Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence
Canadian AI 2026
25–29 May 2026, Vancouver, BC

AdaPrivate-TS: Private Thompson Sampling for Contextual Bandits with Privacy Amplification

Mohammadreza Riyazat[†,*], Eranga Ukwatta[†]
[†] School of Engineering, University of Guelph, Guelph, ON, Canada.

Abstract

We present AdaPrivate-TS, a differentially private contextual bandit algorithm that combines Thompson Sampling with batched zCDP composition. Our key insight is that differential privacy noise inflates the posterior covariance in a structured way—adding $\mathcal{N}(0, \sigma^2 I)$ noise to b yields sampling covariance $v^2 A^{-1} + \sigma^2 A^{-2}$, which Thompson Sampling interprets as increased uncertainty rather than pure corruption. Under event-level privacy (protecting individual interactions) with stochastic contexts, we prove that the privacy cost is only $O\big(\sqrt{d} \cdot \log T/\sqrt{\rho}\big)$—logarithmic in T—because parallel composition amortizes noise across batches. Additionally, we explore privacy amplification via Poisson subsampling, which can reduce effective noise at stringent privacy budgets. Experiments on synthetic and real-world datasets demonstrate: (1) AdaPrivate-TS achieves 93-99% of non-private performance at $\varepsilon \in [0.5, 5]$, outperforming UCB by 0.5–3.7% and up to 18% with tuned adaptive exploration at extreme ε; (2) privacy amplification provides additional 2-5% gains at low ε; (3) on MovieLens and Jester, AdaPrivate-TS achieves the best overall performance among event-level baselines, dominating at $\varepsilon \geq 2$; (4) under DP-SVD private features, TS's advantage over UCB *grows* to +11%, confirming noise-as-uncertainty is not limited to reward privacy. We provide rigorous proofs for privacy guarantees under interactive zCDP composition and comprehensive evaluation including convergence curves, 12-seed CIs, and DP-SVD feature ablation.

Keywords: Differential Privacy, Thompson Sampling, Contextual Bandits, Privacy Amplification, zCDP.

1. Introduction

Online recommender systems face a fundamental tension: learning user preferences requires processing private behavioral data, yet users demand privacy guarantees. Differential privacy (DP) [1] provides rigorous protection but introduces noise that degrades learning. While existing private contextual bandit algorithms [2] treat privacy noise as pure corruption, we identify a key insight: *Thompson Sampling (TS) naturally handles uncertainty through posterior sampling; DP noise inflates the posterior covariance to* $v^2 A^{-1} + \sigma^2 A^{-2}$*, which TS interprets as increased uncertainty rather than corruption.* This leads to **AdaPrivate-TS**, which replaces Upper Confidence Bound (UCB) with TS in the private bandit framework, achieving improved regret by converting noise into exploration. We adopt **event-level DP** where adjacent datasets differ in one user-reward pair (u_t, r_t), with actions being algorithm outputs rather than private inputs (see Definition 2). This aligns with Table-level DP [1] and differs from joint DP [2] (protecting the entire action sequence) and user-level DP (protecting all interactions of one user). Event-level DP is appropriate when users contribute few interactions; for many-interaction users, group privacy scaling applies (see Section 4.6). Under event-level DP with stochastic contexts, the impossibility result of [2] (requiring adversarial contexts and joint DP) does not apply; our batched post-processing approach enables parallel composition.

Contributions:

(1) **Private Thompson Sampling:** TS with DP noise results in covariance $v^2 A^{-1} + \sigma^2 A^{-2}$ (Proposition 4); we prove a tight regret bound where the privacy cost is $O\big(\sqrt{d} \cdot \log T/\sqrt{\rho}\big)$, logarithmic in T.

(2) **Privacy Amplification:** Rigorous RDP-based analysis of Poisson subsampling for

* mriyazat@uoguelph.ca

bandits (Proposition 5), with 2–4% gains at $\varepsilon \leq 1$.

(3) **Comprehensive Experiments:** 93.5–98.7% of non-private TS at $\varepsilon \in [0.5, 5]$, outperforming private UCB by 0.5–3.7% (up to 18% at extreme ε with adaptive tuning) and same-model zCDP baselines by up to 6–10% on real data. Under DP-SVD private features, TS's advantage grows to +11%, and convergence curves confirm stable learning dynamics.

2. Background and Problem Setting

2.1. Contextual Bandits for Recommendation

At round $t \in [T]$, the learner observes a candidate set $\mathcal{A}_t$ with feature vectors $\{x_{t,a} \in \mathbb{R}^d\}_{a \in \mathcal{A}_t}$, selects action $a_t \in \mathcal{A}_t$, and receives stochastic reward $r_t \in [0, 1]$. We assume a linear reward model:

$$\mathbb{E}[r_t | a_t = a] = x_a^\top \theta^* \tag{2.1}$$

for unknown parameter $\theta^* \in \mathbb{R}^d$ with $\|\theta^*\|_2 \leq S$.

Definition 1. *The cumulative regret over T rounds is:*

$$R(T) = \sum_{t=1}^{T} \left(\max_{a \in \mathcal{A}_t} x_a^\top \theta^* - x_{a_t}^\top \theta^* \right) \tag{2.2}$$

The learner maintains sufficient statistics:

$$A_t = \lambda I_d + \sum_{s=1}^{t-1} x_{a_s} x_{a_s}^\top \tag{2.3}$$

$$b_t = \sum_{s=1}^{t-1} r_s x_{a_s} \tag{2.4}$$

with ridge parameter $\lambda > 0$. The regularized least-squares estimate is $\hat{\theta}_t = A_t^{-1} b_t$.

2.2. Privacy Model

Definition 2. *An interaction dataset $D = \{(u_t, r_t)\}_{t=1}^T$ records the user identity u_t and reward r_t at each round. Two datasets D, D' are **adjacent** (written $D \sim D'$) if they differ in exactly one entry $(u_t, r_t) \neq (u_t', r_t')$ for a single index t, with all other entries identical. Actions a_t are* outputs *of the algorithm (determined by the mechanism and public features) and are not part of the adjacency relation.*

We use zero-concentrated DP (zCDP) [3] for tighter composition:

Definition 3. *A randomized mechanism $\mathcal{M}$ satisfies ρ-zCDP if for all adjacent $D \sim D'$ and all $\alpha > 1$:*

$$D_\alpha(\mathcal{M}(D) \| \mathcal{M}(D')) \leq \rho\alpha \tag{2.5}$$

where D_α is the Rényi divergence of order α.

