the analysis, let $a := \log T + d$ and define the parameterized confidence radius (the right-hand side of Lemma I.4) and the auxiliary parameter

$$\gamma(\lambda_+, \lambda_-) := 8RS\left(\sqrt{a} + \frac{Ra}{\sqrt{\lambda_-}}\right) + 16RS^2\sqrt{\lambda_+} + \sqrt{4RS\nu}, \tag{54}$$

$$\phi(\lambda_+, \lambda_-) := \frac{\sqrt{\kappa}e^{3S}}{S}\gamma(\lambda_+, \lambda_-)^2. \tag{55}$$

We claim $\gamma(\lambda_+, \lambda_-) \leq \gamma$. Since $a \leq 2d\iota$, (52) gives $\lambda_- \geq 128R^2d\iota \geq 64R^2a$, that is, $\sqrt{a} \leq \sqrt{\lambda_-}/(8R) \leq \sqrt{\lambda}/(8R)$. Hence $Ra/\sqrt{\lambda_-} = R\sqrt{a}\cdot\sqrt{a}/\sqrt{\lambda_-} \leq \sqrt{a}/8$, and

$$8RS\left(\sqrt{a} + \frac{Ra}{\sqrt{\lambda_-}}\right) \leq 9RS\sqrt{a} \leq \frac{9}{8}S\sqrt{\lambda} \leq \frac{9}{8}RS^2\sqrt{\lambda}.$$

Moreover, $16RS^2\sqrt{\lambda_+} \leq 16\sqrt{2}\,RS^2\sqrt{\lambda}$ by (52). Since $9/8 + 16\sqrt{2} < 24$,

$$\gamma(\lambda_+, \lambda_-) \leq 24RS^2\sqrt{\lambda} + \sqrt{4RS\nu} = \gamma, \qquad \phi(\lambda_+, \lambda_-) \leq \frac{\sqrt{\kappa}e^{3S}}{S}\gamma^2. \tag{56}$$

G.4 Privacy analysis

All user groups and all DOD subphases are fixed before their users' contexts are observed. For $k \geq 2$, a user in D_k contributes to exactly one matrix-summation call inside Algorithm 3; by Theorem E.7, the complete policy π_k is (ε, δ) shuffle differentially private with respect to that user. A user in $\mathcal{B}_1$ or E_k contributes to the shuffled covariance estimate and to the shuffle-private convex optimizer. Assigning $(\varepsilon/2, \delta/2)$ to each of these two mechanisms gives (ε, δ) privacy for the estimation user by sequential composition.

The set $\mathcal{B}_1$ and the sets D_k, E_k for $k \geq 2$ are mutually disjoint. Parallel composition therefore gives (ε, δ) shuffle privacy for the complete adaptive transcript. Subsequent elimination rules, published policies, and actions are postprocessing of this transcript.

G.5 Verifying Assumptions G.1 to G.3

We show that σ and ν in (34) are valid choices in the two assumptions, using the matrix-summation and private-optimization guarantees of Lemmas W.2 and X.3.

Matrix-summation noise. In the warm-up release, $\|x_t x_t^\top\|_{\mathrm{op}} \leq 1$. In every scaled estimator release and every DOD phase, (29) bounds the operator norm of each summand by $4R^2$. Since the variance proxy in Lemma W.2 is quadratic in this bound, it is respectively

$$O\left(\frac{\log^2(d^2/\delta)}{\varepsilon^2}\right) \quad \text{and} \quad O\left(\frac{R^4 \log^2(d^2/\delta)}{\varepsilon^2}\right).$$

Using $(\varepsilon/2, \delta/2)$ for an estimator release changes only universal constants and logarithmic arguments. Because $R \geq 1$, σ^2 from (34) therefore dominates every release, which verifies Assumption G.1. Fresh calls use independent mechanism randomness, which gives the required conditional independence.

Private optimization. Let $N_1 := n_1$ and $N_k := n_k/2$ for $k \geq 2$ be the estimator-group sizes, and let $\Theta := \{\theta : \|\theta\|_2 \leq S\}$, a convex set of diameter $D = 2S$. In batch k, Algorithm 1 runs $\mathcal{P}_{\mathrm{GD}}$ (Algorithm 7) over Θ with budget $(\varepsilon/2, \delta/2)$ on the per-sample loss

$$\tilde{\ell}_t(\theta) := \ell(\theta, x_t, r_t) + \frac{\lambda_+}{2N_k}\|\theta\|_2^2.$$

The empirical mean of $\tilde{\ell}_t$ over the N_k estimator users equals $N_k^{-1}\mathcal{L}_{\mathcal{S}}(\theta; \lambda_+)$, so an empirical-mean error ν_{PGD} is a cumulative-loss error $N_k \nu_{\mathrm{PGD}}$ in Assumption G.2, and the output lies in Θ. The step size η and iteration count are those of Lemma X.3, with $n = N_k$ and failure probability $\zeta = T^{-3}/M$.

Lipschitz constant. Since $\nabla_\theta \ell(\theta, x, r) = (\mu(\langle x, \theta\rangle) - r)\, x$ with $\mu(\cdot), r \in [0, R]$ and $\|x\|_2 \leq 1$, the loss ℓ is R-Lipschitz, so $\tilde{\ell}_t$ is L_k-Lipschitz on Θ with $L_k := R + S\lambda_+/N_k$. For every $k \in [\bar{M} - 1]$ we have

$$\lambda_+ \leq RN_k, \qquad \text{hence} \qquad L_k \leq R + RS \leq 2RS. \tag{57}$$

Indeed, batches $k < \bar{M}$ are not truncated, so $N_k \geq n_2/2 = \alpha/2$ for $2 \leq k \leq \bar{M} - 1$. For $k = 1$, $\sqrt{\kappa} \geq 1/R$, $d, e^{3S} \geq 1$, and $\gamma^2 \geq 576R^2S^4\lambda \geq S\lambda$ give $n_1 \geq (\alpha\lambda)^{2/3}$. Under the standing condition $\alpha \geq 4\lambda$ (implied by $T \geq 16\lambda^2$, since $\alpha \geq \sqrt{T}$), it follows that $Rn_1 \geq (4\lambda^2)^{2/3} \geq 2\lambda$ (as $\lambda \geq 128$) and $R\alpha/2 \geq 2\lambda$, and both are at least λ_+ by (52).

Gradient clipping. In every batch, $\mathcal{P}_{\mathrm{GD}}$ is run with $L = 2RS$ and each user's gradient is clipped to norm at most $2RS$ (Algorithm 7). Every input to the vector-summation protocol then has norm at most $2RS$ deterministically, so the privacy guarantee holds for all batches, including the truncated batch $\bar{M}$, and does not rely on (57). For $k \in [\bar{M} - 1]$, (57) gives $\|\nabla\tilde{\ell}_t\|_2 \leq L_k \leq 2RS$, so clipping is inactive and the utility bound below is unaffected. In batch $\bar{M}$ clipping may bias the output, but $\hat{\theta}_{\bar{M}}$ is not used by any later batch, so the regret analysis is unaffected.

Error bound. For $k \in [\bar{M} - 1]$, Lemma X.3 with $D = 2S$, Lipschitz constant $L_k \leq 2RS$, privacy parameters $(\varepsilon/2, \delta/2)$, and $\Lambda_{\mathrm{shuf}} = \log^2(2N_k d/\delta)\log\big(N_k MT^3\big)\log(2/\delta) \leq 20\log^3(Td/\delta)\log(2/\delta)$ (using $N_k, M \leq T$) gives, with probability at least $1 - T^{-3}/M$ (its requirements $\varepsilon/2 \leq N_k$ and

$\Lambda_{\mathrm{shuf}} \geq 1$ hold trivially),

$$N_k \nu_{\mathrm{PGD}} = O\left(\frac{S \cdot RS\sqrt{d \log^3(Td/\delta) \log(2/\delta)}}{\varepsilon} \right).$$

The group size N_k cancels. Choosing c_ν in (34) as the resulting universal constant and taking a union bound over the at most M batches verifies Assumption G.2.

Design width. Lemma H.2 verifies Assumption G.3: it gives (50) in the warm-up batch and (51) in every batch $k \geq 2$. It uses three events, each of probability at least $1 - T^{-3}$: the DOD event of Lemma F.1, the estimator noise event (47), and the matrix Chernoff event of Lemma H.2. A union bound gives the probability $1 - 3T^{-3}$ stated in Assumption G.3.

Together with (53) and (56), all the hypotheses of Theorem I.1 therefore hold for the concrete protocol. The resulting regret guarantee is stated in Theorem J.1.

H Private Empirical Design Matrices

We transfer the population guarantee of Lemma F.1 to the private matrices $\mathbf{V}, \mathbf{H}_k$ of (42), using the noise event (47) and the parameters $\Gamma, \lambda, \lambda_-, \lambda_+$ from Appendix G.1.

Lemma H.1 (Matrix Chernoff bound Tropp, 2015)**.** *Let $x_1, \ldots, x_n$ be i.i.d. vectors with $\|x_i\|_2 \leq 1$. Then, for every $\varepsilon > 0$,*

$$\mathbb{P}\left[3\varepsilon n I + \sum_{i=1}^{n} x_i x_i^\top \succeq \frac{n}{8} \mathbb{E}\big[xx^\top\big] \right] \geq 1 - 2d \exp(-\varepsilon n/8), \tag{58}$$

where x has the common distribution of the x_i.

Lemma H.2 (Scaled population DOD to private empirical design)**.** *Suppose the events in Lemma F.1 hold. Conditional on the preceding design transcript for each E_k, with probability at least $1 - T^{-3}$ over the estimation samples, the following hold simultaneously on the noise event (47):*

$$\mathbf{V} \succeq \frac{1}{8}\big(n_1 \mathcal{I}_1 + \Gamma I\big), \quad \mathbf{H}_k \succeq \frac{1}{8}\left(\frac{n_k}{2}\mathcal{I}_k + \Gamma I\right), \qquad k \geq 2. \tag{59}$$

Consequently, in the notation of (28),

$$\omega_1\big(\mathbf{V}^{-1}\big) = O\left(\frac{d^2}{n_1}\right), \quad \widetilde{\omega}_k\big(\mathbf{H}_k^{-1}\big) = O\left(\frac{d \log d}{n_k}\right), \qquad k \geq 2, \tag{60}$$

Proof. For a scaled batch, conditional on the design transcript, the vectors $\widetilde{x}_i$ are independent, satisfy $\|\widetilde{x}_i\|_2^2 \leq 4R^2$, and have second moment $\mathcal{I}_k$. Apply Lemma H.1 to $\widetilde{x}_i/(2R)$ for the $|E_k| = n_k/2$ estimation samples with

$$\varepsilon_k := \frac{16 \log\big(2dMT^3\big)}{n_k}.$$

The bound in Lemma H.1 holds for every positive ε_k, so no lower bound on n_k is needed. Moreover,

$$2d \exp\left(-\frac{\varepsilon_k |E_k|}{8}\right) = 2d \exp\big(-\log\big(2dMT^3\big)\big) = \frac{T^{-3}}{M}.$$

After multiplying the resulting matrix inequality by $4R^2$, we obtain, with failure probability at most T^{-3}/M for each batch,

$$96R^2 \log\big(2dMT^3\big) I + \sum_{i \in E_k} \widetilde{x}_i \widetilde{x}_i^\top \succeq \frac{n_k}{16}\mathcal{I}_k. \tag{61}$$

Combining this with (47) and (53) gives

$$\mathbf{H}_k \succeq \frac{1}{8}\left(\frac{n_k}{2}\mathcal{I}_k + \Gamma I\right), \tag{62}$$

which proves (59). Now applying Lemma F.1 proves the $k \geq 2$ bound in (60).

For the warm-up, the n_1 local-design samples are independent with second moment $\mathcal{I}_1$. Applying the same matrix Chernoff bound with $\varepsilon_1 := 8\log\big(2dMT^3\big)/n_1$ gives failure probability T^{-3}/M and $24\log\big(2dMT^3\big)I + \sum_{i\in\mathcal{B}_1} x_i x_i^\top \succeq (n_1/8)\mathcal{I}_1$. On the noise event, $\lambda I + Z_1 \succeq \lambda_- I$, and (53) gives $\lambda_- \geq 24\log\big(2dMT^3\big) + \Gamma/8$, so

$$\mathbf{V} = \lambda I + \sum_{i\in\mathcal{B}_1} x_i x_i^\top + Z_1 \succeq \frac{n_1}{8}\mathcal{I}_1 + \frac{\Gamma}{8}I = \frac{n_1}{8}\Big(\mathcal{I}_1 + \frac{\Gamma}{n_1}I\Big).$$

This is the first inequality in (59). Applying (31) with $\eta = \Gamma/n_1 > 0$ gives $\omega_1(\mathbf{V}^{-1}) \leq (8/n_1)\cdot 2d^2$, the first bound in (60). A union bound over the M batches yields the desired bound. □

I Regret Analysis under the Shuffle-DP Assumptions

Theorem I.1 (Regret under shuffle-DP assumptions)**.** *Suppose $d \geq 2$ and Assumptions G.1 to G.3 hold simultaneously with probability at least $1 - 4T^{-3}$. Assume $\lambda_- \geq 20dR\log T$ and $\gamma(\lambda_+, \lambda_-) \leq \gamma$. Then the expected regret satisfies*

$$R_T = O\Bigg(M\Bigg[R^{1/3}\left(\frac{\sqrt{\kappa}e^{3S}}{S}\,\alpha R^2 d^2\gamma^2\right)^{2/3} + \alpha\gamma\sqrt{\frac{d\log d}{\kappa^*}}\Bigg] + MR\Bigg). \tag{63}$$

The proof follows the argument of Sawarni et al. [2024], with two changes: the confidence radius must accommodate the shuffle perturbation and optimization error, and the design step enters only through Assumption G.3. We first collect the confidence ingredients.

I.1 Concentration inequalities

Lemma I.2 (Bernstein's inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with $|X_t - \mathbb{E}[X_t]| \leq b$. Also, let sum $S := \sum_{t=1}^n (X_t - \mathbb{E}[X_t])$ and let $v = \sum_{t=1}^n Var[X_t]$. Then for any $\delta > 0$, we have*

$$\mathbb{P}\left(S \geq \sqrt{2v\log(1/\delta)} + \frac{2b}{3}\log(1/\delta)\right) \leq \delta \tag{64}$$

We further present the following concentration bound from Vershynin [2018] on the sum of squares of zero-mean sub-Gaussian random variables.

Lemma I.3 (Concentration of squares of sub-Gaussian)**.** *Let $X_1, X_2, \ldots, X_n$ be zero-mean i.i.d. sub-Gaussian random variables with variance proxy σ^2. Then for $n > 64\log(2/\delta)$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n (X_i^2 - \mathbb{E}[X_i^2]) > \sqrt{\frac{8\sigma^4\log(2/\delta)}{n}}\right) \leq \delta \tag{65}$$

We now bound the estimation error of an *approximate* minimizer of the regularized loss. This is a variant of Sawarni et al. [2024, Lemma A.2], in which $\hat{\theta}$ minimizes the loss only up to the error ν guaranteed for the shuffle-private optimizer by Assumption G.2; the optimization error enters the confidence radius additively as $\sqrt{4RS\nu}$.

Lemma I.4 (Confidence bound for an approximate minimizer)**.** *Let $x_1, \ldots, x_t \in \mathbb{R}^d$ with $\|x_s\|_2 \leq 1$, and let $r_1, \ldots, r_t \in [0, R]$ be independent rewards from the canonical exponential family with $\mathbb{E}[r_s] = \mu(\langle x_s, \theta^*\rangle)$, where $\|\theta^*\|_2 \leq S$. Let $\mathcal{N}$ be a symmetric matrix, independent of $r_1, \ldots, r_t$ given $x_1, \ldots, x_t$, with $\lambda_- \boldsymbol{I} \preceq \mathcal{N} \preceq \lambda_+ \boldsymbol{I}$, and set $\mathbf{H}^* := \sum_{s=1}^t \dot{\mu}(\langle x_s, \theta^*\rangle) x_s x_s^\top + \mathcal{N}$. Let $\mathcal{L}_\mathcal{S}(\cdot; \lambda_+)$ be the regularized loss of Assumption G.2 on $\mathcal{S} = \{(r_s, x_s)\}_{s=1}^t$. If $\lambda_- \geq 20dR\log T$ and $\lambda_+ \leq 4\lambda_-$, then with probability at least $1 - T^{-3}$, every $\hat{\theta}$ with $\|\hat{\theta}\|_2 \leq S$ and $\mathcal{L}_\mathcal{S}(\hat{\theta}; \lambda_+) - \min_{\|\theta\|_2 \leq S} \mathcal{L}_\mathcal{S}(\theta; \lambda_+) < \nu$ satisfies*

$$\|\theta^* - \hat{\theta}\|_{\mathbf{H}^*} \leq 8RS\left(\sqrt{\log T + d} + \frac{R(\log T + d)}{\sqrt{\lambda_-}}\right) + 16RS^2\sqrt{\lambda_+} + \sqrt{4RS\nu}. \tag{66}$$

The proof combines Lemma V.1 with the following bound for the unconstrained minimizer $\tilde{\theta} := \arg\min_\theta \mathcal{L}_\mathcal{S}(\theta; \lambda_+)$, stated in terms of $g(\theta) := \sum_{s=1}^t \mu(\langle x_s, \theta\rangle)x_s + \lambda_+\theta$.

Lemma I.5. *In the setting of Lemma I.4, with probability at least* $1 - T^{-3}$,

$$\|g(\tilde{\theta}) - g(\theta^*)\|_{(\mathbf{H}^*)^{-1}} \le \sqrt{\log T + d} + \frac{R(\log T + d)}{\sqrt{\lambda_-}} + 2S\sqrt{\lambda_+}. \tag{67}$$

This lemma differs slightly from Sawarni et al. [2024, Lemma A.2]: here the noisy regularizer $\mathcal{N}$ is sandwiched between $\lambda_- I$ and $\lambda_+ I$, whereas their result uses λI. The same proof applies after making these two substitutions; the only place the two bounds interact is the regularization term, $\lambda_+\|\theta^*\|_{(\mathbf{H}^*)^{-1}} \le \lambda_+ S/\sqrt{\lambda_-} \le 2S\sqrt{\lambda_+}$, which uses $\lambda_+ \le 4\lambda_-$. For our parameters $\lambda_+/\lambda_- \le 17/9$ by (52).

Proof of Lemma I.4. Let $a := \log T + d$ and let $\gamma_0 := \sqrt{a} + Ra/\sqrt{\lambda_-} + 2S\sqrt{\lambda_+}$ be the bound of Lemma I.5; note $\gamma_0 \ge \sqrt{a} \ge 1$. Since $\|x_s\|_2 \le 1$ and $\hat{\theta}, \theta^*$ lie in the ball of radius S, we have $|\langle x_s, \hat{\theta} - \theta^*\rangle| \le 2S$ for all s. Hence Lemma V.1 applies with $c = 2RS$ and $\gamma = \gamma_0$; since $R, S \ge 1$, $2 + c \le 4RS$, and it gives

$$\|\hat{\theta} - \theta^*\|_{\mathbf{H}^*} \le 4RS\gamma_0 + \sqrt{4RS\nu} + \sqrt{\frac{4R^2S}{\sqrt{\lambda_-}}}\,\gamma_0 + \sqrt{4RS\gamma_0},$$

where we split the last square root of Lemma V.1 using $\sqrt{u+v} \le \sqrt{u} + \sqrt{v}$. We absorb the last two terms into $4RS\gamma_0$. First, $\lambda_- \ge 20dR\log T \ge 1$ gives $\sqrt{4R^2S/\sqrt{\lambda_-}} \le 2R\sqrt{S} \le 2RS$. Second, $\gamma_0 \ge 1$ and $RS \ge 1$ give $\sqrt{4RS\gamma_0} \le 2RS\gamma_0$. Therefore

$$\|\hat{\theta} - \theta^*\|_{\mathbf{H}^*} \le 8RS\gamma_0 + \sqrt{4RS\nu} = 8RS\left(\sqrt{a} + \frac{Ra}{\sqrt{\lambda_-}}\right) + 16RS^2\sqrt{\lambda_+} + \sqrt{4RS\nu},$$

on the event of Lemma I.5. Since Lemma V.1 is deterministic, this holds simultaneously for every such $\hat{\theta}$. □

Using this lemma, we can now prove the following corollary.

Corollary I.6. *Let* $\{x_t\}_{t\in\mathcal{B}_1}$ *be the arms pulled during the warm-up batch and let* $\hat{\theta}_1$ *be the estimator of* θ^* *computed at the end of the batch. Then, on the event that Assumptions G.1 and G.2 hold, with probability at least* $1 - T^{-3}$*, simultaneously for every vector* x*,*

$$|\langle x, \theta^* - \hat{\theta}_1\rangle| \le \sqrt{\kappa}\|x\|_{\mathbf{V}^{-1}}\gamma(\lambda_+, \lambda_-) \tag{68}$$

Proof. This corollary is derived from Lemma I.4 and parametrization of $\gamma(\lambda_+, \lambda_-)$ in (54). By applying the lemma, we have[12]

$$|\langle x, \theta^* - \hat{\theta}_1\rangle| \le ||x||_{(\mathbf{H}^*)^{-1}}||\theta^* - \hat{\theta}_1||_{(\mathbf{H}^*)} \le \sqrt{\kappa}||x||_{\mathbf{V}^{-1}}\gamma(\lambda_+, \lambda_-)$$

Considering the definition of κ, we have $\dot{\mu}(\langle x_s, \theta^*\rangle) \ge \frac{1}{\kappa}$, so $\sum_s \dot{\mu}(\langle x_s, \theta^*\rangle)x_s x_s^\top \succeq \frac{1}{\kappa}\sum_s x_s x_s^\top$. Both $\mathbf{H}^*$ and $\mathbf{V}$ contain the same positive-definite ridge $\lambda I + Z_1$, and $\lambda I + Z_1 \succeq \frac{1}{\kappa}(\lambda I + Z_1)$ because $\kappa \ge 1$ (Appendix G.1). Hence $\mathbf{H}^* \succeq \frac{1}{\kappa}\mathbf{V}$, which in turn implies $||x||_{(\mathbf{H}^*)^{-1}} \le \sqrt{\kappa}||x||_{\mathbf{V}^{-1}}$. □

For each $k \ge 2$, define the oracle comparison matrix

$$\mathbf{H}_k^* := \lambda I + Z_k + \sum_{t\in E_k} \dot{\mu}(\langle x_t, \theta^*\rangle)x_t x_t^\top. \tag{69}$$

It differs from $\mathbf{H}_k$ only in that its arm weights use the true curvature $\dot{\mu}(\langle x_t, \theta^*\rangle)$.

The following preliminary lemmas from Section A.3 of Sawarni et al. [2024] also hold in our setup. We restate them for convenience.

[12]Here $\mathbf{H}^* := \lambda I + Z_1 + \sum_{t\in\mathcal{B}_1} \dot{\mu}(\langle x_t, \theta^*\rangle)x_t x_t^\top$, i.e. the matrix of Lemma I.4 instantiated with the warm-up arms and the warm-up ridge $\lambda I + Z_1$.

Lemma I.7. *For each batch $k \geq 2$ and the scaled data matrix $\mathbf{H}_k$ computed at the end of the batch, the following bound holds with probability at least $1 - T^{-3}$:*

$$\mathbf{H}_k \preceq \mathbf{H}_k^* \text{ and } \dot{\mu}(\langle x, \hat{\theta}_1 \rangle) \leq \beta(x)\dot{\mu}(\langle x, \theta^* \rangle)$$

Lemma I.8. *Under the batch schedule in (40), Algorithm 1 uses $O(\log \log T)$ batches.*

Proof. Let $m := 1 + \lceil \log_2 \log_2 T \rceil$. If $M < m$, the claim is immediate. Otherwise, $\alpha = 2\sqrt{T}$. Ignoring integer rounding, the recurrence gives

$$n_m = (2\sqrt{T})^{\sum_{j=0}^{m-2} 2^{-j}} \geq 2T\, T^{-1/2^{m-1}} \geq T, \tag{70}$$

where the last inequality follows from $2^{m-1} \geq \log_2 T$. Thus the nominal mth batch alone reaches the horizon, so the algorithm must stop by that batch. Integer rounding and final-batch truncation can only make it stop earlier. □

The following two results are adaptations, to our notation and private estimators, of Sawarni et al. [2024, Corollary A.7 and Lemma A.9], respectively. Their arguments carry over under Assumptions G.1 and G.2.

Lemma I.9. *Let $\hat{\theta}_k$ be the estimator of θ^* calculated at the end of batch k. Then, for any vector x, the following holds simultaneously for every batch $2 \leq k \leq M - 1$ with probability at least $1 - MT^{-3}$ (one application of Lemma I.4 per batch, each failing with probability at most T^{-3}):*

$$\left| \langle x, \theta^* - \hat{\theta}_k \rangle \right| \leq \|x\|_{\mathbf{H}_k^{-1}} \gamma(\lambda_+, \lambda_-).$$

Lemma I.10. *Let $x \in \mathcal{X}$ be the selected arm in either action group of batch $k \geq 2$, and let x^* be the optimal arm in the arm set $\mathcal{X}$, i.e.,*

$$x^* = \arg\max_{x \in \mathcal{X}} \mu(\langle x, \theta^* \rangle).$$

For $k = 2$, use the convention $\mathbf{H}_1 := \mathbf{V}$. On the confidence events of Corollary I.6 and of the batch-j confidence bounds for $j < k$ (total failure probability at most $(M+1)T^{-3}$), the following inequality holds:

$$\begin{aligned} \mu(\langle x^*, \theta^* \rangle) - \mu(\langle x, \theta^* \rangle) \quad \leq & 6\phi(\lambda_+, \lambda_-) \sum_{\substack{y \in \{x, x^*\} \\ \tilde{y} \in \{\tilde{x}, \tilde{x}^*\}}} \|y\|_{\mathbf{V}^{-1}} \|\tilde{y}\|_{\mathbf{H}_{k-1}^{-1}} \\ & + 2\gamma(\lambda_+, \lambda_-) \sqrt{\dot{\mu}(\langle x^*, \theta^* \rangle)} \left(\|\tilde{x}^*\|_{\mathbf{H}_{k-1}^{-1}} + \|\tilde{x}\|_{\mathbf{H}_{k-1}^{-1}} \right). \end{aligned}$$

We now average the pointwise inequality using Assumption G.3. For $k \geq 2$, let $\mathcal{G}_{k-1}$ denote the joint event on which the guarantees in Assumptions G.1 and G.2, together with the confidence and correct-elimination guarantees, hold through batch $k-1$, and the bounds in Assumption G.3 hold for the warm-up batch and batches $2, \ldots, k-1$. For $k = 2$, the design component of $\mathcal{G}_1$ consists only of the warm-up bound.

Lemma I.11 (Expected one-round regret under the design-width assumption)**.** *Fix $k \geq 2$. Conditioned on $\mathcal{G}_{k-1}$, every action x selected from the surviving set $\mathcal{X}^{(k)} \sim \mathcal{D}_k$ satisfies*

$$\mathbb{E}[\mu(\langle x^*, \theta^* \rangle) - \mu(\langle x, \theta^* \rangle) \,|\, \mathcal{G}_{k-1}] = O\left(\phi(\lambda_+, \lambda_-) \sqrt{\frac{R^2 d^2}{n_1} w_k} + \gamma(\lambda_+, \lambda_-) \sqrt{\frac{w_k}{\kappa^*}} \right), \tag{71}$$

where w_k is the design width (49), which satisfies $w_2 = O(R^2 d^2 / n_1)$ and $w_k = O(d \log d / n_{k-1})$ for $k \geq 3$.

