IMPROVED PRIVATE SPARSE COVARIANCE ESTIMATION WITH MULTISCALE THRESHOLD TESTS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study differentially private covariance estimation in operator norm for mean-zero sub-Gaussian distributions with unknown covariance support and at most k nonzero entries per row. We develop a multiscale random-threshold algorithm with sample complexity $\widetilde{O}(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon))$ for (ε, δ)-differential privacy and error at most $\alpha\sigma^2$, where d is the dimension and σ is a known sub-Gaussian scale. The bound improves the privacy-dependent term of the existing $\widetilde{O}(k^2/\alpha^2 + k^{3/2}\sqrt{d}/(\alpha\varepsilon))$ (Kumar et al., 2026) upper bound by a factor of $\sqrt{k}$, and matches the lower bound of $\widetilde{\Omega}(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon))$ in its applicable parameter regime.

Our key technical ingredient is a direct operator-norm bound on the centered fluctuations of an ideal reconstruction, exploiting conditional independence rather than accumulating entrywise errors across each row. Combined with a multiscale test allocation that balances reconstruction variance against query sensitivity, this bound removes the additional $\sqrt{k}$ factor from the privacy-dependent sample complexity.

1 INTRODUCTION

Covariance estimation is a basic task in multivariate statistics and provides the spectral information used by methods such as principal component analysis. Its difficulty changes substantially when the ambient dimension is comparable to or larger than the number of observations. In this regime, the empirical covariance matrix can be inaccurate in operator norm even when its individual entries are well estimated. Structural assumptions are therefore essential for obtaining useful guarantees from limited data. Covariance sparsity provides one such assumption: each coordinate has nonzero covariance with only a small number of other coordinates, although individual observations may still be dense (Bickel & Levina, 2008; Cai & Zhou, 2012).

A substantial non-private literature shows how to exploit this structure. Bickel & Levina (2008) establish operator-norm guarantees for thresholding the empirical covariance, while Cai & Liu (2011) develop entry-dependent thresholds that adapt to the variability of individual covariance estimates. Cai & Zhou (2012) complement these algorithmic developments with minimax characterizations over sparse covariance classes. These results demonstrate that suitable regularization can replace polynomial dependence on the ambient dimension with dependence primarily on sparsity and logarithmic dimension factors. They also highlight the importance of the loss function: accurate estimation of a covariance matrix in operator norm requires controlling the aggregate effect of errors across its entries, rather than treating those entries as unrelated scalar estimation problems.

Privacy introduces a different constraint. Differential privacy requires the output distribution to remain stable when a single observation is replaced, and this requirement applies to arbitrary neighboring datasets, not only datasets that are typical under a statistical model (Dwork & Roth, 2014). Consequently, sparsity of the population covariance does not directly yield a sparse or low-sensitivity empirical statistic. Even a distribution with diagonal covariance can generate dense sample vectors and dense outer products. For example, if $X \sim \mathcal{N}(0, I_d)$, then $\mathbb{E}\|X\|_2^2 = d$, despite the covariance having only one nonzero entry per row. Thus, bounds based on uniformly small sample norms must be interpreted carefully when assessing dimensional dependence. Private thresholding methods provide an initial way to combine sparsity with privacy (Wang & Xu, 2021), but determining the sample

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cost of privacy under a natural statistical normalization requires a more precise understanding of this distinction.

We consider covariance estimation from n independent and identically distributed observations drawn from a mean-zero, σ-sub-Gaussian distribution on $\mathbb{R}^d$, where the scale $\sigma > 0$ is known. The population covariance Σ is k-row-column sparse: each row and column contains at most k nonzero entries, with diagonal entries counted toward this bound. Its support is unknown, and we impose neither an eigengap condition nor a sparsity assumption on its eigenvectors. Our goal is to construct an (ε, δ)-differentially private estimator $\widehat{\Sigma}$ that approximates Σ in operator norm with high probability:

$$\Pr\Big(\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \le \alpha\sigma^2\Big) \ge 1 - \beta, \tag{1}$$

where the probability is taken jointly over the observations and the internal randomness of the estimator. In the non-private case, thresholding yields the benchmark sample complexity $\widetilde{O}(k^2/\alpha^2)$[1] (Bickel & Levina, 2008). Kumar et al. (2026, Theorem 1) obtain the private upper bound $\widetilde{O}\Big(k^2/\alpha^2 + k^{3/2}\sqrt{d}/(\alpha\varepsilon)\Big)$. Their Theorem 2 gives the lower bound $\widetilde{\Omega}\Big(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon)\Big)$ when $\log(d/\alpha) \lesssim k \lesssim d^{1/2-\lambda}$ for some constant $\lambda > 0$[2] and $\delta \lesssim \varepsilon^2/d^2$. These bounds identify a $\sqrt{d}$ privacy cost in that regime, but leave a $\sqrt{k}$ gap in the privacy term. This raises a natural question: *Is it possible to remove the additional sparsity factor $\sqrt{k}$ from the privacy term and thereby match the lower bound up to logarithmic factors?*

Our contribution. We answer this question affirmatively by developing a multiscale randomized-threshold mechanism that achieves equation 1 using

$$\widetilde{O}\left(\frac{k^2}{\alpha^2} + \frac{k\sqrt{d}}{\alpha\varepsilon}\right) \tag{2}$$

samples. The mechanism satisfies (ε, δ)-differential privacy under replace-one adjacency for every input dataset. Compared with (Kumar et al., 2026), our bound preserves the non-private sample complexity term and improves the privacy term by a factor of $\sqrt{k}$, without imposing additional structural assumptions. Consequently, it closes the $\sqrt{k}$ gap between the upper and lower bounds of Kumar et al. (2026), up to logarithmic factors when $\log(d/\alpha) \lesssim k \lesssim d^{1/2-\lambda}$ for some constant $\lambda > 0$, and $\delta \lesssim \varepsilon^2/d^2$.

Technical ingredients. Let S_{clip} denote the empirical covariance matrix computed from clipped observations, and let $\widehat{\Sigma}$ denote the private estimator. Our key technical observation is that the randomized reconstruction, rather than S_{clip} itself, inherits the row sparsity of the population covariance on a high-probability empirical concentration event. Specifically, conditional on S_{clip}, the ideal estimator Σ^{id} is a sum of independent random matrices whose contributions vanish outside the support of Σ. Writing $\overline{\Sigma}(S_{\mathrm{clip}}) := \mathbb{E}[\Sigma^{\mathrm{id}} \mid S_{\mathrm{clip}}]$, we apply matrix concentration directly to $\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S_{\mathrm{clip}})\|_{\mathrm{op}}$, rather than first bounding the entrywise fluctuations and then using $k\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S_{\mathrm{clip}})\|_{\max}$. This allows us to exploit the centered, independent structure of the random contributions instead of accumulating their absolute errors across each row. We separately control the error of the conditional mean, $\|\overline{\Sigma}(S_{\mathrm{clip}}) - \Sigma\|_{\mathrm{op}}$, using covariance sparsity, and transfer the resulting accuracy guarantee to $\widehat{\Sigma}$ through a high-probability coupling.

Building on this observation, we design our mechanism around three key ingredients.

(i) **A noise-corrected randomized representation.** Rather than releasing covariance entries directly, we construct a multiscale, randomly weighted representation of the clipped empirical covariance. A noise-corrected scalar integral identity gives the corresponding ideal estimator a small truncation bias. Crucially, reweighting compensates in expectation for randomly omitted contributions, without requiring exact support recovery.

[1] Throughout, the tilde notation suppresses logarithmic factors in the problem parameters.

[2] The non-private lower bound holds for $\frac{k^2 \log d}{\alpha^2} \lesssim c_\zeta d^{1/\zeta}$ for any $\zeta > 1$

(ii) **Scale-dependent allocation and direct matrix concentration.** We allocate the expected number of active tests per position quadratically in the threshold scale, while both normalized-query sensitivity and output weights decrease inversely with that scale. This balances the squared-sensitivity and second-moment terms across scales. Conditional on the empirical matrix, the ideal estimator is a sum of independent random matrices, allowing us to control its fluctuations directly in operator norm using matrix Bernstein (Tropp, 2012), rather than accumulating entrywise random-error bounds across each row.

(iii) **Private quotas and high-probability coupling.** To ensure privacy on arbitrary datasets, we enforce public, scale-dependent quotas through private oneshot top-k selection (Qiao et al., 2021). These quotas provide a uniform sensitivity bound for the subsequent query vector for every fixed selected label set. Under the statistical model, we prove that selection retains every potentially nonzero contribution with high probability. Using the same threshold tests and gating noise, the actual output then coincides exactly with the ideal estimator on this event. We can therefore transfer the ideal estimator's accuracy without conditioning its concentration argument on successful selection, while establishing privacy separately for all adjacent datasets.

Related work. Research on covariance estimation spans several complementary directions. Structural regularization and minimax theory characterize the benefits of covariance structure (Cai & Zhou, 2012; Cai et al., 2016). Private matrix release studies privacy–accuracy trade-offs for bounded records and data-dependent error guarantees (Amin et al., 2019; Dong et al., 2022). Private distribution learning and stable estimation address Gaussian and sub-Gaussian mean and covariance estimation through affine-invariant procedures (Kamath et al., 2019; Brown et al., 2023). Research on private sparse covariance estimation combines sparsity with privacy constraints (Wang & Xu, 2021; Kumar et al., 2026). These directions differ in their structural assumptions, input normalizations, privacy definitions, and estimation losses. Appendix A provides a broader discussion.

2 PRELIMINARIES

Sparse covariance matrix. We study the sparse covariance estimation problem of Kumar et al. (2026, Model 1 and Problem 1), without an eigengap assumption. Let $D = \{X_1, \dots, X_n\}$ consist of independent and identically distributed samples from an unknown mean-zero distribution P on $\mathbb{R}^d$. We assume that P is σ-sub-Gaussian: for a known scale $\sigma > 0$,

$$\mathbb{E}_{X\sim P}\big[\exp(u^\top X)\big] \le \exp\left(\frac{\sigma^2\|u\|_2^2}{2}\right) \qquad \text{for every } u \in \mathbb{R}^d. \tag{3}$$

The population covariance $\Sigma = \mathbb{E}_{X\sim P}[XX^\top]$ is *k-row-column sparse*: each row and each column contains at most k nonzero entries, with nonzero diagonal entries counted toward this bound. Since Σ is symmetric, the row and column sparsity conditions are equivalent. The support of Σ is unknown to the learner.