Lemma 1. *For a query $f : \mathcal{D} \to \mathbb{R}^d$ with ℓ_2-sensitivity $\Delta_2 = \max_{D \sim D'} \|f(D) - f(D')\|_2$, the mechanism $\mathcal{M}(D) = f(D) + \mathcal{N}(0, \sigma^2 I_d)$ satisfies ρ-zCDP with:*

$$\rho = \frac{\Delta_2^2}{2\sigma^2} \tag{2.6}$$

Lemma 2. *If $\mathcal{M}$ satisfies ρ-zCDP, then for any $\delta > 0$, $\mathcal{M}$ satisfies (ε, δ)-DP with:*

$$\varepsilon = \rho + 2\sqrt{\rho \log(1/\delta)} \tag{2.7}$$

Remark 1. *Given a target (ε, δ)-DP guarantee, we solve Eq. (2.7) for ρ in closed form. Setting $u = \sqrt{\rho}$ and $c = \sqrt{\log(1/\delta)}$, Eq. (2.7) becomes $u^2 + 2cu - \varepsilon = 0$, results in:*

$$\rho = \left(\sqrt{\varepsilon + \log(1/\delta)} - \sqrt{\log(1/\delta)} \right)^2 \tag{2.8}$$

The noise scale is $\sigma = \Delta_2 / \sqrt{2\rho}$. All experiments compute ρ from Eq. (2.8).

Lemma 3. *If mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_K$ operate on disjoint subsets of the data and each satisfies ρ_k-zCDP, then releasing all outputs satisfies $\max_k \rho_k$-zCDP.*

2.3. Thompson Sampling for Linear Bandits

Thompson Sampling (TS) [4, 5] maintains a posterior distribution over θ^* and selects actions by sampling from this posterior. For linear bandits with Gaussian likelihood, the posterior is:

$$\theta|\mathcal{H}_t \sim \mathcal{N}(\hat{\theta}_t, v^2 A_t^{-1}) \tag{2.9}$$

where $\mathcal{H}_t$ is the history up to time t and $v > 0$ controls exploration.

TS handles uncertainty through sampling where larger posterior variance leads to more exploration, unlike UCB's deterministic bonus.

3. Method: AdaPrivate-TS

3.1. Key Insight: DP Noise as Posterior Inflation

Consider adding Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2 I_d)$ to b_t for privacy:

$$\tilde{b}_t = b_t + \eta \tag{3.1}$$

For UCB, this noise directly corrupts the estimate. For Thompson Sampling, we characterize exactly how it affects the sampling distribution:

Proposition 4. *Let $\eta \sim \mathcal{N}(0, \sigma^2 I_d)$ be independent privacy noise and $\xi \sim \mathcal{N}(0, v^2 A_t^{-1})$ be the TS sampling noise. The TS sample based on $\tilde{b}_t = b_t + \eta$ satisfies:*

$$\tilde{\theta}_t = A_t^{-1}\tilde{b}_t + \xi \sim \mathcal{N}\left(\hat{\theta}_t,\ v^2 A_t^{-1} + \sigma^2 A_t^{-2}\right) \tag{3.2}$$

Proof. Let $\xi \sim \mathcal{N}(0, v^2 A_t^{-1})$ be the TS sampling noise. Then:

$$\begin{aligned}\tilde{\theta}_t &= A_t^{-1}\tilde{b}_t + \xi = A_t^{-1}(b_t + \eta) + \xi \\ &= \hat{\theta}_t + A_t^{-1}\eta + \xi\end{aligned} \tag{3.3}$$

Since $\eta \sim \mathcal{N}(0, \sigma^2 I_d)$ and $\xi \sim \mathcal{N}(0, v^2 A_t^{-1})$ are independent Gaussians:

$$\begin{aligned}\mathrm{Cov}(\tilde{\theta}_t) &= \mathrm{Cov}(A_t^{-1}\eta) + \mathrm{Cov}(\xi) \\ &= A_t^{-1} \cdot \sigma^2 I_d \cdot A_t^{-\top} + v^2 A_t^{-1} \\ &= \sigma^2 A_t^{-2} + v^2 A_t^{-1}\end{aligned} \tag{3.4}$$

where $A_t^{-2} = A_t^{-1}A_t^{-1}$ since A_t is symmetric positive definite. ■

Remark 2. *The covariance is exactly $v^2 A_t^{-1} + \sigma^2 A_t^{-2}$. Since A_t is symmetric positive definite, the Löwner ordering $A_t^{-2} \preceq A_t^{-1}/\lambda_{\min}(A_t)$ holds (each eigenvalue $\lambda_i^{-2} \le \lambda_i^{-1}/\lambda_{\min}$), giving:*

$$Cov(\tilde{\theta}_t) \preceq \left(v^2 + \frac{\sigma^2}{\lambda_{\min}(A_t)}\right) A_t^{-1} \tag{3.5}$$

As data accumulates, $\lambda_{\min}(A_t)$ grows, so the privacy contribution diminishes relative to the base TS variance. TS interprets this inflated covariance as increased uncertainty, exploring more when privacy noise dominates.

3.2. Privacy Amplification via Subsampling

At stringent privacy budgets ($\varepsilon \le 1$), we analyze privacy amplification via *Poisson* subsampling (not without-replacement mini-batching) with rate $q \in (0, 1]$ where each batch reward is included independently with probability q.

The subsampled mechanism operates as follows: (1) First include each reward $r_t x_{a_t}$ independently with probability q, forming $s_k^{\text{sub}} = \sum_{t \in \text{batch } k} Z_t \cdot r_t x_{a_t}$ where $Z_t \sim \text{Bern}(q)$; (2) Then add noise $\tilde{s}_k^{\text{sub}} = s_k^{\text{sub}} + \eta_k$ with $\eta_k \sim \mathcal{N}(0, \sigma^2 I_d)$; (3) Now rescale for unbiasedness $\tilde{b}_{\text{batch}} = \tilde{s}_k^{\text{sub}}/q$. Crucially, the Gaussian noise is added to the subsampled sum *before* rescaling. The rescaling by $1/q$ is a deterministic post-processing step that does not affect privacy. The RDP bound in Eq. (3.6) applies to step (2)—the composition of Poisson subsampling and Gaussian mechanism—and is preserved under the subsequent rescaling

by the post-processing property of RDP. This avoids the sensitivity-rescaling cancellation described below. Poisson sampling is required for Eq. (3.6). After rescaling, sensitivity becomes Δ_2/q, giving ρ/q^2; combined with the q^2 amplification factor, the net cost is ρ—the gain cancels [6]. The resolution uses Rényi divergence directly on the subsampled mechanism without factoring through sensitivity. For Poisson subsampling with rate q before a Gaussian mechanism with noise σ and sensitivity $\Delta_2 = 1$, at RDP order $\alpha \geq 2$ [6, 7]:

$$\varepsilon_{\text{sub}}(\alpha) \leq \frac{1}{\alpha - 1} \log \left(1 + \binom{\alpha}{2} \frac{2q^2}{\sigma^2} + \sum_{j=3}^{\alpha} \binom{\alpha}{j} \frac{(2q^2)^{j/2}}{\sigma^j} \right) \tag{3.6}$$

For $q\alpha/\sigma \ll 1$ (small subsampling rate or large noise), the dominant term gives:

$$\varepsilon_{\text{sub}}(\alpha) \approx \frac{q^2 \alpha}{\sigma^2} = 2q^2 \rho \tag{3.7}$$

This is a factor of $2q^2$ reduction from ρ, achieved because the RDP analysis tracks the privacy loss random variable directly, avoiding the sensitivity-rescaling cancellation.

