On the Curse of Dimensionality in Private Sparse Covariance Estimation and PCA

Syamantak Kumar* Shourya Pandey† Purnamrita Sarkar‡ Kevin Tian§

Abstract

We study high-dimensional *differentially private* (DP) covariance estimation in the operator norm, and principal component analysis (PCA), under *k-row-column sparsity* (k-RCS) of the covariance matrix. In the non-private setting, it is known that $\mathrm{poly}(k, \log d)$ samples suffice to solve both of these problems. However, the only comparable result known under DP [WX21] requires $\Omega(d)$ samples under standard parameterizations of the problem. We investigate when this curse of dimensionality is inherent for sparse covariance estimation tasks under DP.

On the upper bound front, we show that a $\mathrm{poly}(k, \log d)$ sample complexity for PCA is possible under DP, if we also posit sparsity of the leading eigenvector. We complement this result with $\mathrm{poly}(d)$ lower bounds under DP for both sparse covariance estimation and PCA, establishing an *exponential gap* between the private and non-private variants of these problems when $k = \mathrm{polylog}(d)$. To our knowledge, no such separation has previously been demonstrated for any sparse estimation problems in private high-dimensional statistics. Our techniques are flexible enough that they imply stronger lower bounds even for the well-studied problem of standard DP PCA, without sparsity assumptions.

*University of Texas at Austin, syamantak@utexas.edu
†University of Texas at Austin, shouryap@utexas.edu
‡University of Texas at Austin, purna.sarkar@austin.utexas.edu
§University of Texas at Austin, kjtian@cs.utexas.edu

Contents

1 Introduction **1**

1.1 Our results 1

1.2 Related work 4

2 Preliminaries **5**

2.1 Notation 5

2.2 Technical preliminaries 7

3 Private Sparse Covariance Estimation **8**

3.1 Upper bound 9

3.2 Lower bound 12

4 Private Sparse PCA: Algorithms **18**

4.1 Privacy analysis 21

4.2 Utility analysis 23

5 Private Sparse PCA: Lower Bounds **27**

5.1 Pure DP lower bound 27

5.2 Approximate DP lower bound 31

A Utility Results **42**

B Deferred Proofs from Section 3 **44**

C Private Eigenvalue Estimation for RCS Matrices **47**

D Deferred Proofs from Section 5 **49**

E Approximate DP Lower Bound for Standard PCA **50**

F Exponential-mechanism for Sparse PCA **58**

1 Introduction

We study covariance estimation and principal component analysis (PCA) in regimes where the ambient dimension d is comparable to or much larger than the sample size n.[1] In this setting, classical estimators such as the sample covariance matrix and standard PCA are no longer reliable and can be provably inconsistent [JL09, BBAP05, Pau07]. Thus in high dimensions, meaningful inference requires additional structural assumptions on the covariance matrix, such as row sparsity, sparsity of the principal eigenspace, or a combination thereof [BL09, CZZ10]. These assumptions have motivated a rich literature on sparse high-dimensional covariance estimation and sparse PCA, yielding procedures with strong statistical estimation guarantees in settings where classical methods break down [JL09, VL13, BR13, WBS16, AW08, Ma13, DM16, KS24, QLR23].

At the same time, many applications involving sensitive data require rigorous differential privacy (DP) guarantees. Compared to the vast literature on differentially private PCA [CSS13, DTTZ14, LKJO22], there have been surprisingly few works that have investigated private *sparse* PCA. Designing differentially private algorithms for high-dimensional covariance estimation and sparse PCA is particularly challenging, as these tasks inherently induce large sensitivities: when few samples are taken, each contributes more heavily to empirical statistics.

Several recent works [GWWL18, WX21, LW23] have studied differentially private sparse PCA and covariance estimation, under a non-standard parameterization that each data point is uniformly bounded in ℓ_2 norm. This condition enables worst-case sensitivity control and leads naturally to mechanisms that add noise to the empirical covariance matrix, followed by truncation-based post-processing. However, such boundedness assumptions can be overly restrictive in high dimensions. For instance, if $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, then $\|\mathbf{x}_i\|_2$ concentrates on the order of $\sqrt{d}$, so enforcing a unit-norm bound effectively obscures the dependence on d. Under a natural scale-invariant guarantee that factors in the sub-Gaussian parameter (see Model 1), the best prior result by [WX21] has a sample complexity of $\Omega(dk^2)$,[2] substantially larger than the best $\mathrm{poly}(k, \log(d))$ sample complexities for solving the same task without privacy.

This state of affairs suggest that there may be a *curse of dimensionality* that is specific to requiring DP in sparse covariance estimation tasks. Our motivation is precisely to understand whether this gap is inherent under DP, and whether there are additional natural structural assumptions that, when imposed, alleviate the curse of dimensionality. Our main questions are thus as follows.

Can we prove $\Omega(\mathrm{poly}(d))$ *lower bounds for sparse covariance estimation tasks under DP?*
Conversely, can we achieve $\mathrm{poly}(k, \log(d))$ *sample complexities under additional structure?*

1.1 Our results

We develop a suite of new DP algorithms and lower bounds for *sparse covariance estimation* and *sparse PCA*, two canonical problems in high-dimensional statistics, under the following models. We refer the reader to Section 2.2 for preliminaries on differential privacy and matrix concentration.

[1] Throughout, PCA refers to 1-PCA, i.e., the problem of recovering the leading (rank-1) principal component.

[2] Under the scaling discussed near Eq. (2) of [WX21], their distribution is $\sigma^2 = O(\frac{1}{d})$-sub-Gaussian, so achieving error as in Model 1 inflates their sample complexity as stated in their Theorem 2 by a $\frac{1}{\sigma^2} = \Omega(d)$ factor.

Table 1: DP Sparse Covariance Estimation (Problem 1). Logarithmic factors omitted, bounds stated for $\epsilon = \alpha = \beta = \Theta(1)$. All results hold under approximate ((ϵ, δ)) DP.

	Non-Private	Our Results	Prior Results
Upper Bound	k^2 [BL09]	$k^2 + \sqrt{d} \cdot k^{1.5}$ (Thm. 1)	$d \cdot k^2$ [WX21]
Lower Bound	k^2 [CZ12]	$k^2 + \sqrt{d} \cdot k$ (Thm. 2)	None

First, our general Model 1 requires a k-RCS (i.e., with k-sparse rows and columns, see Definition 1) covariance structure without any particular assumptions on the top eigenvector.

Model 1 (k-sparse covariance model)**.** *Fix $(k, d, n) \in \mathbb{N}^3$ with $k \in [d]$, $\gamma \in [0, \frac{1}{2})$, and $\sigma > 0$. In the k-sparse covariance model, there is an unknown k-RCS covariance matrix $\mathbf{\Sigma} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$, with a leading eigenvector $\mathbf{v}_1 \in \mathbb{R}^d$, and eigenvalues satisfying $\lambda_2(\mathbf{\Sigma}) \leq (1 - \gamma)\lambda_1(\mathbf{\Sigma})$. We obtain samples $\{\mathbf{x}_i\}_{i \in [n]} \overset{\text{iid}}{\sim} \mathcal{D}$, a σ-sub-Gaussian distribution with covariance $\mathbf{\Sigma}$.*

Note that under Model 1, the top eigenvector $\mathbf{v}_1$ is well-defined iff the gap parameter satisfies $\gamma > 0$. Our next model, Model 2, additionally enforces sparsity of the top eigenvector $\mathbf{v}_1$.

Model 2 (k-sparse PCA model)**.** *Instantiate Model 1 where $\gamma > 0$, and let $\text{nnz}(\mathbf{v}_1) \leq k$.*

Model 2 is a natural form of additional structure to impose under a k-RCS covariance assumption: indeed, many existing works on sparse PCA phrase the problem with this additional requirement. Given samples from Model 1 or Model 2, we study two different estimation problems regarding $\mathbf{\Sigma}$.

Problem 1 (Sparse covariance estimation)**.** *Let $(\epsilon, \delta, \alpha, \beta) \in (0, 1)^4$. Given σ-sub-Gaussian samples $\{\mathbf{x}_i\}_{i \in [n]}$, the goal is to return an (ϵ, δ)-DP matrix $\widehat{\mathbf{\Sigma}} \in \mathbb{S}^{d \times d}$ satisfying, with probability at least $1 - \beta$,*

$$\left\| \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\|_{\text{op}} \leq \alpha \sigma^2.$$

We note that the normalization by σ^2 in Problem 1 is to maintain scale-invariance of our bounds.

Problem 2 (Sparse PCA)**.** *Let $(\epsilon, \delta, \Delta, \beta) \in (0, 1)^4$. Given samples $\{\mathbf{x}_i\}_{i \in [n]}$, the goal is to return an (ϵ, δ)-DP unit vector $\widehat{\mathbf{v}} \in \mathbb{R}^d$ satisfying, with probability at least $1 - \beta$,*

$$\sin^2 \angle(\widehat{\mathbf{v}}, \mathbf{v}_1) \leq \Delta.$$

Private sparse covariance estimation. In Appendix B, we study Problem 1, where we give new upper and lower bounds that agree in their dependence on d. We summarize our results in Table 1. Note that the assumption in Model 2 does not affect the problem statement, and all of our bounds apply to the more general Model 1, so we do not differentiate the model for these results.