The sub-Gaussian condition implies $v^\top \Sigma v = \mathbb{E}[(v^\top X)^2] \le \sigma^2\|v\|_2^2$ for every $v \in \mathbb{R}^d$, and hence $0 \preceq \Sigma \preceq \sigma^2 I_d$. Importantly, sparsity is imposed only on the population covariance; the observations, their outer products, and the empirical covariance matrix need not be sparse.

Definition 2.1 (Differential privacy)**.** *Let $\mathcal{X}$ be the data domain, and let $\mathcal{Y}$ be a measurable output space. Two datasets $D = \{X_1, \dots, X_n\}$ and $D' = \{X_1', \dots, X_n'\}$ in $\mathcal{X}^n$ are adjacent under replace-one adjacency, denoted $D \sim D'$, if they differ in at most one sample. For $\varepsilon \ge 0$ and $\delta \in [0, 1)$, a randomized mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ satisfies (ε, δ)-differential privacy if, for every adjacent pair $D \sim D'$ and every measurable event $E \subset \mathcal{Y}$,*

$$\Pr(\mathcal{M}(D) \in E) \le e^\varepsilon \Pr(\mathcal{M}(D') \in E) + \delta,$$

where the probabilities are taken over the internal randomness of $\mathcal{M}$, with D and D' held fixed.

Private estimation objective. Let $\|\cdot\|_{\mathrm{op}}$ denote the operator norm, i.e., $\|A\|_{\mathrm{op}} = \max_{\|x\|_2\le 1}\|Ax\|_2$. The goal is to design a randomized mechanism $\mathcal{M} : (\mathbb{R}^d)^n \to \mathbb{S}^d$, where $\mathbb{S}^d$ denotes the space of real symmetric $d \times d$ matrices, that satisfies (ε, δ)-differential privacy under

replace-one adjacency (Definition 2.1) and returns $\widehat{\Sigma} = \mathcal{M}(D)$ satisfying

$$\Pr\Big(\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \le \alpha\sigma^2\Big) \ge 1 - \beta \tag{4}$$

for every distribution P obeying the assumptions above. Here the probability is taken jointly over the sampled dataset and the internal randomness of $\mathcal{M}$. The distributional assumptions are required only for accuracy: the privacy guarantee must hold for every pair of adjacent datasets in $(\mathbb{R}^d)^n$. We seek to minimize the sample size n needed to meet these privacy and accuracy requirements.

Additional notations. Let $\mathcal{E}_d := \{(i,j) : 1 \le i \le j \le d\}$. Thus, $\mathcal{E}_d$ indexes the upper-triangular entries, including the diagonal. For the standard basis vectors $e_1, \dots, e_d$, define

$$B_{ii} := e_i e_i^\top, \qquad B_{ij} := e_i e_j^\top + e_j e_i^\top \quad (i < j).$$

For $e = (i,j) \in \mathcal{E}_d$, write $B_e := B_{ij}$. Then $\{B_e\}_{e \in \mathcal{E}_d}$ forms a basis for the space $\mathbb{S}^d$ of real symmetric $d \times d$ matrices.

We write $\|Q\|_{\max} := \max_{i,j} |Q_{ij}|$ for the entrywise maximum norm. For any symmetric matrix Q with at most k nonzero entries in each row, we have that $\|Q\|_{\mathrm{op}} \le \max_i \sum_j |Q_{ij}| \le k\|Q\|_{\max}$.

Throughout, we assume $d \ge 2$, $1 \le k \le d$, $0 < \alpha \le \frac{1}{4}$, $0 < \varepsilon \le 1$ and $0 < \delta, \beta \le \frac{1}{10}$. All logarithms are natural unless marked $\log_2$. We use $\widetilde{O}(\cdot)$ to hide the logarithmic factors of $(d, k, 1/\varepsilon, 1/\alpha, 1/\beta, 1/\delta)$. We use $[N]$ to denote the set $\{1, 2, \dots, N\}$ for an integer N. For a threshold $z \ge 0$, define the scalar clipping function by $\mathrm{Clip}_z(x) := \max\{\min\{x, z\}, -z\}$, for $x \in \mathbb{R}$. For a vector $X \in \mathbb{R}^d$, we apply clipping coordinatewise: $[\mathrm{Clip}_z(X)](j) := \mathrm{Clip}_z(X(j)), j = 1, \dots, d$. For $b > 0$, let $\mathrm{Lap}(b)$ denote the zero-mean Laplace distribution with scale parameter b and probability density $f_b(x) = \frac{1}{2b}\exp\Big(-\frac{|x|}{b}\Big)$ for $x \in \mathbb{R}$. We write $\mathcal{N}(\mu, \sigma^2)$ for the uni-variate Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$.

3 The Multiscale Threshold Mechanism

In this section, we present our private covariance estimation mechanism, summarized in Algorithm 1, and introduce its parameters L, R, r, and $p_\ell, t_\ell, b_\ell, s_\ell{}_{\ell=0}^{L}$.

As in Kumar et al. (2026, Algorithm 1), the mechanism consists of private selection followed by reconstruction using fresh, independent noise. Their method performs row-wise private top-k selection and constructs a symmetric estimate from independently perturbed entry values, with the diagonal handled separately. Our method instead selects a set J of randomized threshold tests. Each selected, active test produces a value by applying a gate to a noisy normalized query. This value is then multiplied by a public weight and added to the corresponding covariance entry. This design enables direct control of reconstruction fluctuations in operator norm, without converting an entrywise bound into an operator-norm bound.

Pre-process: clipping step. Given the dataset $D = \{X_1, X_2, \dots, X_n\}$, we first clip D up to a proper threshold. Define $R = \sigma\sqrt{2\log(40nd/\beta)}$. We set $Y_i = \mathrm{Clip}_R(X_i)$ and $S(D) = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^\top$. When D is clear from the context, we write S for $S(D)$. The probability that any observation is altered by clipping is accounted for in the overall failure probability β.

Stage I: Private candidate selection. Fix two constants $C_s = 64$ and $C_\rho = 4096$. Define $t_0 = \frac{\alpha\sigma^2}{256k}$ to be the minimal scale, and $L = \big\lceil \log_2 \frac{512k}{\alpha} \big\rceil$ to be total number of scales. Let $t_\ell := 2^\ell t_0$ for $\ell = 0, \dots, L$. By the choice of L, the largest threshold satisfies $2\sigma^2 \le t_L < 4\sigma^2$. We further choose $\rho = \frac{C_\rho L k \log(80dL/\beta)}{\alpha^2\sigma^4}$, $\mu_\ell = \rho t_\ell^2$, $m_\ell = \lceil \mu_\ell \rceil$ and $p_\ell = \mu_\ell / m_\ell$.

We index each candidate threshold test by a tuple $c = (e, \ell, h)$, where $e \in \mathcal{E}_d$ specifies the matrix position, $0 \le \ell \le L-1$ specifies the scale level, and $h \in [m_\ell]$ is the repetition index. Thus, for each position–scale pair (e, ℓ), we create m_ℓ candidates, each using independently generated randomness for activation, threshold generation, and subsequent private selection.

Algorithm 1 Multiscale private covariance estimation

Require: Dataset $D = \{X_1, X_2, \ldots, X_n\}$; public parameters $d, k, \sigma, \alpha, \varepsilon, \delta, \beta$.
1: $Y_i \leftarrow \mathrm{Clip}_R(X_i)$ for $i = 1, 2, \ldots .n$;
2: Set $S \leftarrow \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^\top$.
3: **for** $c = (e, \ell, h) \in \mathcal{C}$ **do**
4: Draw independent public $U_c \sim \mathrm{Bern}(p_\ell)$, $T_c \sim \mathrm{Unif}[t_\ell, 2t_\ell]$.
5: **end for**
6: **for** $\ell = 0, \ldots, L-1$ **do**
7: **if** $s_\ell < M_\ell$ **then**
8: **for** every $c = (e, \ell, h) \in \mathcal{C}_\ell$ **do**
9: Draw independent $\zeta_c \sim \mathrm{Lap}(b_\ell)$
10: $q_c \leftarrow (|S_e|/T_c - 1)U_c - 2(1 - U_c)$
11: **end for**
12: Select $J_\ell \in \arg\max_{\substack{J \subseteq \mathcal{C}_\ell \\ |J| = s_\ell}} \sum_{c \in J}(q_c + \zeta_c)$;
13: **else**
14: $J_\ell \leftarrow \mathcal{C}_\ell$.
15: **end if**
16: **end for**
17: $\widehat{\Sigma} \leftarrow 0_{d \times d}$.
18: **for** each $c = (e, \ell, h) \in J = \bigcup_{\ell=0}^{L-1} J_\ell$, in a fixed public order **do**
19: **if** $U_c = 1$ **then**
20: Draw a fresh independent $\check{Z}_c \sim N(0, r^2)$ and set $Z_c = \mathrm{Clip}_{1/4}(\check{Z}_c)$.
21: $\widehat{\Sigma} \leftarrow \widehat{\Sigma} + \frac{1}{\kappa \rho t_\ell} \cdot \psi\left(\frac{S_e}{T_c} + Z_c\right) B_e$.
22: **end if**
23: **end for**
24: **return** $\widehat{\Sigma}$

Formally, the candidate set is $\mathcal{C} = \{(e, \ell, h) : e \in \mathcal{E}_d, 0 \le \ell \le L-1, 1 \le h \le m_\ell\}$. Let $\mathcal{C}_\ell$ denote the candidate set at level ℓ, i.e., $\mathcal{C}_\ell = \{(e, \ell, h) : e \in \mathcal{E}_d, 1 \le h \le m_\ell\}$. By definition, the size of $\mathcal{C}_\ell$ is $M_\ell = \frac{d(d+1)}{2} \cdot m_\ell$.

For every candidate $c = (e, \ell, h)$, the algorithm generates independent public randomness (see Line 4 Algorithm 1) $U_c \sim \mathrm{Bernoulli}(p_\ell)$ and $T_c \sim \mathrm{Unif}[t_\ell, 2t_\ell]$, and computes $q_c = (|S_e|/T_c - 1)U_c - 2(1 - U_c)$. Here, U_c indicates whether candidate c is active, T_c is its random threshold, and q_c is its unperturbed selection score.