Proof. All expectations below are conditioned on $\mathcal{G}_{k-1}$. By Lemma I.10, the first contribution is at most a constant multiple of

$$\phi(\lambda_+, \lambda_-) \max_{y \in \mathcal{X}^{(k)}} \|y\|_{\mathbf{V}^{-1}} \max_{z \in \mathcal{X}^{(k)}} \|\tilde{z}\|_{\mathbf{H}_{k-1}^{-1}}. \tag{72}$$

By survivor-set monotonicity (Claim 1) with $A = \mathbf{V}^{-1}$ and $j = 1$, (50)in Assumption G.3 gives

$$\omega_k\big(\mathbf{V}^{-1}\big) \le \omega_1\big(\mathbf{V}^{-1}\big) = O\left(\frac{d^2}{n_1}\right). \tag{73}$$

For the second factor when $k = 2$, the scaling multiplier satisfies

$$\frac{\dot\mu(\langle z, \hat\theta_1\rangle)}{\beta(z)} \le 4R^2, \qquad \text{so} \qquad \|\widetilde{z}\|^2_{\mathbf{V}^{-1}} \le 4R^2\|z\|^2_{\mathbf{V}^{-1}}. \tag{74}$$

Applying the same monotonicity claim and (50) in Assumption G.3 therefore gives $w_2 = \widetilde{\omega}_2(\mathbf{V}^{-1}) \le 4R^2\omega_1(\mathbf{V}^{-1}) = O(R^2d^2/n_1)$. For $k \ge 3$, Claim 1 with $j = k-1$ and $A = \mathbf{H}_{k-1}^{-1}$, followed by (51) in Assumption G.3, gives $w_k = \widetilde{\omega}_k(\mathbf{H}_{k-1}^{-1}) \le \widetilde{\omega}_{k-1}(\mathbf{H}_{k-1}^{-1}) = O(d\log d/n_{k-1})$. Thus Cauchy–Schwarz gives

$$\mathbb{E}\left[\max_{y\in\mathcal{X}^{(k)}} \|y\|_{\mathbf{V}^{-1}} \max_{z\in\mathcal{X}^{(k)}} \|\widetilde{z}\|_{\mathbf{H}_{k-1}^{-1}}\right] = O\left(\sqrt{\frac{d^2}{n_1}w_k}\right). \tag{75}$$

For the curvature contribution, the definition of κ^* implies $\dot\mu(\langle x^*, \theta^*\rangle) \le 1/\kappa^*$. Jensen's inequality, followed by (74) and survivor monotonicity when $k = 2$, and by the scaled-width bound with the same monotonicity claim when $k \ge 3$, therefore yields

$$\mathbb{E}\left[\sqrt{\dot\mu(\langle x^*, \theta^*\rangle)}\left(\|\widetilde{x}^*\|_{\mathbf{H}_{k-1}^{-1}} + \|\widetilde{x}\|_{\mathbf{H}_{k-1}^{-1}}\right)\right] = O\left(\sqrt{\frac{w_k}{\kappa^*}}\right). \tag{76}$$

Substituting (75) and (76) into Lemma I.10 proves the lemma. □

J Proof and Shuffle-Private Instantiation

J.1 Proof under the abstract shuffle-DP assumptions

Theorem I.1 (Regret under shuffle-DP assumptions)**.** *Suppose $d \ge 2$ and Assumptions G.1 to G.3 hold simultaneously with probability at least $1 - 4T^{-3}$. Assume $\lambda_- \ge 20dR\log T$ and $\gamma(\lambda_+, \lambda_-) \le \gamma$. Then the expected regret satisfies*

$$R_T = O\left(M\left[R^{1/3}\left(\frac{\sqrt{\kappa}e^{3S}}{S}\,\alpha R^2d^2\gamma^2\right)^{2/3} + \alpha\gamma\sqrt{\frac{d\log d}{\kappa^*}}\right] + MR\right). \tag{63}$$

Proof of Theorem I.1. For every $k \ge 2$, recall the joint past good event $\mathcal{G}_{k-1}$ defined immediately before Lemma I.11. By the simultaneous probability guarantees in Assumptions G.1 to G.3 and the confidence lemmas (the latter fail with probability at most MT^{-3} in total),

$$\Pr\big(\mathcal{G}_{k-1}^c\big) = O(MT^{-3}). \tag{77}$$

Define the conditional expected regret in batch k by

$$R_k := \mathbb{E}\left[\sum_{t\in\mathcal{B}_k}\big(\mu(\langle x_t^*, \theta^*\rangle) - \mu(\langle x_t, \theta^*\rangle)\big)\,\middle|\,\mathcal{G}_{k-1}\right]. \tag{78}$$

The n_1 warm-up users contribute at most $2Rn_1$. For $k \ge 2$, Lemma I.11 applies to each user in $D_k \cup E_k$. Indeed, users in both groups play an arm from the surviving set: a user in D_k plays the original arm of the scaled arm it reports in Algorithm 3, and a user in E_k plays the original arm of its draw from π_k. The bound of Lemma I.11 holds for every surviving arm, so it does not depend on how the arm is selected. Therefore

$$R_k = O\left(n_k\left(\phi(\lambda_+, \lambda_-)\sqrt{\frac{R^2d^2}{n_1}w_k} + \gamma(\lambda_+, \lambda_-)\sqrt{\frac{w_k}{\kappa^*}}\right)\right). \tag{79}$$

For $k = 2$, substituting $n_2 = \alpha$ and $w_2 = O(R^2 d^2/n_1)$ into (79) gives

$$R_2 = O\left(\frac{\alpha\phi(\lambda_+, \lambda_-)R^2 d^2}{n_1} + \frac{\alpha\gamma(\lambda_+, \lambda_-)Rd}{\sqrt{\kappa^* n_1}}\right). \tag{80}$$

For $k \geq 3$, use $w_k = O(d \log d/n_{k-1})$ and $n_k \leq \alpha\sqrt{n_{k-1}}$ in (79). The factor $\sqrt{n_{k-1}}$ then cancels, yielding

$$R_k = O\left(\alpha\phi(\lambda_+, \lambda_-)Rd\sqrt{\frac{d \log d}{n_1}} + \alpha\gamma(\lambda_+, \lambda_-)\sqrt{\frac{d \log d}{\kappa^*}}\right). \tag{81}$$

We now use the warm-up length in (5). The theorem's confidence-radius condition and (55) give

$$\phi(\lambda_+, \lambda_-) \leq \frac{\sqrt{\kappa}e^{3S}}{S}\gamma^2.$$

Because $R \geq 1$, $d \geq 2$, and $\log d \leq d$, the design terms in (80) and (81) are both bounded by

$$O\left(\frac{\sqrt{\kappa}e^{3S}}{S}\,\frac{\alpha\gamma^2 R^2 d^2}{\sqrt{n_1}}\right).$$

Consequently,

$$2Rn_1 + \sum_{k=2}^{M} R_k = O\left(M\left[Rn_1 + \frac{\sqrt{\kappa}e^{3S}}{S}\,\frac{\alpha\gamma^2 R^2 d^2}{\sqrt{n_1}} + \alpha\gamma\sqrt{\frac{d \log d}{\kappa^*}} + \frac{\alpha\gamma Rd}{\sqrt{\kappa^* n_1}}\right]\right). \tag{82}$$

The choice of n_1 balances the first two terms in (82). The ceiling adds at most one warm-up round, whose regret is absorbed by the additive $O(MR)$ term. The choice also absorbs the last term: after substituting (5), the ratio of that term to Rn_1 is, up to a universal constant,

$$\frac{S}{\sqrt{\kappa}e^{3S}\gamma Rd\sqrt{\kappa^*}}. \tag{83}$$

Since $\kappa^* \geq 1/R^2$, we have $Rd\sqrt{\kappa^*} \geq d$. Moreover, $\gamma \geq 24RS^2\sqrt{\lambda} \geq RS$ together with $\kappa \geq 1/R^2$ gives $\sqrt{\kappa}e^{3S}\gamma/S \geq 1$. Thus (83) is at most a universal constant. Consequently,

$$2Rn_1 + \sum_{k=2}^{M} R_k = O\left(M\left[R^{1/3}\left(\frac{\sqrt{\kappa}e^{3S}}{S}\,\alpha\gamma^2 R^2 d^2\right)^{2/3} + \alpha\gamma\sqrt{\frac{d \log d}{\kappa^*}}\right]\right). \tag{84}$$

The regret of any user is at most $2R$. Therefore, by the law of total expectation, (77), and $\sum_{k=1}^{M} n_k \leq T$, the contribution from the complementary events is

$$2R\sum_{k=2}^{M} n_k \Pr\left(\mathcal{G}_{k-1}^c\right) = O(MR/T^2). \tag{85}$$

This term, together with integer-rounding and final-batch effects, is absorbed into the additive $O(MR)$ term in (63). This proves (63). □

J.2 Instantiation with shuffle-private DOD

The verification in Appendix G.5 establishes Assumptions G.1 to G.3 for the concrete protocol, and Appendix G.3 shows that $\gamma(\lambda_+, \lambda_-) \leq \gamma$, so Theorem I.1 applies.

Theorem J.1 (Shuffle-private DOD-GLM regret)**.** *Suppose $d \geq 2$, $0 < \varepsilon \leq 15$, $0 < \delta < 1/2$, and $T \geq \max\{d, 16\lambda^2\}$, and use the parameter settings of Appendix G.1 with $L = 4d$ (ignoring integer rounding, $4L$ divides n_k for $k \geq 2$). Then Algorithm 1 is (ε, δ) shuffle differentially private and its expected regret satisfies*

$$R_T = \widetilde{O}\Bigg(e^{2S}\kappa^{1/3}d^2T^{1/3}(\log T)^{2/3} + \frac{e^{2S}\kappa^{1/3}T^{1/3}\left(d^2(\log T)^{2/3} + d^{5/3}\log T\right)}{\varepsilon^{2/3}} + \frac{d\sqrt{T \log T}}{\sqrt{\kappa^*}} + \frac{\sqrt{T}\left(d\sqrt{\log T} + d^{3/4}(\log T)^{3/4}\right)}{\sqrt{\kappa^*\varepsilon}}\Bigg). \tag{86}$$

Proof. Privacy. The privacy argument in Appendix G.4 shows that the covariance and optimization mechanisms compose sequentially within each estimation group, while parallel composition applies across the disjoint warm-up, design, and estimation groups. Hence the complete transcript of Algorithm 1 is (ε, δ) shuffle differentially private.

Applying the abstract regret bound. As verified in Appendix G.5, the shuffle primitives satisfy Assumptions G.1 and G.2 with the values in (34) (the latter uses $\alpha \geq \sqrt{T} \geq 4\lambda$), and Lemmas F.1 and H.2 verify Assumption G.3, while (56) gives $\gamma(\lambda_+, \lambda_-) \leq \gamma$. Moreover, (53) gives the regularizer condition required by Theorem I.1. We may therefore apply that theorem.

Substituting the parameters. For the schedule of Appendix G.1, $M = O(\log\log T)$ and $\alpha = 2\sqrt{T}$. Here $\widetilde{O}(\cdot)$ treats R and S as constants, apart from the exponential factor e^S; in particular, the factor RS^2 in (39) changes only polynomial factors in R and S. Since $\sigma = \widetilde{O}(1/\varepsilon)$, equations (37), (39), and (34) give, using $d + \log T \leq 2d \log T$,

$$\begin{aligned} \gamma &= \widetilde{O}\left(\sqrt{d \log T} \left(1 + \frac{1}{\sqrt{\varepsilon}} \right) + \frac{d^{1/4} (\log T)^{3/4}}{\sqrt{\varepsilon}} \right), \\ \gamma^2 &= \widetilde{O}\left(d \log T \left(1 + \frac{1}{\varepsilon} \right) + \frac{\sqrt{d} (\log T)^{3/2}}{\varepsilon} \right). \end{aligned} \tag{87}$$

Using $(a+b)^{2/3} \leq a^{2/3} + b^{2/3}$, the two terms in (63) are consequently bounded by

$$\begin{aligned} \left(\frac{\sqrt{\kappa} e^{3S}}{S} d^2 \sqrt{T}\, \gamma^2 \right)^{2/3} &= \widetilde{O}\left(e^{2S} \kappa^{1/3} T^{1/3} \left(d^2 (\log T)^{2/3} \left(1 + \varepsilon^{-2/3} \right) + \frac{d^{5/3} \log T}{\varepsilon^{2/3}} \right) \right), \\ \sqrt{T}\, \gamma \sqrt{\frac{d}{\kappa^*}} &= \widetilde{O}\left(\frac{\sqrt{T}}{\sqrt{\kappa^*}} \left(d \sqrt{\log T} \left(1 + \varepsilon^{-1/2} \right) + \frac{d^{3/4} (\log T)^{3/4}}{\sqrt{\varepsilon}} \right) \right). \end{aligned}$$

Absorbing $M = O(\log\log T)$ and the remaining admissible logarithmic factors into $\widetilde{O}(\cdot)$ gives (86). □

K Properties of Generalized Linear Models

We now state the following key lemmas on GLMs from Appendix C of Sawarni et al. [2024].

Lemma K.1 (Self Concordance for GLMs)**.** *For distributions in the exponential family supported on $[0, R]$, the function $\mu(.)$ satisfies $|\ddot{\mu}(z)| \leq R\dot{\mu}(z)\ \forall z \in \mathbb{R}$.*

The following lemma from Sawarni et al. [2024] is also a modification of the self-concordance result in Faury et al. [2020].

Lemma K.2. *For an exponential distribution with log-partition function $b(\cdot)$, for all $z_1, z_2 \in \mathbb{R}$, letting $\mu(z) := \dot{b}(z)$, the following holds:*

$$\alpha(z_1, z_2) := \int_{v=0}^{1} \dot{\mu}(z_1 + v(z_2 - z_1)) \geq \frac{\dot{\mu}(z)}{1 + R|z_1 - z_2|} \text{ for } z \in \{z_1, z_2\} \tag{25}$$

$$\dot{\mu}(z_2) e^{-R|z_2 - z_1|} \leq \dot{\mu}(z_1) \leq e^{R|z_2 - z_1|} \dot{\mu}(z_2) \tag{26}$$

$$\tilde{\alpha}(z_1, z_2) := \int_{v=0}^{1} (1 - v) \dot{\mu}(z_1 + v(z_2 - z_1)) dv \geq \frac{\dot{\mu}(z_1)}{2 + R|z_1 - z_2|} \tag{88}$$

We also state some nice properties of GLMs.

Lemma K.3 (Properties of GLMs Sawarni et al., 2024)**.** *For any random variable r that is distributed by a canonical exponential family, we have*

1. $\mathbb{E}[r] = \mu(z) = \dot{b}(z)$
2. $\mathbb{V}[r] = \mathbb{E}[(r - \mathbb{E}[r])^2] = \dot{\mu}(z) = \ddot{b}(z)$
3. $\mathbb{E}[(r - \mathbb{E}[r])^3] = \dddot{b}(z)$

The following observation follows from the fact $\dot{b}(.)$ is monotone since $\mathbb{V}[r] = \ddot{b}(z) \geq 0$

Observation 1 (Convexity of $b(.)$)**.** *The function $b(.)$ is a convex function for every exponential family of distributions.*

L Regret and Privacy Analysis of the Joint-DP GLM Algorithm

L.1 Parameter settings

We collect here the explicit parameter settings for Algorithm 2. Define

$$\Lambda_T := \log(T/\zeta), \qquad \Lambda_\delta := \log(1/\delta).$$

The estimator regularizer λ, the confidence-width scaling β (for Criterion II), and the confidence-width scaling γ (for Criterion I) are set as

$$\lambda := \frac{5}{4} d\Lambda_T + \frac{400\sqrt{d}\, R^2\, \Lambda_\delta \log T \left(\sqrt{\Lambda_T} + \sqrt{d}\right)}{e\, \varepsilon}, \tag{89}$$

$$\beta = C \left[R\sqrt{d\Lambda_T} + \frac{R^2 S\, d^{1/4} \Lambda_T^{1/2} \Lambda_\delta^{1/2} \big(d + \Lambda_T\big)^{1/4}}{\sqrt{e\, \varepsilon}} \right] \tag{90}$$

$$\gamma = C \left[R^2 S\sqrt{d\Lambda_T} + \frac{R^3 S^2\, d^{1/4} \Lambda_T^{1/2} \big(d + \Lambda_T\big)^{1/4} \Lambda_\delta^{1/2}}{\sqrt{e\, \varepsilon}} \right]. \tag{91}$$

for some sufficiently large constant C. These choices account for the binary-tree noise, the privacy cost of switching, and the DP-PGD optimizer error; they ensure the confidence bounds remain valid under joint differential privacy.

L.2 Assumptions

We now present two assumptions that our Algorithm 2 satisfies with respect to design matrices $\mathbf{V}$ and $\{\mathbf{H}_k\}_{k=2}^M$ and the parameter estimate $\hat{\theta}_\tau$ at the end of each policy switch which would be necessary to prove the regret bounds.

Assumption L.1. With probability at least $(1 - \zeta/8)$, the following holds true for every time step $t \in [T]$.

$$\lambda_- \boldsymbol{I} \preceq \mathbf{V} - \sum_{s \in \mathcal{T}_o} x_s x_s^\top \preceq \lambda_+ \boldsymbol{I} \text{ and } \lambda_- \boldsymbol{I} \preceq \mathbf{H}_t - \sum_{s \in [t-1] \setminus \mathcal{T}_o} \frac{\dot{\mu}\left(\langle x_s, \hat{\theta}_o \rangle\right)}{e} x_s x_s^\top \preceq \lambda_+ \boldsymbol{I}$$

Assumption L.2. With probability at least $(1 - \zeta/8)$, there exist constants ν_1, ν_2 such that the DP-PGD optimizer minimizes the criterion-specific regularized losses up to errors ν_1 and ν_2 under the first and second policy switches, respectively. Formally,

$$\left\| \mathcal{L}^{(1)}_{\mathcal{T}_o}(\hat{\theta}; \lambda) - \min_{\theta \in \Theta} \mathcal{L}^{(1)}_{\mathcal{T}_o}(\theta; \lambda) \right\| \leq \nu_1 \text{ and } \left\| \mathcal{L}^{(2)}_{[t-1] \setminus \mathcal{T}_o}(\hat{\theta}; \lambda) - \min_{\theta \in \Theta} \mathcal{L}^{(2)}_{[t-1] \setminus \mathcal{T}_o}(\theta; \lambda) \right\| \leq \nu_2 \tag{92}$$

where for any index set $\mathcal{T} \subseteq [T]$,

$$\mathcal{L}^{(i)}_{\mathcal{T}}(\theta; \lambda) := \sum_{s \in \mathcal{T}} \ell(\theta, x_s, r_s) + \frac{\lambda}{2} \|\theta\|_2^2, \quad i \in \{1, 2\}.$$

We first verify that our algorithm satisfies the assumptions stated above with suitable values of ν_1 and ν_2, while ensuring (ε, δ)-JDP as shown in Appendix L.3. Since a user's own context decides whether that user is an exploration round, it also decides which binary tree and which optimizer sequence receive the user's data. We therefore show that each of the two trees (Lemma L.4) and each of the

two optimizer sequences (Lemma L.6) is $\rho_{\text{tot}}/4$-zCDP with respect to a change of *any* single user's data, given the past releases, where

$$\rho_{\text{tot}} := \left(\sqrt{\log(1/\delta) + \varepsilon} - \sqrt{\log(1/\delta)}\right)^2$$

is the zCDP budget that converts to exactly (ε, δ)-DP [Bun and Steinke, 2016]. The switching set $\mathcal{T}_o \setminus \{t\}$ is a post-processing of the tree releases (Appendix T), and adaptive zCDP composition together with the billboard lemma gives (ε, δ)-JDP. Finally, in Appendix M, we bound the regret of Algorithm 2 under these assumptions, thereby completing the joint privacy and regret analysis of our algorithm.

Remark L.3 (Privatization of policy switches). A policy II switch is computed solely from already privatized summaries (the noisy matrices $\mathbf{H}_t$, line 6). A policy I switch at round t also reads the raw context set $\mathcal{X}_t$ (line 9), but only that of user t itself, together with the released $\mathbf{V}_t$. Hence, for any other user $i \neq t$, the decision at round t is a post-processing of the tree releases and of data that does not change between neighbors at i; no separate privacy budget is needed for either switch.

One might also ask why we do not use the sparse vector technique (Section 3.6 of Dwork and Roth [2014]) for policy I switches. Since the switching decisions are already free post-processing, the sparse vector technique would only add cost: the error tolerance in Theorem 3.26 of Dwork and Roth [2014] grows as $\sqrt{c}$, where c is the number of queries above the threshold, which would adversely affect the regret bounds.

L.3 Privacy Analysis of the Joint-DP GLM Algorithm

In this section, we aim to show that our algorithm is (ε, δ) joint differentially private and satisfy the assumptions mentioned above.

Lemma L.4. *Let $\mathbf{V}_t$ denote the value of $\mathbf{V}$ at time step t and let $\mathcal{T}_{o,t}$ denote the value of $\mathcal{T}_o$ in Algorithm 2 at time step t. Given the past releases, each of the two trees releasing $\{\mathbf{V}_t\}_{t=1}^T$ and $\{\mathbf{H}_t\}_{t=1}^T$ is $\rho_{\text{tot}}/4$-zCDP with respect to a change of any single user's data $(r_t, \mathcal{X}_t)$, whichever tree that user's data enters (including a change of the user's leaf between zero and nonzero); hence the two trees are jointly $\rho_{\text{tot}}/2$-zCDP. If, moreover, $\varepsilon \leq 8\log(1/\delta)$, the releases satisfy Assumption L.1.*

Tree-based mechanism [Chan et al., 2011, Dwork et al., 2010]: It is used to compute differentially private prefix sums of a sequence (data-set) $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$. To do this, we first construct a balanced binary tree T with $\{x_1, x_2. \ldots, x_n\}$ as the leaf nodes and each internal node y in tree T stores the sum of all leaf nodes rooted at y. First notice that one can compute any partial sum $\sum_{j=1}^{i} x_j$ using at most $m := \lceil \log(n) + 1 \rceil$ nodes of T. Second, notice that for any two neighboring data sequences D and D' the partial sums stored at no more than m nodes in T are different. Thus, if the count in each node preserves $(\varepsilon_0, \delta_0)$-differential privacy, using the advanced composition of Dwork and Roth [2014] we get that the entire algorithm is $(O(m\varepsilon_0^2 + \varepsilon_0\sqrt{2m\ln(1/\delta')}), m\delta_0 + \delta')$-differentially private. Alternatively, one can ensure that the entire tree is (ε, δ)-differentially private via standard advanced composition by setting $\varepsilon_0 = \varepsilon/\sqrt{8m\ln(2/\delta)}$ and $\delta_0 = \delta/(2m)$, with the composition slack parameter $\delta' = \delta/2$.

However, this choice is somewhat conservative and is best suited to the small-ε regime, where advanced composition gives its sharpest guarantees. Since we do not assume $\varepsilon \leq 1$, we instead use privacy accounting via zero-concentrated differential privacy (zCDP) [Bun and Steinke, 2016], which is particularly convenient for Gaussian mechanisms. Specifically, ρ-zCDP implies $(\rho + 2\sqrt{\rho\log(1/\delta)}, \delta)$-DP for every $\delta > 0$. We allot each of the two trees the budget $\rho_{\text{tot}}/4$ with respect to a change of a single leaf. A leaf enters at most m nodes, and zCDP composes additively, so it is enough that each node be $\rho_{\text{tot}}/(4m)$-zCDP.

We now present how we instantiate the binary tree to compute the noise term $\mathcal{R}_t$ at every step t (Line 10 and Line 14). The construction is almost identical to Section 4.2 of Shariff and Sheffet [2018].

Proof. We first establish the per-leaf sensitivity used to calibrate the Gaussian noise. For any two adjacent datasets differing in user i, only leaf i of each tree can differ given the past releases (for $j \neq i$, leaf j and the tree it enters are functions of the past releases and of $(r_j, \mathcal{X}_j)$). Leaf i of the

V-tree is 0 or $x_i x_i^\top$ and leaf i of the **H**-tree is 0 or $\frac{\dot\mu(\langle x_i,\hat\theta_o\rangle)}{e} x_i x_i^\top$, so any change of leaf i in either tree, including between zero and nonzero, has Frobenius norm at most $\sqrt{2}\max\{1, R^2/(4e)\} \le \frac{4R^2}{e}$ (using $\dot\mu \le R^2/4$ and $R \ge 1$); we use $\Delta := \frac{4R^2}{e}$[13]. This sensitivity bound is deterministic and, in particular, holds without conditioning on the event $\mathcal{E}_o$.

We instantiate the above tree-based algorithm with symmetric Gaussian noise at every node: sample $Z' \in \mathbb{R}^{d\times d}$ with each $Z'_{i,j} \sim N(0, \sigma^2_{\text{noise}})$ i.i.d. and symmetrize to get $Z = (Z' + Z'^\top)/2$ which we add to every node. Further, we construct two separate trees for computing the summation $\sum_{i\in\mathcal{T}_{o,t}} x_i x_i^\top$ and $\sum_{i\in[t]\setminus\mathcal{T}_{o,t}} \frac{\dot\mu(\langle x_i,\hat\theta_o\rangle)}{e} x_i x_i^\top$ for every time step $t \in [T]$ and $\mathcal{T}_{o,t}$ denotes the value of set $\mathcal{T}_o$ at a time-step t. The j^{th} leaf node at the first tree stores $x_j x_j^\top$ if $j \in \mathcal{T}_o$ and 0 otherwise, while the j^{th} leaf node at the second tree stores $\frac{\dot\mu(\langle x_j,\hat\theta_o\rangle)}{e} x_j x_j^\top$ if $j \notin \mathcal{T}_o$ and 0 otherwise. Using the sensitivity bound $\frac{4R^2}{e}$ established above, each node is made $\rho_{\text{tot}}/(4m)$-zCDP by the per-coordinate variance $\sigma^2_{\text{noise}} = \Delta^2/(2\cdot\rho_{\text{tot}}/(4m)) = \frac{32\,m\,R^4}{e^2\rho_{\text{tot}}}$. When $\varepsilon \le 8\log(1/\delta)$, we have $\rho_{\text{tot}} = \varepsilon^2/(\sqrt{\log(1/\delta)+\varepsilon} + \sqrt{\log(1/\delta)})^2 \ge \varepsilon^2/(36\log(1/\delta)) \ge \varepsilon^2/(36(\log(32/\delta))^2)$, hence $\sigma^2_{\text{noise}} \le \frac{1152\,m\,R^4(\log(32/\delta))^2}{e^2\varepsilon^2}$; the error bounds below use this upper bound. Since each leaf $x_j x_j^\top$ (resp. $\frac{\dot\mu(\langle x_j,\hat\theta_o\rangle)}{e} x_j x_j^\top$) depends on earlier releases only through the arm chosen from $\mathcal{X}_j$, each node is a Gaussian mechanism applied to an adaptively chosen input of bounded sensitivity, and the per-node guarantees combine by adaptive composition of zCDP [Bun and Steinke, 2016]. When all entries on Z are sampled from $\mathcal{N}(0,1)$, known concentration results [Vershynin, 2018] on the top singular value of Z give that $\mathbb{P}[\|Z\| > (4\sqrt{d} + \sqrt{2\ln(2T/\alpha)})] < \alpha/2T$. For every time $t \in \mathcal{T}_o$, the cumulative noise matrix $\sum_{i\in\mathcal{T}_o} \mathcal{R}_i$ is a sum of at most m such matrices, thus the variance of each coordinate is $m\sigma^2_{\text{noise}}$. A similar argument can also be presented for time steps $t \notin \mathcal{T}_{o,t}$.

For every time step $t \in [T]$, we can argue with probability at least $1 - \zeta/8$.

$$\left\| \mathbf{V}_t - \lambda \boldsymbol{I} - \sum_{s\in\mathcal{T}_o} x_s x_s^\top \right\| \le \sigma_{\text{noise}}\sqrt{2m}\left(4\sqrt{d} + \sqrt{2\ln(2n/\alpha)}\right) \tag{93}$$

$$\le \frac{4\sqrt{72}\,R^2\log(32/\delta)\log T}{e\,\varepsilon}\left(4\sqrt{d} + 2\sqrt{\log(16T/\zeta)}\right) \tag{94}$$

Similarly,

$$\left\| \mathbf{H}_t - \lambda \boldsymbol{I} - \sum_{s\in[t-1]\setminus\mathcal{T}_o} \frac{\dot\mu\left(\langle x_s,\hat\theta_o\rangle\right)}{e} x_s x_s^\top \right\| \le \frac{4\sqrt{72}\,R^2\log(32/\delta)\log T}{e\,\varepsilon}\left(4\sqrt{d} + 2\sqrt{\log(16T/\zeta)}\right) \tag{95}$$

Using the definition of λ from Equation (89), the equation above proves that Assumption L.1 is satisfied with $\lambda_- = \frac{4\sqrt{d}}{4\sqrt{d}+1}\lambda$ and $\lambda_+ = \frac{4\sqrt{d}+2}{4\sqrt{d}+1}\lambda$.