Our upper bound (Theorem 1) uses a simple thresholding algorithm, a standard strategy for Problem 1 in the non-private setting. Intuitively, the $\sqrt{d}$ factor appears from advanced composition, because we need to privately perform a top-k selection step on each of d rows. Interestingly, our lower bound (Theorem 2) shows that this $\sqrt{d}$ dependence is tight, even under approximate DP. Our

Table 2: DP Sparse PCA (Problem 2), logarithmic factors omitted, bounds stated for $\epsilon = \Delta = \beta = \gamma = \Theta(1)$. Upper bounds hold under (ϵ, δ) DP; lower bounds hold under $(\epsilon, 0)$ DP.

	Non-Private	**Our Results**	**Prior Results**
Upper Bound (Model 2)	k^2 [BL09][3]	k^4 (Thm. 3)	$d \cdot k^2$ [WX21]
Lower Bound (Model 1)	k [VL13]	d (Thm. 4)	None

bound is proven by adapting the fingerprinting techniques of [Nar24] for DP covariance estimation, and extending them to sparse models via a graph-based construction (Model 3). Our results again establish that for small k, there is an exponential separation between the private and non-private variants of Problem 1, highlighting a curse of dimensionality specific to DP.

Private sparse PCA. In Sections 4 and 5, we respectively give new upper and lower bounds for Problem 2, both with and without the eigenvector sparsity assumption in Model 2. Our results summarized in Table 2.

Our main upper bound (Theorem 3) demonstrates that under Model 2, $\mathrm{poly}(k, \log d)$ samples suffices to solve Problem 2 subject to approximate DP (omitting dependences on other parameters). Our algorithm uses the FriendlyCore primitive of [TCK+22] to construct a covariance estimate with d-independent sensitivity in the Frobenius norm. This allows us to add bounded noise entrywise via the Gaussian mechanism to our stable estimate, which negligibly affects the truncation typical in sparse covariance algorithms. By further leveraging the sparsity assumption in Model 2, we estimate the top eigenvector via a private support estimation step, concluding our result.

In light of our Theorem 3, a natural question is to ask whether a similar $\mathrm{poly}(k, \log d)$ sample complexity is attainable without the extra sparsity assumption in Model 2. Indeed, in the non-private setting, no sparsity assumption on $\mathbf{v}_1$ is needed at all, beyond arising from a k-RCS covariance, for $\approx k^2$ samples to solve Problem 2 (say, when $\gamma = \Theta(1)$). Our next result, Theorem 4, dashes these hopes at least with regards to pure DP ($\delta = 0$), by showing that under this restriction, $\gtrsim d$ samples are necessary. Our lower bound is based on a packing argument (Lemma 6.2, [KSU20]) and a coding-theoretic construction of $\exp(\Omega(d))$ covariance matrices that are simultaneously k-RCS and have dense leading eigenvectors. For small values of k (e.g., $k = \mathrm{polylog}(d)$), our lower bound shows an *exponential gap* between the sample complexity of the non-private and (pure) DP versions of this problem.

Previously, exponential separations have been established for several other problems in private statistics, e.g., ℓ_∞ mean estimation and hypothesis selection [BUV14, SU17].[4] However, Theorem 4 is the first such exponential separation for a natural *sparse estimation task in high dimensions*. Our result thus has an important qualitative message: indeed, arguably the central question in sparse estimation is whether a $\mathrm{poly}(d)$ sample complexity is avoidable (when, say, $k = O(\mathrm{polylog}(d))$. Our Theorems 3 and 4 demonstrate that the modeling assumptions can provably shift the sample complexity landscape for DP problems in sparse covariance estimation.

[3] This result holds even under Model 1.

[4] We also mention conceptually-related results by [KLN+11, CU21, NY22], which showed similar exponential gaps under variants of DP, particularly, the local or shuffle DP models.

One notable caveat is that the privacy guarantees in Theorems 3 and 4 do not precisely match, in the sense that our upper bound holds under approximate DP, whereas our lower bound is for pure DP. This potentially creates an opportunity for approximate DP algorithms to solve the general variant of Problem 2 (i.e., under Model 1) using only $\mathrm{poly}(k, \log d)$ samples.

Towards investigating this possibility, in Theorem 5 we give another lower bound, this time under approximate DP. Our bound applies to PCA in a family of k-RCS covariance matrices $\mathbf{\Sigma}$, such that the resulting distribution yields samples with norms $\approx \sqrt{d}$ times $\|\mathbf{\Sigma}\|_{\mathrm{op}}$. We demonstrate that $\gtrsim \frac{d}{\epsilon}$ samples are required to solve the problem even under approximate DP, by adapting the private Assouad's method of [ASZ21]. Unfortunately, the resulting sample distribution is $O(\sqrt{d})$-sub-Gaussian in the sense of Model 1, as enforcing sparsity induces certain spiky directions. Thus, this parameterization does not match our upper bound in Theorem 3. Nonetheless, our results broaden our understanding on the achievability of $\mathrm{poly}(k, \log d)$ sample complexities under different models. We leave proving a $\mathrm{poly}(d)$ lower bound, or excitingly, a $\mathrm{poly}(k, \log d)$ upper bound, under the approximate DP variant of Model 1 as an interesting open problem.

En route to proving our results for Problem 2, we also give a stronger lower bound for DP PCA (in the standard, non-sparse, setting) than prior works. Specifically, Theorem 7 demonstrates that $\gtrsim d$ samples are needed under approximate DP, for a much broader parameter range than previously known: e.g., Section 4.2, [CXZ24] and Theorem 5.4, [LKJO22] only result in comparable bounds in the restrictive setting $\delta = \exp(-\Omega(d))$. We believe this result is of independent interest, as it improves our understanding of the tractability of an extremely well-studied problem.

1.2 Related work

High-dimensional covariance estimation and sparse PCA (non-private). Classical PCA can be statistically inconsistent in modern high-dimensional regimes, motivating structural assumptions such as sparsity of the covariance or of leading eigenvectors. For sparse covariance estimation, a large line of work studies thresholding and related regularization procedures that achieve dimension-free (or near dimension-free) rates under suitable sparsity/regularity conditions, including early thresholding estimators and their refinements [BL09, RLZ09, CL11, DM16]. In parallel, sparse PCA has been extensively studied via optimization-based formulations and algorithmic relaxations, including ℓ_1-penalized or regression-style approaches [ZHT06], semidefinite relaxations [dGJL04, AW08], and iterative schemes such as truncated power methods [YZ13]. A complementary thread establishes statistical limits and computational barriers in sparse PCA, clarifying when polynomial-time methods can (or cannot) attain minimax-optimal rates [BR13, WBS16].

Differential privacy for subspace estimation when $d \leq n$. Differential privacy (DP) provides a rigorous framework for protecting individuals' contributions in statistical analyses [DR14]. A core challenge in private high-dimensional problems is to control sensitivity while preserving spectral structure. For private PCA and related spectral tasks, foundational results include tight privacy-utility analyses for PCA via Gaussian perturbations [DTTZ14, CXZ24] and private iterative methods for dominant subspaces [HP14, LKJO22, DS25]. More broadly, private low-rank approximation and private linear-algebraic primitives have been developed as building blocks for downstream tasks [KT13]. On the distribution-learning side, general-purpose private learners for

high-dimensional structured families illuminate what is achievable when the ambient dimension is large, and boundedness assumptions are undesirable [KLSU19].

Private sparse/structured estimation and selection primitives. The intersection of privacy with sparsity brings additional algorithmic and information-theoretic constraints: even identifying the relevant support (or approximate support) can dominate the privacy budget. For sparse covariance estimation under DP, prior work gives algorithms and rates under high-dimensional sparsity assumptions [GWWL18, WX21, LW23] with $O(1)$ norm bounded assumptions, which may be difficult to satisfy with high-dimensional data. At the level of primitives, differentially private top-k and sparse selection mechanisms—often used to locate large coordinates/entries before estimating magnitudes—have been studied extensively [DR19, QSZ21]. These tools are particularly relevant for sparse PCA pipelines that must privately localize the support of a sparse leading eigenvector (or its projector) prior to accurate recovery. Our results fit into this gap by giving end-to-end private procedures tailored to k-RCS structure (for covariance estimation) and sparse leading components (for PCA), combining structured truncation/thresholding with carefully calibrated noise so that the final guarantees scale primarily with the sparsity level rather than the ambient dimension.

Lower bounds: geometry, private minimax tools, and fingerprinting. On the impossibility side, DP lower bounds for high-dimensional estimation draw on geometric characterizations and packing arguments [HT10], as well as DP analogues of classical minimax techniques (Assouad/Fano/Le Cam) [ASZ21]. Fingerprinting codes and their descendants provide sharp lower bounds for answering many queries and for private statistical estimation, highlighting fundamental gaps between non-private and approximate-DP sample complexity [HT10, BUV14, SU15]. Recent work streamlines and strengthens the fingerprinting approach for modern estimation problems [Nar24]. Our lower bounds build on and adapt these techniques to the covariance/PCA setting under the structural regimes considered here, yielding near-matching (up to logarithmic factors) separations that explain when sparsity-aware private procedures are necessary and when they are sufficient.