Choose $H = \log(80dL/\beta)$, $\Delta = 2R^2/n$, $\varepsilon' = \frac{\varepsilon}{8L}$ and $\delta' = \frac{\delta}{32L}$. Define $s_\ell = \min\{M_\ell, \lceil C_s \left[d\rho \min\{kt_\ell^2, \sigma^4\} + H\right] \rceil\}$. If $s_\ell < M_\ell$, add independent $\zeta_c \sim \mathrm{Lap}(b_\ell)$, where

$$b_\ell = \frac{8\Delta\sqrt{s_\ell \log(M_\ell/\delta')}}{t_\ell \varepsilon'}, \tag{5}$$

and set

$$J_\ell \in \arg\max_{\substack{J \subseteq \mathcal{C}_\ell \\ |J| = s_\ell}} \sum_{c \in J}(q_c + \zeta_c). \tag{6}$$

In words, J_ℓ denotes the set of candidates with the s_ℓ largest noisy scores $q_c + \zeta_c$ among $\mathcal{C}_\ell$[3]. If $s_\ell = M_\ell$, the algorithm directly sets $J_\ell = \mathcal{C}_\ell$. In this way, we obtain $J = \cup_{\ell=0}^{L-1} J_\ell$ as the set of selected private candidate tests.

At each level, the number of active candidates is chosen to balance variance reduction against privacy loss. Distributing the reconstruction weight across more candidates can reduce the estimator's variance conditional on S, but may incur a greater privacy cost during reconstruction. The selec-

[3]The independent, continuous Laplace perturbations ensure that the noisy scores are distinct almost surely, so J_ℓ is uniquely determined. Any ties arising from finite-precision arithmetic can be resolved using a fixed, data-independent ordering of the candidate labels.

tion quotas are therefore calibrated to achieve low conditional variance while remaining within the prescribed privacy budget.

Stage II: Private covariance reconstruction. Define

$$A = \sum_{\ell=0}^{L-1} \frac{s_\ell}{t_\ell^2}, \qquad S_\star = \sum_{\ell=0}^{L-1} s_\ell, \qquad r = \frac{8\Delta\sqrt{A}}{\varepsilon}\sqrt{\log\frac{32}{\delta}}. \tag{7}$$

Let ν_r denote the distribution of $Z = \mathrm{Clip}_{1/4}(\check{Z})$ for $\check{Z} \sim \mathcal{N}(0, r^2)$, and set $\kappa = \mathbb{E}_{Z\sim\nu_r}\left[\frac{1}{1-Z}\right] \in \left[1, \frac{4}{3}\right]$. Define $\psi(x) = \mathbf{1}\{x > 1\} - \mathbf{1}\{x < -1\}$.

For each selected candidate $c \in J$, draw a fresh independent $Z_c \sim \nu_r$ and return

$$\widehat{\Sigma} = \sum_{\ell=0}^{L-1} \sum_{c=(e,\ell,h)\in J_\ell} \frac{\psi(S_e/T_c + Z_c)\cdot U_c}{\kappa\rho t_\ell} B_e. \tag{8}$$

In this reconstruction, for each $e \in \mathcal{E}_d$, $\widehat{\Sigma}_e$ is a noisy weighted combination of the outcomes of active candidate tests across scales. The noises are carefully chosen to ensure that $\widehat{\Sigma}$ concentrates around its conditional mean $\mathbb{E}[\widehat{\Sigma}|S]$ within the desired error tolerance.

4 Analysis of Approximation Error and Privacy

We first present the main result under Algorithm 1. Recall that $L = \lceil \log_2(512k/\alpha) \rceil$ and $H = \log(80dL/\beta)$. Set $M = \sum_{\ell=0}^{L-1} M_\ell \le O\left(\frac{kL^2\log(80dL/\beta)}{\alpha^2\sigma^2}\right)$ and $\Gamma = \log(80LM/(\beta\delta))$.

Theorem 4.1 (Privacy and accuracy)**.** *There exists a universal constant $C > 0$ such that, with the public parameters defined in Section 3, whenever*

$$n \ge C\left[\frac{k^2}{\alpha^2}\log\frac{80d}{\beta} + \frac{k\sqrt{d}}{\alpha\varepsilon}\log\frac{80nd}{\beta}\, L^{3/2}\sqrt{H}\,\Gamma^{3/2}\right], \tag{9}$$

Algorithm 1 is (ε, δ)-differentially private on every input dataset $D \in (\mathbb{R}^d)^n$. Moreover, under the stated distributional assumptions, $\mathbb{P}\left(\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{op}} \le \alpha\sigma^2\right) \ge 1 - \beta$, where the probability is taken over the sampling of D and all randomness of the algorithm.

By hiding the logarithmic terms using $\widetilde{O}$, Algorithm 1 ensures a sample complexity bound of $\widetilde{O}\left(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon)\right)$. This improves the privacy-dependent term by a factor of $\sqrt{k}$ compared to the upper bound in (Kumar et al., 2026), and matches the lower bound of $\widetilde{\Omega}(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon))$ (Kumar et al., 2026) in its applicable parameter regime.

In the rest of this section, we present the proof sketch of Theorem 4.1. We postpone the supporting lemmas and proofs to Appendices B and C. Assume that the condition in equation 9 holds, with C to be specified later.

4.1 Bound of Approximation Error

We work under the sampling model and the sample-size condition of the main theorem. Recall that $S = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^\top$ is the clipped empirical covariance matrix. For the accuracy analysis, introduce an ideal estimator Σ^{id} that includes all candidates in $\mathcal{C}$. Generate independent reconstruction noises $Z_c \sim \nu_r$ for every $c \in \mathcal{C}$, independently of the data, public candidate randomness, and selection noises, and define

$$\Sigma^{\mathrm{id}} := \sum_{\ell=0}^{L-1} \sum_{c=(e,\ell,h)\in\mathcal{C}_\ell} \frac{U_c}{\kappa\rho t_\ell}\psi\left(\frac{S_e}{T_c} + Z_c\right) B_e, \qquad \overline{\Sigma}(S) := \mathbb{E}\left[\Sigma^{\mathrm{id}} \mid S\right]. \tag{10}$$

The conditional expectation averages over both the public candidate randomness and the reconstruction noises. The actual mechanism uses the same contributions, but only for selected candidates in J. Since selection does not inspect the reconstruction noises, generating them in advance and using them on the selected labels preserves the distribution of the actual output.

We decompose the estimation error into four terms:

$$\widehat{\Sigma} - \Sigma = \underbrace{\widehat{\Sigma} - \Sigma^{\mathrm{id}}}_{\text{selection discrepancy}} + \underbrace{\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)}_{\text{reconstruction fluctuation}} + \underbrace{\overline{\Sigma}(S) - S}_{\text{reconstruction bias}} + \underbrace{S - \Sigma}_{\text{sampling and clipping error}}. \tag{11}$$

We analyze the four terms in equation 11 separately.

Sampling and clipping error: $S - \Sigma$. Define $\mathcal{E}_{\mathrm{cov}} = \left\{\|S - \Sigma\|_{\max} \le \frac{t_0}{16}\right\}$ and $\mathcal{S} = \{e \in \mathcal{E}_d : \Sigma_e \neq 0\}$. By Bernstein's inequality (Lemma B.6), the choice of clipping radius R makes the probability that any observed coordinate is clipped at most $\beta/20$. By choosing the constant C in equation 9 large enough, Lemma B.1 therefore gives $\Pr(\mathcal{E}_{\mathrm{cov}}^c) \le \beta/10$. In particular, $\|S - \Sigma\|_{\max} \le t_0/16$ with the required probability.

Selection discrepancy: $\widehat{\Sigma} - \Sigma^{\mathrm{id}}$. By definition, we have that

$$\widehat{\Sigma} - \Sigma^{\mathrm{id}} = \sum_{\ell=0}^{L-1} \sum_{c=(e,\ell,h)\notin J_\ell} \frac{U_c}{\kappa \rho t_\ell} \Psi\left(\frac{S_e}{T_c} + Z_c\right) B_e. \tag{12}$$

Since Z_c is bounded by $[-1/4, 1/4]$, an active candidate $c = (e, \ell, h)$ can contribute to $\widehat{\Sigma}$ only if $|S_e|/T_c > 3/4$. On $\mathcal{E}_{\mathrm{cov}}$, quota coverage and score separation ensure that every such candidate is retained, except on a small-probability event. Under the shared-noise coupling, every omitted ideal contribution is therefore exactly zero. Lemma B.4 gives

$$\Pr\left(\mathcal{E}_{\mathrm{cov}} \cap \{\widehat{\Sigma} \neq \Sigma^{\mathrm{id}}\}\right) \le \frac{\beta}{20}. \tag{13}$$

This is a data-dependent accuracy statement, not a uniform guarantee for every fixed input dataset D. Selection of J remains necessary for the the privacy analysis because it bounds the number of protected queries even on datasets whose empirical covariance S has many large entries.

Reconstruction fluctuation: $\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)$. Fix an empirical matrix S satisfying $\mathcal{E}_{\mathrm{cov}}$. Conditional on this S, the candidate contributions $\{\frac{U_c}{\kappa\rho t_\ell}\Psi\left(\frac{S_e}{T_c} + Z_c\right) B_e\}_{c=(e,\ell,h)\in J}$ are independent and vanish outside $\mathcal{S}$. Since each row has at most k support positions, the conditional matrix variance is at most $kL/\rho = \widetilde{O}\left(\alpha^2\sigma^4\right)$, while each candidate contribution $\frac{U_c}{\kappa\rho t_\ell}\Psi\left(\frac{S_e}{T_c} + Z_c\right) B_e$ has operator norm at most $2/(\rho t_0) = \widetilde{O}(\alpha\sigma^2)$. Lemma B.3 therefore gives

$$\Pr\left(\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)\|_{\mathrm{op}} > \sqrt{\frac{2kL\log(40d/\beta)}{\rho}} + \frac{4\log(40d/\beta)}{3\rho t_0} \,\middle|\, S\right) \le \frac{\beta}{20}. \tag{14}$$

The choices of ρ and t_0 make the displayed threshold at most $\alpha\sigma^2/2$. This argument controls the centered reconstruction directly in operator norm; it does not first bound entrywise fluctuations and then multiply by k.