Since the per-leaf sensitivity Δ holds deterministically (without conditioning on $\mathcal{E}_o$) and covers every change of leaf i, each tree is $\rho_{\text{tot}}/4$-zCDP with respect to a change of any single user's data, given the past releases, by adaptive zCDP composition over the at most m nodes containing leaf i; since both trees can change for the same user, they are jointly $\rho_{\text{tot}}/2$-zCDP. No ζ-slack from the failure event is needed. □

Lemma L.5. *With probability at least $1 - \zeta/8$, Criterion I is triggered at most N_o times and the updates of $\hat\theta_o$ and $\hat\theta_\tau$ are triggered at most count1$_{\text{cutoff}}$ and count2$_{\text{cutoff}}$ times, so the hard caps are not reached.*

[13] Recall that $\dot\mu(\langle x,\theta_1\rangle) = \mathrm{Var}(r \mid x;\theta_1)$. Since the reward lies in $[0,R]$, this variance is bounded by $R^2/4$ (Popoviciu's inequality), hence $\frac{\dot\mu(\langle x,\hat\theta_o\rangle)}{e}\|xx^\top\|_{\text{op}} \le \frac{R^2}{4e} \le \frac{4R^2}{e}$.

This is Lemma T.2, proved in Appendix T: the count1$_{\text{cutoff}}$ and count2$_{\text{cutoff}}$ bounds follow from Lemma S.5, and the N_o bound from the elliptic potential lemma, Lemma S.1. The same section explains why the switching set $\mathcal{T}_o \setminus \{i\}$ is a post-processing of the tree releases and needs no separate privacy budget.

The following lemma establishes privacy and utility guarantees for the DP-PGD optimizer used under Criterion I and Criterion II in Algorithm 2.

Lemma L.6. *Let $\hat{\theta}_{o,t}$ denote the value of estimator $\hat{\theta}_o$ at time step t and $\hat{\theta}_{\tau,t}$ denote the value of estimator $\hat{\theta}_\tau$ at time step t. Assume that R and S are bounded below by positive absolute constants. Then, given the past releases, the sequence of distinct estimators in $\{\hat{\theta}_{o,t}\}_{t=1}^T$ and the sequence of distinct estimators in $\{\hat{\theta}_{\tau,t}\}_{t=1}^T$ are each $\rho_{\text{tot}}/4$-zCDP with respect to a change of any single user's data $(r_t, \mathcal{X}_t)$, whichever sequence that user's data enters. Moreover, if $\varepsilon \le 8\log(1/\delta)$ and every call runs the summed-objective optimizer of Lemma X.5 with public hyper-parameters, the estimators $\hat{\theta}_o$ and $\hat{\theta}_\tau$ satisfy Assumption L.2 with*

$$\nu_1 = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\Lambda_T \log(8/\delta)}\sqrt{d+\Lambda_T}\right), \tag{96}$$

$$\nu_2 = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\log(8/\delta)\log_2\left(1+\frac{TR^3}{d}\right)}\sqrt{d+\Lambda_T}\right). \tag{97}$$

Proof. We prove the claim for $\{\hat{\theta}_{o,t}\}_{t=1}^T$. The proof for $\{\hat{\theta}_{\tau,t}\}_{t=1}^T$ is the same, with count1$_{\text{cutoff}}$ replaced by count2$_{\text{cutoff}}$ and the corresponding loss parameters.

Optimizer instantiation. Every call runs the noisy PGD of Lemma X.5 on the *summed* regularized loss of Assumption L.2. For the call that updates $\hat{\theta}_o$ at round τ_o,

$$F(\theta) = \sum_{s\in\mathcal{T}_o\cap[\tau_o-1]} \ell(\theta, x_s, r_s) + \frac{\lambda_+}{2}\|\theta\|_2^2, \qquad \Theta = \Theta_I.$$

For the call that updates $\hat{\theta}_\tau$ at round τ, the sum runs over $[\tau-1]\setminus\mathcal{T}_o$ and $\Theta = \Theta_{II}$. The size of either dataset depends on whether individual users explored, so it is private. The hyper-parameters (K, σ, η) are therefore set through the *public* bound $N := \tau_o - 1$ (resp. $N := \tau - 1$) on the dataset size. They also depend on the per-example Lipschitz constant $L_0 := R + RS$ of ℓ on $\{\|\theta\|_2 \le S\}$, the Lipschitz constant $G_r := \lambda_+ S$ of the regularizer, the diameter $D = 2S$, the per-call budget, and the per-call confidence level. All of these are public. The constraint set Θ_{II} depends on the data only through $\hat{\theta}_o$ and $\mathbf{V}_{\tau_o}$, which have already been released. So, given the past releases, the same mechanism is applied under any two neighbouring datasets.

Privacy. Let $\rho_\star := \rho_{\text{tot}}/4$. Because of the hard cap in Algorithm 2, for every dataset there are at most

$$\text{count1}_{\text{cutoff}} = 4d\log_2(1+T/d)$$

optimizer calls producing the distinct values in $\{\hat{\theta}_{o,t}\}_{t=1}^T$. Their times are a function of the released $\mathbf{V}_t$ only (line 3). Each call is run with budget $\rho_\star/\text{count1}_{\text{cutoff}}$. Changing user i's data changes the dataset of a call in one of two ways. If i's exploration indicator is unchanged, the point (r_i, x_i) is replaced. If the indicator changes, the point is added to or removed from the dataset. In both cases Lemma X.5 applies, since its sensitivity bound $2L_0$ does not depend on the dataset size. So each call is $\rho_\star/\text{count1}_{\text{cutoff}}$-zCDP with respect to user i, given the past releases. Adaptive composition of zCDP over the at most count1$_{\text{cutoff}}$ calls shows that the sequence is $\rho_\star$-zCDP.

Utility. Run each call with confidence level $\zeta_o := \zeta/(16\text{count1}_{\text{cutoff}})$ (resp. $\zeta_\tau := \zeta/(16\text{count2}_{\text{cutoff}})$). A union bound over at most count1$_{\text{cutoff}}$+count2$_{\text{cutoff}}$ calls shows that all calls succeed simultaneously with probability at least $1-\zeta/8$, as Assumption L.2 requires. Since $d \le T$, $\log(2/\zeta_o)$ and $\log(2/\zeta_\tau)$ are $O(\Lambda_T)$. Lemma X.5 bounds the error of the *summed* objective directly, which is the quantity in Assumption L.2. Since $DL_0 = 2S(R+RS) = O(RS^2)$ when R, S are bounded below by absolute constants,

$$\nu_1 = O\left(RS^2\sqrt{\frac{\text{count1}_{\text{cutoff}}\,(d+\Lambda_T)}{\rho_\star}}\right),$$

since the budget of the $\hat{\theta}_o$ sequence is split over the at most $\text{count1}_{\text{cutoff}}$ optimizer calls allowed by the hard cap. Note that the relevant count is the number of *optimizer calls* $\text{count1}_{\text{cutoff}}$, not the number of exploration rounds N_o: $\hat{\theta}_o$ is recomputed only when the public check $\mathbf{V}_t \not\preceq 2\mathbf{V}_{\tau_o}$ fires, not at every exploration round. Bounding the error with $N_o = 16dR^2\kappa\gamma^2 \log T$ instead would still be valid, but it would put γ and κ into ν_1, make the choice of γ self-referential, and inflate γ by an additive $O(d\sqrt{\kappa}/\varepsilon)$ term. When $\varepsilon \le 8\log(1/\delta)$,

$$\rho_\star = \frac{\varepsilon^2}{4\big(\sqrt{\log(1/\delta)+\varepsilon}+\sqrt{\log(1/\delta)}\big)^2} \ge \frac{\varepsilon^2}{144\log(1/\delta)} \ge \frac{\varepsilon^2}{144\log(8/\delta)}.$$

Substituting this and $\text{count1}_{\text{cutoff}} = 4d\log_2(1+T/d) = O(d\Lambda_T)$ (as $T/d \le T/\zeta$) gives

$$\nu_1 = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\Lambda_T\log(8/\delta)}\sqrt{d+\Lambda_T}\right).$$

For $\{\hat{\theta}_{\tau,t}\}_{t=1}^T$, the hard cap again bounds the number of calls by $\text{count2}_{\text{cutoff}} = 4d\log_2(1+TR^3/d)$, each with budget $\rho_\star/\text{count2}_{\text{cutoff}}$. The diameter-times-Lipschitz quantity is again $O(RS^2)$, so

$$\nu_2 = O\left(RS^2\sqrt{\frac{\text{count2}_{\text{cutoff}}\,(d+\Lambda_T)}{\rho_\star}}\right) = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\log(8/\delta)\log_2\left(1+\frac{TR^3}{d}\right)}\sqrt{d+\Lambda_T}\right).$$

Combining the privacy and utility guarantees for the two estimator sequences proves the lemma. □

We now argue below that Algorithm 2 is (ε,δ) jointly differentially private.

Lemma L.7. *Algorithm 2 is (ε,δ) joint differentially private.*

Proof. Consider two datasets $\mathcal{D},\mathcal{D}'$ over reward–context pairs that differ only at time step i. Order all releases of Algorithm 2 in time: the nodes of both binary trees and the outputs of both optimizer sequences. For $j \ne i$, every input that user j contributes to a release (the leaf j of either tree, which tree it enters, the arm x_j, and the data point (r_j,x_j) entering an optimizer call) is a deterministic function of the preceding releases and of $(r_j,\mathcal{X}_j)$, which coincide under $\mathcal{D}$ and $\mathcal{D}'$. Hence, conditional on the preceding releases, the input of each release differs between $\mathcal{D}$ and $\mathcal{D}'$ only through user i. Note that whether user i is an exploration round depends on $\mathcal{X}_i$, so user i's data may enter *both* trees and *both* optimizer sequences across the two datasets; we therefore charge every piece for user i. By Lemma L.4, the two trees are each $\rho_{\text{tot}}/4$-zCDP with respect to user i, and by Lemma L.6 (whose call counts are capped deterministically), so are the two optimizer sequences. Adaptive composition of zCDP [Bun and Steinke, 2016] shows that the full transcript $\{\mathbf{V}_t,\mathbf{H}_t,\hat{\theta}_{o,t},\hat{\theta}_{\tau,t}\}_{t=1}^T$ is ρ_{tot}-zCDP with respect to user i, hence (ε,δ)-indistinguishable between $\mathcal{D}$ and $\mathcal{D}'$ by the choice of ρ_{tot}. The switching set $\mathcal{T}_o \setminus \{i\}$ is a post-processing of this transcript (Appendix T). Finally, the action at each round $t \ne i$ is a function of the transcript and of user t's own context $\mathcal{X}_t$, so the billboard lemma [Kearns et al., 2014] gives (ε,δ)-joint differential privacy.

□

Now an application of Theorem M.1, we prove the following regret and privacy bound.

Theorem L.8. *Algorithm 2 satisfies (ε,δ) joint differential privacy. Assume further $\zeta < \delta \le \min(1/e, d/R^3)$, $d \le T$, and $\varepsilon \le 8\log(1/\delta)$. Then with probability at least $1-\zeta$, the regret R^T may be bounded as*

$$R^T = O\left(\left[Rd\sqrt{\Lambda_T}\sqrt{\frac{T}{\kappa^*}} + \frac{R^2S\,d^{3/4}\Lambda_T^{1/2}\sqrt{T}\Lambda_\delta^{1/2}\big(d+\Lambda_T\big)^{1/4}}{\sqrt{e\,\varepsilon\,\kappa^*}}\right]\sqrt{\log(RT)}\right)$$
$$+ O\left(\kappa dR^3\left[R^2S\sqrt{d\Lambda_T} + \frac{R^3S^2\,d^{1/4}\Lambda_T^{1/2}\big(d+\Lambda_T\big)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right]^2\log T\right) \quad (98)$$

where $\Lambda_T = \log(T/\zeta)$ and $\Lambda_\delta = \log(1/\delta)$.

Observation 2. *When $\log T \geq d$, the dominant term in the regret is $\widetilde{O}\Big(d\sqrt{T/\kappa^*}\log T + \frac{d^{3/4}\sqrt{T}}{\sqrt{\varepsilon\kappa^*}}(\log T)^{5/4}\Big)$. The latter term in the regret can be thought of as the additive cost for privacy. This term does not match the $\Omega(d/(\varepsilon+\delta))$ lower bound of Appendix U; even for the simpler case of linear contextual bandits we do have a similar additive term for privacy; see Shariff and Sheffet [2018], Azize and Basu [2024b].*

When $\log T < d$, the dominant term in the regret scales as $\widetilde{O}\Big(d\sqrt{T/\kappa^*}\log T + \frac{d\sqrt{T}}{\sqrt{\varepsilon\kappa^*}}(\log T)\Big)$.

Proof. Since $\delta \leq \min\left\{\frac{d}{R^3}, \frac{1}{e}\right\}$, we have $\Lambda_\delta \geq 1$, and $\log\left(1+\frac{TR^3}{d}\right) \leq O(\Lambda_T)$. Indeed, $\delta \leq d/R^3$ implies $TR^3/d \leq T/\delta$, and since $\zeta \leq \delta$, we have $\log(T/\delta) \leq \log(T/\zeta) = \Lambda_T$.

Recall that

$$\lambda_+ = \frac{4\sqrt{d}+2}{4\sqrt{d}+1}\lambda, \qquad \lambda_- = \frac{4\sqrt{d}}{4\sqrt{d}+1}\lambda.$$

Thus $\lambda_+/\lambda_- = O(1)$. From the definition of λ, and absorbing absolute constants inside the $O(\cdot)$ notation, we have

$$\sqrt{\lambda_+} \leq C\left[\sqrt{d\Lambda_T} + \frac{R\,d^{1/4}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right].$$

Here we used $\log T \leq \Lambda_T$, $\log(32/\delta) = O(\Lambda_\delta)$, $\log(16T/\zeta) = O(\Lambda_T)$, $4\sqrt{d}+1 = O(\sqrt{d})$, and

$$\sqrt{\Lambda_T} + \sqrt{d} = O\left(\sqrt{d+\Lambda_T}\right).$$

Moreover, the first term in the definition of λ gives

$$\lambda_- \geq c\,d\Lambda_T$$

for a universal constant $c > 0$.

We first verify $\gamma(\lambda_+) \leq \gamma$. By Equation (99),

$$\gamma(\lambda_+) \leq 4R^2S\sqrt{\lambda_+} + \frac{7Sd}{\sqrt{\lambda_+}}\log\left(\frac{16T}{\zeta R^2 d\lambda_+}\right) + 4RS\sqrt{\nu_1}.$$

Using $\lambda_+/\lambda_- = O(1)$, $\lambda_- \geq cd\Lambda_T$, and the lower bounds on R, S, together with the privacy-noise component $\lambda_+ = \Theta\big(\frac{R^2\sqrt{d}\,\Lambda_\delta\log T(\sqrt{\Lambda_T}+\sqrt{d})}{e\,\varepsilon}\big)$ of Equation (89), the first two terms are bounded by

$$CR^2S\sqrt{d\Lambda_T} + C\frac{R^3S\,d^{1/4}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}.$$

From the optimizer guarantee in Lemma L.6,

$$\nu_1 = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\Lambda_T\log(8/\delta)}\sqrt{d+\Lambda_T}\right).$$

Hence, using $\log(8/\delta) = O(\Lambda_\delta)$,

$$4RS\sqrt{\nu_1} \leq C\frac{R^{3/2}S^2d^{1/4}\Lambda_T^{1/4}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/4}}{\sqrt{\varepsilon}} \leq C\frac{R^3S^2\,d^{1/4}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}},$$

where the last step uses $\Lambda_T, \Lambda_\delta \geq 1$ and that R is bounded below by an absolute constant. This has the same order as the binary-tree term above. Using also $S = O(S^2)$ (as S is bounded below by an absolute constant), we get

$$\gamma(\lambda_+) \leq C\left[R^2S\sqrt{d\Lambda_T} + \frac{R^3S^2\,d^{1/4}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right] \leq \gamma$$

for the choice in (91), with a sufficiently large constant C. The right-hand side does not depend on γ, so no fixed-point argument is needed.

We now verify $\beta(\lambda_+, \lambda_-) \le \beta$. By Equation (100),

$$\beta(\lambda_+, \lambda_-) = 4R\frac{\lambda_+}{\sqrt{\lambda_-}} + \frac{7d}{R\sqrt{\lambda_-}} \log\left(\frac{16T}{\zeta R^2 d\lambda_-}\right) + 10\sqrt{\nu_2}.$$

Using $\lambda_+/\lambda_- = O(1)$, $\lambda_- \ge cd\Lambda_T$, and the bound on $\sqrt{\lambda_+}$ from the updated Equation (89), the first two terms are at most

$$CR\sqrt{d\Lambda_T} + C\frac{R^2\, d^{1/4}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}.$$

The optimizer error gives

$$\nu_2 = O\left(\frac{RS^2}{\varepsilon}\sqrt{d\log(8/\delta)\log\left(1+\frac{TR^3}{d}\right)}\sqrt{d+\log T}\right).$$

Using $\log\big(1+TR^3/d\big) \le O(\Lambda_T)$ and $\log(8/\delta) = O(\Lambda_\delta)$, we get

$$10\sqrt{\nu_2} \le CS\sqrt{\frac{R}{\varepsilon}} d^{1/4}\Lambda_\delta^{1/4}\Lambda_T^{1/4}(d+\Lambda_T)^{1/4}.$$

Since R, S are bounded below by positive absolute constants and $\Lambda_\delta \ge 1$, both the second term above and the contribution from $\sqrt{\nu_2}$ are bounded by

$$C\frac{R^2 S\, d^{1/4}\Lambda_T^{1/2}\Lambda_\delta^{1/2}(d+\Lambda_T)^{1/4}}{\sqrt{e\,\varepsilon}}.$$

Therefore

$$\beta(\lambda_+, \lambda_-) \le C\left[R\sqrt{d\Lambda_T} + \frac{R^2 S\, d^{1/4}\Lambda_T^{1/2}\Lambda_\delta^{1/2}(d+\Lambda_T)^{1/4}}{\sqrt{e\,\varepsilon}}\right],$$

which is the stated choice of β.

Further, we show in Lemma S.5 that Criterion I and Criterion II in Algorithm 2 are invoked at most N_o and count2$_{\text{cutoff}}$ times, respectively. We can now directly instantiate Theorem M.1 for our case to prove the desired regret bound. □

M Regret Analysis under the Joint-DP Assumptions

Before stating the main lemma, we define the parametrized form of β and γ in terms of the regularizer bounds λ_+ and λ_- that we use extensively in the proof and theorem statement.

$$\gamma(\lambda_+) = 4R^2 S\sqrt{\lambda_+} + \frac{7Sd}{\sqrt{\lambda_+}} \log\left(\frac{16T}{\zeta R^2 d\lambda_+}\right) + 4RS\sqrt{\nu_1} \tag{99}$$

$$\beta(\lambda_+, \lambda_-) = 4R\frac{\lambda_+}{\sqrt{\lambda_-}} + \frac{7d}{R\sqrt{\lambda_-}} \log\left(\frac{16T}{\zeta R^2 d\lambda_-}\right) + 10\sqrt{\nu_2} \tag{100}$$

Theorem M.1. *When the events in Assumption L.1 and Assumption L.2 hold, Criterion I and Criterion II are invoked at most N_o and count2$_{\text{cutoff}}$ times, respectively, and the parametric conditions $\gamma(\lambda_+) \le \gamma$ and $\beta(\lambda_+, \lambda_-) \le \beta$ (as defined in Equation* (100) *and Equation* (99)*) are satisfied, the regret R_T can be bounded by some universal constant C with probability at least $1-\zeta$.*

$$R_T \le C\sqrt{d}\beta\sqrt{\log_2\left(\frac{RT}{d\lambda_-}\right)}\sqrt{\sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right)} + 8dR^3\kappa\gamma^2\log T \tag{101}$$

The proof proceeds in several steps. We first establish confidence bounds under both switching and non-switching cases in Appendices N and O. We then use these bounds to control the instantaneous regret (Appendix N), and finally combine all components to prove the main theorem in Appendix Q.

A key technical ingredient is the policy switching criterion, which compares the current noisy design matrix with the one at the last policy update ($\mathbf{H}_t \not\preceq 2\mathbf{H}_\tau$). To analyze this, we prove Lemma S.5, which bounds the total number of switching steps. This result is subsequently applied in the instantaneous regret analysis (Lemma P.3).

We first define the (scaled) data matrix.

$$\mathbf{H}_o^* = \sum_{s \in \mathcal{T}_o \cap [\tau_o - 1]} \left(\dot{\mu}\left(\langle x_s, \theta^* \rangle \right) x_s x_s^\top \right) + \lambda_+ \boldsymbol{I}, \tag{102}$$

where τ_o is the last update time of $\hat{\theta}_o$ (line 3), and observe that $\kappa \mathbf{H}_o^* \succeq \mathbf{V}_{\tau_o}$ under Assumption L.1.

In Theorem L.8, we shall show that for the chosen value of γ and λ in Algorithm 2, this parametrized form is upper bounded by γ.

We now state the main concentration lemma from Theorem 1 of Faury et al. [2020].

Lemma M.2. *Let $\{\mathcal{F}_t\}_{t=1}^\infty$ be a filtration. Let $\{x_t\}_{t=1}^\infty$ be a stochastic process in $\mathcal{B}_2(d)$ such that x_t is $\mathcal{F}_t$ measurable. Let $\{\eta_t\}_{t=1}^\infty$ be a martingale difference sequence such that η_t is $\mathcal{F}_t$ measurable. Furthermore, assume that $|\eta_t| < 1$ almost surely and denote $\sigma_t^2 = Var(\eta_t \mid \mathcal{F}_{t-1})$. Let $\lambda > 0$ and for any $t \geq 1$ define:*

$$S_t = \sum_{s=1}^{t-1} \eta_s x_s \quad \mathbf{H}_t = \sum_{s=1}^{t-1} \sigma_s^2 x_s x_s^\top + \lambda \boldsymbol{I}$$

Then for any $\delta \in (0, 1]$,

$$\mathbb{P}\left(\exists t \geq 1 : ||S_t||_{\mathbf{H}_t^{-1}} \geq \frac{\sqrt{\lambda}}{2} + \frac{2}{\sqrt{\lambda}} \log\left(\frac{det(\mathbf{H}_t)^{1/2}}{\lambda^{d/2}\delta} \right) + \frac{2}{\sqrt{\lambda}} d \log(2) \right) \leq \delta$$

We now state the following concentration bound which is a variant of Lemma B.2 of Sawarni et al. [2024].

N Confidence Bounds under switching criterion I bounds

Lemma N.1. *In any round t, let $\hat{\theta}_o$ be the current estimate, computed at its last update time $\tau_o \leq t$ (line 3) from the rewards in rounds $\mathcal{T}_o \cap [\tau_o - 1]$, and let $\mathbf{H}_o^*$ be defined with the same rounds. Then with probability $1 - \frac{\zeta}{8}$, simultaneously for all t, we have*

$$||\hat{\theta}_o - \theta^*||_{\mathbf{H}_o^*} \leq \gamma(\lambda_+)$$

We now define $\tilde{\theta}_o$ as follows.

$$\tilde{\theta}_o = \arg\min_\theta \sum_{t \in \mathcal{T}_o} \ell(\theta, r_t, x_t) + \frac{\lambda_+}{2} ||\theta||_2^2 \text{ and } g_o(\theta) = \sum_{s \in \mathcal{T}_o} \mu(\langle \theta, x_s \rangle) x_s + \lambda_+ \theta \tag{103}$$

To prove this lemma, we first state the following lemma which is a variant of Lemma B.2 of Sawarni et al. [2024]. We refrain from proving this lemma again as it follows the proof of Lemma B.2 of Sawarni et al. [2024].

Lemma N.2. *At any round t, let $\tilde{\theta}_o$ denote the minimiser computed above. Then with probability at least $(1 - \zeta/16)$, we have*

$$||g_o(\tilde{\theta}_o) - g_o(\theta^*)||_{(\mathbf{H}_o^*)^{-1}} \leq 4R\sqrt{\lambda_+} + \frac{7d}{R\sqrt{\lambda_+}} \log\left(16/\zeta\left(1 + \frac{T}{R^2 d\lambda_+}\right)\right)$$

Using Assumption L.2 and Lemma N.2 and instantiating Lemma V.1 (with $c = 2RS$), we can prove Lemma N.1.

O Confidence Bounds under non-Switching Criterion I rounds

Let $\mathcal{E}_o$ denote the event in Lemma N.1. We now prove the following lemma from Lemma B.3 of Sawarni et al. [2024] in our setup.

Lemma O.1. *If in round t, the switching criterion I is not satisfied, event $\mathcal{E}_o$ holds, and the cap count1$_{cutoff}$ on updates of $\hat{\theta}_o$ has not been reached (Lemma T.2), we have $\left|\langle x, \hat{\theta}_o - \theta^*\rangle\right| \leq \frac{1}{R}$ for all $x \in \mathcal{X}_t$ with probability at least $1 - \zeta/8$.*

Proof.

$$\left|\langle x, \hat{\theta}_o - \theta^*\rangle\right| \leq ||x||_{(\mathbf{H}_o^*)^{-1}} \left|\left|\hat{\theta}_o - \theta^*\right|\right|_{\mathbf{H}_o^*} \quad \text{(Cauchy–Schwarz)} \tag{104}$$

$$\leq ||x||_{(\mathbf{H}_o^*)^{-1}} \gamma(\lambda_+) \quad \text{(via Lemma N.1)} \tag{105}$$

$$\leq ||x||_{\mathbf{V}_{\tau_o}^{-1}} \gamma(\lambda_+)\sqrt{\kappa} \quad (\text{as } \mathbf{V}_{\tau_o} \preceq \kappa \mathbf{H}_o^*) \tag{106}$$

$$\leq \sqrt{2}\, ||x||_{\mathbf{V}_t^{-1}} \gamma(\lambda_+)\sqrt{\kappa} \quad (\text{as } \mathbf{V}_t \preceq 2\mathbf{V}_{\tau_o}, \text{ since line 3 did not trigger after } \tau_o) \tag{107}$$

$$\leq \frac{\sqrt{2}}{\sqrt{2\kappa R^2\gamma^2}} \gamma(\lambda_+)\sqrt{\kappa} \quad \text{(switching criterion I not satisfied)} \tag{108}$$

$$\leq \frac{1}{R} \quad \text{as } \gamma(\lambda_+) \leq \gamma. \tag{109}$$

□

Further, define

$$\mathbf{H}_t^{(u),*} = \sum_{s \in [t-1] \setminus \mathcal{T}_o} \dot{\mu}\left(\langle x_s, \theta^*\rangle\right) x_s x_s^\top + \lambda_+ \boldsymbol{I} \text{ and } \mathbf{H}_t^{(l),*} = \sum_{s \in [t-1] \setminus \mathcal{T}_o} \dot{\mu}\left(\langle x_s, \theta^*\rangle\right) x_s x_s^\top + \lambda_- \boldsymbol{I}$$

We can prove the following corollary using self-concordance property and the lemma above. A similar proof is presented in Corollary B.4 of Sawarni et al. [2024]

Corollary O.2. *Under event $\mathcal{E}_o$, $\mathbf{H}_t \preceq \mathbf{H}_t^{(u),*}$ and $\mathbf{H}_t^{(l),*} \preceq e^2 \mathbf{H}_t$.*

Proof. For a given $s \in [t-1]$, let $\hat{\theta}_o^s$ denote the value of $\hat{\theta}_o$ at that round. Then applying Lemma O.1 and Lemma K.1, we have the following for every $x_s \in \mathcal{X}_s$

$$e^{-1}\dot{\mu}(\langle x_s, \hat{\theta}_o^s\rangle) \leq \dot{\mu}(\langle x_s, \theta^*\rangle) \leq e\dot{\mu}(\langle x_s, \hat{\theta}_o^s\rangle) \tag{110}$$

Now, conclude the proof from Assumption L.1. □

Recall that τ is the round when switching criterion II (Line 6) is satisfied. Now define the following equations.

$$g_\tau(\theta) = \sum_{s \in [\tau-1] \setminus \mathcal{T}_o} \mu\left(\langle x_s, \theta\rangle\right) x_s + \lambda_+ \theta \tag{111}$$

$$\Theta_{II} = \{\theta : \|\theta - \hat{\theta}_o\|_{\mathbf{V}_{\tau_o}} \leq \gamma\sqrt{\kappa} \text{ and } \|\theta\|_2 \leq S\} \tag{112}$$

$$\tilde{\theta}_\tau = \arg\min_\theta \sum_{s\in[\tau-1]\setminus\mathcal{T}_o} \ell(\theta, x_s, r_s) + \frac{\lambda_+}{2}\|\theta\|^2 \tag{113}$$

$$\tag{114}$$

In Theorem L.8 we show that the parametrized form $\beta(\lambda_+, \lambda_-)$ satisfies $\beta(\lambda_+, \lambda_-) \le \beta$ for the β chosen in Algorithm 2.