We evaluate Eq. (3.6) at orders $\alpha \in \{2, \ldots, 6, 8, 16, 32, 64\}$.
Then convert via $\varepsilon(\alpha) = \varepsilon_{\text{sub}}(\alpha) + \log(1/\delta)/(\alpha - 1)$, and minimize over α. Under parallel composition, total cost is $\max_k \varepsilon_{\text{sub},k}$. Optimal α is 16 at ε=1, 3 at ε=5; final ε varies <0.5%. For AdaPrivate-TS-Amp, we recalibrate σ given target (ε, δ) and rate q, we solve for σ_{amp} such that $\min_\alpha[\varepsilon_{\text{sub}}(\alpha; q, \sigma_{\text{amp}}) + \log(1/\delta)/(\alpha - 1)] \leq \varepsilon$. Since amplification reduces privacy cost, $\sigma_{\text{amp}} < \sigma_{\text{base}}$, results in less noise at the same privacy level.

Proposition 5. *The Poisson-subsampled Gaussian mechanism with rate q satisfies $\varepsilon_{sub}(\alpha)$-RDP at order α as given by Eq. (3.6). For $q\alpha/\sigma \ll 1$, the dominant term gives $\varepsilon_{sub}(\alpha) \approx q^2\alpha/\sigma^2$, a factor $2q^2$ reduction from the base ρ-zCDP cost. For $\sigma \geq 1.05$ and $q \leq 0.5$, higher-order terms contribute $< 5\%$.*
Note: the subsampled mechanism's RDP profile is not linear in α, so we state this as an RDP (not zCDP) guarantee and convert to (ε, δ)-DP per the accounting details above.

3.3. Adaptive Exploration Decay

With privacy noise inflating effective variance, early exploration may be excessive. We decay exploration as $v_k = v_0 \cdot \gamma^{k-1}$ (v_0=1.5, γ=0.95), ensuring aggressive exploration when uncertainty is high and refined exploitation as the model converges.

3.4. Complete Algorithm

3.5. Privacy Analysis

Assumption 1. *Features satisfy $\|x_a\|_2 \leq 1$ (enforced via clipping) and rewards satisfy $r_t \in [0, 1]$.*

Theorem 6. *Under Assumption 1 with data-independent batch schedule $\{B_k\}_{k=1}^K$, Algorithm 1 satisfies (ε, δ)-DP (event-level) for the transcript $(a_1, \ldots, a_T)$.*

Proof. We formalize the mechanism as releasing K independent noisy *increments*, then show the full transcript is a deterministic post-processing of these releases and public data.

Step 1: Define the batch-k sufficient statistic increment $s_k = \sum_{t \in \text{batch } k} r_t x_{a_t} \in \mathbb{R}^d$. The mechanism releases the noisy increment:

$$\tilde{s}_k = s_k + \eta_k, \quad \eta_k \sim \mathcal{N}(0, \sigma^2 I_d), \quad k = 1, \ldots, K \tag{3.8}$$

Although Algorithm 1 maintains a running sum $b \leftarrow b + \tilde{s}_k$, this cumulative b is a deterministic function of $\{\tilde{s}_1, \ldots, \tilde{s}_k\}$: namely $b_k = \lambda\theta_0 + \sum_{j=1}^k \tilde{s}_j$. Thus the internal mechanism releases exactly the K independent noisy increments $\{\tilde{s}_k\}_{k=1}^K$.

Step 2: Under Assumption 1, changing one reward (r_t, x_{a_t}) in batch k changes s_k by at most $\|r_t x_{a_t}\|_2 \leq 1$. Thus $\Delta_2 = 1$, and by Lemma 1 each noisy release $\tilde{s}_k$ satisfies ρ-zCDP with $\sigma = 1/\sqrt{2\rho}$.

Step 3: Action a_t in batch k is determined by: (i) public features $\{x_{t,a}\}$; (ii) the design

Algorithm 1 AdaPrivate-TS

Require: Features $\{x_a\}$, prior θ_0, target (ε, δ), batch size B

Compute ρ from (ε, δ) via Eq. (2.8)
Set $\sigma = 1/\sqrt{2\rho}$
$A \leftarrow \lambda I_d$, $b \leftarrow \lambda\theta_0$, batch_b $\leftarrow 0$, count $\leftarrow 0$, $k \leftarrow 1$
for $t = 1, \ldots, T$ **do**
  Observe $\mathcal{A}_t$; clip features: $x_a \leftarrow x_a / \max(1, \|x_a\|_2)$
  Thompson Sampling:
    $L_A \leftarrow \text{Cholesky}(A)$ ▷ $L_A L_A^\top = A$
    $\tilde{\theta} \leftarrow L_A^{-\top}(L_A^{-1} b + v_k \cdot \mathcal{N}(0, I_d))$
    $a_t \leftarrow \arg\max_{a \in \mathcal{A}_t} x_a^\top \tilde{\theta}$
  Observe reward $r_t \in [0, 1]$
  $A \leftarrow A + x_{a_t} x_{a_t}^\top$
  batch_b $\leftarrow$ batch_b $+ r_t \cdot x_{a_t}$
  count $\leftarrow$ count $+1$
  if count $\geq B$ **then** ▷ Batch boundary
    $b \leftarrow b +$ batch_b $+ \mathcal{N}(0, \sigma^2 I_d)$ ▷ Add noise
    Reset batch_b, count; $k \leftarrow k + 1$; $v_k \leftarrow v_{k-1} \cdot \gamma$
  end if
end for

matrix $A_t = \lambda I + \sum_{s<t} x_{a_s} x_{a_s}^\top$ (computed from public features and past actions); (iii) the cumulative noisy reward vector $b_{k-1} = \lambda\theta_0 + \sum_{j=1}^{k-1} \tilde{s}_j$ (from *already-released* increments of prior batches); and (iv) fresh TS randomness ξ_t. Crucially, a_t does *not* depend on within-batch rewards $\{r_s\}_{s \in \text{batch } k}$, since b updates only at batch boundaries.

Step 4: Each rating belongs to exactly one batch. For adjacent D, D' differing in batch j, although later batches adapt to prior releases, this does not increase privacy cost: for any realization of prior outputs $h_{<k}$, the conditional distribution of $\tilde{s}_k | h_{<k}$ is identical under D and D' when $k \neq j$ (since batch k's data is unchanged), with $D_\alpha(P_j(\cdot|h_{<j}) \| Q_j(\cdot|h_{<j})) \leq \rho\alpha$ for $k{=}j$ by Step 2. By the chain rule for Rényi divergence: $D_\alpha(P_{1:K} \| Q_{1:K}) \leq \rho\alpha$, giving ρ-zCDP.