2 Preliminaries

2.1 Notation

We use $\lesssim$, $\eqsim$, and $\gtrsim$ respectively as shorthand for $O(\cdot)$, $\Theta(\cdot)$, $\Omega(\cdot)$, i.e., to hide universal constants. For $n \in \mathbb{N}$ we let $[n] := \{i \in \mathbb{N} \mid i \leq n\}$. For $a, b \in \mathbb{N}$, $a|b$ denotes a is divisible by b. We denote vectors in lowercase boldface letters and matrices in capital boldface letters. We denote the i^{th} canonical basis vector in $\mathbb{R}^d$ by $\mathbf{e}_i$. For $p \in [1, \infty]$ we let $\|\mathbf{v}\|_p$ to denote the ℓ_p norm of $\mathbf{v}$. We use nnz to denote the number of nonzero entries in a vector or matrix, and supp to denote the support (i.e., indices of the nonzero entries). We use $\mathbf{0}_d$ to denote the all-zeroes vector and $\mathbf{1}_d$ to denote the all-ones vector in $\mathbb{R}^d$, and for an event $\mathcal{E}$, we use $\mathbb{1}(\mathcal{E})$ to denote the corresponding 0-1 indicator random variable. For random variables $\mathbf{x}, \mathbf{y}$, we denote statistical independence by $\mathbf{x} \perp\!\!\!\perp \mathbf{y}$.

For $\mathbf{M} \in \mathbb{R}^{m \times n}$ and $S \subseteq [m]$, $T \subseteq [n]$, we use $\mathbf{M}_{S \times T}$ to denote the submatrix indexed by S, T. For matrices $\mathbf{A}, \mathbf{B}$ with the same number of rows, we let $\begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix}$ denote their horizontal concatenation.

We use $\mathbf{A}_{i,:}, \mathbf{A}_{:,i}$ denote the i^{th} row and column of matrix $\mathbf{A}$. We let $\mathbf{I}_d$ be the $d \times d$ identity and $\mathbf{0}_{m\times n}$ be the all-zeroes $m \times n$ matrix and let $\mathbb{1}(\mathcal{E})$ denote the indicator random variable corresponding to the event $\mathcal{E}$. We let $\mathbb{S}^{d\times d}$ be the set of real symmetric $d \times d$ matrices, which we equip with the Loewner partial ordering $\preceq$ and the Frobenius inner product $\langle \mathbf{M}, \mathbf{N}\rangle = \mathrm{Tr}(\mathbf{MN})$. We let $\mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ and $\mathbb{S}^{d\times d}_{\succ \mathbf{0}}$ respectively denote the positive semidefinite and positive definite subsets of $\mathbb{S}^{d\times d}$.

For matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$, we define $\|\mathbf{M}\|_{\text{op}} := \sqrt{\lambda_1(\mathbf{MM}^\top)}, \|\mathbf{M}\|_{\text{F}} := \sqrt{\sum_{i,j\in[d]} \mathbf{M}_{ij}^2}, \|\mathbf{M}\|_{\infty,\infty} := \sup_{i,j\in[d]} |\mathbf{M}_{ij}|$. Denote by $\mathbb{B}_\infty(\mathbf{M}, \tau)$ the set of matrices $\mathbf{M}'$ satisfying $\|\mathbf{M} - \mathbf{M}'\|_{\infty,\infty} \le \tau$.

Our models consider estimation of covariance matrices satisfying the following definition.

Definition 1 (k-RCS)**.** *Let $k \in [\min(m, n)]$. We say that a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ is k-RCS (k-row-column sparse) if for all $i \in [m]$, $\mathrm{nnz}(\mathbf{M}_{i,:}) \le k$ and for all $j \in [n]$, $\mathrm{nnz}(\mathbf{M}_{:,j}) \le k$.*

We also denote the bounded Laplace distribution with parameters $\lambda, \tau \ge 0$ by $\mathsf{BoundedLaplace}(\lambda, \tau)$, which is the distribution of $X \sim \mathsf{Lap}(\lambda)$ conditioned on $|X| \le \tau$.

We finally provide notation for procedures often used in the paper. For a vector $\mathbf{v} \in \mathbb{R}^d$ and $k \in [d]$, we use $\mathrm{top}_k(\mathbf{v})$ to denote the vector in $\mathbb{R}^d$ that zeroes out all but the top-k entries of $\mathbf{v}$ by magnitude (breaking ties arbitrarily). For a vector or matrix argument, and a threshold $\tau > 0$, we use $\mathcal{T}_\tau(\cdot)$ to be the vector or matrix that applies the following thresholding operation entrywise:

$$\mathcal{T}_\tau(c) := \begin{cases} c & |c| \ge \tau \\ 0 & \text{else} \end{cases}. \tag{1}$$

For two unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we define the $\sin^2$ error between them as

$$\sin^2 \angle(\mathbf{u}, \mathbf{v}) := 1 - \frac{\langle \mathbf{u}, \mathbf{v}\rangle^2}{\|\mathbf{u}\|_2^2 \|\mathbf{v}\|_2^2}. \tag{2}$$

Differential privacy. Let $\mathcal{X}$ be some domain, and let $\mathcal{D} \in \mathcal{X}^n$ be a dataset consisting of n elements. We say that two datasets $\mathcal{D}$, $\mathcal{D}' \in \mathcal{X}^n$ are *neighboring* if their symmetric difference has size 1, i.e., they differ in a single element. We use the following definition of differential privacy.

Definition 2 (Differential privacy)**.** *Let $(\epsilon, \delta) \in [0, 1]^2$.*[5] *We say that a randomized algorithm $\mathcal{A} : \mathcal{X}^n \to \Omega$ satisfies (ϵ, δ)-*differential privacy *(or, is (ϵ, δ)-*DP*) if for all events $\mathcal{E} \subseteq \Omega$, and for all neighboring datasets $\mathcal{D}, \mathcal{D}' \in \mathcal{X}^n$, we have*

$$\Pr[\mathcal{A}(\mathcal{D}) \in \mathcal{E}] \le \exp(\epsilon) \Pr\left[\mathcal{A}(\mathcal{D}') \in \mathcal{E}\right] + \delta.$$

Differentially private algorithms obey *basic composition* (Theorem B.1, [DR14]): if $\mathcal{A}_1 : \mathcal{X}^n \to \Omega_1$ is (ϵ_1, δ_1)-DP and $\mathcal{A}_2 : \mathcal{X}^n \times \Omega_1 \to \Omega_2$ is (ϵ_2, δ_2)-DP, then the procedure that runs $\mathcal{A}_1$ on $\mathcal{D}$ and subsequently runs $\mathcal{A}_2$ on $(\mathcal{D}, \mathcal{A}_1(\mathcal{D}))$ is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP. We will also invoke the following advanced composition guarantee for DP algorithms.

[5] In principle, the privacy parameter ϵ can be larger than 1. However, for any $\epsilon \ge 1$, our sample-complexity bounds are unaffected up to constant factors if we instead guarantee $(1, \delta)$-DP (which is weaker than (ϵ, δ)-DP). Therefore, for convenience and to simplify several bounds, we state all results for $\epsilon \in [0, 1]$.

Lemma 1 (Advanced composition, Theorem 3.20 [DR14]). *Fix $k \in \mathbb{N}$ and privacy parameters $\epsilon, \delta, \delta_0 \in [0,1]$. For each $i \in [k]$, let*

$$\mathcal{A}_i : \mathcal{X}^n \times \Omega_1 \times \cdots \times \Omega_{i-1} \to \Omega_i$$

be a (possibly randomized) algorithm such that for every fixed transcript $y_{<i} \in \Omega_1 \times \cdots \times \Omega_{i-1}$, the map $\mathcal{D} \mapsto \mathcal{A}_i(\mathcal{D}; y_{<i})$ is (ϵ, δ)-DP in the sense of Definition 2. Define the adaptive k-fold composition $\mathcal{A}_{1:k}(\mathcal{D}) := (Y_1, \ldots, Y_k)$, $Y_i \sim \mathcal{A}_i(\mathcal{D}; Y_{<i})$. Then $\mathcal{A}_{1:k}$ is $(\epsilon^\star, k\delta + \delta_0)$-DP, where

$$\epsilon^\star := \sqrt{2k \log(1/\delta_0)} \cdot \epsilon \; + \; k\epsilon\,(e^\epsilon - 1).$$

In particular, if $\epsilon \in [0,1]$ then $e^\epsilon - 1 \le 2\epsilon$, and hence $\epsilon^\star \le \sqrt{2k \log(1/\delta_0)} \cdot \epsilon \; + \; 2k\epsilon^2$.

We next state the Gaussian mechanism. Recall that if $\mathbf{v} : \mathcal{X}^n \to \mathbb{R}^k$ is a vector-valued function of a dataset, we say $\mathbf{v}$ has sensitivity Δ if for all neighboring $\mathcal{D}, \mathcal{D}' \in \mathcal{X}^n$, we have $\|\mathbf{v}(\mathcal{D}) - \mathbf{v}(\mathcal{D}')\| \le \Delta$.