One can now prove the following lemma identically to Lemma B.5 of Sawarni et al. [2024].

Lemma O.3. *Under event $\mathcal{E}_o$, with τ_o the last update time of $\hat{\theta}_o$,*

$$\left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{V}_{\tau_o}} \le 2\gamma\sqrt{\kappa} \text{ and } \left|\left|\hat{\theta}_o - \theta^*\right|\right|_{\mathbf{V}_{\tau_o}} \le \gamma(\lambda_+)\sqrt{\kappa}$$

Proof. First, observe that

$$\left|\left|\hat{\theta}_o - \theta^*\right|\right|_{\mathbf{V}_{\tau_o}} \le \sqrt{\kappa}\left|\left|\hat{\theta}_o - \theta^*\right|\right|_{\mathbf{H}^*_o} \le \sqrt{\kappa}\gamma(\lambda_+) \tag{115}$$

The first inequality follows from the fact that $\mathbf{V}_{\tau_o} \preceq \kappa\mathbf{H}^*_o$. And second inequality follows from Lemma N.1.

$$\left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{V}_{\tau_o}} \le \left|\left|\hat{\theta}_\tau - \hat{\theta}_o\right|\right|_{\mathbf{V}_{\tau_o}} + \left|\left|\theta^* - \hat{\theta}_o\right|\right|_{\mathbf{V}_{\tau_o}} \tag{116}$$

$$\le 2\sqrt{\kappa}\gamma \quad \text{(from (115) and } \hat{\theta}_\tau \in \Theta_{II}) \tag{117}$$

□

We now prove the following lemma.

Lemma O.4. *Under event $\mathcal{E}_o$ with probability at least $(1-\zeta/4)$, we have $\left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{H}^{(u),*}_\tau} \le \beta(\lambda_+, \lambda_-)$ and $\left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{H}^{(l),*}_\tau} \le \beta(\lambda_+, \lambda_-)$*

We first define the following equations that globally minimize the regularized loss over $\{r_s, x_s\}_{s\in[\tau-1]\setminus\mathcal{T}_o}$ (the data used at the policy update τ).

$$\tilde{\theta}_\tau = \arg\min_\theta \sum_{s\in[\tau-1]\setminus\mathcal{T}_o} \ell(\theta, r_s, x_s) + \frac{\lambda_+}{2}||\theta||_2^2 \text{ and } g_\tau(\theta) = \sum_{s\in[\tau-1]\setminus\mathcal{T}_o} \mu(\langle\theta, x_s\rangle)x_s + \lambda_+\theta \tag{118}$$

Before we prove this lemma, we state the following lemma. The proof of this lemma is identical to Lemma B.2 of Sawarni et al. [2024] and we refrain from reproving it here.

Lemma O.5. *At any round t, consider the minimisers computed above. Then under event $\mathcal{E}_o$ with probability at least $(1-\zeta/4)$, we have*

$$\left|\left|g_\tau(\tilde{\theta}_\tau) - g_\tau(\theta^*)\right|\right|_{\left(\mathbf{H}^{(u),*}_\tau\right)^{-1}} \le 4R\sqrt{\lambda_+} + \frac{7d}{R\sqrt{\lambda_-}}\log\left(16/\zeta\left(1 + \frac{T}{R^2 d\lambda_-}\right)\right)$$

$$\left|\left|g_\tau(\tilde{\theta}_\tau) - g_\tau(\theta^*)\right|\right|_{\left(\mathbf{H}^{(l),*}_\tau\right)^{-1}} \le 4R\frac{\lambda_+}{\sqrt{\lambda_-}} + \frac{7d}{R\sqrt{\lambda_-}}\log\left(16/\zeta\left(1 + \frac{T}{R^2 d\lambda_-}\right)\right)$$

Proof of Lemma O.4. For all rounds $s \in [\tau-1]\setminus\mathcal{T}_o$, we have

$$\left|\langle x, \hat{\theta}_\tau - \theta^*\rangle\right| \le ||x||_{\mathbf{V}^{-1}}\left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{V}} \quad \text{(Cauchy–Schwarz)} \tag{119}$$

$$\leq 2\,||x||_{\mathbf{V}^{-1}}\,\gamma(\lambda_+)\sqrt{\kappa} \quad \text{(via Lemma O.3)} \tag{120}$$

$$\leq \frac{2}{\sqrt{R^2\kappa\gamma^2}}\gamma(\lambda_+)\sqrt{\kappa} \quad \text{(switching criterion I not satisfied)} \tag{121}$$

$$\leq \frac{2}{R} \tag{122}$$

Using Lemma O.3—which guarantees that θ^* lies in the convex set $\Theta_{II} = \{\theta : \|\theta - \hat{\theta}_o\|_{\mathbf{V}_{\tau_o}} \leq \gamma\sqrt{\kappa},\ \|\theta\|_2 \leq S\}$—and using Lemma O.5, we instantiate Lemma V.1 with $c = 2$ for our setting, thereby establishing Lemma O.4.

□

P Bounding the instantaneous regret at each round t

Let the event in Lemma O.4 be denoted by $\mathcal{E}_\tau$ and now only consider rounds $t \in [T]$ which does not satisfy Switching Criterion I. Let x_t denote the arm played at each step in Line 14. Let x_t^* denote the best arm with the highest reward available at that round.

Corollary P.1. *Under the event $\mathcal{E}_\tau$ for all $x \in \mathcal{X}_t$, we have $|\langle x, \hat{\theta}_\tau - \theta^*\rangle| \leq \beta||x||_{\mathbf{H}_\tau^{-1}}$.*

Proof.

$$|\langle x, \hat{\theta}_\tau - \theta^*\rangle| \leq ||x||_{(\mathbf{H}_\tau^{(u),*})^{-1}} \left|\left|\hat{\theta}_\tau - \theta^*\right|\right|_{\mathbf{H}_\tau^{(u),*}} \quad \text{by Cauchy–Schwarz} \tag{123}$$

$$\leq \beta(\lambda_+, \lambda_-)\,||x||_{(\mathbf{H}_\tau^{(u),*})^{-1}} \quad \text{by Lemma O.4} \tag{124}$$

$$\leq \beta\,||x||_{\mathbf{H}_\tau^{-1}} \quad \text{since } \mathbf{H}_\tau \preceq \mathbf{H}_\tau^{(u),*} \tag{125}$$

□

Lemma P.2. *The arm $\mathcal{X}_t'$ obtained after eliminating arms from $\mathcal{X}_t$ (Line 13), under event $\mathcal{E}_o$ and while the cap count1$_{cutoff}$ is not reached, we have a) $x_t^* \in \mathcal{X}_t'$ and b) $\langle x_t^* - x_t, \theta^*\rangle \leq \frac{4}{R}$*

Proof. Suppose $x' = \arg\max_{x \in \mathcal{X}_t} \text{LCB}_o(x)$. Now, we have for all $x \in \mathcal{X}_t$,

$$\begin{aligned}
\left|\langle x, \hat{\theta}_o - \theta^*\rangle\right| &\leq ||x||_{(\mathbf{H}_o^*)^{-1}} \left|\left|\hat{\theta}_o - \theta^*\right|\right|_{\mathbf{H}_o^*} \quad \text{(Cauchy–Schwarz)} \\
&\leq ||x||_{(\mathbf{H}_o^*)^{-1}}\,\gamma(\lambda_+) \quad \text{(via Lemma N.1)} \\
&\leq ||x||_{\mathbf{V}_{\tau_o}^{-1}}\,\gamma\sqrt{\kappa} \quad \text{(as } \mathbf{V}_{\tau_o} \preceq \kappa\mathbf{H}_o^*\text{)} \\
&\leq ||x||_{\mathbf{V}^{-1}}\,\gamma\sqrt{2\kappa} \quad \text{(as } \mathbf{V} \preceq 2\mathbf{V}_{\tau_o}\text{; line 3 did not trigger after } \tau_o\text{)}
\end{aligned} \tag{126}$$

Thus, $\text{UCB}_o(x_t^*) = \langle x_t^*, \hat{\theta}_o\rangle + \gamma\sqrt{2\kappa}\,||x_t^*||_{\mathbf{V}^{-1}} \geq \langle x_t^*, \theta^*\rangle \geq \langle x', \theta^*\rangle \geq \langle x', \hat{\theta}_o\rangle - \gamma\sqrt{2\kappa}\,||x'||_{\mathbf{V}^{-1}} = \text{LCB}_o(x')$, where the first and last inequalities follow from the bound above and the second is due to optimality of x_t^* (and monotonicity of μ). Thus, x_t^* is not eliminated from $\mathcal{X}_t'$ proving (a).

Since x_t is also in $\mathcal{X}_t'$ (by definition), we have

$$\text{UCB}_o(x_t) = \langle x_t, \hat{\theta}_o\rangle + \sqrt{2\kappa}\gamma\,||x_t||_{\mathbf{V}^{-1}} \geq \langle x', \hat{\theta}_o\rangle - \gamma\sqrt{2\kappa}||x'||_{\mathbf{V}^{-1}} \tag{127}$$

$$\geq \langle x_t^*, \hat{\theta}_o\rangle - \gamma\sqrt{2\kappa}||x_t^*||_{\mathbf{V}^{-1}} \quad (x' \text{ has max } \text{LCB}_o(.)) \tag{128}$$

Again, using the fact that $\langle x_t^*, \hat{\theta}_o\rangle \geq \langle x_t^*, \theta^*\rangle - \gamma\sqrt{2\kappa}\,||x_t^*||_{\mathbf{V}^{-1}}$ and $\langle x_t, \hat{\theta}_o\rangle \leq \langle x_t, \theta^*\rangle + \gamma\sqrt{2\kappa}\,||x_t||_{\mathbf{V}^{-1}}$, we obtain $\langle x_t, \theta^*\rangle + 2\gamma\sqrt{2\kappa}\,||x_t||_{\mathbf{V}^{-1}} \geq \langle x_t^*, \theta^*\rangle - 2\gamma\sqrt{2\kappa}\,||x_t^*||_{\mathbf{V}^{-1}}$ invoking (126)

This gives us $\langle x_t^* - x_t, \theta^* \rangle \leq 2\gamma\sqrt{2\kappa}\,||x_t||_{\mathbf{V}^{-1}} + 2\gamma\sqrt{2\kappa}\,||x_t^*||_{\mathbf{V}^{-1}}$. However, since the switching criterion I is not satisfied, we have $||x||_{\mathbf{V}^{-1}} \leq \frac{1}{R\gamma\sqrt{2\kappa}}$ for all $x \in \mathcal{X}_t$. Plugging in above, we prove (b).

□

Lemma P.3. *Under event $\mathcal{E}_\tau$, $\langle x_t^* - x_t, \theta^* \rangle \leq 2\sqrt{2}\beta||x_t||_{\mathbf{H}_t^{-1}}$*

Proof.

$$\begin{aligned}
\langle x_t^*, \theta^* \rangle - \langle x_t, \theta^* \rangle \leq& \left(\langle x_t^*, \hat{\theta}_\tau \rangle + \beta \left\|x_t^*\right\|_{\mathbf{H}_\tau^{-1}}\right) - \left(\langle x_t, \hat{\theta}_\tau \rangle - \beta \left\|x_t\right\|_{\mathbf{H}_\tau^{-1}}\right) \quad \text{(Corollary P.1)} \\
\overset{(a)}{\leq}& \left(\langle x_t, \hat{\theta}_\tau \rangle + \beta \left\|x_t\right\|_{\mathbf{H}_\tau^{-1}}\right) - \left(\langle x_t, \hat{\theta}_\tau \rangle - \beta \left\|x_t\right\|_{\mathbf{H}_\tau^{-1}}\right) \\
\leq& 2\sqrt{2}\,\beta \left\|x_t\right\|_{\mathbf{H}_t^{-1}} \quad \text{(as } \mathbf{H}_t \preceq 2\mathbf{H}_\tau \text{ since Line 6 not triggered)} \qquad (129)
\end{aligned}$$

(a) holds true since optimistic x_t is chosen at and x_t^* not eliminated (Lemma P.2). □

Q Proof of the Joint-DP Regret Theorem

Theorem M.1. *When the events in Assumption L.1 and Assumption L.2 hold, Criterion I and Criterion II are invoked at most N_o and count2$_{cutoff}$ times, respectively, and the parametric conditions $\gamma(\lambda_+) \leq \gamma$ and $\beta(\lambda_+, \lambda_-) \leq \beta$ (as defined in Equation* (100) *and Equation* (99)*) are satisfied, the regret R_T can be bounded by some universal constant C with probability at least $1 - \zeta$.*

$$R_T \leq C\sqrt{d}\beta\sqrt{\log_2\left(\frac{RT}{d\lambda_-}\right)}\sqrt{\sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}\left(\langle x_t^*, \theta^* \rangle\right)} + 8dR^3\kappa\gamma^2 \log T \qquad (101)$$

Proof. First, we assume that the events of Assumption L.1 ($\zeta/8$), Assumption L.2 ($\zeta/8$), $\mathcal{E}_o$ (Lemma N.1, $\zeta/8$), $\mathcal{E}_\tau$ (Lemma O.4, $\zeta/4$), and Lemma T.2 ($\zeta/8$) hold throughout the algorithm; the failure probabilities are in parentheses, so by a union bound this happens with probability at least $1 - 3\zeta/4 \geq 1 - \zeta$. Further, the regret of algorithm can be bounded as

$$\begin{aligned}
R_T =& \sum_{t\in[T]} \mu\left(\langle x_t^*, \theta^* \rangle\right) - \mu\left(\langle x_t, \theta^* \rangle\right) \qquad (130) \\
\leq& R|\mathcal{T}_o| + \sum_{t\in[T]\setminus\mathcal{T}_o} \mu\left(\langle x_t^*, \theta^* \rangle\right) - \mu\left(\langle x_t, \theta^* \rangle\right) \qquad \text{Upper bound of } R \text{ on rewards} \qquad (131) \\
\leq& R|\mathcal{T}_o| + \sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}(z)\langle x_t^* - x_t, \theta^* \rangle \quad \text{(some } z \in [\langle x_t, \theta^* \rangle, \langle x_t^*, \theta^* \rangle]) \qquad (132)
\end{aligned}$$

One can bound $|\mathcal{T}_o|$ now we denote $R_\tau(T) = \sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}(z)\langle x_t^* - x_t, \theta^* \rangle$.

Similar to previous sections $\mathbf{H}_\tau$ denotes the matrix last updated before time step $t \in [T]$. Now we upper bound as follows for some constant C.

$$\begin{aligned}
R_\tau(T) \leq& \sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}(z) C\beta \left\|x_t\right\|_{\mathbf{H}_t^{-1}} \quad \text{(by Lemma P.3)} \qquad (133) \\
\leq& C\beta \sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}(z) 2\sqrt{2} \left\|x_t\right\|_{\left(\mathbf{H}_t^{(l),*}\right)^{-1}} \quad \text{by Corollary O.2} \qquad (134) \\
\leq& C\beta \sum_{t\in[T]\setminus\mathcal{T}_o} \sqrt{\dot{\mu}\left(\langle x_t^*, \theta^* \rangle\right) \dot{\mu}\left(\langle x_t, \theta^* \rangle\right)} \exp\left(R\langle x_t^* - x_t, \theta^* \rangle\right) \left\|x_t\right\|_{\left(\mathbf{H}_t^{(l),*}\right)^{-1}} \quad \text{by Lemma K.2} \\
& \qquad (135)
\end{aligned}$$

$$\leq C\beta \sum_{t\in[T]\setminus\mathcal{T}_o} \sqrt{\dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right)\dot{\mu}\left(\langle x_t, \theta^*\rangle\right)} e^4 \left|\left|x_t\right|\right|_{\left(\mathbf{H}_t^{(l),*}\right)^{-1}} \quad \text{by Lemma P.2} \tag{136}$$

$$= C\beta \sum_{t\in[T]\setminus\mathcal{T}_o} \sqrt{\dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right)} \left|\left|\tilde{x}_t\right|\right|_{\left(\mathbf{H}_t^{(l),*}\right)^{-1}} \quad \tilde{x}_t = \sqrt{\dot{\mu}\left(\langle x_t, \theta^*\rangle\right)} x_t \tag{137}$$

$$\leq C\beta \sqrt{\sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right) \left(\sum_{t\in[T-1]\setminus\mathcal{T}_o} \left|\left|\tilde{x}_t\right|\right|^2_{\left(\mathbf{H}_t^{(l),*}\right)^{-1}}\right)} \quad \text{by Cauchy–Schwarz} \tag{138}$$

$$\leq Ce^2\beta \sqrt{\sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right) \cdot 2d\log\left(1 + \frac{T}{d\kappa^*\lambda_-}\right)} \quad \text{(Lemma S.1, with } \|\tilde{x}_t\|^2 \leq \tfrac{1}{\kappa^*}\text{)} \tag{139}$$

$$\leq C\sqrt{d}\beta \sqrt{\log_2\left(\frac{RT}{d\lambda_-}\right)} \sqrt{\sum_{t\in[T]\setminus\mathcal{T}_o} \dot{\mu}\left(\langle x_t^*, \theta^*\rangle\right)} \tag{140}$$

Now substituting things back and bounding $\mathcal{T}_o$, we can prove the desired lemma.

□

R Proofs for Shuffle-Private Population Optimal Design

This section proves the accuracy and privacy guarantees stated in Theorems E.6 and E.7, using the notation of Appendix E and Algorithm 3. We recall only the cumulative quantities needed in the proof.

Set $b = m/(2L)$ and let $D_0, \ldots, D_L$ be the predetermined phases in Algorithm 3. Define

$$M_G := \mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}}[G_{\mathcal{X}}], \qquad M_r := \mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}}\left[z_r(\mathcal{X}) z_r(\mathcal{X})^\top \mid \mathcal{F}_{r-1}\right], \tag{141}$$

where $\mathcal{F}_{r-1}$ is the public transcript through W_{r-1}. Let S_r be the cumulative sum of the nonprivate matrix reports through phase r, let Z_r be the error of the shuffle summation in that phase, and set $Z_{0:r} := \sum_{s=0}^{r} Z_s$. We call

$$\widehat{B}_r := \Gamma I + S_r \tag{142}$$

the *empirical design matrix* (the nonprivate counterpart of the released matrix). Thus the public matrix in the algorithm satisfies

$$W_r = \widehat{B}_r + Z_{0:r}, \tag{143}$$

Thus W_r is the *public design matrix*: the empirical design matrix perturbed by the cumulative shuffle noise. Set

$$q := 2\Delta \log\frac{4d(L+1)}{\beta}, \qquad \eta_L := 2\sigma\sqrt{L+1}\left(4\sqrt{d} + 2\sqrt{\log\frac{4(L+1)}{\beta}}\right), \tag{144}$$

R.1 Concentration of the nonprivate design matrices

The policy in phase r depends on W_{r-1}, so the matrices collected across different phases are not identically distributed. Conditional on the previous phases, one could apply the ordinary matrix Chernoff bound separately within phase r. Summing the resulting phasewise comparisons, however, would also sum their additive deviations and introduce an unnecessary factor of L. Instead, we flatten all users into one adapted sequence and inspect the cumulative sum only at the $L+1$ predetermined phase boundaries. The matrix supermartingale below then gives one additive deviation, while the union bound costs only $\log(L+1)$. This is also why we need a predictable-sum comparison rather than only the fixed-policy, i.i.d. specialization in Lemma H.1.

We derive the comparison directly from the large-deviation supermartingale of Tropp [2011, Section 2.2, Lemma 2.1]. We spell out Tropp's increments, conditional-cgf comparators, and cumulative processes because they differ from the public design matrices denoted by W_r in our algorithm.

Lemma R.1 (Two-sided matrix predictable-sum comparison)**.** *Let $(Y_t, \mathcal{F}_t)_{t\ge 1}$ be an adapted symmetric-matrix process satisfying $0 \preceq Y_t \preceq \Delta I$, and set $A_n = \sum_{t\le n} \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}]$ and $S_n = \sum_{t\le n} Y_t$. For every deterministic n and $x > 0$,*

$$\mathbb{P}\left[S_n \not\preceq 2A_n + xI\right] \le d \exp\left(-\frac{(\log 2)x}{\Delta}\right), \tag{145}$$

$$\mathbb{P}\left[A_n \not\preceq 2S_n + xI\right] \le d \exp\left(-\frac{(1-e^{-1})x}{\Delta}\right). \tag{146}$$

Proof. Write $\mu_t := \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}]$. We first prove the upper comparison. In the notation of Tropp [2011, Lemma 2.1], take

$$X_t^+ := Y_t/\Delta, \qquad V_t^+ := \mu_t/\Delta, \qquad g_+(\theta) := e^\theta - 1. \tag{147}$$

Here X_t^+ is adapted and V_t^+ is predictable. Since the graph of the scalar exponential lies below its chord on $[0,1]$, functional calculus gives

$$e^{\theta Y_t/\Delta} \preceq I + (e^\theta - 1)Y_t/\Delta. \tag{148}$$

Taking conditional expectations, using monotonicity of the matrix logarithm, and then $\log(I+H) \preceq H$ for $H \succ -I$, we obtain the conditional-cgf bound

$$\log \mathbb{E}[e^{\theta X_t^+} \mid \mathcal{F}_{t-1}] \preceq (e^\theta - 1)V_t^+ = g_+(\theta)V_t^+. \tag{149}$$

Tropp denotes the cumulative increment and comparator by

$$\mathsf{Y}_k^+ := \sum_{t=1}^k X_t^+ = S_k/\Delta, \qquad \mathsf{W}_k^+ := \sum_{t=1}^k V_t^+ = A_k/\Delta. \tag{150}$$

These symbols are local to this proof; in particular, W_k^+ is not the released matrix W_k. Lemma 2.1 of Tropp [2011] applied to (149) states that

$$L_k^+(\theta) := \operatorname{tr} \exp\left(\frac{\theta S_k - (e^\theta - 1)A_k}{\Delta}\right) \tag{151}$$

is a positive supermartingale with $L_0^+(\theta) = d$. Hence $\mathbb{E}\, L_n^+(\theta) \le d$.

Set $\theta = \log 2$, so $e^\theta - 1 = 1$ and $2\theta - (e^\theta - 1) > 0$. On the event $S_n \not\preceq 2A_n + xI$, we have a unit vector v for which $v^\top(S_n - 2A_n - xI)v > 0$. Therefore

$$v^\top\big(\theta S_n - (e^\theta - 1)A_n\big)v > \big(2\theta - (e^\theta - 1)\big)v^\top A_n v + \theta x \ge \theta x. \tag{152}$$

Thus $L_n^+(\theta) > e^{\theta x/\Delta}$. Markov's inequality yields

$$\mathbb{P}\left[S_n \not\preceq 2A_n + xI\right] \le d e^{-\theta x/\Delta}, \tag{153}$$

which is (145).

For the reverse comparison, use the same lemma with

$$X_t^- := -Y_t/\Delta, \qquad V_t^- := -\mu_t/\Delta, \qquad g_-(\theta) := 1 - e^{-\theta}. \tag{154}$$

The chord bound for $e^{-\theta x}$ on $[0,1]$ gives

$$\log \mathbb{E}[e^{\theta X_t^-} \mid \mathcal{F}_{t-1}] \preceq -(1-e^{-\theta})\mu_t/\Delta = g_-(\theta)V_t^-. \tag{155}$$

The cumulative processes are now $\mathsf{Y}_k^- = -S_k/\Delta$ and $\mathsf{W}_k^- = -A_k/\Delta$, so Tropp's positive supermartingale is

$$L_k^-(\theta) := \operatorname{tr} \exp\left(\frac{-\theta S_k + (1-e^{-\theta})A_k}{\Delta}\right), \qquad \mathbb{E}\, L_n^-(\theta) \le d. \tag{156}$$

Take $\theta = 1$ and write $g = 1 - e^{-1}$. Since $2g - 1 > 0$, the event $A_n \not\preceq 2S_n + xI$ implies

$$\lambda_{\max}(-S_n + gA_n) > gx. \tag{157}$$

Consequently $L_n^-(1) > e^{gx/\Delta}$, and Markov's inequality gives

$$\mathbb{P}\left[A_n \not\preceq 2S_n + xI\right] \le d e^{-gx/\Delta}. \tag{158}$$

This is (146). □

Define the predictable population matrix A_r and the corresponding *population design matrix* B_r by

$$A_r = \frac{m}{2} M_G + b \sum_{s=1}^{r} M_s, \qquad B_r = \Gamma I + A_r. \tag{159}$$

Thus $\widehat{B}_r$ and B_r are the empirical and population counterparts of the same cumulative design, and W_r is the released version of $\widehat{B}_r$.

Lemma R.2 (Design matrix comparison)**.** *With probability at least $1 - \beta/2$, simultaneously for every $r = 0, \dots, L$,*

$$\widehat{B}_r \preceq 2B_r - (\Gamma - q)I, \qquad B_r \preceq 2\widehat{B}_r - (\Gamma - q)I, \qquad q = 2\Delta \log \frac{4d(L+1)}{\beta}. \tag{160}$$

Proof. Flatten the reports into an adapted sequence Y_t. During the baseline phase, $Y_t = z_0(\mathcal{X}_t) z_0(\mathcal{X}_t)^\top$ with $z_0(\mathcal{X}_t) \sim \pi_G(\mathcal{X}_t)$, and hence $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = M_G$. During phase D_r, the public matrix W_{r-1} is fixed before that phase's users arrive, so $Y_t = z_r(\mathcal{X}_t) z_r(\mathcal{X}_t)^\top$ and $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = M_r$. In every case, $0 \preceq Y_t \preceq \Delta I$ by the bounded-context assumption in Theorem E.6.

Apply Lemma R.1 at each of the $L + 1$ phase boundaries. A union bound over both tails and all boundaries shows that, with probability at least $1 - \beta/2$, simultaneously for every r,

$$S_r \preceq 2A_r + qI, \qquad A_r \preceq 2S_r + qI, \tag{161}$$

For each tail and phase boundary, this choice makes the larger of the two bounds in (145)–(146) at most $\beta/(4(L+1))$: indeed, $2(1 - e^{-1}) > 1$ and $2 \log 2 > 1$. Adding ΓI to the first inequality gives

$$\widehat{B}_r = \Gamma I + S_r \preceq \Gamma I + 2A_r + qI = 2B_r - (\Gamma - q)I, \tag{162}$$

and the second inequality gives $B_r \preceq 2\widehat{B}_r - (\Gamma - q)I$ in the same way. □

R.2 Cumulative shuffle noise and the public sandwich

The phasewise noise matrices $Z_0, \dots, Z_L$ of one DOD invocation are not independent: the inputs of phase r are selected using W_{r-1}, which contains $Z_0, \dots, Z_{r-1}$, and the randomized rounding in the matrix-summation protocol has an input-dependent law. Hence the entries of $Z_{0:r}$ are not independent, and a random-matrix bound for matrices with independent entries does not apply to $Z_{0:r}$ directly. We instead combine the within-call independence of Assumption G.1 with a martingale moment-generating-function bound and a net argument.