Step 5: Given $\{\tilde{s}_k\}_{k=1}^K$ and public data (features, TS randomness), the entire action transcript $(a_1, \ldots, a_T)$ is a deterministic function—each a_t is computed from b_{k-1}, A_t, and ξ_t as in Step 3. By the post-processing property of zCDP, the transcript inherits ρ-zCDP, converting to (ε, δ)-DP via Lemma 2. ■

3.6. **Regret Analysis**

Assumption 2. *At each round t, the candidate set $\mathcal{A}_t$ contains arms whose features are drawn i.i.d. from a distribution satisfying $\lambda_{\min}(\mathbb{E}[x_a x_a^\top]) \geq \lambda_0 > 0$ and $\|x_a\|_2 \leq 1$.*

Theorem 7. *Under Assumptions 1 and 2, AdaPrivate-TS with K uniform batches of size $B = T/K$ achieves, for any $\delta_r \in (0, 1)$, with probability at least $1 - \delta_r$:*

$$R(T) \leq \underbrace{C_1 d\sqrt{T \log(T/\delta_r)}}_{\textit{TS + staleness}} + \underbrace{\frac{C_2 \sigma \sqrt{d} \log(T/\delta_r)}{\lambda_0}}_{\textit{privacy (amortized)}} \tag{3.9}$$

where $C_1 = O(vS)$, $C_2 = O(S)$, and $S = \|\theta^*\|_2$. In expectation:

$$\mathbb{E}[R(T)] = O\left(d\sqrt{T \log T} + \frac{\sqrt{d} \log T}{\sqrt{\rho}\, \lambda_0} \right) \tag{3.10}$$

*The privacy cost is $O\big(\sqrt{d} \cdot \log T/\sqrt{\rho}\big)$—**logarithmic** in T, not $O(\sqrt{T})$—because parallel composition ensures total privacy cost ρ (not $K\rho$), and each batch's noise is amortized over B rounds while $\lambda_{\min}(A_k)$ grows linearly.*

Proof. We decompose the regret into TS exploration (including staleness) and privacy noise.

Step 1: At round t in batch k, the TS sample is $\tilde{\theta}_k = A_k^{-1}(b_k + \eta_k) + \xi_k = \hat{\theta}_k + \zeta_k + \xi_k$,

where $\zeta_k = A_k^{-1}\eta_k$ is the privacy bias (fixed within batch k) and $\xi_k \sim \mathcal{N}(0, v_k^2 A_k^{-1})$. Since $a_t = \arg\max_a x_a^\top \tilde{\theta}_k$:

$$r_t \le 2\max_a |x_a^\top(\tilde{\theta}_k - \theta^*)| \le 2\max_a |x_a^\top(\hat{\theta}_k - \theta^* + \xi_k)| + 2\|\zeta_k\|_2 \tag{3.11}$$

using $|x_a^\top \zeta_k| \le \|\zeta_k\|_2$ for $\|x_a\|_2 \le 1$.

Step 2: The first term in (3.11)—estimation error plus TS sampling noise—is handled by the standard analysis of [5]. The privacy noise *increases* posterior variance (Proposition 4), which aids TS anti-concentration. We now account for staleness. Within batch k, the estimate $\hat{\theta}_k$ is fixed for B rounds while A_t continues to update. The per-round instantaneous regret due to estimation error and TS sampling satisfies $r_t^{\mathrm{TS}} \le 2\|\tilde{\theta}_k - \theta^*\|_{A_t} \cdot \|x_{a_t}\|_{A_t^{-1}}$, and by the elliptical potential lemma [5], $\sum_{t=1}^T \|x_{a_t}\|^2_{A_t^{-1}} \le 2d\log(1+T/(d\lambda))$. Since $\hat{\theta}_k$ uses data up to batch $k-1$, the staleness gap $\|\hat{\theta}_k - \hat{\theta}_t\|$ for t in batch k is bounded by $O(\sqrt{B/\lambda_{\min}(A_k)})$, which contributes at most $O(\sqrt{B}\cdot\sqrt{d\log T/\lambda_{\min}(A_k)})$ per batch. Summing over K batches, this staleness overhead is $O(\sqrt{dTB})$ for constant B, which is $O(d\sqrt{T})$ and absorbed into C_1:

$$R_{\mathrm{TS}}(T) \le C_1 \cdot d\sqrt{T\log(T/\delta_r)} \tag{3.12}$$

with probability $\ge 1 - \delta_r/2$, where $C_1 = O(vS + S\sqrt{B/\lambda})$.

Step 3: The privacy bias $\zeta_k = A_k^{-1}\eta_k$ is drawn once per batch and contributes $2B\|\zeta_k\|_2$ to the batch regret. Since $\eta_k \sim \mathcal{N}(0, \sigma^2 I_d)$, the bias ζ_k follows $\zeta_k \sim \mathcal{N}(0, \sigma^2 A_k^{-2})$. Its expected squared norm satisfies:

$$\mathbb{E}[\|\zeta_k\|_2^2] = \sigma^2 \operatorname{tr}(A_k^{-2}) = \sigma^2 \sum_{i=1}^d \lambda_i(A_k)^{-2} \tag{3.13}$$

To obtain a high-probability bound on $\|\zeta_k\|_2$, we apply the Hanson–Wright concentration inequality for Gaussian quadratic forms. Since $\zeta_k = A_k^{-1}\eta_k$ with $\eta_k \sim \mathcal{N}(0, \sigma^2 I_d)$, for any $u > 0$:

$$\Pr\left[\|\zeta_k\|_2 > \sigma\sqrt{\operatorname{tr}(A_k^{-2})} + \sigma\|A_k^{-1}\|_{\mathrm{op}}\sqrt{2u}\right] \le e^{-u} \tag{3.14}$$

Setting $u = \log(2K/\delta_r)$ and applying a union bound over K batches, with probability $\ge 1 - \delta_r/2$, simultaneously for all k:

$$\|\zeta_k\|_2 \le \sigma\sqrt{\operatorname{tr}(A_k^{-2})} + \frac{\sigma}{\lambda_{\min}(A_k)}\sqrt{2\log(2K/\delta_r)} \tag{3.15}$$

Using the cruder bound $\operatorname{tr}(A_k^{-2}) \le d/\lambda_{\min}(A_k)^2$, this simplifies to:

$$\|\zeta_k\|_2 \le \frac{\sigma\sqrt{d}}{\lambda_{\min}(A_k)} \cdot c(\delta_r) \tag{3.16}$$

where $c(\delta_r) = 1 + \sqrt{2\log(2K/\delta_r)/d} = O(\sqrt{\log(K/\delta_r)})$. Under Assumption 2, $\lambda_{\min}(A_k) \ge kB\lambda_0$ for $k \ge 1$ (absorbing λ). The per-batch privacy regret is:

$$R_{\mathrm{priv}}^{(k)} \le \frac{2B\sigma\sqrt{d}\, c(\delta_r)}{kB\lambda_0} = \frac{2\sigma\sqrt{d}\, c(\delta_r)}{k\lambda_0} \tag{3.17}$$