Fact 1 (Theorem A.1, [DR14]). *Let $\mathbf{v} : \mathcal{X}^n \to \mathbb{R}^k$ have sensitivity Δ, and let $(\epsilon, \delta) \in [0,1]^2$. Then, drawing a sample from $\mathcal{N}(\mathbf{v}(\mathcal{D}), \sigma^2 \mathbf{I}_k)$ is (ϵ, δ)-DP, for any $\sigma \ge \frac{2\Delta}{\epsilon} \cdot \sqrt{\log(\frac{2}{\delta})}$.*

We also require the bounded Laplace mechanism, which is known to give the following guarantee.

Fact 2 (Lemma 9, [ALT24]). *Let $s : \mathcal{X}^n \to \mathbb{R}$ have sensitivity Δ, and let $(\epsilon, \delta) \in [0,1]^2$. Then, drawing $\xi \sim \mathsf{BoundedLaplace}(\frac{\Delta}{\epsilon}, \tau)$ and outputting $s(\mathcal{D}) + \xi$ is (ϵ, δ)-DP for any $\tau \ge \frac{\Delta}{\epsilon} \log(\frac{4}{\delta})$.*

Fact 2 is established in [ALT24] via a coupling argument, leveraging the observation that $\mathsf{BoundedLaplace}(\lambda)$ and $\mathsf{Lap}(\lambda)$ produce identical samples except with some probability.

2.2 Technical preliminaries

We start by defining sub-Gaussianity for multivariate distributions.

Definition 3 (σ-sub-Gaussianity). *A mean-zero random vector $\mathbf{x} \in \mathbb{R}^d$ is said to be σ-sub-Gaussian for proxy $\sigma > 0$ if, for any vector $\mathbf{u} \in \mathbb{R}^d$,*

$$\mathbb{E}\left[\exp\left(\mathbf{u}^\top \mathbf{x}\right)\right] \le \exp\left(\frac{\sigma^2 \|\mathbf{u}\|_2^2}{2}\right)$$

Matrix concentration and Linear Algebra. We require the following standard facts to bound the approximation error of random sampling, under our Models 1 and 2.

Fact 3 (Lemma B.3, [ZL16]). *For some $\mu > 0$ and $\tau > 0$, let orthonormal $\mathbf{U}$ have columns spanning the eigenspace of $\mathbf{A} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ corresponding to eigenvalues $\le \mu$, and let orthonormal $\mathbf{V}$ have columns spanning the eigenspace of $\mathbf{B} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ corresponding to eigenvalues $\ge \mu + \tau$. Then,*

$$\left\|\mathbf{U}^\top \mathbf{V}\right\|_{\mathrm{op}} \le \frac{\|\mathbf{A} - \mathbf{B}\|_{\mathrm{op}}}{\tau}.$$

Fact 4 (Lemma 6.26, [Wai19]). *Let $\mathcal{D}$ be a σ-sub-Gaussian distribution with covariance $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, let $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$, and let $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$. There exists a universal constant $C > 0$ such that for all $\delta \in (0,1)$,*

$$\Pr\left[\max_{(i,j)\in[d]\times[d]} \left|\widehat{\mathbf{\Sigma}}_{ij} - \mathbf{\Sigma}_{ij}\right| \geq C\sigma^2\left(\sqrt{\frac{\log(\frac{d}{\delta})}{n}} + \frac{\log(\frac{d}{\delta})}{n}\right)\right] \leq \delta.$$

Fact 5 (Exercise 4.7.3, [Ver18]). *Let $\mathcal{D}$ be a σ-sub-Gaussian distribution with covariance $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, let $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$, and let $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$. There exists a universal constant $C > 0$ such that for all $\delta \in (0,1)$,*

$$\Pr\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\text{op}} \geq C\sigma^2\left(\sqrt{\frac{d+\log(\frac{1}{\delta})}{n}} + \frac{d+\log(\frac{1}{\delta})}{n}\right)\right] \leq \delta.$$

As a simple corollary of Fact 5, we derive a concentration bound for all submatrices.

Corollary 1. *In the setting of Fact 5, let $s \in [d]$. There exists a universal constant $C > 0$ such that for all $\delta \in (0,1)$,*

$$\Pr\left[\max_{\substack{S\subseteq[d]\\|S|\leq s}} \left\|\widehat{\mathbf{\Sigma}}_{S\times S} - \mathbf{\Sigma}_{S\times S}\right\|_{\text{op}} \geq C\sigma^2\left(\sqrt{\frac{s\log(d)+\log(\frac{1}{\delta})}{n}} + \frac{s\log(d)+\log(\frac{1}{\delta})}{n}\right)\right] \leq \delta.$$

Proof. Observe that there are $\binom{d}{s} \leq d^s$ such submatrices. Thus, it is enough to apply Fact 5 to each submatrix with failure probability set to $\delta \leftarrow \delta/d^s$ and dimension replaced by $d \leftarrow s$, and then conclude by a union bound. □

Fact 6. *Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d\times d}$ satisfying $\forall i,j \in [d]^2, 0 \leq [\mathbf{B}]_{ij} \leq [\mathbf{A}]_{ij}$. Then $\|\mathbf{B}\|_{\text{op}} \leq \|\mathbf{A}\|_{\text{op}}$.*

Fact 7 (Theorem 6.2.3, [Wai19]). *Let $\{\mathbf{x}_i\}_{i=1}^n$ be an i.i.d. sequence of zero-mean σ-sub-Gaussian random vectors with covariance matrix $\mathbf{\Sigma}$ with adjacency pattern $[\mathbf{A}]_{ij} = \mathcal{I}(\mathbf{\Sigma}_{ij} \neq 0)$. Let $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$. If $n > \log d$, then for any $\delta > 0$, the thresholded sample covariance matrix $\mathcal{T}_{\lambda_n}(\widehat{\mathbf{\Sigma}})$ with $\frac{\lambda_n}{\sigma^2} = 8\sqrt{\frac{\log d}{n}} + \delta$ satisfies*

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\infty,\infty} \leq 2\lambda_n,\ \mathcal{T}_{\lambda_n}(\widehat{\mathbf{\Sigma}}) \text{ is } k\text{-RCS, and } \left\|\mathcal{T}_{\lambda_n}(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_{\text{op}} \leq 2\|\mathbf{A}\|_{\text{op}}\lambda_n \tag{3}$$

with probability at least $1 - 8\exp\left(-\frac{n}{16}\min\{\delta,\delta^2\}\right)$.

3 Private Sparse Covariance Estimation

In this section, we describe our algorithm to obtain a differentially private covariance estimate under Model 1. Prior work [WX21] has studied this problem under the additional assumption

Algorithm 1 Row-wise one-shot Top-k private covariance estimation : $\mathsf{PrivCov}(\mathcal{D}, k, \epsilon, \delta, \sigma^2, \beta)$

Input: Dataset $\mathcal{D} = \{\mathbf{x}_t\}_{t\in[n]} \subseteq \mathbb{R}^d$, sparsity $k \in [d]$, privacy $(\epsilon, \delta) \in (0,1]^2$, failure $\beta \in (0,1)$, sub-Gaussian proxy $\sigma > 0$.

Set truncation level $R \leftarrow \sigma\sqrt{2\log(6nd/\beta)}$.
For all $t \in [n]$, $\mathbf{y}_t \leftarrow \mathsf{Trunc}_R(\mathbf{x}_t)$, i.e. $(\mathbf{y}_t)_j := \mathrm{sign}((\mathbf{x}_t)_j) \cdot \min\{|(\mathbf{x}_t)_j|, R\}$ for each $j \in [d]$.
$\widehat{\mathbf{\Sigma}} \leftarrow (1/n)\sum_{t=1}^n \mathbf{y}_t\mathbf{y}_t^\top$.
Set $\delta_0 \leftarrow \delta/2, \quad \delta_{\mathrm{row}} \leftarrow \delta/(2d), \quad \epsilon_{\mathrm{row}} \leftarrow \epsilon/4\sqrt{2d\log(1/\delta_0)}$.
Set entrywise sensitivity proxy $\Delta \leftarrow 2R^2/n$ and noise scale $b \leftarrow (2\Delta/\epsilon_{\mathrm{row}}) \cdot \sqrt{k\log(d/\delta_{\mathrm{row}})}$.
for $i = 1$ **to** d **do**
    Sample $\mathbf{z}_i^{\mathrm{sel}} \in \mathbb{R}^d$ as $(\mathbf{z}_i^{\mathrm{sel}})_j \overset{i.i.d}{\sim} \mathsf{Lap}(b)$, $\widetilde{\mathbf{\Sigma}}_{i,:}^{\mathrm{sel}} \leftarrow \widehat{\mathbf{\Sigma}}_{i,:} + (\mathbf{z}_i^{\mathrm{sel}})^\top$.
    $S_i \leftarrow \mathrm{supp}(\mathrm{top}_k(\widetilde{\mathbf{\Sigma}}_{i,:}^{\mathrm{sel}}))$.
    Sample $\mathbf{z}_i^{\mathrm{val}} \in \mathbb{R}^d$ with $(\mathbf{z}_i^{\mathrm{val}})_j \overset{i.i.d}{\sim} \mathsf{Lap}(b)$, $\widetilde{\mathbf{\Sigma}}_{i,:}^{\mathrm{val}} \leftarrow \widehat{\mathbf{\Sigma}}_{i,:} + (\mathbf{z}_i^{\mathrm{val}})^\top$.
end for
Initialize $\mathbf{M} \leftarrow \mathbf{0}_{d\times d}$.
for $1 \le i < j \le d$ **do**
    if $j \in S_i$ **and** $i \in S_j$ **then**
        $\mathbf{M}_{ij} \leftarrow \mathbf{M}_{ji} \leftarrow \frac{1}{2}(\widetilde{\mathbf{\Sigma}}_{ij}^{\mathrm{val}} + \widetilde{\mathbf{\Sigma}}_{ji}^{\mathrm{val}})$.
    end if
end for
return $\mathbf{M}$.