Lemma R.3 (Cumulative DOD noise)**.** *Under Assumption G.1, for every fixed $r \in \{0, \dots, L\}$ and $t \geq 0$, conditional on the transcript preceding the DOD invocation,*

$$\mathbb{P}\left[\|Z_{0:r}\|_{\mathrm{op}} > 2\sigma\sqrt{r+1}\left(4\sqrt{d} + 2\sqrt{t}\right) \right] \leq 2e^{-t}. \tag{163}$$

Proof. For $s = 0, \dots, r$, let $\mathcal{G}_s$ be the σ-field generated by the transcript preceding the invocation, the releases of phases $0, \dots, s-1$, and the inputs submitted in phase s. Then $Z_{0:s-1}$ is $\mathcal{G}_s$-measurable.

One call. Fix a deterministic unit vector v. Then

$$v^\top Z_s v = \sum_i v_i^2 (Z_s)_{ii} + \sum_{i<j} 2 v_i v_j (Z_s)_{ij}.$$

Conditional on $\mathcal{G}_s$, the upper-triangular entries of Z_s are independent, mean-zero and sub-Gaussian with variance proxy at most σ^2 by Assumption G.1. Hence $v^\top Z_s v$ is conditionally sub-Gaussian with variance proxy at most

$$\sigma^2 \Big(\sum_i v_i^4 + 4 \sum_{i<j} v_i^2 v_j^2 \Big) \leq 2\sigma^2 \Big(\sum_i v_i^2 \Big)^2 = 2\sigma^2,$$

that is, $\mathbb{E}[e^{\theta v^\top Z_s v} \mid \mathcal{G}_s] \leq e^{\theta^2 \sigma^2}$ for every $\theta \in \mathbb{R}$. This bound does not depend on $\mathcal{G}_s$.

Across calls. Since $Z_{0:s-1}$ is $\mathcal{G}_s$-measurable, the tower property gives

$$\mathbb{E}\big[e^{\theta v^\top Z_{0:s}v}\big] = \mathbb{E}\Big[e^{\theta v^\top Z_{0:s-1}v}\ \mathbb{E}\big[e^{\theta v^\top Z_s v} \mid \mathcal{G}_s\big]\Big] \le e^{\theta^2\sigma^2}\ \mathbb{E}\big[e^{\theta v^\top Z_{0:s-1}v}\big].$$

Iterating over $s = 0, \dots, r$ shows that $v^\top Z_{0:r}v$ is sub-Gaussian with variance proxy $2(r+1)\sigma^2$. No independence across calls is used.

Net. Let $\mathcal{N}$ be a $\frac{1}{4}$-net of the unit sphere with $|\mathcal{N}| \le 9^d$, fixed in advance. For a symmetric matrix Z, $\|Z\|_{\mathrm{op}} \le 2\max_{v\in\mathcal{N}} |v^\top Z v|$ [Vershynin, 2018, Exercise 4.4.3]. A union bound over $\mathcal{N}$ and both tails gives

$$\mathbb{P}\left[\|Z_{0:r}\|_{\mathrm{op}} > 4\sigma\sqrt{r+1}\,u\right] \le 2\cdot 9^d e^{-u^2}, \qquad u \ge 0.$$

Take $u = \sqrt{d\log 9 + t} \le \sqrt{\log 9}\,\sqrt{d} + \sqrt{t}$. Since $4\sqrt{\log 9} \le 6 \le 8$, the threshold is at most $2\sigma\sqrt{r+1}(4\sqrt{d} + 2\sqrt{t})$, and the failure probability is $2e^{-t}$. □

Lemma R.4 (Uniform cumulative shuffle-noise bound)**.** *With probability at least $1 - \beta/2$, simultaneously for every $r = 0, \dots, L$,*

$$\|Z_{0:r}\|_{\mathrm{op}} \le \eta_L. \tag{164}$$

Proof. Fix the current DOD invocation and identify $Z_{0:r}$ with its cumulative error $Z^{\mathrm{DOD}}_{k,0:r}$ from Assumption G.1. Apply Lemma R.3 at each fixed phase index r. Take $t = \log(4(L+1)/\beta)$ and union bound over all phase indices $r = 0, \dots, L$. Since $r + 1 \le L + 1$, this proves the claim. □

Lemma R.5 (Public/population matrix sandwich)**.** *On the intersection of the events in Lemmas R.2 and R.4, if $\Gamma \ge q + 2\eta_L$, then, for every $r = 0, \dots, L$,*

$$\frac{1}{2}B_r \preceq W_r \preceq 2B_r, \qquad \frac{1}{2}B_r^{-1} \preceq W_r^{-1} \preceq 2B_r^{-1}. \tag{165}$$

Proof. The reverse inequality in (160) yields $\widehat{B}_r \succeq \frac{1}{2}B_r + \frac{1}{2}(\Gamma - q)I$. Since $W_r = \widehat{B}_r + Z_{0:r}$ and $\|Z_{0:r}\|_{\mathrm{op}} \le \eta_L$,

$$W_r \succeq \frac{1}{2}B_r + \left(\frac{\Gamma - q}{2} - \eta_L\right) I \succeq \frac{1}{2}B_r. \tag{166}$$

Similarly, the forward inequality gives

$$W_r \preceq 2B_r - (\Gamma - q - \eta_L)I \preceq 2B_r. \tag{167}$$

Since $B_r \succeq \Gamma I$, the lower bound also gives $W_r \succeq (\Gamma/2)I$. Inverting the two PSD-order inequalities reverses their order and proves the inverse sandwich. □

R.3 Population argmax and bounded increments

Lemma R.6 (Public argmax is a population approximate argmax)**.** *On the sandwich event, for every $r = 1, \dots, L$,*

$$\mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}} \max_{x\in\mathcal{X}} x^\top B_{r-1}^{-1} x \le 4\,\mathrm{tr}\big(B_{r-1}^{-1}M_r\big). \tag{168}$$

Proof. For every fixed $\mathcal{X}$, the inverse sandwich and the definition of $z_r(\mathcal{X})$ give

$$\max_{x\in\mathcal{X}} x^\top B_{r-1}^{-1} x \le 2\max_{x\in\mathcal{X}} x^\top W_{r-1}^{-1} x \tag{169}$$

$$= 2 z_r(\mathcal{X})^\top W_{r-1}^{-1} z_r(\mathcal{X}) \tag{170}$$

$$\le 4 z_r(\mathcal{X})^\top B_{r-1}^{-1} z_r(\mathcal{X}). \tag{171}$$

Take a fresh population expectation conditional on $\mathcal{F}_{r-1}$ and use (141). □

Lemma R.7 (Local-G coverage bounds every population increment)**.** *For every $r = 1, \dots, L$,*

$$M_r \preceq 2d\,M_G. \tag{172}$$

Consequently, with $H_r = bM_r$,

$$H_r \preceq \frac{1}{2}B_0 \preceq \frac{1}{2}B_{r-1}. \tag{173}$$

Proof. Pointwise, $z_r(\mathcal{X}) \in \mathcal{X}$, so Corollary E.2 gives $z_r(\mathcal{X})z_r(\mathcal{X})^\top \preceq 2dG_{\mathcal{X}}$. Taking a fresh population expectation proves (172). Since $m/2 = Lb$ and $L \geq 4d$,

$$H_r = bM_r \preceq 2bdM_G \preceq \frac{1}{2}LbM_G \preceq \frac{1}{2}B_0. \tag{174}$$

Finally $B_{r-1} \succeq B_0$. □

R.4 Mixture identity and log-determinant potential

We apply the PSD-increment form of the elliptical-potential lemma stated in Corollary S.2. It is the direct full-rank extension of Lemma S.1; the same log-determinant pattern appears in Ruan et al. [2021, Lemma 29].

Lemma R.8 (Exact population mixture identity)**.** *The returned policy satisfies*

$$m\mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}}) = A_L, \qquad B_L = \Gamma I + m\mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}}). \tag{175}$$

Proof. By (22),

$$\mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}}) = \frac{1}{2}M_G + \frac{1}{2L}\sum_{r=1}^{L} M_r. \tag{176}$$

Multiplication by $m = 2Lb$ gives

$$m\mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}}) = LbM_G + b\sum_{r=1}^{L} M_r = A_L. \tag{177}$$

Adding ΓI proves the second identity. □

Proof of Theorem E.6. The intersection of Lemmas R.2 and R.4 has probability at least $1-\beta$. The condition (23) is exactly $\Gamma \geq q + 2\eta_L$, so Lemma R.5 holds.

By Lemma R.8,

$$\mathcal{V}_{\mathcal{D}}^{(\Gamma/m)}(\pi_{\mathrm{DOD}}) = m \mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}} \max_{x\in\mathcal{X}} x^\top B_L^{-1} x. \tag{178}$$

Because $B_L \succeq B_{r-1}$, Lemma R.6 implies, for every r,

$$\mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}} \max_{x\in\mathcal{X}} x^\top B_L^{-1} x \leq 4\,\mathrm{tr}\big(B_{r-1}^{-1}M_r\big). \tag{179}$$

Average these L inequalities and use $m/L = 2b$ and $H_r = bM_r$ to obtain

$$\mathcal{V}_{\mathcal{D}}^{(\Gamma/m)}(\pi_{\mathrm{DOD}}) \leq 8\sum_{r=1}^{L} \mathrm{tr}\big(B_{r-1}^{-1}H_r\big). \tag{180}$$

By (173), the sequence $B_r = B_{r-1} + H_r$ satisfies the hypotheses of Corollary S.2 with $\alpha = 1/2$. Therefore

$$\sum_{r=1}^{L} \mathrm{tr}\big(B_{r-1}^{-1}H_r\big) \leq 2\log\frac{\det B_L}{\det B_0}. \tag{181}$$

Substitution into (180) yields

$$\mathcal{V}_{\mathcal{D}}^{(\Gamma/m)}(\pi_{\mathrm{DOD}}) = O\left(\log\frac{\det B_L}{\det B_0}\right). \tag{182}$$

Finally, (172) gives

$$B_L = B_0 + b\sum_{r=1}^{L} M_r \tag{183}$$

$$\preceq B_0 + 2d(LbM_G) \preceq (1+2d)B_0. \tag{184}$$

Therefore

$$\log \frac{\det B_L}{\det B_0} = O(d \log d), \tag{185}$$

which proves (24).

For the remaining assertion, apply Jensen's inequality to the concave map $s \mapsto \sqrt{s}$:

$$\mathop{\mathbb{E}}_{\mathcal{X}\sim\mathcal{D}} \sqrt{\max_{x\in\mathcal{X}} x^\top \big((\Gamma/m)I + \mathcal{I}_{\mathcal{D}}(\pi_{\mathrm{DOD}})\big)^{-1} x} \le \sqrt{\mathcal{V}_{\mathcal{D}}^{(\Gamma/m)}(\pi_{\mathrm{DOD}})} \tag{186}$$

$$= O(\sqrt{d \log d}). \tag{187}$$

□

R.5 Privacy

We use adaptive parallel composition on disjoint user phases and postprocessing; see, e.g., Dwork and Roth [2014].

Proof of Theorem E.7. Fix neighboring design datasets that differ in the record of one user. That user belongs to exactly one predetermined phase, say D_j. Releases before phase D_j have identical distributions under the two datasets. Conditional on any fixed preceding transcript, the shuffled release for D_j is (ε, δ)-differentially private by assumption.

All later matrices, argmax rules, and the final policy may depend adaptively on the release from D_j. Conditional on the transcript already released, however, later phases use only records of other users and fresh mechanism randomness. They therefore define a common Markov postprocessing kernel under the two neighboring datasets. Adaptive postprocessing preserves the privacy guarantee of the release from D_j. Thus the likelihood-ratio inequality for the complete transcript has the same (ε, δ) parameters.

Since the changed user can occur in only one phase, there is no composition factor $L + 1$. The returned policy is a function of the complete transcript and is covered by postprocessing. This proves user-level (ε, δ) shuffle privacy. □

S Useful lemmas for adversarial contexts

We first recall standard design matrix inequalities from Abbasi-yadkori et al. [2011], including the elliptic potential lemma and determinant bounds, which are key tools in linear contextual bandit analyses. Beyond these, Lemma S.5 adapts the rarely-switching argument to our setting, where policy switches are based on directional growth criterion (see line 6 of Algorithm 2)

Lemma S.1 (Elliptic Potential Lemma (Lemma 10 of Abbasi-yadkori et al., 2011))**.** *Let $x_1, x_2, \ldots, x_t$ be a sequence of vectors in $\mathbb{R}^d$ and let $||x_s||_2 \le L$ for all $s \in [t]$. Further, let $\mathbf{T}_s = \sum_{m=1}^{s-1} x_m x_m^\top + \lambda \boldsymbol{I}$. Suppose $\lambda \ge L^2$. Then,*

$$\sum_{s=1}^{t} ||x_s||^2_{\mathbf{T}_s^{-1}} \le 2d \log\left(1 + \frac{L^2 t}{\lambda d}\right) \tag{188}$$

The preceding lemma adds one rank-one matrix $x_s x_s^\top$ at a time. The next corollary isolates its underlying log-determinant argument and allows each update to be an arbitrary positive-semidefinite matrix H_r. Indeed, when $H_r = x_r x_r^\top$, its left-hand summand becomes $\mathrm{tr}(V_{r-1}^{-1} H_r) = \|x_r\|^2_{V_{r-1}^{-1}}$, recovering the rank-one quantity in Lemma S.1. The additional condition $H_r \preceq \alpha V_{r-1}$ ensures that no normalized matrix increment is too large.

Corollary S.2 (PSD-increment elliptical potential)**.** *Let $V_0 \succ 0$ and $V_r = V_{r-1} + H_r$, where $H_r \succeq 0$ and $H_r \preceq \alpha V_{r-1}$ for every r, with $0 < \alpha \le 1$. Then*

$$\sum_{r=1}^{L} \mathrm{tr}(V_{r-1}^{-1} H_r) \le (1+\alpha) \log \frac{\det V_L}{\det V_0} \le 2 \log \frac{\det V_L}{\det V_0}. \tag{189}$$

Proof. Since $V_0 \succ 0$ and every $H_r \succeq 0$, all the matrices V_r are positive definite. Thus their symmetric inverse square roots are well-defined. Set

$$C_r := V_{r-1}^{-1/2} H_r V_{r-1}^{-1/2}. \tag{190}$$

Congruencing $0 \preceq H_r \preceq \alpha V_{r-1}$ by $V_{r-1}^{-1/2}$ gives $0 \preceq C_r \preceq \alpha I$. Consequently, if $\lambda_1(C_r), \ldots, \lambda_d(C_r)$ are its eigenvalues, then each lies in $[0, \alpha]$. By cyclicity of the trace,

$$\operatorname{tr}(V_{r-1}^{-1} H_r) = \operatorname{tr}\Big(V_{r-1}^{-1/2} H_r V_{r-1}^{-1/2}\Big) = \operatorname{tr}(C_r). \tag{191}$$

For $x \geq 0$, $\log(1+x) \geq x/(1+x)$. Hence, when $x \in [0, \alpha]$, we have $x \leq (1+x)\log(1+x) \leq (1+\alpha)\log(1+x)$. Applying this scalar inequality to each eigenvalue of C_r yields

$$\operatorname{tr}(C_r) = \sum_{j=1}^{d} \lambda_j(C_r) \tag{192}$$

$$\leq (1+\alpha) \sum_{j=1}^{d} \log(1 + \lambda_j(C_r)) \tag{193}$$

$$= (1+\alpha) \log\det(I + C_r). \tag{194}$$

It remains to identify this determinant. From $V_r = V_{r-1} + H_r$, we have

$$I + C_r = V_{r-1}^{-1/2} V_r V_{r-1}^{-1/2}. \tag{195}$$

Therefore, by multiplicativity of the determinant,

$$\det(I + C_r) = \det\Big(V_{r-1}^{-1/2}\Big)^2 \det(V_r) = \frac{\det V_r}{\det V_{r-1}}. \tag{196}$$

Combining (191)– (196) and summing over r gives

$$\sum_{r=1}^{L} \operatorname{tr}(V_{r-1}^{-1} H_r) \leq (1+\alpha) \sum_{r=1}^{L} \log \frac{\det V_r}{\det V_{r-1}} \tag{197}$$

$$= (1+\alpha) \log \frac{\det V_L}{\det V_0}, \tag{198}$$

where the final equality is the telescoping identity. Finally, $1 + \alpha \leq 2$ because $\alpha \leq 1$. □

Lemma S.3 (Lemma 12 of Abbasi-yadkori et al., 2011)**.** *Let $A \succeq B \succeq 0$. Then* $\sup_{x \neq 0} \frac{x^\top A x}{x^\top B x} \leq \frac{det(A)}{det(B)}$

Lemma S.4 (Lemma 10 of Abbasi-yadkori et al., 2011)**.** *Let $\{x_s\}_{s=1}^t$ be a set of vectors. Define the sequence $\{\mathbf{V}_s\}_{s=1}^t$ as $\mathbf{T}_1 = \lambda \boldsymbol{I}$, $\mathbf{T}_{s+1} = \mathbf{T}_s + x_s x_s^\top$ for $s \in [t-1]$ Further, let $||x_s||_2 \leq L \forall s \in [t]$. Then, det$(\mathbf{T}_t) \leq (\lambda + tL^2/d)^d$*

The noise level in the following lemma is chosen based on the bounds $\lambda_- = \frac{4\sqrt{d}}{4\sqrt{d}+1}\lambda$ and $\lambda_+ = \frac{4\sqrt{d}+2}{4\sqrt{d}+1}\lambda$ (see the proof of Lemma L.4), together with the assumption that the noisy matrices satisfy Assumption L.1.

Lemma S.5. *Let $\{x_s\}_{s=1}^t$ be a set of vectors in $\mathbb{R}^d$. Consider the sequence $\{\mathbf{T}_s\}_{s=1}^t$ such that $\mathbf{T}_s = \sum_{i=1}^s x_i x_i^\top + \lambda \boldsymbol{I}$ with $\mathbf{T}_0 = \lambda \boldsymbol{I}$. Further, let $||x_s||_2 \leq L \forall s \in [t]$ and let $\hat{\mathbf{T}}_s$ be a noisy estimate of $\mathbf{T}_s$ satisfying $||\hat{\mathbf{T}}_s - \mathbf{T}_s||_2 \leq \frac{\lambda}{4\sqrt{d}+1}$ for every $s \in [t]$. Define the set $\{0 = \tau_1, \tau_2, \ldots, \tau_m \leq t\}$ such that: $\hat{\mathbf{T}}_s \preceq 2\hat{\mathbf{T}}_{\tau_i}$ for $\tau_i \leq s < \tau_{i+1}$ but $\hat{\mathbf{T}}_{\tau_{i+1}} \not\preceq 2\hat{\mathbf{T}}_{\tau_i}$ for $i \in \{1, \ldots, m-1\}$. The following properties hold*

- *The number of times switching happens (m) is at most $1 + d \log_{11/8}\Big(1 + \frac{tL^2}{\lambda d}\Big)$.*

Proof. At switch time τ_j, the criterion gives $v^\top \hat{\mathbf{T}}_{\tau_j} v \geq 2v^\top \hat{\mathbf{T}}_{\tau_{j-1}} v$ for some unit vector v. Using the criterion $\|\hat{\mathbf{T}}_s - \mathbf{T}_s\|_2 \leq \eta := \frac{\lambda}{4\sqrt{d}+1}$,

$$\begin{aligned} v^\top \mathbf{T}_{\tau_j} v &\geq v^\top \hat{\mathbf{T}}_{\tau_j} v - \eta \\ &\geq 2v^\top \hat{\mathbf{T}}_{\tau_{j-1}} v - \eta \\ &\geq 2(v^\top \mathbf{T}_{\tau_{j-1}} v - \eta) - \eta \\ &= 2v^\top \mathbf{T}_{\tau_{j-1}} v - 3\eta \end{aligned}$$

Thus,

$$\frac{v^\top \mathbf{T}_{\tau_j} v}{v^\top \mathbf{T}_{\tau_{j-1}} v} \geq 2 - \frac{3\eta}{v^\top \mathbf{T}_{\tau_{j-1}} v}$$

By the switching criterion, $v^\top \mathbf{T}_{\tau_{j-1}} v \geq \lambda$, so

$$\frac{v^\top \mathbf{T}_{\tau_j} v}{v^\top \mathbf{T}_{\tau_{j-1}} v} \geq 2 - \frac{3\eta}{\lambda}$$

With $\eta = \frac{\lambda}{4\sqrt{d}+1}$, we get

$$\frac{v^\top \mathbf{T}_{\tau_j} v}{v^\top \mathbf{T}_{\tau_{j-1}} v} \geq 2 - \frac{3}{4\sqrt{d}+1}$$

Thus, $\frac{v^\top \mathbf{T}_{\tau_j} v}{v^\top \mathbf{T}_{\tau_{j-1}} v} \geq 11/8$ for all $d \geq 1$.

By Lemma S.3,

$$\frac{\det(\mathbf{T}_{\tau_j})}{\det(\mathbf{T}_{\tau_{j-1}})} \geq \max_{\|v\|=1} \frac{v^\top \mathbf{T}_{\tau_j} v}{v^\top \mathbf{T}_{\tau_{j-1}} v} \geq 11/8$$

Now, we use the determinant-based doubling argument to bound the number of switches explicitly: At each switch j, we have from above:

$$\frac{\det(\mathbf{T}_{\tau_j})}{\det(\mathbf{T}_{\tau_{j-1}})} \geq 11/8$$

Applying this recursively for m switches,

$$\det(\mathbf{T}_{\tau_m}) \geq (11/8)^{m-1} \det(\mathbf{T}_{\tau_1})$$

By Lemma S.4, $\det(\mathbf{T}_{\tau_m}) \leq \det(\mathbf{T}_t) \leq (\lambda + tL^2/d)^d$ and $\det(\mathbf{T}_{\tau_1}) = \lambda^d$. Thus,

$$(11/8)^{m-1} \lambda^d \leq (\lambda + tL^2/d)^d$$

Taking logarithms,

$$(m-1)\log(11/8) \leq d \log\left(1 + \frac{tL^2}{\lambda d}\right)$$

So,

$$m \leq 1 + \frac{d}{\log(11/8)} \log\left(1 + \frac{tL^2}{\lambda d}\right)$$

which is $O(d \log t)$.

□

Remark S.6. One can observe that for monotone *noiseless* design matrices $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_t$ switching according to the criterion in Lemma S.5 ensures that $\frac{\det(\hat{\mathbf{T}}_{\tau_{i+1}})}{\det(\hat{\mathbf{T}}_{\tau_i})} \geq 2$ from Lemma S.4. Thus, the number of policy switches under our criterion of spectral directions is bounded by the policy switches under the determinant doubling criterion used commonly under rarely switching algorithms [Abbasi-yadkori et al., 2011, Sawarni et al., 2024].

S.1 Issues with determinant based switching criterion for noisy design matrices

Consider the setting in Lemma S.5, but now suppose the switching criterion is based on determinant doubling, as in [Sawarni et al., 2024, Abbasi-yadkori et al., 2011]. For noisy, non-monotone matrices, the classical determinant doubling argument breaks down: we can no longer guarantee that the determinant of the noisy matrix at the last policy switch is upper bounded by the determinant at the final time step. As a result, new technical challenges arise:

In this setting, two central technical challenges arise:

- **Bounding the number of switches m:** Even with noisy, non-monotone design matrices, it is important to control how often the policy switches under the determinant doubling criterion.
- **Controlling norm inflation:** We must ensure that for any $\tau_{j-1} \le s < \tau_j$, the norm $||y||_{\hat{\mathbf{T}}_s^{-1}}$ can be bounded by a manageable multiple of $||y||_{\hat{\mathbf{T}}_{\tau_{j-1}}^{-1}}$. This is crucial for bounding the final term in the per-timestep regret in Lemma P.3.

However, one can alternately try to argue by bounding $|\log\det\hat{\mathbf{T}}_s - \log\det\mathbf{T}_s|$ at each time step s. This can be bounded as follows:

Note that $\nabla \log\det(X) = X^{-1}$ for any positive definite matrix X. Thus by mean value theorem

$$|\log\det\hat{\mathbf{T}}_s - \log\det\mathbf{T}_s| \le \sup_{t\in[0,1]} \left\| \left(\hat{\mathbf{T}}_s + t(\mathbf{T}_s - \hat{\mathbf{T}}_s)\right)^{-1} \right\|_F \cdot \|\hat{\mathbf{T}}_s - \mathbf{T}_s\|_F \tag{199}$$

$$\le \sqrt{d}\sqrt{d} \sup_{t\in[0,1]} \left\| \left(\hat{\mathbf{T}}_s + t(\mathbf{T}_s - \hat{\mathbf{T}}_s)\right)^{-1} \right\|_2 \cdot \|\hat{\mathbf{T}}_s - \mathbf{T}_s\|_2 \tag{200}$$

$$\overset{(a)}{\le} d\frac{1}{\lambda_-} \cdot \|\hat{\mathbf{T}}_s - \mathbf{T}_s\|_2 \tag{201}$$

$$= d \cdot \frac{4\sqrt{d}+1}{4\sqrt{d}\,\lambda} \cdot \frac{\lambda}{4\sqrt{d}+1} \tag{202}$$

$$= \Omega(\sqrt{d}) \tag{203}$$

Justification for (a): For all $t \in [0,1]$, $\hat{\mathbf{T}}_s + t(\mathbf{T}_s - \hat{\mathbf{T}}_s)$ is a convex combination of $\hat{\mathbf{T}}_s$ and $\mathbf{T}_s$, so its minimum eigenvalue is at least $\min\{\lambda_{\min}(\hat{\mathbf{T}}_s), \lambda_{\min}(\mathbf{T}_s)\}$. Using $\mathbf{T}_s \succeq \lambda \boldsymbol{I}$ and $\|\hat{\mathbf{T}}_s - \mathbf{T}_s\|_2 \le \frac{\lambda}{4\sqrt{d}+1}$, we get $\lambda_{\min}(\hat{\mathbf{T}}_s), \lambda_{\min}(\mathbf{T}_s) \ge \lambda - \frac{\lambda}{4\sqrt{d}+1} = \lambda\frac{4\sqrt{d}}{4\sqrt{d}+1}$. Also, we use $\|A\|_F \le \sqrt{d}\|A\|_2$ for any $d \times d$ matrix A to bound the Frobenius norm by the operator norm.

This implies that the best bound we can get on $|\log\det\mathbf{T}_s - \log\det\mathbf{T}_{\tau_i}|$ scales as $O(\sqrt{d})$ for $s \in [\tau_i, \tau_{i+1})$.

Matrix sandwich bounds for noisy design matrices.

$$\hat{\mathbf{T}}_s \preceq \mathbf{T}_s + \frac{\lambda}{4\sqrt{d}+1}\boldsymbol{I} \preceq \mathbf{T}_s + \frac{\mathbf{T}_s}{4\sqrt{d}+1} \preceq \left(1 + \frac{1}{4\sqrt{d}+1}\right)\mathbf{T}_s$$

$$\hat{\mathbf{T}}_s \succeq \mathbf{T}_s - \frac{\lambda}{4\sqrt{d}+1}\boldsymbol{I} \succeq \mathbf{T}_s - \frac{\mathbf{T}_s}{4\sqrt{d}+1} \succeq \left(1 - \frac{1}{4\sqrt{d}+1}\right)\mathbf{T}_s$$

Norm inequalities for inverse matrices. For any $y \in \mathbb{R}^d$ and $\tau_{j-1} \le s \le \tau_j$, we have

$$\begin{aligned}
\|y\|_{\hat{\mathbf{T}}_{\tau_{j-1}}^{-1}} &\le \left(1 - \frac{1}{4\sqrt{d}+1}\right)^{-1} \|y\|_{\mathbf{T}_{\tau_{j-1}}^{-1}} \\
&\le e^{O(\sqrt{d})} \left(1 - \frac{1}{4\sqrt{d}+1}\right)^{-1} \|y\|_{\mathbf{T}_s^{-1}} \\
&\le e^{\sqrt{d}} \left(1 - \frac{1}{4\sqrt{d}+1}\right)^{-1} \left(1 + \frac{1}{4\sqrt{d}+1}\right) \|y\|_{\hat{\mathbf{T}}_s^{-1}}
\end{aligned}$$

$$\leq e^{\sqrt{d}}\|y\|_{\hat{\mathbf{T}}_s^{-1}}$$

These ratios would substantially degrade the per-step regret bound in Lemma P.3. A way to avoid this might be to scale the private term in λ by a factor of $\sqrt{d}$ but this would degrade the private regret by square root of this factor too.