Summing over $K = T/B$ batches:

$$R_{\mathrm{priv}}(T) \le \frac{2\sigma\sqrt{d}\, c(\delta_r)}{\lambda_0}\sum_{k=1}^K \frac{1}{k} = \frac{2\sigma\sqrt{d}\, c(\delta_r)}{\lambda_0} \cdot H_K \tag{3.18}$$

where $H_K \le 1 + \ln K \le 1 + \ln(T/B)$. Substituting $\sigma = 1/\sqrt{2\rho}$:

$$R_{\mathrm{priv}}(T) = O\left(\frac{\sqrt{d}\log(T/B)}{\sqrt{\rho}\,\lambda_0}\right) \tag{3.19}$$

Step 4: Adding (3.12) and (3.18):

$$R(T) \leq C_1 d\sqrt{T\log(T/\delta_r)} + \frac{C_2\sigma\sqrt{d}\log(T/\delta_r)}{\lambda_0} \tag{3.20}$$

In expectation, $\mathbb{E}[R(T)] = O(d\sqrt{T\log T} + \sqrt{d}\log T/(\sqrt{\rho}\lambda_0))$. ■

Remark 3. *The $O\big(\sqrt{d}\cdot\log T\big)$ privacy term contrasts sharply with $\tilde{O}(T^{3/4})$ under joint DP [2] or $O(\sqrt{d}\cdot\sqrt{T}/\sqrt{\rho})$ under sequential composition. The difference is that parallel composition charges $\max_k \rho_k = \rho$ (not $\sum_k$), so each batch independently uses full noise scale $\sigma = 1/\sqrt{2\rho}$. Batch size B does not appear in the privacy term—it can be chosen freely based on computational or practical constraints. The $O(\sqrt{T})$ empirical scaling observed in Section 4.9 is dominated by the TS exploration term, not privacy.*

4. Experiments

4.1. Experimental Setup

We evaluate the algorithms in a synthetic environment where $d = 20$ dimensional features for $n = 100$ arms with $\theta^* \sim \mathcal{N}(0, I_d)$ normalized to $\|\theta^*\|_2 = 2$. Features are sampled from $\mathcal{N}(0, I_d)$ and normalized to unit norm. Rewards follow $\mathbb{E}[r|a] = \sigma(\theta^{*\top} x_a)$ where σ is the sigmoid function. Our theory assumes rewards (Eq. 2.1), while synthetic experiments use a logistic link—standard in bandit practice [5] as the sigmoid is approximately linear near zero. We verified on a purely linear environment that the TS advantage is consistent (1.2–1.8%). We use T=10,000, 5 arms/round, cold-start, and **12 seeds**, $\lambda = 1.0$, $v = 1.0$ (TS exploration), $B = 300$ (batch size), $\delta = 10^{-5}$ and feature clipping at $\|x\|_2 = 1$.

We also test the algorithms on MovieLens-25M [8] with 50-dimensional SVD features, T=60,000, warm-start prior and Jester joke-rating dataset [9].

We compare against the following baselines, all using the same (ε, δ) computed via Eq. (2.8) where applicable:

- **LinUCB / Linear TS**: Non-private (α=1, v=1)
- **AdaPrivate-UCB**: Our batched zCDP with UCB (α=1, same B, σ)
- **JDP-LinUCB** [2]: Tree-based aggregation, depth=$\lceil\log_2 T\rceil$, noise calibrated to joint DP
- **zCDP-Episodic** [10]: $\hat{\theta}$-perturbation, confidence-based episodes (c_{trigger}=2, geometric growth, σ_θ=$\Delta_\theta/\sqrt{2\rho}$)
- **AdaC-OFUL-zCDP / RarelySwitching**: Reproduced per [10], same ρ, same features, grid-searched $\alpha \in \{0.5, 1, 2\}$
- **DGS-LinUCB** [11]: Per-step $\hat{\theta}$-perturbation with dynamic sensitivity $\|A_t^{-1}\|_2$, sequential composition

All baselines were tuned with comparable effort. For UCB-based methods (LinUCB, AdaPrivate-UCB, AdaC-OFUL, RarelySwitching), we grid-searched $\alpha \in \{0.1, 0.5, 1.0, 2.0\}$ and report the best. For TS methods, we searched $v \in \{0.5, 1.0, 1.5, 2.0\}$. All batched methods used $B \in \{100, 300, 500, 1000\}$. JDP-LinUCB used tree depth $\lceil\log_2 T\rceil$ (no free parameter beyond α). zCDP-Episodic used $c_{\text{trigger}} \in \{1, 2, 4\}$ with geometric growth.

Our **AdaPrivate-TS** uses TS with sufficient statistics perturbation (noise on b only; A is from public features). **AdaPrivate-TS-Amp** adds Poisson subsampling (Section 3.2). **AdaPrivate-TS-Adaptive** adds exploration decay (v_0=1.5, γ=0.95).

4.2. Main Results: Thompson Sampling vs UCB

TS consistently outperforms UCB by 0.5–1.5% under default hyperparameters (Table 1), with the largest *default-setting* gap at $\varepsilon = 1$ (+1.5%). Paired t-tests confirm significance: $p < 0.01$ at $\varepsilon \in \{0.5, 1, 2\}$ and $p = 0.04$ at $\varepsilon = 5$. With *tuned adaptive exploration decay* (v_0=1+$\sigma/(d\sqrt{\lambda_0})$, γ=0.95)—a separate configuration from the default—the TS advantage increases to 7–9% at low $\varepsilon \in [0.1, 0.5]$. The largest absolute gap (up to 18%) occurs at

Table 1. Synthetic environment results. Private methods are reported as percentage of the reward achieved by their non-private counterparts.