that datapoints $\{\mathbf{x}_i\}_{i\in[n]} \overset{\mathsf{iid}}{\sim} \mathcal{D}$ additionally satisfy the constraint $\|\mathbf{x}_i\|_2 \le 1$. Scaling their result in Theorem 2 appropriately to transfer to Model 1 yields a sample complexity of $\Omega(dk^2\log(d/\delta)/\alpha^2\epsilon^2)$ to achieve an (ϵ,δ)-DP estimate with operator norm at most α. Their algorithm first adds noise to make the sample covariance $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$ private and then analyzes the utility by exploiting the k-RCS property of the population covariance $\mathbf{\Sigma}$. However, making the entire covariance matrix private leads to a suboptimal dimension dependence.

In Section 3.1, we propose Algorithm 1 and show that it improves the dimension dependence from d to $\sqrt{d}$ in Theorem 1, and subsequently in Section 3.2 we show that this dependence is tight via a matching lower bound in Theorem 2.

3.1 Upper bound

Theorem 1 (Private k-sparse Covariance Estimation)**.** *Under Model 1, for Problem 1, there is an algorithm (Algorithm 1) which, on input $\{\mathbf{x}_i\}_{i\in[n]} \overset{\mathsf{iid}}{\sim} \mathcal{D}$, outputs an (ϵ,δ)-DP matrix $\mathbf{M} \in \mathbb{S}^{d\times d}$ such that for n satisfying,*

$$n \ge \Omega\left(\frac{k^2}{\alpha^2}\log\left(\frac{d}{\beta}\right) + \frac{k\sqrt{dk}}{\alpha\epsilon}\log\left(\frac{d}{\beta}\right)\log\left(\frac{d}{\delta}\right)\log\left(\frac{nd}{\beta}\right)\right)$$

we have, $\|\mathbf{M} - \mathbf{\Sigma}\|_{\mathrm{op}} \le \alpha\sigma^2$ with probability at least $1-\beta$.

Proof. We work with the notation in Algorithm 1. In particular, recall the definitions,

$$R := \sigma\sqrt{2\log\left(\frac{6nd}{\beta}\right)},\quad \Delta := \frac{2R^2}{n},\quad \epsilon_{\text{row}} := \frac{\epsilon}{4\sqrt{2d\log(1/\delta_0)}} \text{ and } b := \frac{2\Delta}{\epsilon_{\text{row}}}\sqrt{k\log\left(\frac{d}{\delta_{\text{row}}}\right)}.$$

For the universal constant $C > 0$ from Fact 4, we further define,

$$\lambda_{\text{samp}} := C\sigma^2\left(\sqrt{\frac{\log(6d/\beta)}{n}} + \frac{\log(6d/\beta)}{n}\right),\quad \lambda_{\text{priv}} := b\cdot\log\left(\frac{12d^2}{\beta}\right),\quad \lambda_n := \lambda_{\text{samp}} + \lambda_{\text{priv}},$$

Privacy. We first note that $\widehat{\mathbf{\Sigma}}$ has entrywise sensitivity at most $\Delta = 2R^2/n$; for adjacent datasets $\mathcal{D} \sim \mathcal{D}'$,

$$\left|\widehat{\mathbf{\Sigma}}_{ij}(\mathcal{D}) - \widehat{\mathbf{\Sigma}}_{ij}(\mathcal{D}')\right| = \frac{1}{n}\left|(\mathbf{y}_t)_i(\mathbf{y}_t)_j - (\mathbf{y}'_t)_i(\mathbf{y}'_t)_j\right| \le \frac{2R^2}{n} = \Delta.$$

Fix a row i. The row mechanism uses $\widetilde{\mathbf{\Sigma}}^{\text{sel}}_{i,:} = \widehat{\mathbf{\Sigma}}_{i,:} + (\mathbf{z}_i^{\text{sel}})^\top$ only to compute the set S_i of Top-k largest entries in magnitude, and then releases values using *fresh* independent noise $\widetilde{\mathbf{\Sigma}}^{\text{val}}_{i,:} = \widehat{\mathbf{\Sigma}}_{i,:} + (\mathbf{z}_i^{\text{val}})^\top$. By Theorem 2.2 of [QSZ21], the row release is $(\epsilon_{\text{row}}, \delta_{\text{row}})$-DP. The final output $\mathbf{M}$ is obtained by post-processing the row releases, so it remains $(\epsilon_{\text{row}}, \delta_{\text{row}})$-DP per row. Composing the d rows via Lemma 1 with parameter $\delta_0 = \delta/2$ yields overall $(\epsilon^\star, d\delta_{\text{row}} + \delta_0)$-DP with

$$\epsilon^\star \le \sqrt{2d\log(1/\delta_0)}\,\epsilon_{\text{row}} + 2d\epsilon_{\text{row}}^2 \le \epsilon, \qquad d\delta_{\text{row}} + \delta_0 = \delta.$$

Hence $\mathbf{M}$ is (ϵ, δ)-DP.

Utility. Define the events

$$\mathcal{E}_{\text{clip}} := \left\{\max_{t\in[n]}\max_{j\in[d]}|(\mathbf{x}_t)_j| \le R\right\},\ \mathcal{E}_{\text{samp}} := \left\{\max_{i,j\in[d]}\left|\widehat{\mathbf{\Sigma}}_{ij} - \mathbf{\Sigma}_{ij}\right| \le \lambda_{\text{samp}}\right\},$$
$$\mathcal{E}_{\text{noise}} := \left\{\max_{i,j\in[d]}\left|(\mathbf{z}^{\text{sel}})_{ij}\right| \le \lambda_{\text{priv}} \bigcap \max_{i,j\in[d]}\left|(\mathbf{z}^{\text{val}})_{ij}\right| \le \lambda_{\text{priv}}\right\},$$

and set $\mathcal{E} := \mathcal{E}_{\text{clip}} \cap \mathcal{E}_{\text{samp}} \cap \mathcal{E}_{\text{noise}}$, i.e $\mathcal{E}$ denotes the event that (i) no truncation occurs, (ii) the sample covariance concentrates entrywise, and (iii) both Laplace noise matrices are uniformly bounded. Let $\widehat{\mathbf{\Sigma}}^{\text{untr}} := \frac{1}{n}\sum_{t=1}^n \mathbf{x}_t\mathbf{x}_t^\top$. On $\mathcal{E}_{\text{clip}}$ we have $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{\text{untr}}$, hence on $\mathcal{E}$,

$$\max_{i,j}\left|\widehat{\mathbf{\Sigma}}_{ij} - \mathbf{\Sigma}_{ij}\right| \le \lambda_{\text{samp}},\qquad \max_{i,j}\left|\widetilde{\mathbf{\Sigma}}^{\text{sel}}_{ij} - \mathbf{\Sigma}_{ij}\right| \le \lambda_{\text{samp}}+\lambda_{\text{priv}},\qquad \max_{i,j}\left|\widetilde{\mathbf{\Sigma}}^{\text{val}}_{ij} - \mathbf{\Sigma}_{ij}\right| \le \lambda_{\text{samp}}+\lambda_{\text{priv}}.$$

Therefore for $\lambda_n := \lambda_{\text{samp}} + \lambda_{\text{priv}}$,

$$\left\{\max_{i,j}\left|\widetilde{\mathbf{\Sigma}}^{\text{sel}}_{ij} - \mathbf{\Sigma}_{ij}\right| > \lambda_n\right\} \cup \left\{\max_{i,j}\left|\widetilde{\mathbf{\Sigma}}^{\text{val}}_{ij} - \mathbf{\Sigma}_{ij}\right| > \lambda_n\right\} \subseteq \mathcal{E}^c.$$

By a union-bound, $\Pr(\mathcal{E}^c) \le \Pr(\mathcal{E}^c_{\text{clip}}) + \Pr(\mathcal{E}^c_{\text{samp}}) + \Pr(\mathcal{E}^c_{\text{noise}})$. Using Fact 4 with failure $\beta/3$, we have $\Pr(\mathcal{E}^c_{\text{samp}}) \le \frac{\beta}{3}$. Next, with $R = \sigma\sqrt{2\log\left(\frac{6nd}{\beta}\right)}$, using sub-Gaussianity of $\{\mathbf{x}_i\}_{i\in[n]}$,

$$\Pr(\mathcal{E}^c_{\text{clip}}) \le \sum_{t=1}^n\sum_{j=1}^d \Pr\left(|(\mathbf{x}_t)_j| > R\right) \le nd\cdot 2\exp\Big(-\frac{R^2}{2\sigma^2}\Big) = nd\cdot 2\exp\Big(-\log\Big(\frac{6nd}{\beta}\Big)\Big) = \frac{\beta}{3}.$$