S.2 Determinant doubling algorithm with $d^{1/4}$ higher factor in private regret

We now propose an algorithm by tweaking λ, β and γ slightly so that the determinant doubling criterion still works (however at a cost of $\sqrt{d}$ factor increase in the private term). Formally we change line 6 from $\mathbf{H}_t \not\preceq 2\mathbf{H}_\tau$ to $\det(\mathbf{H}_t) \geq 2\det(\mathbf{H}_\tau)$

$$\lambda := \frac{5}{4}d\Lambda_T + \frac{400d\,R^2\,\Lambda_\delta \log T\left(\sqrt{\Lambda_T}+\sqrt{d}\right)}{e\,\varepsilon}, \tag{204}$$

$$\beta = C\left[R\sqrt{d\Lambda_T} + \frac{R^2 S\, d^{1/2}\Lambda_T^{1/2}\Lambda_\delta^{1/2}\big(d+\Lambda_T\big)^{1/4}}{\sqrt{e\,\varepsilon}}\right] \tag{205}$$

$$\gamma = C\left[R^2 S\sqrt{d\Lambda_T} + \frac{R^3 S^2\, d^{1/2}\Lambda_T^{1/2}\big(d+\Lambda_T\big)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right]. \tag{206}$$

for some sufficiently large constant C. We would also need to redefine $\lambda_+ = \lambda\frac{4d+2}{4d+1}$ and $\lambda_- = \lambda\frac{4d}{4d+1}$ to ensure that Assumption L.1 holds true given the error bounds on design matrix from the binary tree mechanism (see Equation (95) and Equation (93)).

Based on the calculation above, we state the following results changing the design matrix error from $\lambda/(4\sqrt{d}+1)$ to $\lambda/(4d+1)$ with new definition of λ

We now state the following lemma.

Lemma S.7. *Let $\{x_s\}_{s=1}^t$ be a set of vectors in $\mathbb{R}^d$. Consider the sequence $\{\mathbf{T}_s\}_{s=1}^t$ such that $\mathbf{T}_{s+1} = \sum_{i=1}^s x_i x_i^\top + \lambda \boldsymbol{I}$. Further, let $||x_s||_2 \leq L \forall s \in [t]$ and let $\hat{\mathbf{T}}_s$ be a noisy estimate of $\mathbf{T}_s$ satisfying $||\hat{\mathbf{T}}_s - \mathbf{T}_s||_2 \leq \frac{\lambda}{4d+1}$ for every $s \in [t]$. Define the set $\{0 = \tau_1, \tau_2, \ldots, \tau_m \leq t\}$ such that: $\det\left(\hat{\mathbf{T}}_s\right) < 2\det\left(\hat{\mathbf{T}}_{\tau_i}\right)$ for $\tau_i \leq s < \tau_{i+1}$ but $\det\left(\hat{\mathbf{T}}_{\tau_{i+1}}\right) \geq 2\det\left(\hat{\mathbf{T}}_{\tau_i}\right)$ for $i \in \{1, \ldots, m-1\}$. The following properties hold*

- *The number of times switching happens (m) is at most $d\log_2\left(1+\frac{tL^2}{\lambda d}\right)$.*
- *For any $j \in [m]$ and any $y \in \mathbb{R}^d$, we have $||y||_{\hat{\mathbf{T}}_{\tau_{j-1}}^{-1}} \leq (5/3)\sqrt{2}\,e^{1/4}||y||_{\hat{\mathbf{T}}_s^{-1}}$ whenever $\tau_{j-1} \leq s < \tau_j$.*

The last statement would be useful for Equation (129) in Lemma P.3 where we need to change the matrix norm from $\mathbf{H}_t^{-1}$ to $\mathbf{H}_\tau^{-1}$.

Proof. Given the new definition of λ in Equation (204) with the new bound in terms of λ as follows: $||\hat{\mathbf{T}}_s - \mathbf{T}_s||_2 \leq \frac{\lambda}{4d+1}$,

we have the following

$$\hat{\mathbf{T}}_s \preceq \mathbf{T}_s + \frac{\lambda}{4d+1}\boldsymbol{I} \preceq \mathbf{T}_s + \frac{\mathbf{T}_s}{4d+1} \preceq \left(1+\frac{1}{4d+1}\right)\mathbf{T}_s \tag{207}$$

$$\hat{\mathbf{T}}_s \succeq \mathbf{T}_s - \frac{\lambda}{4d+1}\boldsymbol{I} \succeq \mathbf{T}_s - \frac{\mathbf{T}_s}{4d+1} \succeq \left(1-\frac{1}{4d+1}\right)\mathbf{T}_s \tag{208}$$

and the calculation in section Appendix S.1 would lead to $|\log\det\hat{\mathbf{T}}_s - \log\det\mathbf{T}_s| \leq 1/4$.

We now have

$$\begin{aligned}
\|y\|_{\hat{\mathbf{T}}^{-1}_{\tau_{j-1}}} &\le \left(1-\frac{1}{4d+1}\right)^{-1} \|y\|_{\mathbf{T}^{-1}_{\tau_{j-1}}} \\
&\le \sqrt{\tfrac{\det \mathbf{T}_s}{\det \mathbf{T}_{\tau_{j-1}}}} \left(1-\frac{1}{4d+1}\right)^{-1} \|y\|_{\mathbf{T}_s^{-1}} \quad \text{(Lemma S.3, since } \mathbf{T}_s \succeq \mathbf{T}_{\tau_{j-1}}\text{)} \\
&\le \sqrt{2}\, e^{1/4} \left(1-\frac{1}{4d+1}\right)^{-1} \left(1+\frac{1}{4d+1}\right) \|y\|_{\hat{\mathbf{T}}_s^{-1}} \\
&\le (5/3)\sqrt{2}\, e^{1/4} \|y\|_{\hat{\mathbf{T}}_s^{-1}},
\end{aligned}$$

where the second step uses $\det \mathbf{T}_s / \det \mathbf{T}_{\tau_{j-1}} \le e^{1/2} \det \hat{\mathbf{T}}_s / \det \hat{\mathbf{T}}_{\tau_{j-1}} < 2e^{1/2}$, as $|\log\det \hat{\mathbf{T}}_r - \log\det \mathbf{T}_r| \le 1/4$ and no switch occurs before τ_j.

proving the second point of the lemma above. We now argue bound the number of switches by considering the running maximum of the noisy design matrices.

As $\mathbf{T}_t = \lambda I + \sum_{s=1}^{t-1} x_s x_s^\top$, by Lemma S.4, we have

$$\det(\mathbf{T}_t) \le \left(\lambda + \frac{tL^2}{d}\right)^d.$$

Further, from monotonicity one can argue that

$$\det(\mathbf{T}_s) \le \left(\lambda + \frac{tL^2}{d}\right)^d \qquad \text{for every } s \in [t].$$

From Equation (207), we can observe that

$$\max_{s\le t} \det\left(\hat{\mathbf{T}}_s\right) \le \left(1+\frac{1}{4d}\right)^d \left(\lambda + \frac{tL^2}{d}\right)^d \le 4/3 \left(\lambda + \frac{tL^2}{d}\right)^d.$$

We have the following set of equations from the multiplicative step:

$$\begin{aligned}
\max_{s\le t} \det\left(\hat{\mathbf{T}}_s\right) &\ge \det\left(\hat{\mathbf{T}}_{\tau_{m-1}}\right) && (209)\\
&\ge 2\det\left(\hat{\mathbf{T}}_{\tau_{m-2}}\right) && (210)\\
&\vdots \\
&\ge 2^{m-2}\det\left(\hat{\mathbf{T}}_0\right) = 2^{m-2}\lambda^d. && (211)
\end{aligned}$$

Now applying the bound $\max_{s\le t} \det\left(\hat{\mathbf{T}}_s\right)$ we get the desired bound on number of switches m.

□

With these new definitions, we verify the parametric conditions $\gamma(\lambda_+) \le \gamma$ and $\beta(\lambda_+, \lambda_-) \le \beta$, following the same algebra as in the proof of Theorem L.8 but adapted to the redefined λ_+, λ_- and the $\sqrt{d}$-larger privacy term in λ.

Verification of $\gamma(\lambda_+) \le \gamma$ and $\beta(\lambda_+, \lambda_-) \le \beta$. Since $\lambda_+/\lambda_- = O(1)$, $\lambda_- \ge c\, d\Lambda_T$, and the privacy term in Equation (204) is $\sqrt{d}$ larger than in Equation (89), we obtain

$$\sqrt{\lambda_+} \;\le\; C\left[\sqrt{d\Lambda_T} \;+\; \frac{R\, d^{1/2}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right],$$

i.e., the $d^{1/4}$ in the corresponding bound for Theorem L.8 is replaced by $d^{1/2}$.

$\gamma(\lambda_+) \le \gamma$**.** By Equation (99), the first two terms $4R^2S\sqrt{\lambda_+} + \frac{7Sd}{\sqrt{\lambda_+}}\log(\cdot)$ are bounded (using $R, S \ge c$ and $\lambda_- \ge c\,d\Lambda_T$) by

$$A \;:=\; C\left[R^2S\sqrt{d\Lambda_T} \;+\; \frac{R^3S\,d^{1/2}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right].$$

Since count1$_{\text{cutoff}}$ does not depend on the policy II switching criterion, the optimizer guarantee in Lemma L.6 is unchanged, giving $\nu_1 = O\big(\frac{RS^2}{\varepsilon}\sqrt{d\Lambda_T\log(8/\delta)}\sqrt{d+\Lambda_T}\big)$, which does not depend on γ. Hence $4RS\sqrt{\nu_1} \le C\,\frac{R^{3/2}S^2d^{1/4}\Lambda_T^{1/4}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/4}}{\sqrt{\varepsilon}}$, which is dominated by the second term of A after replacing S by S^2 there (as $R, S, \Lambda_T, \Lambda_\delta \ge c$). So $\gamma(\lambda_+) \le \gamma$ for the choice Equation (206), and no fixed-point argument is needed.

$\beta(\lambda_+, \lambda_-) \le \beta$**.** By Equation (100), the first two terms $4R\,\lambda_+/\sqrt{\lambda_-} + \frac{7d}{R\sqrt{\lambda_-}}\log(\cdot)$ are bounded by

$$C\left[R\sqrt{d\Lambda_T} \;+\; \frac{R^2\,d^{1/2}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right].$$

Under Lemma S.7, count2$_{\text{cutoff}}$ keeps the same $O(d\Lambda_T)$ order, so ν_2 retains the form $O\big(\frac{RS^2}{\varepsilon}\sqrt{d\Lambda_\delta\Lambda_T}\sqrt{d+\Lambda_T}\big)$, and $10\sqrt{\nu_2}$ is absorbed into the second term using $R, S, \Lambda_T, \Lambda_\delta \ge c$. This matches Equation (205). □

Theorem S.8. *Assume $\zeta < \delta \le \min(1/e, d/R^3)$, $d \le T$, and $\varepsilon \le 8\log(1/\delta)$. The doubling-criterion variant of Algorithm 2 (with line 6 replaced by $\det(\mathbf{H}_t) \ge 2\det(\mathbf{H}_\tau)$ and λ, β, γ as in Equations* (204) *to* (206)*) satisfies (ε, δ) joint differential privacy (for every ε) and, with probability at least $1-\zeta$, the regret R^T may be bounded as*

$$R^T = O\left(\left[Rd\sqrt{\Lambda_T}\sqrt{\tfrac{T}{\kappa^*}} + \frac{R^2S\,d\,\Lambda_T^{1/2}\sqrt{T}\,\Lambda_\delta^{1/2}(d+\Lambda_T)^{1/4}}{\sqrt{e\,\varepsilon\,\kappa^*}}\right]\sqrt{\log(RT)}\right)$$
$$+\, O\left(\kappa dR^3\left[R^2S\sqrt{d\Lambda_T} + \frac{R^3S^2\,d^{1/2}\Lambda_T^{1/2}(d+\Lambda_T)^{1/4}\Lambda_\delta^{1/2}}{\sqrt{e\,\varepsilon}}\right]^2\log T\right) \quad (212)$$

where $\Lambda_T = \log(T/\zeta)$ and $\Lambda_\delta = \log(1/\delta)$.

Suppressing the dependence on $R, S, \kappa^*, \log(1/\delta)$ and absolute constants, the parametric scalings in Equations (204) to (206) simplify to

$$\lambda = \widetilde{O}\Big(d\log T + \tfrac{d}{\varepsilon}(\log T)\max\Big(\sqrt{\log T}, \sqrt{d}\Big)\Big), \quad \beta = \widetilde{O}\Big(\sqrt{d\log T} + \tfrac{d^{1/2}(\log T)^{1/2}(d+\log T)^{1/4}}{\sqrt{\varepsilon}}\Big), \quad (213)$$

$$\gamma = \widetilde{O}\Big(\sqrt{d\log T} + \tfrac{d^{1/2}(\log T)^{1/2}(d+\log T)^{1/4}}{\sqrt{\varepsilon}}\Big), \quad (214)$$

i.e., compared with Equation (10), the privacy term in λ is $\sqrt{d}$ larger and the privacy term in β, γ has $d^{1/2}$ in place of $d^{1/4}$. The corresponding concise regret bound is

Theorem S.9 (Concise version of Theorem S.8)**.** *The doubling-criterion variant of Algorithm 2 satisfies (ε, δ) joint differential privacy and, with probability at least $1-\zeta$ (under the assumptions of Theorem S.8),*

$$R^T \;\le\; \widetilde{O}\left(\sqrt{T}\left(\frac{d\log T}{\sqrt{\kappa^*}} + \frac{d\,\log T\,(\log T + d)^{1/4}}{\sqrt{\varepsilon\,\kappa^*}}\right)\right)\,\mathrm{poly}(R, S).$$

The bound above (with the $(\log T + d)^{1/4}$ factor) holds for all T and plays the role of Theorem 6.1 in the doubling-criterion variant. Compared with Theorem 6.1, the privacy term carries a $d^{1/4}$ overhead — the price for using the simpler determinant-doubling switching criterion in line 6, which propagates from the $\sqrt{d}$ inflation of the privacy component of λ (halved by $\sqrt{\cdot}$ when entering β, γ).

Observation 3. *When $T = \Omega(e^d)$, the dominant term in the regret scales as $\widetilde{O}\Big(d\sqrt{T/\kappa^*}\log T + \frac{d\sqrt{T}}{\sqrt{\varepsilon\kappa^*}}(\log T)^{5/4}\Big)$; the latter term is the additive privacy cost.*

When $T = O(e^d)$, the dominant term in the regret scales as $\widetilde{O}\Big(d\sqrt{T/\kappa^*}\log T + \frac{d^{5/4}\sqrt{T}}{\sqrt{\varepsilon\kappa^*}}\log T\Big)$.

T Switching Times and Update Counts

Let A_t denote the indicator that Criterion I is triggered at round t, so that $t \in \mathcal{T}_o$ iff $A_t = 1$.

Remark T.1 (Switching set is post-processing). Consider two datasets $\mathcal{D}$ and $\mathcal{D}'$ over reward, context set pairs which differ only at index i. Then $\{\mathcal{T}_{o,t} \setminus \{i\}\}_{t=1}^T$ is a post-processing of the binary-tree releases $\{\mathbf{V}_t\}_{t=1}^T$ under the same map for $\mathcal{D}$ and $\mathcal{D}'$; in particular it is at least as indistinguishable between $\mathcal{D}$ and $\mathcal{D}'$ as those releases (Lemma L.4), at no additional privacy cost.

Justification. For every $t \neq i$,

$$A_t = \mathbb{1}\Big\{\max_{x \in \mathcal{X}_t} \|x\|^2_{\mathbf{V}_t^{-1}} \geq \frac{1}{2\gamma^2 \kappa R^2}\Big\},$$

where $\mathbf{V}_t$ is the matrix released by the binary tree at the start of round t (Criterion I is not capped). Hence $(A_t)_{t \neq i} = G\big((\mathbf{V}_t)_{t=1}^T\big)$ for a map G that is determined by $(\mathcal{X}_t)_{t \neq i}$, which coincide under $\mathcal{D}$ and $\mathcal{D}'$, and $\mathcal{T}_{o,t} \setminus \{i\} = \{s \leq t : s \neq i,\ A_s = 1\}$. The claim follows by post-processing. Note that the update times of $\hat{\theta}_o$ and $\hat{\theta}_\tau$ (lines 3 and 6) are likewise functions of the released matrices and public counters only, so neither reveals A_i. □

Lemma T.2. *With probability at least $1 - \zeta/8$, Criterion I in Algorithm 2 is triggered at most N_o times, and the update conditions for $\hat{\theta}_o$ (line 3) and $\hat{\theta}_\tau$ (line 6) are met at most count1$_{cutoff}$ and count2$_{cutoff}$ times, respectively; in particular, the hard caps are never reached.*

Proof. We now define $\mathbf{V}_t$ as the value of $\mathbf{V}$ at time step t and thus we have, $\mathbf{V}_t \succeq \lambda_- \boldsymbol{I} + \sum_{t \in \mathcal{T}_o} x_t x_t^\top$ (from Assumption L.1).

Thus, wherever the switching criterion is satisfied, we have $||x||^2_{\mathbf{V}^{-1}} \geq \frac{1}{8\gamma^2 \kappa R^2}$ which further implies that

$$\sum_{t \in \mathcal{T}_o} ||x_t||^2_{\mathbf{V}^{-1}} \geq \frac{|\mathcal{T}_o|}{8\gamma^2 \kappa R^2} \tag{215}$$

However, applying the elliptic potential lemma (lemma S.1), we have $\sum_{t \in \mathcal{T}_o} ||x_t||^2_{\mathbf{V}^{-1}} \leq 2d \log\left(1 + \frac{|\mathcal{T}_o|}{\lambda_- d}\right)$ [14]

Thus, we get $|\mathcal{T}_o| \leq 16 d R^2 \kappa \gamma^2 \log T = N_o$ as $\lambda_- > 1$.

The bound on the number of updates of $\hat{\theta}_o$ follows from Lemma S.5 applied to the released matrices $\mathbf{V}_t$ (whose leaves $x_t x_t^\top$ satisfy $\|x_t\|_2 \leq 1$), using Assumption L.1: the update condition $\mathbf{V}_t \not\preceq 2\mathbf{V}_{\tau_o}$ is met at most $1 + d \log_{11/8}(1 + T/(\lambda d)) \leq 4d \log_2(1 + T/d) =$ count1$_{\text{cutoff}}$ times, since $\lambda \geq 1$ and $T \geq d$.

The bound on count2$_{\text{cutoff}}$ follows from a natural application of Lemma S.5 on application of Assumption L.1 and Lemma S.4. Note that the bound L in Lemma S.5 is $R\sqrt{R}$ as $\frac{\dot{\mu}(x, \hat{\theta})}{e} \leq \dot{\mu}(x, \theta^*) \leq \frac{1}{\kappa^*} \leq R^2$ from (110) which holds under event $\mathcal{E}_o$ (the event in Lemma N.1) which is shown to hold with high probability.

□

U A Privacy Lower Bound for JDP GLM Bandits

We extend the privacy lower bound of He et al. [2022, Proposition 4.3] from linear contextual bandits with Gaussian noise to canonical GLM bandits, using their notation and hard instance. The only model-specific changes are: (i) the per-sample total variation bound, where $\ddot{b} \leq M_0$ replaces the

[14] One could bound $\mathbf{V} \succeq \lambda_- \boldsymbol{I} + \sum_{t \in \mathcal{T}_o} x_t x_t^\top$ and then apply the lemma.

Gaussian computation; and (ii) the per-round regret of a wrong action, which gains a factor m_0 from $\dot\mu \ge m_0$. Both constants are needed only *locally*. For $r_0 > 0$ let

$$m_0 := \inf_{|z|\le r_0} \dot\mu(z) > 0, \qquad M_0 := \sup_{|z|\le r_0} \dot\mu(z) < \infty.$$

For instance, for the logistic link, $m_0 = \dot\mu(r_0) \to 1/4$ and $M_0 = 1/4$ as $r_0 \to 0$, irrespective of κ.

Lower-bound framework. We follow the formulation of He et al. [2022, Section 2]. A problem instance is a pair (θ^*, F), where $\theta^* \in \Theta^*$ is the unknown parameter and F is the distribution from which the arm contexts $X_t = (x_{t1}, \dots, x_{tK})$ are drawn independently in each round t. The per-round regret of an algorithm π is $R_t^\pi(\theta^*, F) = \max_a \mu(x_{ta}^\top \theta^*) - \mu(x_{t,a_t^\pi}^\top \theta^*)$, and the minimax regret over a class Π of algorithms is

$$R_{[T]}(\Theta^*, \{F\}) := \inf_{\pi\in\Pi} \sup_{\theta^*\in\Theta^*} \sum_{t=1}^{T} \mathbb{E}\big[R_t^\pi(\theta^*, F)\big].$$

To lower bound it, we place a prior ν_r on Θ^* and use $\sup_{\theta^*\in\Theta^*} \ge \mathbb{E}_{\theta^*\sim\nu_r}$.

Privacy notion. In Definition C.4, the data of user t is $U_t = (\mathcal{X}_t, r_t)$, its context set and *observed* reward; accordingly, in the privacy analysis of Algorithm 2 (Appendix L.3), neighbouring datasets differ in one round's context set and observed reward. For the lower bound we use a slightly stronger notion, in which a user's data consists of their context set together with the *potential* rewards of all arms, whether or not those arms are pulled. In the terminology of Azize and Basu [2024a], these are the *View* and *Table* versions of DP, respectively: Table DP implies View DP with the same parameters, the two coincide for $\delta = 0$, and for $\delta > 0$ Table DP is strictly stronger [Azize and Basu, 2024a, Proposition 1]. The lower-bound argument of He et al. [2022, Proposition 4.3] likewise operates on full-information data containing the rewards of all arms.

Definition U.1 (Table JDP)**.** An algorithm is (ε, δ)-Table JDP if it satisfies Definition C.4 with the data of user t replaced by $U_t = (\mathcal{X}_t, (y_{t,a})_{a\in[K]})$, so that neighbouring user sequences differ in one user's context set and potential rewards.

This is not a restriction for our algorithm. The privacy analysis of Algorithm 2 conditions round by round on previously released outputs, and user u enters only through the arm chosen from $\mathcal{X}_u$ and its reward y_{u,a_u}. The analysis therefore applies verbatim when users are U_u, so Algorithm 2 is (ε, δ)-Table JDP, and our upper and lower bounds concern the same class of algorithms.

Problem instance. Following He et al. [2022, Section 4.1.1], we specify (F, ν_r) for $K = 2$ as follows. Write $\theta = (\theta^1, \dots, \theta^{\lceil d/2\rceil})$ with $\theta^i \in \mathbb{R}^2$ for $i \le \lfloor d/2\rfloor$, and $\theta^{\lceil d/2\rceil} = \theta_d$ if d is odd.

- F: Draw $z_t = (0, \dots, 0, z_t^i, 0, \dots, 0)$, where $1 \le i \le \lfloor d/2\rfloor$ is chosen with equal probability and z_t^i is uniform on the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$, so that $\|z_t^i\|_2 = 1$.[15] Assign $(x_{t1}, x_{t2}) = (0, -z_t)$ or $(z_t, 0)$ with equal probability. The reward of arm a is drawn from the GLM (Definition 4.1) with predictor $x_{ta}^\top \theta^*$.
- ν_r: The uniform distribution on the product of circles $\Theta_r^* = \{\theta^* : \|\theta^{*i}\|_2 = r,\ 1 \le i \le \lfloor d/2\rfloor,\ \theta_d = 0 \text{ if } d \text{ is odd}\}$.

We write $\overline{\mathcal{H}}_t = \{(x_{sa}, y_{sa}) : a = 1, 2,\ 1 \le s \le t-1\}$ for the full-information history, which contains the rewards of both arms, and $N_t^i = \sum_{s=1}^{t-1} \mathbf{1}\{z_s^i \ne 0\}$, so that $\mathbb{E}[N_t^i] = (t-1)/\lfloor d/2\rfloor$. We write $\mathcal{C}_t = \{(x_{s1}, x_{s2}) : s \le t-1\}$ for the past contexts. Every predictor satisfies $|x_{ta}^\top \theta^*| \le \|z_t^i\|_2 \|\theta^{*i}\|_2 = r$, so all predictors lie in $[-r_0, r_0]$ whenever $r \le r_0$. Also, $\Theta_r^* \subseteq \{\|\theta\|_2 \le S\}$ whenever $r\sqrt{\lfloor d/2\rfloor} \le S$.

Proposition U.2. *Let $d \ge 2$, let $\Theta^* = \{\theta^* : \|\theta^*\|_2 \le S\}$, and let Π be the set of (ε, δ)-Table JDP algorithms (Definition U.1). For $r \le r_{\max} := \min\{r_0,\ S/\sqrt{\lfloor d/2\rfloor}\}$,*

$$R_{[T]}(\Theta^*, \{F\}) = \Omega\Big(\sum_{t=1}^{T} m_0 \min\Big\{r,\ \frac{d^2}{M_0\, r\, t^2(\varepsilon+\delta)^2}\Big\}\Big).$$

[15] He et al. [2022] use a truncated Gaussian z_t^i, which they need for their Theorem 3.2. Our argument uses only that $\|z_t^i\|_2 \le 1$ and that the direction of z_t^i is uniform and independent of $\|z_t^i\|_2$. Hence any rotation-invariant law on the unit disk, including theirs, gives the same bound up to the constant $\mathbb{E}\,\|z_t^i\|_2$.

Letting $r = \frac{d}{\sqrt{M_0}\,T(\varepsilon+\delta)}$, which is admissible when $T(\varepsilon+\delta) \geq d/(\sqrt{M_0} r_{\max})$, we obtain $R_{[T]}(\Theta^*, \{F\}) = \Omega\big(\frac{m_0}{\sqrt{M_0}} \cdot \frac{d}{\varepsilon+\delta}\big)$. This is $\Omega(\frac{d}{\varepsilon+\delta})$ when $m_0, M_0 = \Theta(1)$ on $[-r_0, r_0]$, matching He et al. [2022, Proposition 4.3]. Since stochastic contexts are a special case of adversarial ones, the bound applies to setting **M2**.

Two ingredients. *(i) Per-sample TV.* For predictors $\eta, \eta' \in [-r_0, r_0]$, let P_η denote the reward law. Then $D_{\mathrm{KL}}(P_\eta \| P_{\eta'}) = b(\eta') - b(\eta) - \dot{b}(\eta)(\eta' - \eta) = \frac{1}{2}\ddot{b}(\xi)(\eta' - \eta)^2 \leq \frac{M_0}{2}(\eta - \eta')^2$ for some ξ between η and η'. Pinsker's inequality then gives

$$d_{\mathrm{TV}}(P_\eta, P_{\eta'}) \leq \tfrac{\sqrt{M_0}}{2}|\eta - \eta'|. \tag{216}$$

For $\theta, \theta' \in \Theta_r^*$ that differ only in block i, conditional on $\mathcal{C}_t$, the record of round s has the same law under θ and θ' unless $z_s^i \neq 0$. In that case, by (216), its TV is at most $\frac{\sqrt{M_0}}{2}|z_s^{i\top}(\theta^i - \theta'^i)| \leq \frac{\sqrt{M_0}}{2}\|\theta^i - \theta'^i\|_2$.

(ii) Comparison of DP outputs. We use the following linearized form of Karwa and Vadhan [2018, Lemma 6.1]. Their lemma is stated for i.i.d. samples; this version allows independent, non-identically distributed records, which is what conditioning on $\mathcal{C}_t$ produces.