Method	ε=0.5	ε=1	ε=2	ε=5
LinUCB (NP)		100% (baseline)		
Linear TS (NP)		101.2 ± 1.1%		
AdaPrivate-UCB	92.2±2.1	95.2±2.3	96.9±1.9	98.2±1.5
AdaPrivate-TS	**93.5±1.5**	**96.7±1.7**	**98.2±1.9**	**98.7±2.0**
TS Improvement	+1.3%	+1.5%	+1.3%	+0.5%

Table 2. Privacy amplification provides gains at low ε

Method	ε=0.5	ε=1	ε=2	ε=5
AdaPrivate-UCB	92.2±2.1	95.2±2.3	96.9±1.9	98.2±1.5
AdaPrivate-TS	93.5±1.5	96.7±1.7	98.2±1.9	98.7±2.0
TS-Amp (q=0.3)	**95.9±1.6**	96.5±1.3	96.7±1.3	96.7±1.4
TS-Amp (q=0.5)	95.4±1.5	**97.5±1.3**	97.7±1.5	97.7±1.5
TS-Adaptive	93.6±1.5	96.8±1.7	**98.3±1.9**	**98.7±2.0**
Best Gain vs UCB	**+3.7%**	**+2.3%**	+1.4%	+0.5%

ε=5 with adaptive tuning, because with moderate privacy noise TS fully exploits posterior structure while UCB's fixed confidence bound cannot adapt similarly. To be precise: the 0.5–1.5% gaps are for default TS vs. default UCB (both v=1, α=1); the 7–18% gaps are for TS-Adaptive vs. default UCB (matched B and σ, different exploration strategy). On MovieLens at ε=1 (200,000 replay steps), AdaPrivate-UCB suffers an initial CTR drop from 0.73 to 0.67 when the first batch of privacy noise arrives, then slowly recovers to 0.71. AdaPrivate-TS avoids this dip entirely—climbing steadily from 0.66 to 0.72—and overtakes UCB after ∼800 effective steps, confirming that TS's noise-as-uncertainty interpretation yields more stable learning under privacy.

4.3. **Privacy Amplification Results**

Amplification helps most at low ε, with gains persisting on real data. We tested $q \in \{0.1, 0.3, 0.5, 0.7\}$; q=0.3 is optimal at ε=1 and q=0.3–0.5 at ε=5. On MovieLens (T=60,000, B=300, 5 seeds), amplification with q=0.3 yields +4.6% at ε=1 and +1.3% at ε=5 over the non-amplified baseline. Too-aggressive subsampling (q=0.1) increases variance (±6% std vs. ±2%), while $q \geq 0.7$ offers negligible amplification. Results are not sensitive to tighter mechanism-specific bounds since evaluating at $\alpha \leq 64$ captures >99.5% of the optimal RDP bound.

4.4. **Comparison with State-of-the-Art**

Within event-level zCDP, AdaPrivate-TS achieves the highest performance at $\varepsilon \geq 2$ on MovieLens (up to +6.5% over AdaC-OFUL-zCDP at ε=10) and at every ε on Jester; at ε=1 on MovieLens, all event-level methods perform comparably (71.7–72.6%). JDP-LinUCB operates under joint DP (a *stronger* privacy model that also protects actions) and uses tree-based composition.

AdaC-OFUL-zCDP and RarelySwitching-zCDP both use event-level zCDP with parallel composition, enabling direct comparison. AdaPrivate-TS outperforms by up to +6.5% on MovieLens and +10.5% on Jester, confirming TS's posterior inflation provides utility beyond batching alone.

To test generalization, we also evaluate on the Jester joke-rating dataset [9] (d=50 SVD features, T=30,000). Table 3 shows AdaPrivate-TS dominates all baselines at every ε: +2.9% over zCDP-Episodic and +3.6% over AdaPrivate-UCB at ε=1. This confirms the TS advantage generalizes beyond MovieLens.

Table 3. Real-data results (% of non-private LinUCB) across multiple ε.

Method	ε=1	ε=2	ε=5	ε=10
MovieLens-25M *(NP: CTR=.700±.012; LinTS: 94.5%)*				
zCDP-Episodic	71.9±5.2	74.2±4.9	80.8±5.0	84.2±7.6
AdaC-OFUL-zCDP	72.3±11.9	74.4±12.6	79.8±12.1	84.1±11.2
RarelySwitching-zCDP	72.6±10.8	75.2±9.9	81.9±10.5	84.7±10.2
AdaPrivate-UCB	71.7±6.3	74.8±6.5	80.7±5.0	84.6±5.0
DGS-LinUCB	70.9±2.5	71.0±2.5	71.1±2.3	71.2±2.0
AdaPrivate-TS	71.7±6.3	**75.2±5.8**	**84.4±6.3**	**90.6±4.9**
JDP-LinUCB†	66.5±4.0	67.8±5.2	69.4±5.8	71.8±2.8
Jester *(NP: CTR=.363±.014; LinTS: 94.4%)*				
zCDP-Episodic	70.8±3.3	71.5±3.4	73.3±3.5	75.7±3.7
AdaC-OFUL-zCDP	72.3±3.2	73.1±3.8	74.5±3.4	76.3±4.3
RarelySwitching-zCDP	67.4±6.3	68.0±6.8	70.9±7.5	74.6±7.1
AdaPrivate-UCB	70.1±6.2	70.7±6.8	76.3±4.1	82.8±2.3
DGS-LinUCB	70.1±2.2	70.1±2.2	70.5±2.2	70.0±2.2
AdaPrivate-TS	**73.7±2.7**	**78.5±2.7**	**81.3±3.5**	**83.3±2.8**
JDP-LinUCB†	73.4±2.4	74.0±2.8	77.4±1.9	81.6±3.5

†Joint DP, tree-based composition (stronger privacy model). DGS: per-step sequential composition [11].

Table 4. Component ablation on synthetic data at $\varepsilon = 1$.

Configuration	**% NP**	**Gain**
Baseline (UCB + A&b noise)	91.2±2.4	–
+ B_ONLY noise	93.5±2.1	+2.3%
+ Thompson Sampling	95.2±1.9	+1.7%
+ Privacy Amplification (q=0.5)	96.8±1.5	+1.6%
+ Adaptive Exploration	**97.5±1.3**	+0.7%

Table 5. Off-policy evaluation on synthetic data with known ground truth (d=50, T=50,000).

Method	**True Value**	**Naive Error**	**IPS Error**	**DR Error**
LinUCB (NP)	0.534	0.229	0.083	0.089
AdaPrivate (ε=1)	0.765	0.005	0.149	0.144
AdaPrivate (ε=5)	0.766	0.002	0.148	0.142
AdaPrivate (ε=10)	0.764	<0.001	0.147	0.141

4.5. **Ablation Study**

Each component contributes incrementally, totaling $+$**6.3%** over the baseline. Performance is robust at ε=1: across $B \in \{100, 300, 500, 1000\}$, all achieve ≥94.1% (best at B=300: 96.7±1.7%); across $v \in [0.5, 2.0]$, all achieve ≥95.8%; adaptive decay is stable (±0.4%) over $(v_0, \gamma) \in \{1.0, 1.5, 2.0\} \times \{0.9, 0.95, 0.99\}$.

4.6. **User-Level Privacy Analysis**

Under zCDP group privacy, protecting all k ratings from one user scales as $k^2\rho$. For user-level $\varepsilon_{\text{user}}$=5: k=1 gives ρ=0.450 (80.4% NP), k=5 gives ρ=0.018 (26.0%), k=10 gives ρ=0.0045 (14.2%). Event-level DP suits users with $k \leq 5$; dedicated user-level mechanisms are needed for heavy users.