And finally, for $\lambda_{\text{priv}} := b \log\left(\frac{12d^2}{\beta}\right)$ and $Z \sim \mathsf{Lap}(b)$,

$$\Pr(\mathcal{E}_{\text{noise}}^c) \leq 2d^2 \Pr(|Z| > \lambda_{\text{priv}}) = 2d^2 e^{-\lambda_{\text{priv}}/b} = 2d^2 \cdot \frac{\beta}{12d^2} = \frac{\beta}{6} \leq \frac{\beta}{3}.$$

Thus $\Pr(\mathcal{E}) \geq 1 - \beta$. On $\mathcal{E}$, for all $i, j \in [d]$ we have the two uniform bounds

$$\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}} - \mathbf{\Sigma}_{ij}\right| \leq \lambda_n \qquad \text{and} \qquad \left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{val}} - \mathbf{\Sigma}_{ij}\right| \leq \lambda_n. \tag{4}$$

Fix a row i and work on $\mathcal{E}$. If $\mathbf{\Sigma}_{ij} = 0$, then $\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| = \left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}} - \mathbf{\Sigma}_{ij}\right| \leq \lambda_n$. Hence any index j satisfying $\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| > \lambda_n$ must have $\mathbf{\Sigma}_{ij} \neq 0$. Since $\mathbf{\Sigma}$ is k-RCS, row i has at most k nonzero entries, so there are at most k indices with $\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| > \lambda_n$. Therefore Top-$k$ must include all such indices, i.e. $\{j \in [d] : \left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| > \lambda_n\} \subseteq S_i$. Equivalently, if $j \notin S_i$ then $\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| \leq \lambda_n$, and thus

$$|\mathbf{\Sigma}_{ij}| \leq \left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}}\right| + \left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{sel}} - \mathbf{\Sigma}_{ij}\right| \leq \lambda_n + \lambda_n = 2\lambda_n, \tag{5}$$

Fix $j \neq i$. We then divide our analysis into two cases:

(i) Mutually selected entries. If $j \in S_i$ and $i \in S_j$, then $\mathbf{M}_{ij} = \mathbf{M}_{ji} = \frac{1}{2}(\widetilde{\mathbf{\Sigma}}_{ij}^{\text{val}} + \widetilde{\mathbf{\Sigma}}_{ji}^{\text{val}})$. Using symmetry $\mathbf{\Sigma}_{ij} = \mathbf{\Sigma}_{ji}$ and (4), $|\mathbf{M}_{ij} - \mathbf{\Sigma}_{ij}| \leq \frac{1}{2}\left|\widetilde{\mathbf{\Sigma}}_{ij}^{\text{val}} - \mathbf{\Sigma}_{ij}\right| + \frac{1}{2}\left|\widetilde{\mathbf{\Sigma}}_{ji}^{\text{val}} - \mathbf{\Sigma}_{ji}\right| \leq \lambda_n$.
(ii) Not mutually selected. In this case, $\mathbf{M}_{ij} = 0$. We further have,

1. If $\mathbf{\Sigma}_{ij} = 0$, then $|\mathbf{M}_{ij} - \mathbf{\Sigma}_{ij}| = 0$
2. If $\mathbf{\Sigma}_{ij} \neq 0$, then either $j \notin S_i$ or $i \notin S_j$; by (5) and symmetry, $|\mathbf{\Sigma}_{ij}| \leq 2\lambda_n$, hence $|\mathbf{M}_{ij} - \mathbf{\Sigma}_{ij}| = |\mathbf{\Sigma}_{ij}| \leq 2\lambda_n$.

Let $\mathbf{E} := \mathbf{M} - \mathbf{\Sigma}$. Consider an adjacency matrix $\mathbf{A}$ such that $\mathbf{A}_{i,j} = \mathcal{I}\left(\{\mathbf{M}_{i,j} \neq 0\} \cup \{\mathbf{\Sigma}_{i,j} \neq 0\}\right)$. Consider the matrix $\mathbf{B} := |\mathbf{M} - \mathbf{\Sigma}|$, where the $|.|$ operator is applied entrywise. Then, by the above two cases, we have,

$$\mathbf{B}_{i,j} \leq 2\lambda_n \mathbf{A}_{i,j}$$

Then, since both $\mathbf{B}$ and $2\lambda_n\mathbf{A}$ have non-negative entries, using Fact 6, we have

$$\|\mathbf{M} - \mathbf{\Sigma}\|_{\text{op}} \leq \|\mathbf{B}\|_{\text{op}} \leq 2\lambda_n \|\mathbf{A}\|_{\text{op}}$$

Since each row of $\mathbf{A}$ has at most $2k$ non-zero entries, we have $\|\mathbf{A}\|_{\text{op}} \leq 2k$. Finally, substituting the definitions of λ_{samp} and λ_{priv} (and using $\Delta = 2R^2/n = \widetilde{O}(\sigma^2 \log(nd/\beta)/n)$ together with $\epsilon_{\text{row}}^{-1} = \widetilde{O}(\sqrt{d\log(1/\delta)}/\epsilon)$) gives,

$$\|\mathbf{M} - \mathbf{\Sigma}\|_{\text{op}} \;\leq\; O\left(\sigma^2 k\sqrt{\frac{\log(d/\beta)}{n}} \;+\; \frac{\sigma^2 k\sqrt{dk}}{n\epsilon}\,\log\left(\frac{nd}{\beta}\right)\log\left(\frac{d}{\beta}\right)\log\left(\frac{d}{\delta}\right)\right).$$

The result then follows the setting the RHS smaller than α and rearranging the inequality. □

3.2 Lower bound

In this section, we provide our lower bound in Theorem 2 for private sparse covariance estimation, complementing the upper bound proved in Theorem 1 and showing a $\sqrt{d}$ factor is necessary in the sample complexity.

Theorem 2. *Let $c \in (0,1)$ be a universal constant. Let $\alpha \leq c$, $\epsilon \leq 1$ and $\delta \leq \frac{\epsilon^2}{d^2}$. Let $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$ with any covariance matrix $\mathbf{\Sigma}$ satisfying Model 1 with $\gamma = 0$, $k \geq \Omega(\log(d))$ and $d \geq 10$. Let $M : \mathcal{X} \to \mathbb{S}^{d\times d}$ be any (ϵ,δ)-DP mechanism that takes $\mathcal{X}$ as input and satisfies, $\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathsf{op}} \leq \alpha\sigma^2$ with probability at least $2/3$. Then, we must have*

$$n \geq \widetilde{\Omega}\left(\min\left\{\frac{k^2}{\alpha^2} + \frac{k\sqrt{d}}{\alpha\epsilon},\ \frac{d\exp(k)}{\epsilon}\right\}\right)$$

where $\widetilde{\Omega}$ hides poly-logarithmic factors in $(d, \frac{1}{\alpha})$.

Our proof follows the fingerprinting-based strategy proposed in [Nar24], adapted to k-RCS matrices. We will require the following useful facts about the Inverse Wishart distribution and its posterior estimation properties.

Fact 8. *Let $\mathbf{\Sigma} \sim \mathsf{InvWishart}(\mathbf{\Psi}_0, \nu_0)$ for $\mathbf{\Psi}_0 \in \mathbb{R}^{d\times d}$ and $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma})$ then letting p be the density of $\mathbf{\Sigma}$, the following hold.*

1. $p(\mathbf{\Sigma}) \propto \det(\mathbf{\Sigma})^{-(\nu_0+d+1)/2} \exp\left(-\mathrm{Tr}\left(\mathbf{\Psi}_0\mathbf{\Sigma}^{-1}\right)/2\right)$
2. *For $\nu_0 \geq d+2$, $\mathbb{E}[\mathbf{\Sigma}] = \mathbf{\Psi}_0/(\nu_0 - d - 1)$*
3. *For $\widehat{\mathbf{\Sigma}} := (1/n)\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$, $\mathbb{E}[\mathbf{\Sigma}|\mathcal{X}] = w_n\widehat{\mathbf{\Sigma}} + (1-w_n)\mathbb{E}[\mathbf{\Sigma}]$ where $w_n := n/(\nu_0 + n - d - 1)$*

We also define the following *graph projection* operation, useful for defining the fingerprining scores subsequently. Let $G := (V, E)$ be a fixed graph on d nodes. We define a linear projection operator $P_G := \mathbb{S}^{d\times d} \to \mathbb{S}^{d\times d}$ as, for all $i, j \in [d]^2$,

$$[P_G(\mathbf{M})]_{i,j} := \begin{cases} [\mathbf{M}]_{i,j}, & \text{if } i = j \text{ or } (i,j) \in E \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

We now define a particular graph G over $[d]$ nodes along with an associated prior distribution that will be used in our lower bound instance. We assume d is a multiple of k for simplicity.