Reconstruction bias: $\overline{\Sigma}(S) - S$. By the design of activation rates, random thresholds and public weights, we have that

$$\overline{\Sigma}_e(S) = \mathbb{E}[\Sigma_e^{\mathrm{id}} | S] = \frac{1}{\kappa}\int_{t_0}^{t_L} \mathbb{E}_Z \psi(S_e/t + Z)\, dt.$$

The noise-corrected identity

$$\frac{1}{\kappa}\int_0^\infty \mathbb{E}_Z \psi(a/t + Z)\, dt = a, \qquad a \in \mathbb{R},$$

shows that random activation and noisy gating introduce no additional uncorrected multiplicative bias. On $\mathcal{E}_{\mathrm{cov}}$, the largest threshold t_L is sufficiently large that the integrand vanishes above t_L. The discrepancy from S_e therefore comes entirely from omitting thresholds below t_0. Lemma B.2 yields

$$\|\overline{\Sigma}(S) - S\|_{\max} \le t_0. \tag{15}$$

Combining the bounds. To convert the entrywise bounds for the last two terms into an operator-norm bound, we first add them and then use support preservation. On $\mathcal{E}_{\mathrm{cov}}$, any position e with zero $\Sigma_e = 0$ satisfies $|S_e/T_c + Z_c| \le 1/16 + 1/4 < 1$, so the contribution at position e always vanishes. Thus $\overline{\Sigma}(S) - \Sigma$ has at most k nonzero entries per row, even though $\overline{\Sigma}(S) - S$ and $S - \Sigma$ may individually be dense. It follows that

$$\begin{aligned}\|\overline{\Sigma}(S) - \Sigma\|_{\mathrm{op}} &\le k\|\overline{\Sigma}(S) - \Sigma\|_{\max} \\ &\le k\left(\|\overline{\Sigma}(S) - S\|_{\max} + \|S - \Sigma\|_{\max}\right) \le \frac{17}{16}kt_0 = \frac{17}{4096}\alpha\sigma^2.\end{aligned} \tag{16}$$

Consequently, whenever the empirical, coupling, and matrix concentration conclusions hold,

$$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \le \|\widehat{\Sigma} - \Sigma^{\mathrm{id}}\|_{\mathrm{op}} + \|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)\|_{\mathrm{op}} + \|\overline{\Sigma}(S) - \Sigma\|_{\mathrm{op}} < \alpha\sigma^2. \tag{17}$$

Define $\mathcal{E}_{\mathrm{mat}} = \left\{\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)\|_{\mathrm{op}} \le \frac{\alpha\sigma^2}{2}\right\}$. By equation 14,

$$\Pr(\mathcal{E}_{\mathrm{cov}} \cap \mathcal{E}_{\mathrm{mat}}^c) = \mathbb{E}[\mathbf{1}_{\mathcal{E}_{\mathrm{cov}}} \Pr(\mathcal{E}_{\mathrm{mat}}^c \mid S)] \le \frac{\beta}{20}.$$

A union bound therefore gives

$$\Pr\left(\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} > \alpha\sigma^2\right) \le \Pr(\mathcal{E}_{\mathrm{cov}}^c) + \Pr\left(\mathcal{E}_{\mathrm{cov}} \cap \{\widehat{\Sigma} \ne \Sigma^{\mathrm{id}}\}\right) + \Pr(\mathcal{E}_{\mathrm{cov}} \cap \mathcal{E}_{\mathrm{mat}}^c) \le \frac{\beta}{5} < \beta.$$

4.2 The Differential Privacy Guarantee

We then establish the privacy guarantee of our mechanism for every neighboring pair $D \sim D'$ under replace-one adjacency. We first fix all public activations and thresholds $\{(U_c, T_c)\}_{c\in\mathcal{C}}$ and prove a privacy guarantee uniform over their realizations.

Recall

$$\Delta = \frac{2R^2}{n}, \qquad A = \sum_{\ell=0}^{L-1} \frac{s_\ell}{t_\ell^2}, \qquad S_\star = \sum_{\ell=0}^{L-1} s_\ell, \qquad r = \frac{8\Delta\sqrt{A}}{\varepsilon}\sqrt{\log\frac{32}{\delta}}.$$

Write $\eta := 2S_\star \exp\left(-\frac{1}{32r^2}\right)$. For a sufficiently large choice of C in equation 9, Lemma C.6 yields $r \le \frac{1}{8\sqrt{1}}$ and $(1 + e^{\varepsilon/4})\eta \le \frac{\delta}{16}$.

Private selection. By the clipping step, we have that $|S_e(D) - S_e(D')| \le \Delta$. Consequently, the score q_c for any candidate c has sensitivity at most Δ/t_ℓ. Thus the score vector $\{q_c\}_{c\in C_\ell}$ at scale ℓ has L_∞ sensitivity at most Δ/t_ℓ. Applying the oneshot Laplace selection theorem with the prescribed scale b_ℓ makes the unordered selected label set J_ℓ to be (ε', δ')-differentially private, where $\varepsilon' = \varepsilon/(8L)$ and $\delta' = \delta/(32L)$. A scale with $s_\ell = M_\ell$ returns all labels deterministically. Composition over the scales therefore gives

$$J = (J_0, \ldots, J_{L-1}) \quad \text{is} \quad (\varepsilon/8, \delta/32)\text{-differentially private.} \tag{18}$$

Lemma C.1 provides the full proof of equation 18.

Uniform sensitivity after selection. For any fixed possible label collection $j = (j_0, \ldots, j_{L-1})$ with $|j_\ell| = s_\ell$, arrange its labels in a public order and set

$$F_j(D) := (U_c S_e(D)/T_c)_{c=(e,\ell,h)\in j} \in \mathbb{R}^{S_\star}.$$

By Lemma C.3, the public quotas bound the squared sensitivity of this entire vector:

$$\|F_j(D) - F_j(D')\|_2^2 \le \sum_{\ell=0}^{L-1} \sum_{c\in j_\ell} \frac{\Delta^2}{t_\ell^2} = \Delta^2 \sum_{\ell=0}^{L-1} \frac{s_\ell}{t_\ell^2} = \Delta^2 A. \tag{19}$$

Gaussian protection and bounded reconstruction noise. For a random variable X, let $\mathcal{L}(X)$ denote the distribution of X. Let $G \sim \mathcal{N}(0, r^2 I_{S_\star})$ and $P_D^j := \mathcal{L}(F_j(D) + G)$.

The preceding sensitivity bound and the prescribed choice of r ensure that the ordinary Gaussian mechanism P_D^j is $(\varepsilon/4, \delta/16)$-differentially private for every fixed j (see Lemma C.7).

Define $Z = \mathrm{Clip}_{1/4}(G)$ and $Q_D^j := \mathcal{L}(F_j(D) + Z)$. Under this coupling, the Gaussian and clipped-noise outputs agree unless some coordinate of G leaves $[-1/4, 1/4]$. Hence Lemma C.4 gives the uniform comparison

$$\mathrm{TV}(P_D^j, Q_D^j) \le \Pr\left(\max_{a \le S_\star} |G_a| > \frac{1}{4}\right) \le \eta. \tag{20}$$

Replacing the two Gaussian laws in the privacy inequality by their clipped-noise counterparts costs at most $(1 + e^{\varepsilon/4})\eta$ in the additive privacy parameter. Consequently,

$$D \longmapsto F_j(D) + Z \quad \text{is} \quad (\varepsilon/4, \delta/8)\text{-differentially private}, \tag{21}$$

uniformly over all fixed j (see Lemma C.4).

Composition and postprocessing. Selection of J does not inspect the fresh reconstruction noises. Therefore, after the first-stage transcript equals j, the second stage executes exactly the fixed-label mechanism just analyzed. Adaptive composition gives privacy for the auxiliary transcript

$$\mathcal{T}(D) := (J, F_J(D) + Z), \qquad \mathcal{T} \text{ is } (3\varepsilon/8, 5\delta/32)\text{-differentially private.} \tag{22}$$

Writing $V = F_J(D) + Z$, the final matrix is recovered as

$$\widehat{\Sigma} = \sum_{\ell=0}^{L-1} \sum_{c=(e,\ell,h)\in J_\ell} \frac{U_c}{\kappa \rho t_\ell} \psi(V_c) B_e. \tag{23}$$

This map uses only the protected transcript (J, V) and public quantities. Thus the matrix inherits the privacy guarantee from (J, V). Finally, the bound is uniform over the public activations and thresholds $\{U_c, T_c\}_{c\in\mathcal{C}}$, so averaging over their data-independent law preserves it, even if those public variables are revealed. Lemma C.5 formalizes these steps. Since $3\varepsilon/8 \le \varepsilon$ and $5\delta/32 \le \delta$, the mechanism in Algorithm 1 is (ε, δ)-differentially private on every input dataset D.

5 Discussion

We developed a multiscale randomized-threshold mechanism for private sparse covariance estimation. The proposed mechanism achieves operator-norm error $\alpha\sigma^2$ with probability at least $1 - \beta$ using $\widetilde{O}(k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon))$ samples under the sub-Gaussian assumption, while satisfying (ε, δ)-differential privacy under replace-one adjacency for arbitrary input datasets. Our result removes the additional $\sqrt{k}$ factor from the previous privacy term, matching the lower bound of Kumar et al. (2026) up to logarithmic factors in its applicable parameter regime. The key insight is to represent covariance entries through noise-corrected, randomly weighted threshold tests, enabling direct control of reconstruction fluctuations at the matrix level without requiring exact support recovery.

Three directions merit further investigation. First, there is still a gap between the upper and lower bounds when $\delta > \varepsilon^2/d^2$. It remains open whether the lower bound of $k^2/\alpha^2 + k\sqrt{d}/(\alpha\varepsilon)$ still applies in this regime. Second, extending the mechanism to approximately sparse covariance matrices would broaden its applicability. This extension would require replacing the exact-support argument with quantitative bounds on the bias and variance contributed by small but nonzero entries. Thirdly, a direct privacy analysis of the final aggregated matrix may yield tighter guarantees than the current analysis, which establishes privacy for the more informative auxiliary transcript containing the selected labels and perturbed queries. Understanding how aggregation conceals intermediate information could enable sharper noise calibration, potentially improving the logarithmic factors or practical accuracy of the mechanism.

AI USE STATEMENT

Generative AI tools were used to improve the writing and exposition of this manuscript, assist with mathematical discussion and checking derivations and proofs, support literature lookup and LaTeX preparation, and search for and refine candidate choices of algorithmic parameters. The parameter search included the initial candidate-set size, the clipping thresholds, the size of the selected candidate set J, and the per-scale repetition counts m_ℓ. This assistance therefore included technical exploration of the construction and was not limited to language editing. AI-assisted proof checking does not constitute formal verification or an independent guarantee of correctness. The authors take full responsibility for the final manuscript, including the mathematical claims and proofs, the algorithm and parameter choices, and the accuracy and originality of the text and references.