4.7. **Counterfactual Evaluation with IPS/DR**

We validate using IPS and doubly robust (DR) estimators on synthetic data with known ground truth.

AdaPrivate achieves higher true policy value than LinUCB across all protocols, confirming main results are not replay artifacts. On Jester (T=15,000, 5 seeds), all three OPE estimators (IPS, DR, SNIPS) produce consistent rankings: AdaPrivate-TS improves mono-

Table 6. Event-level vs. central-model DP baselines (synthetic, d=20, T=10,000, 12 seeds). AdaPrivate-TS uses parallel composition; central baselines use sequential composition.

Method	**Privacy**	ε=0.5	ε=1	ε=2	ε=5
LinUCB (NP)	None	100%	100%	100%	100%
AdaPrivate-TS	Event-level	**93.5**	**96.7**	**98.2**	**98.7**
CoreSet-Elim	Central	86.9	91.1	95.2	98.2
Batched-Agg	Central	83.3	84.9	87.5	92.1

tonically from ε=1 to ε=10 (DR: 0.168 $\rightarrow$ 0.178, ESS $\approx$ 730). DR and SNIPS are more stable than raw IPS (CV <8% vs. <12%), confirming DR is preferred for private bandit evaluation. Absolute regret is reported on synthetic data (Section 4.9) where ground truth is available.

4.8. Comparison with Central-Model DP Baselines

We compare against two central-model DP baselines using sequential composition with a trusted curator: (1) **CoreSet-Elimination-DP**, phased elimination with per-phase Gaussian noise, and (2) **Batched-Aggregation-DP**, batched LinUCB with sequential zCDP. Both split the privacy budget across phases ($\rho_{\text{per-batch}} = \rho_{\text{total}}/K$).

AdaPrivate-TS outperforms both central-model baselines despite a weaker privacy model (event-level, no trusted curator): +6.6% over CoreSet-Elimination and +10.2% over Batched-Aggregation at ε=0.5. Under sequential composition, central baselines must split the budget (ρ/K per batch), requiring $\sqrt{K}\times$ more noise; AdaPrivate-TS's parallel composition keeps full noise scale ($\max_k \rho_k = \rho$). At ε=5, the gap narrows (98.2% vs. 98.7%) as the splitting penalty diminishes.

4.9. Absolute Regret and T-Scaling

On synthetic data (d=20, 12 seeds), $R(T)/\sqrt{T}$ is near-constant over $T \in \{2500, \ldots, 20000\}$: 1.56–1.75 at ε=1, 0.84–0.88 at ε=5, confirming $O(\sqrt{T})$ scaling.

5. Related Work

Shariff and Sheffet [2] introduce JDP-LinUCB with $\tilde{O}(T^{3/4})$ regret under joint DP with adversarial contexts. Azize and Basu [10] give an $\Omega(d\sqrt{T}/\sqrt{\rho})$ lower bound under sequential composition; our $O(\sqrt{d}\log T/\sqrt{\rho})$ cost uses parallel composition with stochastic contexts. Hu and Hegde [12] study DP-TS for finite arms with lazy/doubling designs; we extend the noise-as-uncertainty principle to linear contextual bandits via sufficient-statistics perturbation. Balle et al. [6] and Mironov et al. [7] provide tight RDP bounds for subsampled Gaussian mechanisms. Our B_ONLY strategy adds noise on b only (d dimensions vs. d^2+d). Recent central-model DP linear bandits achieve $\tilde{O}(\sqrt{T} + \text{poly}(d)/\varepsilon)$ via core-set elimination and batched aggregation [13, 14], requiring a central curator or shuffle protocol. Our approach differs: TS enables per-round action selection (vs. epoch-based elimination) with $O(d^2)$ per-round cost (vs. $O(d^3)$), and our event-level parallel composition achieves comparable privacy cost without centralized infrastructure. The covariance-inflation insight (Proposition 4) is specific to posterior sampling and does not apply to elimination-based methods.

6. Discussion and Limitations

6.1. Privacy Notion and User-Level Extension

We use event-level DP (Definition 2), weaker than joint DP [2] or user-level DP. The naive group privacy scaling ($k^2\rho$ for a user with k interactions) is impractical for $k > 5$. We propose a **per-user contribution clipping** extension that provides meaningful user-level guarantees for moderate k:

Mechanism: Maintain per-user accumulators $c_u = \sum_{t:u_t=u} r_t x_{a_t}$. Before adding to the

batch sum, clip each user's contribution: $\tilde{c}_u = c_u \cdot \min(1, C/\|c_u\|_2)$ where $C > 0$ is the clipping threshold. The batch sum becomes $s_k = \sum_u \tilde{c}_u^{(k)}$. Since each user's clipped contribution has $\|\tilde{c}_u^{(k)}\|_2 \leq C$, changing all interactions of one user changes s_k by at most C. The sensitivity is $\Delta_2 = C$ (instead of k without clipping), giving $\sigma_{\text{user}} = C/\sqrt{2\rho}$.

Privacy guarantee: For target $(\varepsilon_{\text{user}}, \delta)$-user-level DP, set $\sigma_{\text{user}} = C/\sqrt{2\rho_{\text{user}}}$ where $\rho_{\text{user}} = \rho(\varepsilon_{\text{user}}, \delta)$. The clipping threshold C trades off bias (from clipping large contributions) against noise (larger C requires more noise). We set $C = \sqrt{k_{\max}} \approx \sqrt{5}$ as a practical default, targeting users with $\leq k_{\max}$ interactions.

Regret impact: Clipping introduces bias $\leq \sum_u \max(0, \|c_u\| - C)/B$ per batch, which is small when most users have few interactions ($k \leq k_{\max}$). The noise increases by a factor of C (since $\sigma_{\text{user}} = C \cdot \sigma_{\text{event}}$), so the privacy regret term becomes $O(C\sqrt{d}\log T/(\sqrt{\rho_{\text{user}}}\lambda_0))$.

On synthetic data (d=20, T=10,000, $\varepsilon_{\text{user}}$=5, 12 seeds), clipping with C=$\sqrt{k}$ achieves 96.4% of non-private performance for k=10 interactions per user, vs. 90.6% for naive group privacy—a +5.8% improvement. At k=20, the gap widens to +7.3% (94.8% vs. 87.4%). Clipping makes user-level DP practical for users with moderate interaction counts ($k \leq 20$), whereas group privacy degrades rapidly beyond k=5.