Model 3 (Fingerprinting Graph Prior)**.** *Fix $(d,k) \in \mathbb{N}^2$ with d divisible by k and $k \geq \Omega(\log(d/k))$. Take d vertices, and partition them into $B := d/k$ disjoint blocks of size k with $V := \bigcup_{b=1}^{B} C_b$, $|C_b| = k$. Let G be the graph whose connected components are cliques on each C_b and no edges between components. Let $r := 2k$. For each block $b \in [B]$, draw an independent covariance block*

$$\forall b \in [B], \qquad \mathbf{\Sigma}_b \ \sim\ \mathsf{InvWishart}\left((r-k-1)\mathbf{I}_k,\ r\right), \tag{7}$$

and define the full covariance matrix $\mathbf{\Sigma} := \mathbf{diag}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \cdots \mathbf{\Sigma}_B) \in \mathbb{R}^{d\times d}$ as the block diagonal matrix formed by arranging $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \cdots \mathbf{\Sigma}_B$ on the diagonals. We obtain samples $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

Since the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ satisfies Definition 3 with $\sigma^2 := \|\mathbf{\Sigma}\|_{\text{op}}$, and the graph G defined in Model 3 has, by definition, a k-RCS adjacency matrix, then Model 3 generates a covariance matrix $\mathbf{\Sigma}$ satisfying Model 1 with $\sigma^2 := \|\mathbf{\Sigma}\|_{\text{op}}$. We now provide some useful properties for $\mathbf{\Sigma}$ sampled from Model 3, with the proof deferred to Appendix B.

Lemma 2 (Properties of $\mathbf{\Sigma}$ from Model 3)**.** *Let $\mathbf{\Sigma}$ and $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ be generated from Model 3 with associated graph G. Let for all $i \in [n]$, $\mathbf{x}_{i,b} \in \mathbb{R}^{|C_b|}$ denotes the i^{th} vector with coordinates in the block C_b. Let $a := \frac{r-k+1}{2}$ and P_G denote the graph projection operator in Definition 6, then the following properties hold.*

1. $P_G(\widehat{\mathbf{\Sigma}}) = \mathbf{diag}\left(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_2, \cdots, \widehat{\mathbf{\Sigma}}_B\right)$ *for* $\widehat{\mathbf{\Sigma}}_b := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_{i,b}\mathbf{x}_{i,b}^\top$.
2. $\forall\, i \in [n],\ b \in [B],\ \mathbf{x}_{i,b} \sim \mathcal{N}(0, \mathbf{\Sigma}_b)$, *and* $\forall\, b_1 \neq b_2 \in [B],\ \mathbf{x}_{i,b_1} \perp\!\!\!\perp \mathbf{x}_{i,b_2}$.
3. $\forall t > 0,\ \mathbb{P}(\|\mathbf{\Sigma}\|_{\text{op}} > t) \leq \frac{d}{k}(\frac{e^2}{t})^a$ *and if* $q \leq \frac{a}{2}$, *then* $\mathbb{E}[\|\mathbf{\Sigma}\|_{\text{op}}^q] \leq 2\exp(4q)$.
4. $\forall b \in [B],\ \ g_1\frac{k^2}{n} \leq \mathbb{E}\big[\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\|_{\text{F}}^2\big] \leq g_2\frac{k^2}{n}$.
5. $g_2 \cdot dk/n \leq \mathbb{E}[\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\|_{\text{F}}^2] \leq g_3 \cdot dk/n$.

Here $g_1 > 0$, $0 < g_2 < g_3$ are universal constants.

We now establish a posterior concentration result, following the proof of Lemma 6 in [Nar24].

Lemma 3. *Let $\mathbf{\Sigma}, \mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ be generated from Model 3 and $\widehat{\mathbf{\Sigma}} := \sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top/n$. Then,*

$$\mathbb{E}_{\mathcal{X}}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}|X\right] - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\text{F}}^2\right] \leq O\left(\frac{dk^2}{n^2} + \frac{dk^3}{n^3}\right)$$

Proof. Let for all $i \in [n]$, $\mathbf{x}_{i,b} \in \mathbb{R}^{|C_b|}$ denotes the i^{th} vector with coordinates in the block C_b and define $\mathcal{X}_b := \{\mathbf{x}_{i,b}\}_{i\in[n]}$. Note that the prior is independent over the blocks $\{C_b\}_{b\in[B]}$, i.e $p\left(\mathbf{\Sigma}\right) := \prod_{b\in[B]} p\left(\mathbf{\Sigma}_b\right)$. Furthermore, given a particular draw, $\mathbf{\Sigma}$, from the prior, the likelihood of the data also factorises over the blocks, i.e $p\left(\mathcal{X}|\mathbf{\Sigma}\right) \propto \prod_{b=1}^B p\left(\mathcal{X}_b|\mathbf{\Sigma}_b\right)$. Therefore, the posterior factorises as

$$p\left(\mathbf{\Sigma}|X\right) \propto p\left(\mathbf{\Sigma}\right).p\left(X|\mathbf{\Sigma}\right) = \left(\prod_{b=1}^B p\left(\mathbf{\Sigma}_b\right)\right).\left(\prod_{b=1}^B p\left(\mathcal{X}_b|\mathbf{\Sigma}_b\right)\right) \propto \prod_{b=1}^B p\left(\mathbf{\Sigma}_b|\mathcal{X}_b\right)$$

Now consider $\mathbb{E}\left[\mathbf{\Sigma}_b|X\right]$. We have

$$\begin{aligned}
\mathbb{E}\left[\mathbf{\Sigma}_b|X\right] &= \int_{\mathbf{\Sigma}_1,\cdots\mathbf{\Sigma}_b} \mathbf{\Sigma}_b p\left(\mathbf{\Sigma}|X\right) \mathrm{d}\mathbf{\Sigma}_1 \mathrm{d}\mathbf{\Sigma}_2 \cdots \mathrm{d}\mathbf{\Sigma}_b \\
&= \int_{\mathbf{\Sigma}_1} \cdots \int_{\mathbf{\Sigma}_b} \mathbf{\Sigma}_b \prod_{b=1}^{B} p\left(\mathbf{\Sigma}_b|\mathcal{X}_b\right) \mathrm{d}\mathbf{\Sigma}_1 \mathrm{d}\mathbf{\Sigma}_2 \cdots \mathrm{d}\mathbf{\Sigma}_b \\
&= \left(\int_{\mathbf{\Sigma}_b} \mathbf{\Sigma}_b p\left(\mathbf{\Sigma}_b|\mathcal{X}_b\right) \mathrm{d}\mathbf{\Sigma}_b\right) \left(\prod_{j\neq b} \int_{\mathbf{\Sigma}_j} p\left(\mathbf{\Sigma}_j|X_{\cdot,j}\right) \mathrm{d}\mathbf{\Sigma}_j\right) \\
&= \int_{\mathbf{\Sigma}_b} \mathbf{\Sigma}_b p\left(\mathbf{\Sigma}_b|\mathcal{X}_b\right) \mathrm{d}\mathbf{\Sigma}_b = \mathbb{E}\left[\mathbf{\Sigma}_b|\mathcal{X}_b\right] \qquad (8)
\end{aligned}$$

Then, since the partitions are disjoint, we have using Items 1, 2 from Lemma 2,

$$\mathbb{E}_{\mathcal{X}}\left[\|\mathbb{E}\left[\mathbf{\Sigma}|\mathcal{X}\right] - P_G(\widehat{\mathbf{\Sigma}})\|_F^2\right] = \sum_{b=1}^{B} \mathbb{E}_{\mathcal{X}}\left[\|\mathbb{E}\left[\mathbf{\Sigma}_b|\mathcal{X}\right] - \widehat{\mathbf{\Sigma}}_b\|_F^2\right] = \sum_{b=1}^{B} \mathbb{E}_{\mathcal{X}_b}\left[\|\mathbb{E}\left[\mathbf{\Sigma}_b|\mathcal{X}_b\right] - \widehat{\mathbf{\Sigma}}_b\|_F^2\right] \qquad (9)$$

Note that $\mathbf{\Sigma}_b \sim \mathsf{InvWishart}\left((r-k-1)\mathbf{I}_k, r\right)$, $\mathcal{X}_b \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \mathbf{\Sigma}_b\right)$. Using Fact 8 and Item 4 from Lemma 2,

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}_b}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}_b|\mathcal{X}_b\right] - \widehat{\mathbf{\Sigma}}_b\right\|_{\mathrm{F}}^2\right] &= \mathbb{E}_{\mathcal{X}_b}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}_b|\mathcal{X}_b\right] - \widehat{\mathbf{\Sigma}}_b\right\|_{\mathrm{F}}^2\right] \\
&= (1-w_n)^2 \mathbb{E}_{\mathcal{X}_b}\left[\left\|\widehat{\mathbf{\Sigma}}_b - \mathbb{E}\left[\mathbf{\Sigma}_b\right]\right\|_{\mathrm{F}}^2\right], \quad \text{for } w_n := \frac{n}{r+n-k-1} \\
&\leq 3(1-w_n)^2 \mathbb{E}_{\mathcal{X}_b}\left[\left\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\right\|_{\mathrm{F}}^2 + \|\mathbf{\Sigma}_b\|_{\mathrm{F}}^2 + \|\mathbb{E}\left[\mathbf{\Sigma}_b\right]\|_{\mathrm{F}}^2\right] \\
&\leq \left(\frac{r-k-1}{n+r-k-1}\right)^2 O\left(\frac{k^2}{n} + k\right) \leq O\left(\frac{r^2k^2}{n^2} + \frac{r^2k}{n^3}\right) = O\left(\frac{k^3}{n^2} + \frac{k^4}{n^3}\right)
\end{aligned}$$

Substituting in (9), we have,

$$\mathbb{E}_X\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}|X\right] - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2\right] \leq O\left(\frac{d}{k} \cdot \left(\frac{k^3}{n^2} + \frac{k^4}{n^3}\right)\right) = O\left(\frac{dk^2}{n^2} + \frac{dk^3}{n^3}\right)$$

□

Following the arguments in the proof of Lemma 7 of [Nar24], we first establish a non-private error lower bound in Lemma 4.