ETHICS STATEMENT

This work is a theoretical study of differentially private sparse covariance estimation. It does not involve human participants, empirical experiments on personal or sensitive datasets, or the release of such data.

REPRODUCIBILITY STATEMENT

This paper reports theoretical results and does not rely on empirical experiments or external datasets. Section 2 specifies the distributional assumptions, estimation objective, and privacy definition. Algorithm 1 and Section 3 describe the mechanism and its public parameters. Section 4 gives the proof overview. Appendices B and C contain the detailed accuracy and privacy proofs, including the sufficient sample-size bounds. These specifications and proofs provide the basis for reproducing the theoretical derivations.

REFERENCES

Kareem Amin, Travis Dick, Alex Kulesza, Andres Munoz, and Sergei Vassilvitskii. Differentially private covariance estimation. *Advances in Neural Information Processing Systems*, 32, 2019.

Peter J Bickel and Elizaveta Levina. Covariance regularization by thresholding. 2008.

Gavin Brown, Samuel B Hopkins, and Adam Smith. Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. *arXiv preprint arXiv:2301.12250*, 2023.

T Tony Cai and Harrison H Zhou. Optimal rates of convergence for sparse covariance matrix estimation. 2012.

T Tony Cai, Zhao Ren, and Harrison H Zhou. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. 2016.

Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

Wei Dong, Yuting Liang, and Ke Yi. Differentially private covariance revisited. *Advances in Neural Information Processing Systems*, 35:850–861, 2022.

Tommaso d'Orsi and Gleb Novikov. On purely private covariance estimation. *arXiv preprint arXiv:2510.26717*, 2025.

David Durfee and Ryan M Rogers. Practical differentially private top-k selection with pay-what-you-get composition. *Advances in Neural Information Processing Systems*, 32, 2019.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3-4):211–487, 2014.

Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pp. 1853–1902. PMLR, 2019.

Syamantak Kumar, Shourya Pandey, Purnamrita Sarkar, and Kevin Tian. On the curse of dimensionality in private sparse covariance estimation and PCA. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 4349–4400. PMLR, 29 Jun–03 Jul 2026.

Shyam Narayanan. Better and simpler lower bounds for differentially private statistical estimation. *IEEE Transactions on Information Theory*, 71(2):1376–1388, 2024.

Gang Qiao, Weijie Su, and Li Zhang. Oneshot differentially private top-k selection. In *International Conference on Machine Learning*, pp. 8672–8681. PMLR, 2021.

Adam J Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

Di Wang and Jinhui Xu. Differentially private high dimensional sparse covariance matrix estimation. *Theoretical Computer Science*, 865:119–130, 2021.

A Additional Related Works

Structured covariance estimation. Non-private research studies how covariance structure reduces statistical complexity and which rates are minimax optimal. For sparse covariance matrices, hard, generalized, and adaptive thresholding provide operator-norm guarantees (Bickel & Levina, 2008; Rothman et al., 2009; Cai & Liu, 2011). Cai & Zhou (2012) establish minimax rates under specified sparsity and high-dimensional conditions, while Cai et al. (2016) survey structured covariance and precision-matrix estimation. Operator-norm guarantees control the aggregate effect of estimation errors across entries, rather than treating entries as unrelated scalar estimation problems. Here, sparsity concerns the population covariance, not individual observations; sparse covariance therefore need not imply sparse empirical outer products or low sensitivity.

Private matrix release. This direction studies empirical covariance release under bounded-input assumptions, with guarantees depending on the loss and privacy definition. Amin et al. (2019) study pure-private covariance release with Frobenius-error guarantees. Dong et al. (2022) obtain trace-sensitive and tail-sensitive Frobenius-error bounds under concentrated differential privacy. d'Orsi & Novikov (2025) study pure-private empirical covariance estimation, including spectral-norm guarantees. The distinction between empirical matrix release and population estimation also matters, since the latter includes sampling error. Comparisons must preserve input normalization: bounds for unit-norm records do not directly translate into dimension-independent sample complexity for population covariance estimation to operator-norm error $\alpha\sigma^2$ under a sub-Gaussian model.

Private distribution learning and stable estimation. A complementary direction estimates distributional parameters without strong prior bounds on Gaussian parameters. Kamath et al. (2019) use recursive private preconditioning for high-dimensional distribution learning, including Gaussian learning in total variation distance. Brown et al. (2023) develop affine-invariant private mean and covariance estimators for sub-Gaussian distributions using stable estimation. These approaches address general distributional estimation rather than specifically exploiting unknown row-sparse covariance support. Here, the sub-Gaussian scale is known, the covariance support is unknown, and the loss is absolute operator norm.

Private sparse estimation and lower bounds. Wang & Xu (2021) develop private thresholding for high-dimensional sparse covariance estimation. Under the same sub-Gaussian normalization as this paper, Kumar et al. (2026) combine private row-wise selection with noisy entry reconstruction, obtaining $\widetilde{O}(k^2/\alpha^2 + k^{3/2}\sqrt{d}/(\alpha\varepsilon))$ sample complexity. On the other side, their privacy lower bound of $\tilde{\Omega}(k\sqrt{d}/(\alpha\varepsilon))$ draws on techniques related to Narayanan (2024), while Cai & Zhou (2012) provide the non-private benchmark. Our rate matches this lower bound up to logarithmic factors when $\log d \lesssim k \lesssim d^{1/2-\lambda}$ for some constant $\lambda > 0$ and $\delta \le \epsilon^2/d^2$. These additional restrictions are not required for the upper-bound guarantee.

Private selection. Private top-k selection is studied by Durfee & Rogers (2019) and Qiao et al. (2021); the latter provide the oneshot Laplace guarantee used here. Our mechanism selects threshold-test labels under public, scale-dependent quotas rather than recovering the population support. These quotas bound the sensitivity of the reconstruction queries for every fixed selected label set on arbitrary datasets.

B Supporting Lemmas For the Approximation Error

All notation is inherited from the main argument. For $c = (e, \ell, h)$, define

$$W_c := \frac{U_c}{\kappa\rho t_\ell}\psi(S_e/T_c + Z_c)B_e, \qquad \overline{W}_c := W_c - \mathbb{E}[W_c \mid S], \qquad c = (e, \ell, h).$$

In particular, $\Sigma^{\mathrm{id}} = \sum_{c\in\mathcal{C}} W_c$ and $\Sigma^{\mathrm{id}} - \overline{\Sigma}(S) = \sum_{c\in\mathcal{C}} \overline{W}_c$.

For any candidate test $c = (e, \ell, h)$ and any candidate-indexed quantity X, we use the notations X_c and $X_{e,\ell,h}$ interchangeably.

Lemma B.1 (Empirical covariance concentration)**.** *There is a universal constant C such that*

$$n \geq C\frac{k^2}{\alpha^2}\log\frac{80d}{\beta} \quad \Longrightarrow \quad \Pr(\mathcal{E}_{\mathrm{cov}}^c) \leq \frac{\beta}{10}.$$

Proof. Recall $R = \sigma\sqrt{2\log(40nd/\beta)}$. The coordinatewise sub-Gaussian tail bound and a union bound give

$$\Pr\left(\max_{a,j}|X_{a,j}| > R\right) \leq 2nd\exp\left(-\frac{R^2}{2\sigma^2}\right) = \frac{\beta}{20}.$$

On the complementary event $\mathcal{E}_{\mathrm{clip}}$, clipping does not alter any record, and hence

$$S = S^{\mathrm{raw}} := \frac{1}{n}\sum_{a=1}^{n} X_a X_a^\top.$$

For fixed i, j, the centered products $X_{a,i}X_{a,j} - \Sigma_{ij}$ are independent across the samples and have sub-exponential scale bounded by $C\sigma^2$. Indeed, sub-Gaussian moment bounds and Cauchy–Schwarz give, for every integer $p \geq 1$,

$$\mathbb{E}|X_{a,i}X_{a,j}|^p \leq \left(\mathbb{E}|X_{a,i}|^{2p}\,\mathbb{E}|X_{a,j}|^{2p}\right)^{1/2} \leq (C\sigma^2 p)^p,$$

and centering changes this bound by at most a universal constant. Then, by Bernstein's inequality (Lemma B.6),

$$\Pr\left(|S_{ij}^{\mathrm{raw}} - \Sigma_{ij}| > \tau\right) \leq 2\exp\left[-cn\min\left\{\frac{\tau^2}{\sigma^4}, \frac{\tau}{\sigma^2}\right\}\right].$$

Taking $\tau = t_0/16 = \alpha\sigma^2/(4096k)$ and applying a union bound over at most d^2 entries gives

$$\Pr\left(\|S^{\mathrm{raw}} - \Sigma\|_{\max} > \frac{t_0}{16}\right) \leq 2d^2\exp\left(-c'n\frac{\alpha^2}{k^2}\right) \leq \frac{\beta}{20}$$

under the stated sample-size condition.

Finally,

$$\mathcal{E}_{\mathrm{cov}}^c \subseteq \mathcal{E}_{\mathrm{clip}}^c \cup \left\{\|S^{\mathrm{raw}} - \Sigma\|_{\max} > \frac{t_0}{16}\right\}.$$

Adding the two failure probabilities proves the claim. □

Lemma B.2. *Recall that $\mathcal{S} = \{e \in \mathcal{E}_d : \Sigma_e \neq 0\}$. For every fixed S satisfying $\mathcal{E}_{\mathrm{cov}}$,*

$$\overline{\Sigma}_e(S) = \frac{1}{\kappa}\int_{t_0}^{t_L} \mathbb{E}_Z\psi(S_e/t + Z)\,dt, \qquad \|\overline{\Sigma}(S) - S\|_{\max} \leq t_0.$$

Moreover, if $e \notin \mathcal{S}$, every ideal contribution at position e vanishes, so $\Sigma_e^{\mathrm{id}} = \overline{\Sigma}_e(S) = 0$. In particular,

$$\|\overline{\Sigma}(S) - \Sigma\|_{\mathrm{op}} \leq \frac{17}{16}kt_0.$$

Proof. The symmetry and bounded support of ν_r imply

$$1 \leq \kappa = \mathbb{E}(1 - Z)^{-1} \leq \frac{4}{3}.$$

The lower bound follows from Jensen's inequality and $\mathbb{E}Z = 0$; the upper bound follows from $Z \leq 1/4$.