6.2. **Stochastic Contexts and Forced Exploration**

Assumption 2 ($\lambda_0 > 0$) is essential for the $O(\log T)$ privacy cost; under adversarial contexts, $\lambda_{\min}$ may stagnate. We address this with **forced exploration**: with probability $p_t = \min(1, d\log(t)/(t\gamma))$ for tuning parameter $\gamma > 0$, play a uniformly random arm, ensuring:

$$\lambda_{\min}(A_t) \geq \lambda + \Omega\left(\frac{d\log^2 t}{\gamma|\mathcal{A}|}\right) \tag{6.1}$$

even under adversarial contexts, provided candidate sets span $\mathbb{R}^d$. The additional regret from forced exploration is $O(d\log^2 T/\gamma)$. Setting $\gamma = d$, the total regret becomes:

$$R_{\text{forced}}(T) = O\left(d\sqrt{T\log T} + \frac{\sqrt{d}\log T}{\sqrt{\rho}} \cdot \gamma|\mathcal{A}| + \frac{d\log^2 T}{\gamma}\right) \tag{6.2}$$

removing dependence on λ_0 at the cost of a weaker $O(\log^2 T)$ privacy term. On synthetic data (d=20, 12 seeds) with adversarial rank-deficient contexts, AdaPrivate-TS maintains 94–96% of non-private performance; forced exploration (γ=d) adds $< 1\%$ random rounds.

6.3. **Public Contexts and DP-SVD Feature Privatization**

B_ONLY assumes public item features. If features must be private, A requires noise, adding $O(d^2)$ dimensions. We ablated this using a DP-SVD mechanism on MovieLens (T=60,000, 5 seeds, $\varepsilon_{\text{reward}}$=1) as follows.

DP-SVD mechanism: Given the $n_{\text{users}} \times n_{\text{items}}$ rating matrix R, we first clip each user's row to unit ℓ_2-norm: $\tilde{r}_u = r_u/\max(1, \|r_u\|_2)$, ensuring $\|\tilde{r}_u\|_2 \leq 1$. We then compute the item covariance $C = \tilde{R}^\top\tilde{R}/n_{\text{users}}$. Replacing one user's clipped row changes C in Frobenius norm by at most $\|\tilde{r}_u\tilde{r}_u^\top\|_F/n_{\text{users}} \leq 1/n_{\text{users}}$, giving ℓ_2-sensitivity $\Delta_2 = 1/n_{\text{users}}$. We add Gaussian noise $\tilde{C} = C + \mathcal{N}(0, \sigma_{\text{feat}}^2 I)$ with $\sigma_{\text{feat}} = \Delta_2/\sqrt{2\rho_{\text{feat}}}$, clip negative eigenvalues to zero (PSD projection), and extract the top-d eigenvectors as private features. This is a *one-shot offline* computation independent of the online bandit interaction.

Privacy composition: DP-SVD is computed offline from training data while the reward-DP mechanism operates on separate test interactions, so the two budgets do *not* compose ($\varepsilon_{\text{feat}}$ protects co-rating patterns; $\varepsilon_{\text{reward}}$ protects online rewards). If applied to overlapping data, sequential composition would give $\rho_{\text{feat}} + \rho_{\text{reward}}$; here data disjointness removes the need for joint accounting.

We tested at $\varepsilon_{\text{feat}} \in \{1, 5, 10\}$ (all with δ=10^{-5}): overall degradation is $\leq 3.2\%$ vs. public features. Crucially, TS's advantage over UCB *grows* under private features: from -2.5%

(public) to +4.3% ($\varepsilon_{\text{feat}}$=5) and +11.1% ($\varepsilon_{\text{feat}}$=2), confirming that TS interprets feature noise as additional exploration uncertainty rather than corruption.

6.4. Tighter Trace-Based Regret Bound

The $\lambda_{\min}$-based bound in Theorem 7 is worst-case over the spectrum of A_k. Using $\operatorname{tr}(A_k^{-2})$ directly yields a tighter bound $R_{\text{priv}}(T) = O(\sigma\sqrt{d_{\text{eff}}}\log T/B)$, where $d_{\text{eff}} = \sum_i \mu_i^{-2}$ is the *effective privacy dimension* determined by the eigenvalues μ_i of $\mathbb{E}[xx^\top]$. For rapidly decaying spectra (e.g., SVD features with $\mu_i \propto i^{-2}$), $d_{\text{eff}} = O(1)$ and the privacy cost becomes dimension-free. On MovieLens (50-dim SVD features), $d_{\text{eff}} \approx 8.3$, yielding a 6× tighter bound.

7. Conclusion

AdaPrivate-TS demonstrates that TS is better suited than UCB for private contextual bandits: DP noise inflates the posterior covariance to $v^2A^{-1} + \sigma^2A^{-2}$, interpreted as uncertainty rather than corruption, yielding a privacy cost of $O\big(\sqrt{d}\cdot\log T/\sqrt{\rho}\big)$ via parallel composition. On MovieLens and Jester, AdaPrivate-TS achieves the best performance among event-level zCDP baselines at $\varepsilon \geq 2$; under DP-SVD private features, TS's advantage over UCB grows to +11%. Poisson amplification provides additional gains at $\varepsilon \leq 1$, offering a principled tool for practitioners who need to tighten privacy without sacrificing recommendation quality.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith. "Calibrating noise to sensitivity in private data analysis". In: *TCC*. 2006, pp. 265–284.

[2] R. Shariff and O. Sheffet. "Differentially private contextual linear bandits". In: *NeurIPS*. Vol. 31. 2018.

[3] M. Bun and T. Steinke. "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *TCC*. 2016, pp. 635–658.

[4] W. R. Thompson. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

[5] S. Agrawal and N. Goyal. "Thompson sampling for contextual bandits with linear payoffs". In: *ICML*. 2013, pp. 127–135.

[6] B. Balle, G. Barthe, and M. Gavin. "Privacy amplification by subsampling: Tight analyses via couplings and divergences". In: *NeurIPS*. Vol. 31. 2018.

[7] I. Mironov, K. Talwar, and L. Zhang. "Rényi differential privacy of the sampled Gaussian mechanism". In: *arXiv preprint arXiv:1908.10530*. 2019.

[8] F. M. Harper and J. A. Konstan. "The movielens datasets: History and context". In: *ACM TIIS* 5.4 (2015), pp. 1–19.

[9] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. "Eigentaste: A constant time collaborative filtering algorithm". In: *Information Retrieval* 4.2 (2001), pp. 133–151.

[10] A. Azize and D. Basu. "Concentrated Differential Privacy for Bandits". In: *IEEE SaTML*. arXiv:2309.00557. 2024.

[11] H. Wang, D. Zhao, and H. Wang. "Dynamic Global Sensitivity for Differentially Private Contextual Bandits". In: *arXiv preprint arXiv:2208.14555* (2022).

[12] B. Hu and N. Hegde. "Near-optimal Thompson sampling-based algorithms for differentially private stochastic bandits". In: *UAI*. 2022, pp. 844–852.

[13] K. Zheng, T. Cai, W. Huang, Z. Li, and L. Wang. "Locally differentially private (contextual) bandits learning". In: *NeurIPS*. Vol. 33. 2020.

[14] S. R. Chowdhury and X. Zhou. "Shuffle Private Linear Contextual Bandits". In: *ICML*. 2022, pp. 3984–4009.