Lemma U.3. *Let $D = (u_1, \ldots, u_n)$ and $D' = (u'_1, \ldots, u'_n)$ have independent entries with $d_{\mathrm{TV}}(u_s, u'_s) \leq \rho_s$, and let $\mathcal{M}$ be (ε, δ)-DP. Then for every event E, $\big|\Pr(\mathcal{M}(D) \in E) - \Pr(\mathcal{M}(D') \in E)\big| \leq (\varepsilon + \delta)\sum_s \rho_s$.*

Proof. Let

$$\alpha_{\varepsilon,\delta} \coloneqq \frac{e^\varepsilon - 1 + 2\delta}{e^\varepsilon + 1}.$$

We first claim that, for every pair of neighboring deterministic datasets x, x', we have

$$d_{\mathrm{TV}}\big(\mathcal{M}(x), \mathcal{M}(x')\big) \leq \alpha_{\varepsilon,\delta} \leq \varepsilon + \delta. \tag{217}$$

Indeed, fix an event E, and write

$$p \coloneqq \Pr(\mathcal{M}(x) \in E), \qquad q \coloneqq \Pr(\mathcal{M}(x') \in E).$$

Suppose without loss of generality that $p \geq q$. By differential privacy, applied respectively to E and to E^c with the roles of x, x' reversed,

$$p \leq e^\varepsilon q + \delta, \qquad 1 - q \leq e^\varepsilon(1 - p) + \delta.$$

Equivalently,

$$p - q \leq (e^\varepsilon - 1)q + \delta,$$

and

$$p - q \leq \frac{(e^\varepsilon - 1)(1 - q) + \delta}{e^\varepsilon}.$$

Maximizing the minimum of these two upper bounds over $q \in [0, 1]$ gives

$$p - q \leq \frac{e^\varepsilon - 1 + 2\delta}{e^\varepsilon + 1} = \alpha_{\varepsilon,\delta}.$$

This proves the first inequality in (217). The second follows since

$$\frac{e^\varepsilon - 1}{e^\varepsilon + 1} \leq \varepsilon \qquad \text{and} \qquad \frac{2\delta}{e^\varepsilon + 1} \leq \delta.$$

For each $s \in [n]$, take a maximal coupling (U_s, U'_s) of u_s, u'_s, so that

$$\Pr(U_s \neq U'_s) = d_{\mathrm{TV}}(u_s, u'_s) \leq \rho_s.$$

Since the entries are independent, we may take these couplings independently across s. Define

$$X \coloneqq (U_1, \ldots, U_n), \qquad X' \coloneqq (U'_1, \ldots, U'_n).$$

Then $X \overset{d}{=} D$ and $X' \overset{d}{=} D'$.

Conditional on X, X', change the coordinates on which they differ one at a time. By the triangle inequality and (217),

$$|\Pr(\mathcal{M}(X) \in E \mid X) - \Pr(\mathcal{M}(X') \in E \mid X')| \le (\varepsilon + \delta) d_H(X, X').$$

Taking expectations over the coupling gives

$$\begin{aligned} |\Pr(\mathcal{M}(D) \in E) - \Pr(\mathcal{M}(D') \in E)| &\le (\varepsilon + \delta)\mathbb{E}[d_H(X, X')] \\ &= (\varepsilon + \delta) \sum_{s=1}^{n} \Pr(U_s \neq U'_s) \\ &\le (\varepsilon + \delta) \sum_{s=1}^{n} \rho_s, \end{aligned}$$

as desired. □

In the proof of He et al. [2022, Proposition 4.3], the privacy comparison is first used to lower bound the estimation error, and the regret bound then follows from their Theorem 3.4, which requires the π-induced estimator θ_t^π to lie in the class $\hat{\Theta}_t$ of all (ε, δ)-DP functions that map $\overline{\mathcal{H}}_t$ to a vector in $\mathbb{R}^d$. However, θ_t^π is a deterministic functional of the action probabilities of π, and it need not be (ε, δ)-DP. We therefore apply the same two-point comparison directly to the action a_t^π, which is private by Definition U.1.

Proof of Proposition U.2. Actions are DP. Under Definition U.1, user s is $(x_{s1}, x_{s2}, y_{s1}, y_{s2})$, so $\overline{\mathcal{H}}_t$ is the data of users $1, \dots, t-1$. The action a_t^π depends on $\overline{\mathcal{H}}_t$, (x_{t1}, x_{t2}), and internal randomness, but not on later users. Hence, for fixed (x_{t1}, x_{t2}), the map $\overline{\mathcal{H}}_t \mapsto a_t^\pi$ is (ε, δ)-DP.

GLM regret of a wrong action. Since μ is increasing, arm 1 is optimal iff $z_t^\top \theta^* > 0$, as in the linear case. Both predictors of round t lie in $[-r, r]$, where $\dot{\mu} \ge m_0$. By the mean value theorem, the linear identity in the proof of He et al. [2022, Lemma 3.1] becomes

$$R_t^\pi(\theta^*, F) \;\ge\; m_0 \Big(-z_t^\top \theta^* \, \mathbf{1}\{a_t^\pi = 1,\; z_t^\top \theta^* \le 0\} + z_t^\top \theta^* \, \mathbf{1}\{a_t^\pi = 2,\; z_t^\top \theta^* > 0\} \Big). \tag{218}$$

Two-point argument. Fix t, i, $\varphi \in (0, \pi/2]$, and $\theta \in \Theta_r^*$. Let $\theta' \in \Theta_r^*$ be obtained from θ by rotating block θ^i counterclockwise by φ while keeping all other blocks θ^j, $j \neq i$, unchanged, so that $\|\theta^i - \theta'^i\|_2 = 2r \sin(\varphi/2) \le r\varphi$. Write $\theta^i = r(\cos\alpha, \sin\alpha)$ and $z_t^i = (\cos\beta, \sin\beta)$, and let $\psi = \beta - \alpha - \pi/2$. Then $z_t^\top \theta = -r \sin\psi$ and $z_t^\top \theta' = r \sin(\varphi - \psi)$. Let $A_i = \{z_t^i \neq 0,\; \psi \in [\frac{\varphi}{4}, \frac{3\varphi}{4}] \cup [\pi + \frac{\varphi}{4}, \pi + \frac{3\varphi}{4}]\}$. On A_i, $z_t^\top \theta$ and $z_t^\top \theta'$ have opposite signs, so the optimal arms under θ and θ' differ. Moreover, $\min\{|z_t^\top \theta|, |z_t^\top \theta'|\} \ge r \sin\frac{\varphi}{4} \ge \frac{r\varphi}{2\pi}$. Consider $z_t^\top \theta > 0 > z_t^\top \theta'$ (the other case is symmetric). By (218), ingredient (i), and Lemma U.3,

$$\begin{aligned} &\mathbb{E}\big[R_t^\pi(\theta^*, F) \;\big|\; \theta^* = \theta, \mathcal{C}_{t+1}\big] + \mathbb{E}\big[R_t^\pi(\theta^*, F) \;\big|\; \theta^* = \theta', \mathcal{C}_{t+1}\big] \\ &\quad \ge m_0 \frac{r\varphi}{2\pi} \Big[\Pr\big(a_t^\pi = 2 \;\big|\; \theta^* = \theta, \mathcal{C}_{t+1}\big) + \Pr\big(a_t^\pi = 1 \;\big|\; \theta^* = \theta', \mathcal{C}_{t+1}\big) \Big] \\ &\quad \ge m_0 \frac{r\varphi}{2\pi} \Big(1 - (\varepsilon + \delta) \tfrac{\sqrt{M_0}}{2} N_t^i \, r\varphi \Big). \end{aligned}$$

Here $\mathcal{C}_{t+1}$ also contains the current contexts (x_{t1}, x_{t2}). Write $\gamma(N) := (\varepsilon + \delta)\frac{\sqrt{M_0}}{2} N \, r\varphi$. Since the left-hand side is nonnegative, we may replace $1 - \gamma(N_t^i)$ by $(1 - \gamma(N_t^i))^+$. We now remove the conditioning in four steps.

(a) Integrate over contexts. Let $f_i(\theta) := \mathbb{E}[R_t^\pi(\theta^*, F)\mathbf{1}\{z_t^i \neq 0\} \mid \theta^* = \theta]$ be the expected regret in round t on the event that block i is active. Since $R_t^\pi \ge 0$ and $A_i \subseteq \{z_t^i \neq 0\}$, we have $f_i(\vartheta) \ge \mathbb{E}[R_t^\pi \mathbf{1}\{A_i\} \mid \theta^* = \vartheta]$ for $\vartheta \in \{\theta, \theta'\}$. The contexts $\mathcal{C}_{t+1}$ have the same law under θ and θ', so we may integrate the display above over $\{A_i\}$. Given $z_t^i \neq 0$, the angle β is uniform and independent of $\mathcal{C}_t$, so A_i is independent of N_t^i and $\Pr(A_i) = \frac{\varphi}{2\pi\lfloor d/2 \rfloor}$. Hence

$$f_i(\theta) + f_i(\theta') \;\ge\; m_0 \frac{r\varphi}{2\pi} \cdot \frac{\varphi}{2\pi\lfloor d/2 \rfloor} \cdot \mathbb{E}\big[(1 - \gamma(N_t^i))^+\big] \;\ge\; \frac{m_0 \, r\varphi^2}{4\pi^2 \lfloor d/2 \rfloor} \Big(1 - \gamma\big(\tfrac{t-1}{\lfloor d/2 \rfloor}\big) \Big)^+ =: B,$$

where the last step is Jensen's inequality, since $N \mapsto (1-\gamma(N))^+$ is convex, combined with $\mathbb{E}[N_t^i] = \frac{t-1}{\lfloor d/2 \rfloor}$.

(b) Average over the prior. The bound B does not depend on θ. Since $\theta' \sim \nu_r$ when $\theta \sim \nu_r$, we have $\mathbb{E}_{\nu_r} f_i(\theta') = \mathbb{E}_{\nu_r} f_i(\theta)$. Taking $\mathbb{E}_{\nu_r}$ in (a) therefore gives $2\ \mathbb{E}_{\nu_r} f_i(\theta) \geq B$.

(c) Sum over blocks. In every round z_t has exactly one nonzero block, i.e. $\sum_i \mathbf{1}\{z_t^i \neq 0\} = 1$, so $\sum_i f_i(\theta) = \mathbb{E}[R_t^\pi(\theta^*, F) \mid \theta^* = \theta]$. Summing (b) over the $\lfloor d/2 \rfloor$ blocks gives

$$\mathop{\mathbb{E}}_{\nu_r} \mathbb{E}\big[R_t^\pi(\theta^*, F)\big] \ \geq\ \frac{\lfloor d/2 \rfloor}{2} B \ =\ \frac{m_0\, r\varphi^2}{8\pi^2}\Big(1 - (\varepsilon+\delta)\frac{\sqrt{M_0}(t-1)\, r\varphi}{2\lfloor d/2 \rfloor}\Big)^+.$$

(d) Choose the angle. Choosing $\varphi = \min\{\frac{\pi}{2},\ \lfloor d/2 \rfloor / ((\varepsilon+\delta)\sqrt{M_0}(t-1)r)\}$ (with $\varphi = \frac{\pi}{2}$ if $t=1$) makes the bracket at least $\frac{1}{2}$. Hence

$$\begin{aligned}\mathop{\mathbb{E}}_{\nu_r} \mathbb{E}\big[R_t^\pi(\theta^*, F)\big] \ &\geq\ \frac{m_0}{16\pi^2} \min\Big\{\frac{\pi^2 r}{4},\ \frac{\lfloor d/2 \rfloor^2}{M_0\, r\, (t-1)^2(\varepsilon+\delta)^2}\Big\} \\ &= \Omega\Big(m_0 \min\Big\{r,\ \frac{d^2}{M_0\, r\, t^2(\varepsilon+\delta)^2}\Big\}\Big),\end{aligned}$$

for every $\pi \in \Pi$

Finally, $R_{[T]}(\Theta^*, \{F\}) \geq \inf_{\pi\in\Pi} \mathbb{E}_{\nu_r} \sum_t \mathbb{E}[R_t^\pi(\theta^*, F)]$, since $\Theta_r^* \subseteq \Theta^*$ and supremum exceeds the expectation. Summing the regret bound over t proves the proposition. □

Remark U.4 (Comparison with He et al. [2022]). The following are taken from He et al. [2022]: the instance (F, ν_r), N_t^i, the use of Karwa and Vadhan [2018, Lemma 6.1], the two-point comparison between nearby parameters on the circle, the Jensen step over N_t^i, and the resulting rate. The GLM-specific changes are (216) and (218).

V Bounding parameter estimation error under approximate convex minimization

Fix $0 < \lambda_- \leq \lambda_+$. Denote $\mathcal{L}_t(\theta) := \sum_{s=1}^t \ell(\theta, x_s, r_s) + \frac{\lambda_+}{2}\|\theta\|_2^2$, let $\hat{\theta}$ minimise this loss over a convex set Θ up to an error of ν, and let $\tilde{\theta}$ denote the global minimiser of $\mathcal{L}_t$, i.e.,

$$\mathcal{L}_t(\hat{\theta}) - \min_{\theta\in\Theta} \mathcal{L}_t(\theta) < \nu \text{ for some } \hat{\theta} \in \Theta \text{ and } \tilde{\theta} = \arg\min_\theta \mathcal{L}_t(\theta). \tag{219}$$

Further, for any symmetric matrix $\mathcal{N}$ with $\lambda_- \boldsymbol{I} \preceq \mathcal{N} \preceq \lambda_+ \boldsymbol{I}$, let

$$\mathbf{H}(\theta) = \sum_{s=1}^t \dot{\mu}\left(\langle x_s, \theta\rangle\right) x_s x_s^\top + \mathcal{N} \quad \text{and} \quad g(\theta) = \sum_{s=1}^t \mu\left(\langle x_s, \theta\rangle\right) x_s + \lambda_+ \theta.$$

When $\mathcal{N} = \lambda \boldsymbol{I}$, take $\lambda_- = \lambda_+ = \lambda$.

We now have the following lemma.

Lemma V.1. *Suppose* $\|g(\tilde{\theta}) - g(\theta^*)\|_{\mathbf{H}(\theta^*)^{-1}} \leq \gamma$. *If* $\theta^* \in \Theta$ *and* $\max_{s\in[t]} |\langle x_s, \hat{\theta} - \theta^*\rangle| \leq \frac{c}{R}$, *then*

$$\|\hat{\theta} - \theta^*\|_{\mathbf{H}(\theta^*)} \leq (2+c)\gamma + \sqrt{(2+c)\nu} + \sqrt{(2+c)\left(\frac{R}{\sqrt{\lambda_-}}\gamma^2 + \gamma\right)}.$$

Proof. Write $\Delta := \hat{\theta} - \theta^*$ and $\mathbf{H} := \mathbf{H}(\theta^*)$. By Lemma K.2 and the hypothesis on $|\langle x_s, \Delta\rangle|$, $\tilde{\alpha}(\langle x_s, \theta^*\rangle, \langle x_s, \hat{\theta}\rangle) \geq \frac{\dot{\mu}(\langle x_s, \theta^*\rangle)}{2+c}$ for every s.

Following Abeille et al. [2021], Sawarni et al. [2024], the exact second-order Taylor expansion of $\mathcal{L}_t$ around θ^* gives

$$\mathcal{L}_t(\hat{\theta}) - \mathcal{L}_t(\theta^*) - \langle \nabla\mathcal{L}_t(\theta^*), \Delta\rangle = \|\Delta\|_{\tilde{\mathbf{G}}}^2, \qquad \tilde{\mathbf{G}} := \sum_{s=1}^t \tilde{\alpha}\left(\langle x_s, \theta^*\rangle, \langle x_s, \hat{\theta}\rangle\right) x_s x_s^\top + \frac{\lambda_+}{2}\boldsymbol{I}.$$

Since $c \geq 0$ and $\mathcal{N} \preceq \lambda_+ \boldsymbol{I}$, we have $\tilde{\mathbf{G}} \succeq \frac{1}{2+c}\mathbf{H}$. Hence, by Cauchy–Schwarz,

$$\|\Delta\|_{\mathbf{H}}^2 \leq (2+c)\left(\left|\mathcal{L}_t(\hat{\theta}) - \mathcal{L}_t(\theta^*)\right| + \|\nabla\mathcal{L}_t(\theta^*)\|_{\mathbf{H}^{-1}}\|\Delta\|_{\mathbf{H}}\right). \tag{220}$$

Gradient term. Since $\tilde{\theta}$ is the unconstrained minimiser, $\nabla\mathcal{L}_t(\tilde{\theta}) = 0$, i.e., $\sum_{s\in[t]} r_s x_s = g(\tilde{\theta})$. Therefore $\nabla\mathcal{L}_t(\theta^*) = g(\theta^*) - \sum_{s\in[t]} r_s x_s = g(\theta^*) - g(\tilde{\theta})$, and $\|\nabla\mathcal{L}_t(\theta^*)\|_{\mathbf{H}^{-1}} \leq \gamma$.

Loss term. Since $\theta^* \in \Theta$, (219) gives $\mathcal{L}_t(\hat{\theta}) - \mathcal{L}_t(\theta^*) \leq \nu$, and since $\tilde{\theta}$ is the global minimiser, $\mathcal{L}_t(\hat{\theta}) - \mathcal{L}_t(\theta^*) \geq -(\mathcal{L}_t(\theta^*) - \mathcal{L}_t(\tilde{\theta}))$. Hence

$$\left|\mathcal{L}_t(\hat{\theta}) - \mathcal{L}_t(\theta^*)\right| \leq \left(\mathcal{L}_t(\theta^*) - \mathcal{L}_t(\tilde{\theta})\right) + \nu.$$

Following the algebraic manipulations of Appendix E.1 of Sawarni et al. [2024] and Appendix B of Abeille et al. [2021], and using $\mathbf{H} \succeq \lambda_- \boldsymbol{I}$,

$$\mathcal{L}_t(\theta^*) - \mathcal{L}_t(\tilde{\theta}) \leq \frac{R}{\sqrt{\lambda_-}}\|g(\theta^*) - g(\tilde{\theta})\|_{\mathbf{H}^{-1}}^2 + \|g(\theta^*) - g(\tilde{\theta})\|_{\mathbf{H}^{-1}} \leq \frac{R}{\sqrt{\lambda_-}}\gamma^2 + \gamma.$$

Conclusion. Substituting both bounds into (220) gives $x^2 \leq bx + b'$ with $x = \|\Delta\|_{\mathbf{H}}$, $b = (2+c)\gamma$, and $b' = (2+c)\left(\frac{R}{\sqrt{\lambda_-}}\gamma^2 + \gamma + \nu\right)$. Since $x^2 \leq bx + b'$ implies $x \leq b + \sqrt{b'}$, and $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$, the claim follows. □

W Shuffled Vector Summation Protocols

We now formally describe the vector summation protocols that have been described in Algorithm 1. The local randomizer, shuffler and aggregator have been formally defined in Algorithm 4, Algorithm 5 and Algorithm 6 respectively.

This protocol operates independently on each entry of vector $r_t x_t$ and $x_t x_t^\top$, transmits only 0/1 bits, and the number of bits sent is sampled from a binomial distribution. This implementation has three parameters g (accuracy parameter), $b \in \mathbb{N}$ and $p \in [0,1]$ where the latter two parameters control the accuracy. For an entry $x \in [0,1]$, it is first encoded as $\hat{x} = \bar{x} + \gamma_1$ where $\bar{x} = \lfloor xg \rfloor$ and $\gamma_1 \sim \text{Bern}(xg - \bar{x})$. Then a binomial noise $\gamma_2 \sim \text{Bin}(b,p)$ is generated. The output of the one-dimensional randomiser is simply a collection of $g+b$ bits where $\hat{x} + \gamma_2$ bits are one and the rest are zero. As discussed in Cheu et al. [2022], the vector is transmitted by biasing each entry by Δ [16] and then applying the mechanism described above. The shuffler shuffles the messages received from all users uniformly at random and then sums them pointwise and finally debiases it by subtracting $n\Delta$ from the sum where n denotes the total number of users.

$$b = \frac{24 \cdot 10^4 g^2 \cdot \left(\log\left(\frac{4(d^2+1)}{\delta}\right)\right)^2}{\varepsilon^2 n}, p = \frac{1}{4} \text{ and } g = \max\{2\sqrt{n}, d, 4\} \tag{221}$$

We now state the following lemma on privacy and utility from Cheu et al. [2022] and Chowdhury and Zhou [2022] respectively. While the privacy guarantee follows from the former, the sub-Gaussianity guarantee follows from Section C.3 of Chowdhury and Zhou [2022].

Lemma W.1. *For any $0 < \varepsilon \leq 15$, $0 < \delta < 1/2$, $d, n \in \mathbb{N}$, and $\Delta > 0$, for inputs $\tilde{X} = (x_1, \ldots, x_n)$ of vectors with maximum norm $\|x_i\|_2 \leq \Delta$, let the shuffle protocol $\mathcal{P}_1$ denote $A_1 \circ R_1 \circ S_1$. Now choosing the values of b, g and p from (221), we have*

- *The shuffle protocol $\mathcal{P}_1$ is (ε, δ)-shuffle private.*
- *Further, the shuffle protocol $\mathcal{P}_1$ provides an unbiased estimate of the sum $\sum_i x_i$. Further, the difference is a random vector with independent entries where each entry is a sub-Gaussian random variable with zero mean and variance-proxy $O\left(\frac{\Delta^2 \log^2(d^2/\delta)}{\varepsilon^2}\right)$.*

[16] The bias term is chosen as the bound on $r_t x_t$ and $x_t x_t^\top$.

Algorithm 4 Local Randomizer $\mathcal{R}_{\text{Vec}}$ [Cheu et al., 2022]

function $\mathcal{R}^*(x, g, b, p, \Delta)$
  $\bar{x} \leftarrow \lfloor xg/\Delta \rfloor$
  Sample $\gamma_1 \sim \text{Ber}(xg/\Delta - \bar{x})$
  $\hat{x} \leftarrow \bar{x} + \gamma_1$
  Sample $\gamma_2 \sim \text{Bin}(b, p)$
  Let m be a multiset with $\hat{x} + \gamma_2$ copies of 1 and $(g + b) - (\hat{x} + \gamma_2)$ copies of 0
  return m
end function
function $R_1(rx, \Delta = R, \varepsilon, \delta)$
  Set g, b, p as in (221)
  for $k \leftarrow 1$ to d **do**
    $w_k \leftarrow [rx]_k + \Delta$
    $m_k \leftarrow \mathcal{R}^*(w_k, g, b, p, 2\Delta)$
  end for
  $M_1 \leftarrow \{(k, m_k)\}_{k \in [d]}$ {Labeled outputs}
  return M_1
end function
function $R_2(xx^\top, \Delta = 1, \varepsilon, \delta)$
  Set g, b, p as in (221) with ε, δ.
  for $i \leftarrow 1$ to d **do**
    for $j \leftarrow i$ to d **do**
      $w_{(i,j)} \leftarrow [xx^\top]_{(i,j)} + \Delta$
      $m_{(i,j)} \leftarrow \mathcal{R}^*(w_{(i,j)}, g, b, p, 2\Delta)$
      $m_{(j,i)} \leftarrow m_{(i,j)}$
    end for
  end for
  $M_2 \leftarrow \{((i, j), m_{(i,j)})\}_{(i,j) \in [d] \times [d]}$ {Labeled outputs}
  return M_2
end function

Algorithm 5 Shuffler $\mathcal{S}_{\text{Vec}}$ [Cheu et al., 2022]

Require: $[n]$ is the set of users
function $S_1(\{M_{\tau,1}\}_{\tau \in [n]})$
  Uniformly permute all messages $\{(g + b) \cdot n \cdot d$ bits total$\}$
  Set y_k to be the collection of bits labeled by $k \in [d]$
  Set $Y_1 \leftarrow \{y_1, \ldots, y_d\}$
  return Y_1
end function
function $S_2(\{M_{\tau,2}\}_{\tau \in [n]})$
  Uniformly permute all messages $\{(g + b) \cdot n \cdot d^2$ bits total$\}$
  Set $y_{(i,j)}$ to be the collection of bits labeled by $(i, j) \in [d] \times [d]$
  Set $Y_2 \leftarrow \{y_{(i,j)}\}_{(i,j) \in [d] \times [d]}$
  return Y_2
end function

Lemma W.2. *For any $0 < \varepsilon \leq 15$, $0 < \delta < 1/2$, $d, n \in \mathbb{N}$, and $\Delta > 0$, for inputs $\tilde{X} = (X_1, \ldots, X_n)$ of symmetric matrices with Frobenius norm $\|X_i\|_F \leq \Delta$ (for the rank-one inputs used in this paper this equals the operator norm), let the shuffle protocol $\mathcal{P}_2$ denote $A_2 \circ R_2 \circ S_2$. Now choosing the values of b, g and p from (221), we have*

- *The shuffle protocol $\mathcal{P}_2$ is (ε, δ)-shuffle private.*
- *Further, the shuffle protocol $\mathcal{P}_2$ provides an unbiased estimate of the sum $\sum_i X_i$ and the difference is a random symmetric matrix with independent upper-triangular entries with sub-Gaussian zero mean and variance-proxy $O\left(\frac{\Delta^2 \log^2(d^2/\delta)}{\varepsilon^2}\right)$.*

Algorithm 6 Analyzer $\mathcal{A}_{\text{Vec}}$ [Cheu et al., 2022]

Require: Privacy parameters ε, δ and number of users n

function $\mathcal{A}^*(y, \Delta)$
  return $\frac{2\Delta}{g}\left(\left(\sum_{i=1}^{(g+b)\cdot n} y_i\right) - p \cdot b \cdot n\right)$
end function
function $A_1(Y_1)$
  Set g, b, p as in Equation (221)
  for $k \leftarrow 1$ to d **do**
    $z_k \leftarrow \mathcal{A}^*(y_k, g, b, p, \Delta)$
    $o_k \leftarrow z_k - n \cdot \Delta$ {Re-center}
  end for
  return $\{o_1, \ldots, o_d\}$
end function
function $A_2(Y_2)$
  Set g, b, p as in Equation (221)
  for $i \leftarrow 1$ to d **do**
    for $j \leftarrow i$ to d **do**
      $z_{(i,j)} \leftarrow \mathcal{A}^*(y_{(i,j)}, g, b, p, \Delta)$
      $o_{(i,j)} \leftarrow z_{(i,j)} - n \cdot \Delta$ {Re-center}
      $o_{(j,i)} \leftarrow o_{(i,j)}$
    end for
  end for
  return $\{o_{(i,j)}\}_{(i,j)\in[d]\times[d]}$
end function

X Private optimization algorithms

Notation. Throughout this section, T denotes the projected-gradient iteration count of the optimizer, an internal parameter of $\mathcal{P}_{\text{GD}}$ and unrelated to the bandit horizon (which we never reference in this section).

We now describe the shuffle convex optimizer $\mathcal{P}_{GD}$, which uses the vector summation protocols $(\mathcal{R}_{\text{Vec}}, \mathcal{S}_{\text{Vec}}, \mathcal{A}_{\text{Vec}})$. Before presenting the algorithm, we first define convexity and Lipschitzness of the loss function with respect to θ over Θ.

Definition X.1. A loss function $\ell : \Theta \times X \to \mathbb{R}$ is convex in Θ if, for all $x \in X$, all $t \in [0, 1]$, and all $\theta, \theta' \in \Theta$,

$$\ell(t\theta + (1-t)\theta', x) \le t\ell(\theta, x) + (1-t)\ell(\theta', x).$$

It is L-Lipschitz in Θ if, for all $x \in X$ and all $\theta, \theta' \in \Theta$,

$$|\ell(\theta, x) - \ell(\theta', x)| \le L\|\theta - \theta'\|_2.$$

For a dataset $\mathcal{D} = \{x_i\}_{i=1}^n$, define the empirical loss

$$\ell(\theta, \mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, x_i).$$

The following lemma is included only for completeness. It is a standard utility analysis for projected gradient descent with noisy gradients, following the usual projected-gradient descent proof of Bubeck [2015]. Our contribution is not the optimization analysis itself, but its use inside the private optimization subroutine required by our algorithm.