For $a > 0$, the negative gate is impossible, while

$$\psi(a/t + Z) = \mathbf{1}\left\{t < \frac{a}{1 - Z}\right\}.$$

Integrating first over t gives

$$\int_0^\infty \mathbb{E}_Z\psi(a/t + Z)\,dt = \mathbb{E}_Z\frac{a}{1 - Z} = a\kappa.$$

For $a = -b < 0$, the positive gate is impossible, and the integral equals $-b\,\mathbb{E}(1+Z)^{-1} = a\kappa$ by symmetry. The case $a = 0$ is immediate. Thus

$$\frac{1}{\kappa}\int_0^\infty \mathbb{E}_Z\psi(a/t+Z)\,dt = a, \qquad a \in \mathbb{R}. \tag{24}$$

For the uniform-threshold construction, independence within a candidate and $m_\ell p_\ell = \rho t_\ell^2$ give

$$\begin{aligned}\mathbb{E}\left[\sum_{h=1}^{m_\ell}\frac{U_{e,\ell,h}}{\kappa\rho t_\ell}\psi(a/T_{e,\ell,h}+Z_{e,\ell,h})\,\middle|\,S\right] &= \frac{m_\ell p_\ell}{\kappa\rho t_\ell^2}\int_{t_\ell}^{2t_\ell}\mathbb{E}_Z\psi(a/t+Z)\,dt \\ &= \frac{1}{\kappa}\int_{t_\ell}^{2t_\ell}\mathbb{E}_Z\psi(a/t+Z)\,dt.\end{aligned}$$

Taking $a = S_e$ and summing over the dyadic scales proves the conditional-mean formula.

On $\mathcal{E}_{\mathrm{cov}}$,

$$|S_e| \le \sigma^2 + \frac{t_0}{16}, \qquad \frac{4}{3}|S_e| < 2\sigma^2 \le t_L.$$

A nonzero gate requires $t < 4|S_e|/3$, so the integrand vanishes for $t \ge t_L$. Subtracting equation 24 from the conditional-mean formula yields

$$\overline{\Sigma}_e(S) - S_e = -\frac{1}{\kappa}\int_0^{t_0}\mathbb{E}_Z\psi(S_e/t+Z)\,dt.$$

Since $|\psi| \le 1$ and $\kappa \ge 1$,

$$|\overline{\Sigma}_e(S) - S_e| \le \frac{t_0}{\kappa} \le t_0.$$

If $\Sigma_e = 0$, then $|S_e| \le t_0/16$, and every candidate at that position satisfies

$$|S_e/T_c + Z_c| \le \frac{1}{16} + \frac{1}{4} < 1.$$

Every such contribution is zero. Hence $\overline{\Sigma}(S) - \Sigma$ is supported on $\mathcal{S}$ and has at most k nonzero entries per row. Its entries satisfy

$$|\overline{\Sigma}_e(S) - \Sigma_e| \le |\overline{\Sigma}_e(S) - S_e| + |S_e - \Sigma_e| \le t_0 + \frac{t_0}{16}.$$

The row-sum bound for a symmetric matrix now gives

$$\|\overline{\Sigma}(S) - \Sigma\|_{\mathrm{op}} \le k\left(t_0 + \frac{t_0}{16}\right) = \frac{17}{16}kt_0.$$

□

Lemma B.3 (Conditional matrix concentration)**.** *For every fixed S satisfying $\mathcal{E}_{\mathrm{cov}}$ and every $u > 0$,*

$$\Pr\left(\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)\|_{\mathrm{op}} > \sqrt{\frac{2kLu}{\rho}} + \frac{4u}{3\rho t_0}\,\middle|\,S\right) \le 2de^{-u}.$$

In particular, for $u = \log(40d/\beta) \le H$ and the specified public parameters,

$$\Pr\left(\|\Sigma^{\mathrm{id}} - \overline{\Sigma}(S)\|_{\mathrm{op}} > \frac{\alpha\sigma^2}{2}\,\middle|\,S\right) \le \frac{\beta}{20}.$$

Proof. Conditional on S, each W_c depends only on its own independent triple (U_c, T_c, Z_c). Thus the centered matrices $\overline{W}_c$ are independent and self-adjoint, with $\mathbb{E}[\overline{W}_c \mid S] = 0$.

Fix e, ℓ and consider $c = (e,\ell,h)$ over different h. Note that $U_c^2 = U_c$ and $\psi^2 \le 1$ imply

$$\sum_{h=1}^{m_\ell}\mathbb{E}[W_{e,\ell,h}^2 \mid S] \preceq \frac{m_\ell p_\ell}{\kappa^2\rho^2 t_\ell^2}B_e^2 = \frac{1}{\kappa^2\rho}B_e^2 \preceq \frac{1}{\rho}B_e^2. \tag{25}$$

By Lemma B.2, positions outside $\mathcal{S}$ contribute nothing. Direct multiplication gives

$$B_{ii}^2 = e_i e_i^\top, \qquad B_{ij}^2 = e_i e_i^\top + e_j e_j^\top \quad (i < j).$$

Consequently, $\sum_{e \in \mathcal{S}} B_e^2 \preceq k I_d$, because each diagonal entry of this sum counts the nonzeros in the corresponding row of Σ. Summing equation 25 over positions and scales yields

$$\sum_{c \in \mathcal{C}} \mathbb{E}[W_c^2 \mid S] \preceq \frac{kL}{\rho} I_d.$$

Centering cannot increase the second-moment matrix:

$$\mathbb{E}[\overline{W}_c^2 \mid S] = \mathbb{E}[W_c^2 \mid S] - \big(\mathbb{E}[W_c \mid S]\big)^2 \preceq \mathbb{E}[W_c^2 \mid S].$$

Thus the conditional matrix variance satisfies

$$v := \left\| \sum_{c \in \mathcal{C}} \mathbb{E}[\overline{W}_c^2 \mid S] \right\|_{\mathrm{op}} \le \frac{kL}{\rho}.$$

Since $\|B_e\|_{\mathrm{op}} = 1$, $t_\ell \ge t_0$, and $\kappa \ge 1$,

$$\|W_c\|_{\mathrm{op}} \le \frac{1}{\rho t_0}, \qquad \|\overline{W}_c\|_{\mathrm{op}} \le \frac{2}{\rho t_0} =: B.$$

The self-adjoint matrix Bernstein inequality (Tropp, 2012, Theorem 6.1), applied to the sum and its negative, gives for any $x > 0$,

$$\Pr\left(\left\| \sum_{c \in \mathcal{C}} \overline{W}_c \right\|_{\mathrm{op}} > x \;\middle|\; S \right) \le 2d \exp\left(-\frac{x^2}{2(v + Bx/3)} \right).$$

For $x = \sqrt{2vu} + 2Bu/3$ for some $u \ge 0$, one has $x^2 \ge 2u(v + Bx/3)$, so

$$\Pr\left(\left\| \sum_{c \in \mathcal{C}} \overline{W}_c \right\|_{\mathrm{op}} > \sqrt{2vu} + \frac{2Bu}{3} \;\middle|\; S \right) \le 2de^{-u}.$$

Substituting the bounds on v and B proves the first claim. If $v = 0$, all centered summands vanish almost surely and the conclusion is immediate.

For $u = \log(40d/\beta) \le H$, recall $t_0 = \alpha\sigma^2/(256k)$ and $\rho = C_\rho kLH/(\alpha^2\sigma^4)$. Then

$$\sqrt{\frac{2kLu}{\rho}} + \frac{4u}{3\rho t_0} = \alpha\sigma^2 \left(\sqrt{\frac{2u}{C_\rho H}} + \frac{1024u}{3C_\rho LH} \right) \le \alpha\sigma^2 \left(\sqrt{\frac{2}{C_\rho}} + \frac{1024}{3C_\rho L} \right) \le \frac{1}{2}\alpha\sigma^2,$$

where the last inequality uses $C_\rho = 4096$ and $L \ge 11$. Since $2de^{-u} = \beta/20$, the second claim follows. □

Lemma B.4 (Coupling lemma)**.** *Under the sample-size condition of the main theorem, the actual and ideal estimators admit a coupling such that*

$$\Pr\Big(\mathcal{E}_{\mathrm{cov}} \cap \{\widehat{\Sigma} \ne \Sigma^{\mathrm{id}}\} \Big) \le \frac{\beta}{20}.$$

Proof. Generate the reconstruction noises Z_c for all candidates $c \in \mathcal{C}$ before selection, and use the same noise on each selected label in both estimators. Selection does not inspect these noises, so this construction preserves the distribution of the actual mechanism.

Define

$$\mathcal{R}_\ell = \{c \in \mathcal{C}_\ell : q_c \ge -3/8\}, \qquad K_\ell = \{e \in \mathcal{E}_d : |\Sigma_e| \ge t_\ell/2\},$$
$$B_\ell = \sum_{e \in K_\ell} \sum_{h=1}^{m_\ell} U_{e,\ell,h}, \qquad D_\ell = d\rho \min\{k t_\ell^2, \sigma^4\}.$$

If $c = (e, \ell, h) \in \mathcal{R}_\ell$, then $U_c = 1$ and $|S_e| \geq 5T_c/8$. On $\mathcal{E}_{\text{cov}}$,

$$|\Sigma_e| \geq \frac{5}{8}t_\ell - \frac{t_0}{16} \geq \frac{9}{16}t_\ell > \frac{t_\ell}{2}.$$

Hence $e \in K_\ell$ and, by summing the corresponding indicator inequality, $|\mathcal{R}_\ell| \leq B_\ell$.

The k-sparsity and $\sum_j \Sigma_{ij}^2 \leq \sigma^4$ give $|K_\ell| \leq d \min\left\{k, \frac{4\sigma^4}{t_\ell^2}\right\}$. For the uniform-threshold construction, $m_\ell p_\ell = \rho t_\ell^2$, so

$$\mathbb{E}B_\ell = |K_\ell| m_\ell p_\ell \leq d\rho \min\{kt_\ell^2, 4\sigma^4\} \leq 4D_\ell.$$

Since K_ℓ is fixed once Σ is fixed, B_ℓ is a sum of independent Bernoulli variables. Therefore,

$$\mathbb{E}e^{B_\ell} \leq \exp\big((\mathrm{e}-1)\mathbb{E}B_\ell\big).$$

For a layer with $s_\ell < M_\ell$, the quota satisfies $s_\ell \geq 64(D_\ell + H)$, and Markov's inequality gives

$$\begin{aligned}\Pr(B_\ell > s_\ell) &\leq \exp\big(-s_\ell + (\mathrm{e}-1)\mathbb{E}B_\ell\big) \\ &\leq \exp\big(-64(D_\ell + H) + 4(\mathrm{e}-1)D_\ell\big) \leq e^{-H}.\end{aligned}$$

If $s_\ell = M_\ell$, then $B_\ell \leq s_\ell$ deterministically. Thus, defining $\mathcal{E}_{\text{quota}} = \bigcap_{\ell=0}^{L-1}\{B_\ell \leq s_\ell\}$, a union bound gives

$$\Pr(\mathcal{E}_{\text{quota}}^c) \leq Le^{-H} = \frac{\beta}{80d} \leq \frac{\beta}{40}.$$

On $\mathcal{E}_{\text{cov}} \cap \mathcal{E}_{\text{quota}}$, we consequently have $|\mathcal{R}_\ell| \leq s_\ell$ at every scale.

Recall that $M = \sum_{\ell=0}^{L-1} M_\ell$ and $\Gamma = \log(80LM/(\beta\delta))$. Then $b_\ell \leq 1/(64\Gamma)$ for any $0 \leq \ell \leq L-1$. The Laplace tail bound therefore implies

$$\Pr(|\zeta_c| > 1/32) = \exp\left(-\frac{1}{32b_\ell}\right) \leq e^{-2\Gamma}.$$

There are at most M generated selection noises. For the event

$$\mathcal{E}_{\text{noise}} = \{\, |\zeta_c| \leq 1/32 \text{ for every } c \in C \,\},$$

by Lemma B.5, we obtain $b_\ell \leq \frac{1}{64\Gamma}$ and

$$\Pr(\mathcal{E}_{\text{noise}}^c) \leq Me^{-2\Gamma} \leq \frac{\beta}{40}.$$

We continue the analysis conditioned on $\mathcal{E}_{\text{noise}} \cap \mathcal{E}_{\text{cov}} \cap \mathcal{E}_{\text{quota}}$. For an active candidate c with $|S_e|/T_c > 3/4$, its noisy score satisfies

$$q_c + \zeta_c > -\frac{1}{4} - \frac{1}{32} = -\frac{9}{32}.$$

By contrast, every c' such that $c' \notin \mathcal{R}_\ell$,

$$q_{c'} + \zeta_{c'} < -\frac{3}{8} + \frac{1}{32} = -\frac{11}{32} < -\frac{9}{32}.$$

Since the size of $\mathcal{R}_\ell$ is at most s_ℓ, the rank of c is at most s_ℓ. Therefore, c must be included by J_ℓ.

As a result, every unselected active candidate therefore satisfies $|S_e|/T_c \leq 3/4$, and hence

$$|S_e/T_c + Z_c| \leq \frac{3}{4} + \frac{1}{4} = 1.$$

Its gate is zero, and inactive candidates contribute zero through $U_c = 0$. Under the shared-noise coupling,

$$\Sigma^{\text{id}} - \widehat{\Sigma} = \sum_{\ell=0}^{L-1} \sum_{c \in \mathcal{C}_\ell \setminus J_\ell} W_c = 0$$

on $\mathcal{E}_{\text{cov}} \cap \mathcal{E}_{\text{quota}} \cap \mathcal{E}_{\text{noise}}$. Consequently,

$$\mathcal{E}_{\text{cov}} \cap \{\widehat{\Sigma} \neq \Sigma^{\text{id}}\} \subseteq \mathcal{E}_{\text{quota}}^c \cup \mathcal{E}_{\text{noise}}^c,$$

and the desired probability bound follows by a union bound. □

Lemma B.5. *For every $0 \le \ell < L$ with $s_\ell < M_\ell$, it holds that $b_\ell \le \frac{1}{64\Gamma}$.*

Proof. Since $H \ge 1$, the definition of s_ℓ gives $s_\ell \le \lceil 64(d\rho k t_\ell^2 + H) \rceil \le 65(d\rho k t_\ell^2 + H)$. Using $t_\ell \ge t_0$, the definitions of t_0 and ρ, and $dL \ge 22$, we obtain

$$\frac{s_\ell}{t_\ell^2} \le 65\left(d\rho k + \frac{H}{t_0^2}\right) = 65 d\rho k\left(1 + \frac{16}{dL}\right) \le 130 d\rho k.$$

Moreover, $\log \frac{M_\ell}{\delta'} = \log \frac{32LM_\ell}{\delta} \le \Gamma$. Substituting $\varepsilon' = \frac{\varepsilon}{8L}$,$\Delta = \frac{4\sigma^2}{n}\log\frac{40nd}{\beta}$ and $\rho = \frac{4096LkH}{\alpha^2\sigma^4}$ into the definition of b_ℓ yields

$$b_\ell = \frac{64L\Delta}{\varepsilon}\sqrt{\frac{s_\ell}{t_\ell^2}\log\frac{M_\ell}{\delta'}} \le \frac{64L\Delta}{\varepsilon}\sqrt{130 d\rho k\Gamma} \le C_0\,\frac{k\sqrt{d}}{n\alpha\varepsilon}\log\frac{80nd}{\beta}\,L^{3/2}\sqrt{H\Gamma},$$

where $C_0 > 0$ is universal. For a sufficiently large universal constant C in equation 9, we have $b_\ell \le \frac{1}{64\Gamma}$. □

Lemma B.6 (Bernstein's inequality for sub-exponential variables)**.** *Let $X_1, \dots, X_n$ be independent, mean-zero, real-valued random variables satisfying $\|X_i\|_{\psi_1} \le K$ for all i, where $\|X\|_{\psi_1} := \inf\{s > 0 : \mathbb{E}\exp(|X|/s) \le 2\}$. There exists a universal constant $c > 0$ such that, for every $t > 0$,*

$$\Pr\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) \le 2\exp\left[-c\min\left\{\frac{t^2}{nK^2}, \frac{t}{K}\right\}\right].$$

Equivalently, for every $\tau > 0$,

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge \tau\right) \le 2\exp\left[-cn\min\left\{\frac{\tau^2}{K^2}, \frac{\tau}{K}\right\}\right].$$

Consequently, there is a universal constant $C > 0$ such that, for every $\eta \in (0,1)$, with probability at least $1 - \eta$,

$$\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \le CK\left(\sqrt{\frac{\log(2/\eta)}{n}} + \frac{\log(2/\eta)}{n}\right).$$

C Supporting Lemmas for the Privacy Guarantee

The arguments below fix the public activations and thresholds until the final averaging step. The notation and public parameters are the same as in Section 4.2.

Lemma C.1 (Clipped-entry sensitivity and private selection)**.** *For every neighboring pair $D \sim D'$ and every $e \in \mathcal{E}_d$, $|S_e(D) - S_e(D')| \le \Delta$. For every fixed realization of the public activations $\{U_c\}_{c\in\mathcal{C}}$ and thresholds $\{T_c\}_{c\in\mathcal{C}}$, the set J of selected candidates is $(\varepsilon/8, \delta/32)$-differentially private.*

Proof. Suppose record a is replaced. For $e = (i, j)$, all other terms in the empirical average cancel, and hence

$$S_e(D) - S_e(D') = \frac{Y_{a,i}Y_{a,j} - Y'_{a,i}Y'_{a,j}}{n}.$$

Each clipped coordinate lies in $[-R, R]$, so each product lies in $[-R^2, R^2]$. It follows that

$$|S_e(D) - S_e(D')| \le \frac{2R^2}{n} = \Delta.$$

For $c = (e, \ell, h)$,

$$q_c(D) = \left(\frac{|S_e(D)|}{T_c} - 1\right)U_c - 2(1 - U_c).$$

For an active candidate c, the triangle inequality and $T_c \ge t_\ell$ give

$$|q_c(D) - q_c(D')| = \frac{\big||S_e(D)| - |S_e(D')|\big|}{T_c} \le \frac{|S_e(D) - S_e(D')|}{T_c} \le \frac{\Delta}{t_\ell}.$$

For an inactive candidate, both scores equal -2. Therefore the whole score vector at scale ℓ has L_∞ sensitivity at most Δ/t_ℓ.

We use the one-shot Laplace calibration invoked in (Qiao et al., 2021, Theorem 2.2):

Lemma C.2. *Assume f is a score vector of dimension M. Assume $\varepsilon_0 \le 0.2$, $\delta_0 \le 0.05$ and $M \ge 2$. Define the L_∞ sensitivity $g_f := \max_{D\sim D'} \|f(D)-f(D')\|_\infty$. Let $\{\zeta_i\}_{i=1}^M$ be i.i.d. random variable obeying $\mathrm{Lap}(b)$ with $b = \frac{8g_f\sqrt{s\log(M/\delta_0)}}{\varepsilon_0}$. Then the mechanism of selecting s indices with top scores in $\{f_i+\zeta_i\}_{i=1}^M$ is (ε_0,δ_0)-differential private.*

Selecting the largest scores is equivalent to selecting the smallest negated scores; only the unordered set of selected labels is released.