Lemma X.2 (Utility of noisy projected gradient descent)**.** *Suppose $\ell(\cdot, \mathcal{D})$ is convex and L-Lipschitz over a closed convex set $\Theta \subset \mathbb{R}^d$ of diameter D. Let $\theta^*_{\mathcal{D}} \in \arg\min_{\theta\in\Theta} \ell(\theta, \mathcal{D})$. Consider the update $\theta_{t+1} = \pi_\Theta(\theta_t - \eta\bar{g}_t)$, where $\bar{g}_t = g_t + n_t$ and $g_t \in \partial\ell(\theta_t, \mathcal{D})$, for $t = 0, \ldots, T-1$. Let $\bar{\theta} = \frac{1}{T}\sum_{t=0}^{T-1}\theta_t$. If, with probability at least $1 - \zeta$,*

$$\frac{1}{T}\sum_{t=0}^{T-1}\|n_t\|_2^2 \le B_2 \qquad \textit{and} \qquad \left|\frac{1}{T}\sum_{t=0}^{T-1}\langle n_t, \theta_t - \theta^*_{\mathcal{D}}\rangle\right| \le B_1,$$

Algorithm 7 $\mathcal{P}_{\text{GD}}$, multi-round shuffle-private gradient descent [Cheu et al., 2022]

Require: Data points $\{x_i\}_{i=1}^n$, privacy parameters ε, δ, failure probability ζ, pointwise loss function $\ell(\cdot)$ with Lipschitz parameter L, convex set Θ.

Set the number of iterations $T = \frac{\varepsilon^2 n^2}{d\Lambda_{\text{shuf}}}$ with $\Lambda_{\text{shuf}} = \log^2(nd/\delta)\log(n/\zeta)\log(1/\delta)$, and step size $\eta = \frac{D}{L\sqrt{T}}$.

1: Initialize parameter estimate $\theta_0 \leftarrow 0$.
2: **for** $t = 0, 1, \ldots, T-1$ **do**
3: Compute the noisy full-batch gradient

$$\bar{g}_t \leftarrow \frac{1}{n} P_1 \left(\text{clip}_L(\nabla_\theta \ell(\theta_t, x_1)), \ldots, \text{clip}_L(\nabla_\theta \ell(\theta_t, x_n)); \frac{\varepsilon}{2\sqrt{2T\log(1/\delta)}}, \frac{\delta}{T+1}, L \right),$$

where $P_1 = A_1 \circ S_1 \circ R_1$ and $\text{clip}_L(g) := g \min\{1, L/\|g\|_2\}$.

4: Compute and store the parameter

$$\theta_{t+1} \leftarrow \pi_\Theta(\theta_t - \eta \bar{g}_t).$$

5: **end for**
6: **return** final estimate $\bar{\theta} \leftarrow \frac{1}{T}\sum_{t=0}^{T-1} \theta_t$.

then, with probability at least $1 - \zeta$,

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le \frac{D^2}{2\eta T} + \eta L^2 + \eta B_2 + B_1.$$

Proof. By convexity and $g_t \in \partial \ell(\theta_t, \mathcal{D})$,

$$\ell(\theta_t, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le \langle g_t, \theta_t - \theta^*_{\mathcal{D}} \rangle = \langle \bar{g}_t, \theta_t - \theta^*_{\mathcal{D}} \rangle - \langle n_t, \theta_t - \theta^*_{\mathcal{D}} \rangle.$$

This is the only point at which one must be careful: the noisy direction $\bar{g}_t$ is not itself a subgradient, so the additional inner-product term with n_t must be retained.

Let $\alpha_{t+1} = \theta_t - \eta \bar{g}_t$, so $\theta_{t+1} = \pi_\Theta(\alpha_{t+1})$ and $\eta \bar{g}_t = \theta_t - \alpha_{t+1}$. Using the identity $2\langle a-b, a-c\rangle = \|a-c\|_2^2 + \|a-b\|_2^2 - \|b-c\|_2^2$ with $(a, b, c) = (\theta_t, \alpha_{t+1}, \theta^*_{\mathcal{D}})$ gives

$$\langle \bar{g}_t, \theta_t - \theta^*_{\mathcal{D}} \rangle = \frac{\|\theta_t - \theta^*_{\mathcal{D}}\|_2^2 - \|\alpha_{t+1} - \theta^*_{\mathcal{D}}\|_2^2}{2\eta} + \frac{\eta}{2}\|\bar{g}_t\|_2^2.$$

Since projection onto a closed convex set is non-expansive and $\theta^*_{\mathcal{D}} \in \Theta$, $\|\theta_{t+1} - \theta^*_{\mathcal{D}}\|_2 \le \|\alpha_{t+1} - \theta^*_{\mathcal{D}}\|_2$. Therefore,

$$\langle \bar{g}_t, \theta_t - \theta^*_{\mathcal{D}} \rangle \le \frac{\|\theta_t - \theta^*_{\mathcal{D}}\|_2^2 - \|\theta_{t+1} - \theta^*_{\mathcal{D}}\|_2^2}{2\eta} + \frac{\eta}{2}\|\bar{g}_t\|_2^2.$$

Combining the last two displays yields the one-step inequality

$$\ell(\theta_t, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le \frac{\|\theta_t - \theta^*_{\mathcal{D}}\|_2^2 - \|\theta_{t+1} - \theta^*_{\mathcal{D}}\|_2^2}{2\eta} + \frac{\eta}{2}\|\bar{g}_t\|_2^2 - \langle n_t, \theta_t - \theta^*_{\mathcal{D}} \rangle.$$

Averaging over $t = 0, \ldots, T-1$, the squared-distance terms telescope to at most D^2, since Θ has diameter D. Using convexity again, $\ell(\bar{\theta}, \mathcal{D}) \le \frac{1}{T}\sum_{t=0}^{T-1} \ell(\theta_t, \mathcal{D})$, we get

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le \frac{D^2}{2\eta T} + \frac{\eta}{2T}\sum_{t=0}^{T-1} \|\bar{g}_t\|_2^2 - \frac{1}{T}\sum_{t=0}^{T-1} \langle n_t, \theta_t - \theta^*_{\mathcal{D}} \rangle.$$

Finally, $\|\bar{g}_t\|_2^2 = \|g_t + n_t\|_2^2 \le 2\|g_t\|_2^2 + 2\|n_t\|_2^2 \le 2L^2 + 2\|n_t\|_2^2$, because ℓ is L-Lipschitz. Substituting this and the assumed bounds on the two noise terms gives

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le \frac{D^2}{2\eta T} + \eta L^2 + \eta B_2 + B_1.$$

□

We rely on the shuffle-private optimizer guarantees developed in Lemma B.6 of Cheu et al. [2022]. The lemma below can be viewed as a high-probability version of Lemma B.6 in Cheu et al. [2022]. While Cheu et al. [2022] gives the corresponding utility guarantee in expectation, our analysis requires a high-probability bound; we therefore include the short adaptation for completeness. The privacy guarantee follows from the shuffle vector-summation protocol in Lemma W.2 and its multi-round composition. The utility guarantee follows by applying the generic noisy-PGD analysis in Lemma X.2.

Lemma X.3 (Shuffle-private convex optimizer)**.** *The protocol $\mathcal{P}_{GD}$ is multi-round (ε, δ) shuffle DP. Suppose $\ell(\cdot, \mathcal{D})$ is convex and L-Lipschitz over a convex set $\Theta \subset \mathbb{R}^d$ of diameter D. For any dataset $\mathcal{D} = \{x_i\}_{i=1}^n$ and any $\eta > 0$, the output of $\mathcal{P}_{GD}$ satisfies, with probability at least $1 - \zeta$,*

$$\ell(\bar{\theta}, \mathcal{D}) - \min_{\theta \in \Theta} \ell(\theta, \mathcal{D}) \leq O\left(\frac{D^2}{\eta T} + \eta L^2 + \frac{\eta d T L^2 \log^2(d^2 T/\delta) \log(dT/\zeta) \log(1/\delta)}{n^2 \varepsilon^2} + \frac{DL \log\big(d^2 T/\delta\big) \sqrt{\log(1/\zeta) \log(1/\delta)}}{n\varepsilon} \right)$$

Under the choice

$$\eta = \frac{D}{L\sqrt{T}}, \qquad T = \frac{\varepsilon^2 n^2}{d \Lambda_{\mathrm{shuf}}}, \qquad \Lambda_{\mathrm{shuf}} = \log^2\left(\frac{nd}{\delta}\right) \log\left(\frac{n}{\zeta}\right) \log\left(\frac{1}{\delta}\right),$$

and assuming $0 < \varepsilon \leq n$ and $\Lambda_{\mathrm{shuf}} \geq 1$, the output of $\mathcal{P}_{GD}$ satisfies, with probability at least $1 - \zeta$,

$$\ell(\bar{\theta}, \mathcal{D}) - \min_{\theta \in \Theta} \ell(\theta, \mathcal{D}) \leq O\left(\frac{DL \sqrt{d \log^2\big(\frac{nd}{\delta}\big) \log\Big(\frac{n}{\zeta}\Big) \log\big(\frac{1}{\delta}\big)}}{\varepsilon n} \right).$$

Proof. The privacy statement is the multi-round shuffle-DP guarantee of the shuffle vector-summation protocol used in Algorithm 7; see Lemma B.4 of Cheu et al. [2022]. We only prove the utility bound.

Write the noisy gradient returned by the shuffle protocol as $\bar{g}_t = g_t + n_t$. By the utility guarantee of the shuffle vector-summation protocol, each coordinate of n_t is mean-zero sub-Gaussian with variance proxy

$$\tau^2 = O\left(\frac{T L^2 \log^2(d^2 T/\delta) \log(1/\delta)}{n^2 \varepsilon^2} \right).$$

We verify the two noise conditions in Lemma X.2. Since $n_{t,j}^2$ is sub-exponential with scale $O(\tau^2)$, concentration over all dT coordinates gives, with probability at least $1 - \zeta/2$,

$$\frac{1}{T} \sum_{t=0}^{T-1} \|n_t\|_2^2 = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j=1}^{d} n_{t,j}^2 \leq O\left(d\tau^2 \log(dT/\zeta) \right).$$

Thus $B_2 = O(d\tau^2 \log(dT/\zeta))$.

Next, conditioned on the history before round t, θ_t is fixed and $\|\theta_t - \theta_{\mathcal{D}}^*\|_2 \leq D$. Hence $\langle n_t, \theta_t - \theta_{\mathcal{D}}^* \rangle$ is mean-zero sub-Gaussian with variance proxy $O(\tau^2 D^2)$. Summing over t gives, with probability at least $1 - \zeta/2$,

$$\left| \frac{1}{T} \sum_{t=0}^{T-1} \langle n_t, \theta_t - \theta_{\mathcal{D}}^* \rangle \right| \leq O\left(D\tau \sqrt{\frac{\log(1/\zeta)}{T}} \right).$$

Thus $B_1 = O(D\tau\sqrt{\log(1/\zeta)/T})$. Applying Lemma X.2 and substituting the expression for τ^2 yields

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta_{\mathcal{D}}^*, \mathcal{D}) \leq O\left(\frac{D^2}{\eta T} + \eta L^2 + \frac{\eta d T L^2 \log^2(d^2 T/\delta) \log(dT/\zeta) \log(1/\delta)}{n^2 \varepsilon^2} + \frac{DL \log\big(d^2 T/\delta\big) \sqrt{\log(1/\zeta) \log(1/\delta)}}{n\varepsilon} \right).$$

Finally, set $\eta = D/(L\sqrt{T})$ and choose

$$T = \frac{\varepsilon^2 n^2}{d \Lambda_{\mathrm{shuf}}}, \qquad \Lambda_{\mathrm{shuf}} = \log^2\left(\frac{nd}{\delta}\right) \log\left(\frac{n}{\zeta}\right) \log\left(\frac{1}{\delta}\right).$$

Assume $0 < \varepsilon \le n$ and $\Lambda_{\mathrm{shuf}} \ge 1$. Then the first two terms satisfy

$$\frac{D^2}{\eta T} + \eta L^2 = O\left(\frac{DL}{\sqrt{T}}\right) = O\left(\frac{DL\sqrt{d\Lambda_{\mathrm{shuf}}}}{\varepsilon n}\right).$$

For the third term, since

$$d^2T = \frac{\varepsilon^2 n^2 d}{\Lambda_{\mathrm{shuf}}} \le n^4 d, \qquad dT = \frac{\varepsilon^2 n^2}{\Lambda_{\mathrm{shuf}}} \le n^4,$$

we have, up to absolute constants,

$$\log\bigl(d^2T/\delta\bigr) = O\biggl(\log\biggl(\frac{nd}{\delta}\biggr)\biggr), \qquad \log(dT/\zeta) = O\biggl(\log\biggl(\frac{n}{\zeta}\biggr)\biggr).$$

Hence

$$\frac{\eta dT L^2 \log^2(d^2T/\delta)\log(dT/\zeta)\log(1/\delta)}{n^2\varepsilon^2} = O\left(\frac{DL\sqrt{d\Lambda_{\mathrm{shuf}}}}{\varepsilon n}\right).$$

Finally, the martingale/noise-inner-product term is

$$O\left(\frac{DL\log(nd/\delta)\sqrt{\log(1/\zeta)\log(1/\delta)}}{n\varepsilon}\right),$$

which is dominated by $O(DL\sqrt{d\Lambda_{\mathrm{shuf}}}/(\varepsilon n))$ since $d \ge 1$ and $n \ge 1$. Combining the terms gives

$$\ell(\bar{\theta}, \mathcal{D}) - \min_{\theta\in\Theta} \ell(\theta, \mathcal{D}) \le O\left(\frac{DL\sqrt{d\Lambda_{\mathrm{shuf}}}}{\varepsilon n}\right) = O\left(\frac{DL\sqrt{d\log^2\bigl(\frac{nd}{\delta}\bigr)\log\bigl(\frac{n}{\zeta}\bigr)\log\bigl(\frac{1}{\delta}\bigr)}}{\varepsilon n}\right).$$

□

We now present the analysis for the full-batch DP-PGD optimizer in terms of zero-concentrated differential privacy [Bun and Steinke, 2016]. This gives better composition guarantees without assuming that the total privacy budget ε is small.

Lemma X.4 (Noisy full-batch PGD under zCDP)**.** *Suppose $\ell(\cdot, \mathcal{D})$ is convex and L-Lipschitz over a closed convex set $\Theta \subset \mathbb{R}^d$ of diameter D. Let $\theta^*_{\mathcal{D}} \in \arg\min_{\theta\in\Theta} \ell(\theta, \mathcal{D})$. Consider full-batch projected gradient descent with Gaussian perturbations, $\theta_{t+1} = \pi_\Theta(\theta_t - \eta(g_t + n_t))$, where $g_t \in \partial\ell(\theta_t, \mathcal{D})$ and $n_t \sim \mathcal{N}(0, \sigma^2 I_d)$ independently across $t = 0, \dots, T-1$. If*

$$\sigma^2 = \frac{2L^2T}{n^2\rho},$$

then, under replacement adjacency, the sequence of T noisy gradients is ρ-zCDP. Moreover, with probability at least $1 - \zeta$,

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le O\left(\frac{D^2}{\eta T} + \eta L^2 + \eta d\sigma^2 + \frac{\eta\sigma^2\log(1/\zeta)}{T} + D\sigma\sqrt{\frac{\log(1/\zeta)}{T}}\right).$$

In particular, choosing

$$\eta = \frac{D}{L\sqrt{T}}, \qquad T = \frac{n^2\rho + \log(1/\zeta)}{d},$$

and assuming $\log(1/\zeta) \le n^2\rho$, we have

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le O\left(\frac{DL\sqrt{d}}{n\sqrt{\rho}} + \frac{DL\sqrt{\log(1/\zeta)}}{n\sqrt{\rho}}\right).$$

Proof. We first prove privacy. Since $\ell(\theta, \mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, x_i)$ and each per-example loss is L-Lipschitz, each per-example subgradient has norm at most L. Under replacement adjacency, changing one data point changes the full-batch gradient by at most $\Delta = 2L/n$ in ℓ_2 norm. A Gaussian mechanism with sensitivity Δ and covariance $\sigma^2 I_d$ is $\Delta^2/(2\sigma^2)$-zCDP, so each noisy gradient release is

$$\rho_0 = \frac{(2L/n)^2}{2\sigma^2} = \frac{2L^2}{n^2\sigma^2}.$$

By adaptive composition of zCDP mechanisms [Bun and Steinke, 2016], the T releases are $T\rho_0 = 2TL^2/(n^2\sigma^2)$-zCDP. Substituting $\sigma^2 = 2L^2T/(n^2\rho)$ gives total privacy ρ-zCDP.

For utility, apply Lemma X.2. Since $\sum_{t=0}^{T-1} \|n_t\|_2^2/\sigma^2 \sim \chi^2_{dT}$, chi-square concentration gives, with probability at least $1 - \zeta/2$,

$$\frac{1}{T}\sum_{t=0}^{T-1} \|n_t\|_2^2 \le O\left(d\sigma^2 + \frac{\sigma^2 \log(1/\zeta)}{T}\right).$$

Thus $B_2 = O(d\sigma^2 + \sigma^2 \log(1/\zeta)/T)$. Next, conditioned on the history before round t, $\langle n_t, \theta_t - \theta^*_{\mathcal{D}}\rangle$ is mean-zero Gaussian with variance at most $\sigma^2 D^2$, because $\|\theta_t - \theta^*_{\mathcal{D}}\|_2 \le D$. Hence, with probability at least $1 - \zeta/2$,

$$\left|\frac{1}{T}\sum_{t=0}^{T-1} \langle n_t, \theta_t - \theta^*_{\mathcal{D}}\rangle\right| \le O\left(D\sigma\sqrt{\frac{\log(1/\zeta)}{T}}\right),$$

so $B_1 = O(D\sigma\sqrt{\log(1/\zeta)/T})$. The first displayed utility bound follows from Lemma X.2.

Finally, substitute $\eta = D/(L\sqrt{T})$ and $\sigma^2 = 2L^2T/(n^2\rho)$. The first two terms are $O(DL/\sqrt{T})$, and the remaining noise terms become

$$\eta d\sigma^2 = O\left(\frac{DLd\sqrt{T}}{n^2\rho}\right), \qquad \frac{\eta\sigma^2 \log(1/\zeta)}{T} = O\left(\frac{DL\log(1/\zeta)}{n^2\rho\sqrt{T}}\right),$$

and

$$D\sigma\sqrt{\frac{\log(1/\zeta)}{T}} = O\left(\frac{DL\sqrt{\log(1/\zeta)}}{n\sqrt{\rho}}\right).$$

Thus the T-dependent terms are

$$O\left(\frac{DL}{\sqrt{T}}\left(1 + \frac{\log(1/\zeta)}{n^2\rho}\right) + \frac{DLd\sqrt{T}}{n^2\rho}\right).$$

Choose

$$T = \frac{n^2\rho + \log(1/\zeta)}{d}.$$

To verify the balancing, let $A = n^2\rho$ and $B = \log(1/\zeta)$. Then the T-dependent terms become

$$O\left(DL\sqrt{d}\frac{1 + B/A}{\sqrt{A+B}} + DL\sqrt{d}\frac{\sqrt{A+B}}{A}\right).$$

Since

$$\frac{1 + B/A}{\sqrt{A+B}} = \frac{A+B}{A\sqrt{A+B}} = \frac{\sqrt{A+B}}{A},$$

both terms are of the same order. Hence the T-dependent terms are

$$O\left(DL\sqrt{d}\frac{\sqrt{A+B}}{A}\right) = O\left(\frac{DL\sqrt{d}}{n\sqrt{\rho}}\sqrt{1 + \frac{\log(1/\zeta)}{n^2\rho}}\right).$$

Using $\sqrt{1+x} \le 1 + \sqrt{x}$, this is

$$O\left(\frac{DL\sqrt{d}}{n\sqrt{\rho}} + \frac{DL\sqrt{d\log(1/\zeta)}}{n^2\rho}\right).$$

Under the assumption $\log(1/\zeta) \le n^2\rho$, the second term is dominated by the first. Adding the martingale/noise-inner-product term gives

$$\ell(\bar{\theta}, \mathcal{D}) - \ell(\theta^*_{\mathcal{D}}, \mathcal{D}) \le O\left(\frac{DL\sqrt{d}}{n\sqrt{\rho}} + \frac{DL\sqrt{\log(1/\zeta)}}{n\sqrt{\rho}}\right).$$

This proves the stated bound. □

Summed objective with public hyper-parameters. Lemma X.4 calibrates the noise σ, the step size η and the number of iterations through the dataset size n. In Algorithm 2 the size of each optimizer dataset ($|\mathcal{T}_o \cap [\tau_o - 1]|$ or $|[\tau - 1] \setminus \mathcal{T}_o|$) depends on whether a given user explored, so n itself is private and must not enter the mechanism. The next lemma runs noisy PGD on the *summed* objective, whose per-step sensitivity does not depend on n, and sets every hyper-parameter from a public upper bound $N \ge n$.

Lemma X.5 (Noisy PGD on a summed objective with public parameters)**.** *Let $\Theta \subset \mathbb{R}^d$ be a closed convex set of diameter D. Let $\mathcal{D} = \{z_1, \dots, z_n\}$ be a dataset whose size n is unknown to the mechanism, and let $N \ge n$ be a public bound. Let*

$$F_{\mathcal{D}}(\theta) := \sum_{i=1}^{n} f(\theta, z_i) + r(\theta),$$

where each $f(\cdot, z)$ is convex and L_0-Lipschitz on Θ, and r is a data-independent convex function that is G_r-Lipschitz on Θ. Then $F_{\mathcal{D}}$ is L_F-Lipschitz on Θ for the public constant $L_F := NL_0 + G_r$. Fix $\rho > 0$ and $\zeta \in (0, 1)$, and run $\theta_{k+1} = \pi_\Theta\big(\theta_k - \eta(g_k + n_k)\big)$ for $k = 0, \dots, K-1$, with $g_k \in \partial F_{\mathcal{D}}(\theta_k)$, $n_k \sim \mathcal{N}(0, \sigma^2 I_d)$ i.i.d., and output $\bar{\theta} = \frac{1}{K}\sum_{k=0}^{K-1} \theta_k$, where

$$K = \left\lceil \frac{\rho L_F^2}{dL_0^2} + \frac{\log(2/\zeta)}{d} \right\rceil, \qquad \sigma^2 = \frac{2L_0^2 K}{\rho}, \qquad \eta = \frac{D}{\sqrt{K\,(L_F^2 + d\sigma^2)}}.$$

These parameters depend only on $(N, L_0, G_r, D, d, \rho, \zeta)$, and not on n or the data.

1. (Privacy) *For any two datasets that differ by adding, removing, or replacing a single point, $\bar{\theta}$ is ρ-zCDP.*
2. (Utility) *With probability at least $1 - \zeta$,*

$$F_{\mathcal{D}}(\bar{\theta}) - \min_{\theta \in \Theta} F_{\mathcal{D}}(\theta) \le O\left(DL_0\sqrt{\frac{d + \log(2/\zeta)}{\rho}}\right).$$

Proof. Privacy. The regularizer r does not depend on the data. If $\mathcal{D}'$ is obtained from $\mathcal{D}$ by adding or removing one point z, then $\|\nabla F_{\mathcal{D}}(\theta) - \nabla F_{\mathcal{D}'}(\theta)\|_2 = \|\nabla f(\theta, z)\|_2 \le L_0$ (for subgradients, choose them consistently on the common points). Replacing z by z' gives at most $2L_0$. So every iteration is a Gaussian mechanism with ℓ_2-sensitivity at most $\Delta = 2L_0$, *whatever* n is, and is therefore $\Delta^2/(2\sigma^2) = 2L_0^2/\sigma^2$-zCDP. The iterate θ_k, at which the next gradient is evaluated, is a function of the earlier noisy releases, so adaptive composition of zCDP [Bun and Steinke, 2016] over K iterations gives $2KL_0^2/\sigma^2 = \rho$. The output $\bar{\theta}$ is post-processing. All of K, σ, η and Θ are the same under both datasets, so the mechanism being compared is the same.

Utility. Apply Lemma X.2 with Lipschitz constant L_F. Exactly as in the proof of Lemma X.4, with probability at least $1-\zeta$ we have $B_2 = O\big(d\sigma^2 + \sigma^2\log(2/\zeta)/K\big)$ and $B_1 = O\big(D\sigma\sqrt{\log(2/\zeta)/K}\big)$. Since $K \ge \log(2/\zeta)/d$, we get $B_2 = O(d\sigma^2)$. Hence

$$F_{\mathcal{D}}(\bar{\theta}) - \min_{\Theta} F_{\mathcal{D}} \le O\left(\frac{D^2}{\eta K} + \eta\big(L_F^2 + d\sigma^2\big)\right) + B_1 = O\left(D\sqrt{\frac{L_F^2 + d\sigma^2}{K}}\right) + B_1.$$

By the choice of K, $L_F^2/K \le dL_0^2/\rho$, and $d\sigma^2/K = 2dL_0^2/\rho$. So the first term is $O(DL_0\sqrt{d/\rho})$. Also $B_1 = O\big(DL_0\sqrt{2\log(2/\zeta)/\rho}\big)$. Adding the two bounds proves the claim. □

Remark X.6. The bound is n times the averaged-objective bound of Lemma X.4, so using the summed objective with public parameters costs nothing in the rate. Replacing n by the public bound N only increases the iteration count $K \approx \rho N^2/d$, which is a computational cost. It also removes the condition $\log(1/\zeta) \le n^2\rho$ of Lemma X.4.

NeurIPS Paper Checklist

1. **Claims**
   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?
   Answer: [Yes]
   Justification: The two regret bounds stated in the abstract are formalized in Theorem 5.1 (shuffle-DP, stochastic contexts) and Theorem 6.1 (joint-DP, adversarial contexts), and the technical contributions are summarized in Section 1.1.
   Guidelines:
   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.
2. **Limitations**
   Question: Does the paper discuss the limitations of the work performed by the authors?
   Answer: [Yes]
   Justification: Limitations are stated in the Conclusion paragraph at the end of the main paper: the gap between the privacy-dependent term of our joint-DP bound and the corresponding lower bound (see also the remark following Theorem 6.1), and that our experiments evaluate only the joint-DP algorithm.
   Guidelines:
   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.
3. **Theory assumptions and proofs**
   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?
   Answer: [Yes]

Justification: All theorem statements include their assumptions, and full proofs are deferred to the appendix; specifically, the proofs of Theorem 5.1 and Theorem 6.1 appear in Appendix J and Appendix Q respectively, with supporting lemmas cross-referenced throughout.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The full experimental setup—model, dimension, number of arms, horizon, parameter sampling, and number of repetitions—is given in Section 7 and Appendix B. The Remark in Appendix B additionally provides the explicit closed-form expressions used for the hyperparameters λ, $\lambda_{\max}$, $\lambda_{\min}$. Code is included in the supplemental material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide code reproducing all experiments in the supplemental material, with hyperparameters and configurations documented in Appendix B.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: All experimental settings—privacy parameters (ε, δ), dimension d, number of arms K, horizon T, parameter sampling, and the explicit closed-form expressions for the hyperparameters λ, $\lambda_{\max}$, $\lambda_{\min}$—are detailed in Appendix B.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Table 2 reports $\pm$ standard errors of cumulative regret over 10 independent runs for every algorithm and every value of κ.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The Compute paragraph at the end of Appendix B states that all experiments are low-compute and run on an Apple M1 chip within a few hours per run.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes, we do follow it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Our work is primarily theoretical in nature and does not involve deployment, human subjects, or direct real-world decision-making. As such, we do not anticipate immediate societal impacts beyond the broader downstream use of theoretical advances in machine learning.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to

any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We do not release any data or models in this paper.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use any external datasets or pretrained models; baseline algorithms used in our experiments are reimplemented from their published descriptions and properly cited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: Yes, we do provide documentation in the code and proofs for the theory.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve any crowdsourcing or research with human subjects; all experiments are on synthetic data.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve any human subjects research, so IRB approval is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: By providing targeted hints and guidance, we prompted GPT-5.6 Sol to help develop the privatized distributional optimal design in Algorithm 1. In a subsequent iteration, Claude Opus 5.5 identified a bug in our earlier proof of the joint-DP guarantee for Algorithm 2, and proposed updating the estimators $\hat{\theta}_o, \hat{\theta}_\tau$ outside Criterion I, only at publicly triggered rounds, which fixes the issue and simplifies the privacy accounting. In both cases, we fully verified the resulting proofs, rewriting and simplifying them to make them easier to follow.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.