Note that $L \ge 11$, so $\varepsilon' = \frac{\varepsilon}{8L} \le \frac{1}{88} < 0.2$, $\delta' = \frac{\delta}{32L} \le \frac{1}{3520} < 0.05$ and $M_\ell \ge \frac{d(d+1)}{2} \ge 3$. For a scale with $s_\ell < M_\ell$, substituting $g_f = \Delta/t_\ell$, $s = s_\ell$, and $M = M_\ell$ yields the prescribed noise scale

$$b_\ell = \frac{8(\Delta/t_\ell)\sqrt{s_\ell \log(M_\ell/\delta')}}{\varepsilon'}.$$

Thus this scale is (ε',δ')-differentially private. A scale with $s_\ell = M_\ell$ returns all labels independently of the data and has zero privacy cost for the selection. Basic composition over at most L scales gives $(L\varepsilon', L\delta') = (\varepsilon/8, \delta/32)$. The proof is finished. □

Lemma C.3. *For every fixed label collection $j = \{j_\ell\}_{\ell=0}^{L-1}$ with $|j_\ell| = s_\ell$, every fixed realization of the public activations and thresholds, and every neighboring pair $D \sim D'$, the query F_j satisfies*

$$\|F_j(D) - F_j(D')\|_2 \le \Delta\sqrt{A}.$$

Proof. Using Lemma C.1 and the fact that $U_c \in \{0,1\}$ and $T_c \ge t_\ell$ for any c, we obtain

$$\|F_j(D) - F_j(D')\|_2^2 = \sum_{\ell=0}^{L-1} \sum_{c=(e,\ell,h)\in j_\ell} \frac{U_c^2\big(S_e(D) - S_e(D')\big)^2}{T_c^2} \le \sum_{\ell=0}^{L-1}\sum_{c\in j_\ell} \frac{\Delta^2}{t_\ell^2}$$
$$= \Delta^2 \sum_{\ell=0}^{L-1} \frac{s_\ell}{t_\ell^2} = \Delta^2 A.$$

The proof is finished. □

Lemma C.4. *For every fixed j and every fixed realization of the public activations and thresholds, let P_D^j and Q_D^j be the Gaussian and clipped-noise output laws in Section 4.2. Then*

$$\mathrm{TV}(P_D^j, Q_D^j) \le \eta \qquad \textit{for every } D.$$

Moreover, letting $Z_c \sim \nu_r$ for each $c \in \mathcal{C}$, the clipped-noise query mechanism $D \mapsto F_j(D) + Z$ is $(\varepsilon/4, \delta/8)$-differentially private.

Proof. By the standard Gaussian-mechanism calibration result Lemma C.7 (see also (Dwork & Roth, 2014)), we learn that for every measurable output event F and every neighboring pair $D \sim D'$,

$$P_D^j(F) \le e^{\varepsilon/4} P_{D'}^j(F) + \frac{\delta}{16}. \tag{26}$$

Recall that $S_\star = \sum_{\ell=0}^{L-1} s_\ell$. Generate one vector $G \sim \mathcal{N}(0, r^2 I_{S_\star})$ and set $Z = \mathrm{Clip}_{1/4}(G)$ coordinatewise. Define

$$A_G = \left\{ \max_{a\le S_\star} |G_a| \le \frac{1}{4} \right\}.$$

On A_G, $Z = G$, and hence $F_j(D) + Z = F_j(D) + G$ for every fixed D. For any measurable F,

$$|Q_D^j(F) - P_D^j(F)| \le \mathbb{E}\,|\mathbf{1}\{F_j(D) + Z \in F\} - \mathbf{1}\{F_j(D) + G \in F\}| \le \Pr(A_G^c).$$

Taking the supremum over F, then applying a union bound and the Gaussian tail bound, gives

$$\mathrm{TV}(P_D^j, Q_D^j) \le \Pr(A_G^c) \le \sum_{a=1}^{S_\star} 2\exp\left(-\frac{(1/4)^2}{2r^2}\right) = 2S_\star \exp\left(-\frac{1}{32r^2}\right) = \eta. \tag{27}$$

Using equation 26 and the total-variation bound on both neighboring datasets, we obtain

$$Q_D^j(F) \le P_D^j(F) + \eta \le e^{\varepsilon/4} P_{D'}^j(F) + \frac{\delta}{16} + \eta \le e^{\varepsilon/4} Q_{D'}^j(F) + \frac{\delta}{16} + (1 + e^{\varepsilon/4})\eta.$$

Noting that $(1 + e^{\varepsilon/4})\eta \le \delta/16$, the last two additive terms are at most $\delta/8$, which proves the asserted $(\varepsilon/4, \delta/8)$ guarantee. □

Lemma C.5 (Privacy of the final output)**.** *The auxiliary transcript (J, V) and the released matrix are $(3\varepsilon/8, 5\delta/32)$-differentially private for every fixed public seed. Therefore, the mechanism in Algorithm 1 is (ε, δ)-differentially private for every input dataset.*

Proof. For every fixed public seed, Lemma C.1 protects the first-stage transcript J. For every possible value j of that transcript, Lemma C.4 protects the second-stage query with the same privacy parameters. Thus, after selection returns j, the second-stage conditional mechanism is precisely $F_j(D) + Z$, with independent coordinates of Z having law ν_r. The uniform statement for all j, including values not realized on a particular dataset, is the condition required by adaptive composition.

The adaptive basic composition theorem adds the two budgets Dwork & Roth (2014), giving

$$\left(\frac{\varepsilon}{8} + \frac{\varepsilon}{4}, \frac{\delta}{32} + \frac{\delta}{8}\right) = \left(\frac{3\varepsilon}{8}, \frac{5\delta}{32}\right)$$

for $(J, F_J(D) + Z)$.

For $V = F_J(D) + Z$, the reconstruction formula equation 23 uses only the transcript and public quantities. For $c = (e, \ell, h)$, when $U_c = 1$, the coordinate is $V_c = S_e(D)/T_c + Z_c$, exactly as required by the algorithm. When $U_c = 0$, it is $V_c = Z_c \in [-1/4, 1/4]$, so the gate is zero; these dummy noise coordinates can be added for the proof and omitted in the implementation. Thus the output is a data-independent postprocessing of the auxiliary transcript and inherits its privacy guarantee.

To remove the conditioning on the public seed, let F be any event in the joint space of public seeds and released matrices, and let $F_{U,T}$ denote its section at a fixed seed. The conditional privacy bound is uniform, so

$$\begin{aligned}
\Pr_D\big((U, T, \widehat{\Sigma}) \in F\big) &= \mathbb{E}_{U,T}\Big[\Pr_D\big(\widehat{\Sigma} \in F_{U,T} \mid U, T\big)\Big] \\
&\le \mathbb{E}_{U,T}\left[e^{3\varepsilon/8} \Pr_{D'}\big(\widehat{\Sigma} \in F_{U,T} \mid U, T\big) + \frac{5\delta}{32}\right] \\
&= e^{3\varepsilon/8} \Pr_{D'}\big((U, T, \widehat{\Sigma}) \in F\big) + \frac{5\delta}{32}.
\end{aligned}$$

Therefore, the privacy guarantee holds even when all public candidate randomness is revealed. The proof is finished since $3\varepsilon/8 \le \varepsilon$ and $5\delta/32 \le \delta$, □

Lemma C.6. *Recall $M = \sum_{\ell=0}^{L-1} M_\ell$ and $\Gamma = \log(80LM/(\beta\delta))$. There is a universal constant C such that*

$$n \ge C\frac{k\sqrt{d}}{\alpha\varepsilon} \log\frac{80nd}{\beta}\, L^{3/2}\sqrt{H}\,\Gamma^{3/2} \tag{28}$$

implies $r \le 1/(8\sqrt{\Gamma})$ and $(1 + e^{\varepsilon/4})\eta \le \frac{\delta}{16}$.

Proof. Recall $D_\ell = d\rho \min\{kt_\ell^2, \sigma^4\}$. Since $H \ge 1$, the ceiling and cap in the quota imply

$$s_\ell \le \lceil 64(D_\ell + H) \rceil \le 65(D_\ell + H) \le 65(d\rho k t_\ell^2 + H).$$

Also, the geometric scale sequence gives

$$\sum_{\ell=0}^{L-1} \frac{1}{t_\ell^2} = \frac{1}{t_0^2} \sum_{\ell=0}^{L-1} 4^{-\ell} \le \frac{4}{3t_0^2}.$$

Consequently,

$$A \le 65 d\rho k L + \frac{260H}{3t_0^2} \le C\left(d\rho k L + \frac{H}{t_0^2}\right) \le \frac{Cdk^2L^2H}{\alpha^2\sigma^4}. \tag{29}$$

The last substitution uses $\rho = C_\rho LkH/(\alpha^2\sigma^4)$ and $t_0 = \alpha\sigma^2/(256k)$.

Substituting equation 29 and $\Delta = 2R^2/n$ into the prescribed Gaussian scale gives

$$r \le C\frac{k\sqrt{d}}{n\alpha\varepsilon}\frac{R^2}{\sigma^2} L\sqrt{H}\sqrt{\log(32/\delta)}.$$

Since

$$\frac{R^2}{\sigma^2} = 2\log(40nd/\beta) \le 2\log(80nd/\beta),$$

it suffices for $r \le 1/(8\sqrt{\Gamma})$ that

$$n \ge C\frac{k\sqrt{d}}{\alpha\varepsilon}\log\frac{80nd}{\beta}\, L\sqrt{H}\sqrt{\log(32/\delta)}\sqrt{\Gamma}.$$

Now $\log(32/\delta) \le \Gamma$ and $L, \Gamma \ge 1$, so $\sqrt{\log(32/\delta)}\sqrt{\Gamma} \le \Gamma \le \sqrt{L}\,\Gamma^{3/2}$.

Finally, $r \le 1/(8\sqrt{\Gamma})$ and $S_\star \le M$ imply

$$\eta = 2S_\star e^{-1/(32r^2)} \le 2Me^{-2\Gamma} = \frac{\beta^2\delta^2}{3200L^2M}.$$

Therefore,

$$\frac{16(1+e^{\varepsilon/4})\eta}{\delta} \le \frac{\beta^2\delta(1+e^{\varepsilon/4})}{200L^2M} < 1$$

under the stated parameter ranges. □

Lemma C.7 (Gaussian protection for fixed selected labels)**.** *Fix the public activations and thresholds. For every fixed label collection j, the mechanism with output law*

$$P_D^j := \mathcal{L}\big(F_j(D) + G\big), \qquad G \sim \mathcal{N}(0, r^2 I_{S_\star}),$$

is $(\varepsilon/4, \delta/16)$-differentially private, where we recall that $r = \frac{8\Delta\sqrt{A}}{\varepsilon}\sqrt{\log\frac{32}{\delta}}$.

Proof. By Lemma C.3, F_j has L_2 sensitivity at most $s_2 = \Delta\sqrt{A}$. A sufficient Gaussian-mechanism calibration (see Dwork & Roth (2014)) for $(\varepsilon_0, \delta_0)$-differential privacy is

$$r \ge \frac{2s_2}{\varepsilon_0}\sqrt{\log\frac{2}{\delta_0}}, \qquad 0 < \varepsilon_0 \le 1.$$

Taking $\varepsilon_0 = \varepsilon/4$ and $\delta_0 = \delta/16$, the required noise scale is

$$\frac{2\Delta\sqrt{A}}{\varepsilon/4}\sqrt{\log\frac{2}{\delta/16}} = \frac{8\Delta\sqrt{A}}{\varepsilon}\sqrt{\log\frac{32}{\delta}} = r.$$

The proof is finished. □