Lemma 4. *Let $\rho < c_1\sqrt{d}$ for a sufficiently small absolute constant c_1. Let $\mathbf{\Sigma}, \mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ be generated using Model 3. Then any (possibly randomized) estimator $f : (\mathbb{R}^d)^n \to \mathbb{S}^{d\times d}$ satisfying $\mathbb{E}_{\mathbf{\Sigma},\mathcal{X}}[\|f(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2] \leq \rho^2$ must have*

$$n \geq \Omega\left(\frac{dk}{\rho^2}\right).$$

Proof. Fix any deterministic f (randomized follows by conditioning on its coins). For any random matrix $\mathbf{V}$ with mean $\mathbf{U}$, $\mathbb{E}\|\mathbf{V}\|_F^2 \geq \mathbb{E}\|\mathbf{V}-\mathbf{U}\|_F^2$. Applying this to $\mathbf{V} = f(\mathcal{X}) - \mathbf{\Sigma}$ conditional on $\mathcal{X}$ gives

$$\mathbb{E}\left[\|f(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2 \mid \mathcal{X}\right] \geq \mathbb{E}\left[\|\mathbb{E}[\mathbf{\Sigma}|\mathcal{X}] - \mathbf{\Sigma}\|_{\mathrm{F}}^2 \mid \mathcal{X}\right].$$

Removing conditioning,

$$\mathbb{E}[\|f(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2] \geq \mathbb{E}[\|\mathbb{E}[\mathbf{\Sigma}|\mathcal{X}] - \mathbf{\Sigma}\|_{\mathrm{F}}^2].$$

Next use $\|\mathbf{A}-\mathbf{B}\|_F^2 \leq 2\|\mathbf{A}-\mathbf{C}\|_F^2 + 2\|\mathbf{C}-\mathbf{B}\|_F^2$ with $\mathbf{A} = P_G(\widehat{\mathbf{\Sigma}})$, $\mathbf{B} = \mathbf{\Sigma}$, $\mathbf{C} = \mathbb{E}[\mathbf{\Sigma}|X]$:

$$\mathbb{E}\left\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_{\mathrm{F}}^2 \leq 2\mathbb{E}\left\|\mathbb{E}[\mathbf{\Sigma}|\mathcal{X}] - \mathbf{\Sigma}\right\|_{\mathrm{F}}^2 + 2\mathbb{E}\left\|\mathbb{E}[\mathbf{\Sigma}|\mathcal{X}] - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2.$$

Rearranging and using Lemma 2, Item 5, and Lemma 3 yields, for absolute constants $g_2, g_4 > 0$

$$\mathbb{E}[\|f(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2] \geq g_2\frac{dk}{n} - g_4\Big(\frac{dk^2}{n^2} + \frac{dk^3}{n^3}\Big).$$

Let $n := \frac{g_2}{3}\cdot\frac{dk}{\rho^2}$, then for $\rho \leq c_1\sqrt{d}$ with sufficiently small c_1, we have $n \geq \max\{1, \frac{g_2}{3c_1^2}\}k$. Then, $g_4\Big(\frac{dk^2}{n^2} + \frac{dk^3}{n^3}\Big) \leq 2g_4\frac{dk^2}{n^2}$. Then again for sufficiently small c_1,

$$\mathbb{E}[\|f(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2] \geq g_2\frac{dk}{n} - 2g_4\frac{dk^2}{n^2} \geq \frac{g_2}{2}\frac{dk}{n} = \frac{3\rho}{2}$$

Thus, the RHS can be forced to exceed $3\rho/2$ unless the stated lower bound on n holds. □

Let $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$ drawn from Model 3 and further draw iid samples $\mathcal{X}' := \{\mathbf{x}_i'\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$. Let $\mathcal{X}_{\sim i}$ be formed by replacing $\mathbf{x}_i$ from $\mathcal{X}$ by $\mathbf{x}_i'$. For any (ϵ,δ)-DP mechanism $M : \mathcal{X} \to \mathbb{S}^{d\times d}$, we define, $\forall i \in [n]$,

$$Z_i := \left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G\left(\mathbf{x}_i\mathbf{x}_i^\top\right) - P_G(\mathbf{\Sigma})\right\rangle, Z_i' := \left\langle M(\mathcal{X}_{\sim i}) - \mathbf{\Sigma}, P_G\left(\mathbf{x}_i\mathbf{x}_i^\top\right) - P_G(\mathbf{\Sigma})\right\rangle \tag{10}$$

Our next result provides upper bounds on these fingerprinting scores, following Lemma 4, [Nar24].

Lemma 5. *Let $c_1 < 1 < c_3$ be a universal constants. Under Model 3, consider the fingerprinting scores defined in (10). Let $\mathbb{E}_{\mathbf{\Sigma},\mathcal{X},M}[\|M(\mathcal{X}) - \mathbf{\Sigma}\|_F^4] \leq \rho^4$ for $\rho < c_1\sqrt{d}$. Let $\epsilon \in (0,1)$, $\delta < \frac{\epsilon^2}{d^2}$ and $d > c_3$. Then, for some consant $C > 0$, $\mathbb{E}[\sum_{i\in[n]} Z_i] \leq C\,\rho\,\epsilon\,n$.*

Proof. The proof is immediate from the proof of Lemma 4 and Corollary 1 in [Nar24], by noting that $\|P_G(\mathbf{x}_i\mathbf{x}_i^\top)\|_F \leq \|\mathbf{x}_i\mathbf{x}_i^\top\|_F \leq \|\mathbf{x}_i\|_2^2$ and using Lemma 2, Item 3 for the moment bounds on $\|\mathbf{\Sigma}\|_{\mathrm{op}}$. □

We next establish the following lower bound on the average score,

Lemma 6. *Let $\mathbf{\Sigma}, \mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ be generated using Model 3. Let $M : \mathcal{X} \to \mathbb{S}^{d\times d}$ take $\mathcal{X}$ as input with randomness independent of $\mathbf{\Sigma}$ conditioned on $\mathcal{X}$. Then for $\widehat{\mathbf{\Sigma}} := \sum_{i\in[n]}\mathbf{x}_i\mathbf{x}_i/n$, we have*

$$\mathbb{E}\left[\left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G\left(\widehat{\mathbf{\Sigma}}\right) - \mathbf{\Sigma}\right\rangle\right]$$
$$\geq \mathbb{E}\left[\left\|P_G\left(\widehat{\mathbf{\Sigma}}\right) - \mathbf{\Sigma}\right\|_{\mathrm{F}}^2\right] - \sqrt{\mathbb{E}\left[\left\|M(\mathcal{X}) - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2\right]\cdot\mathbb{E}\left[\left\|\mathbb{E}[\mathbf{\Sigma}|X] - P_G\left(\widehat{\mathbf{\Sigma}}\right)\right\|_{\mathrm{F}}^2\right]}.$$

Proof. The proof follows directly from Lemma 5 of [Nar24] with $\widehat{\mathbf{\Sigma}} \leftarrow P_G(\widehat{\mathbf{\Sigma}})$. □

Lemma 7. *Let $c_1, c_2 < 1 < c_3$ be constants such that $d > c_3$, $\rho < c_1\sqrt{d}$, $\epsilon < 1$, and $\delta < \frac{\epsilon^2}{d^2}$. Let $\mathbf{\Sigma}$, $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$ be generated from Model 3. Let $M : \mathcal{X} \to \mathbb{S}^{d\times d}$ be any (ϵ, δ)-DP mechanism that takes $\mathcal{X}$ as input and satisfies $\mathbb{E}_{\mathbf{\Sigma},\mathcal{X},M}\left[\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^4\right] \le \rho^4$, then we must have,*

$$n \ge c_2 . \max\left\{\frac{dk}{\rho^2}, \frac{dk}{\rho\epsilon}\right\}$$

Proof. We start with the non-private error term. By the Cauchy-Schwarz inequality,

$$\mathbb{E}\,\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2 \le \sqrt{\mathbb{E}\,\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^4} \le \rho^2.$$

Applying Lemma 4 to the estimator $f = M$ yields

$$n \ge \Omega\left(\frac{dk}{\rho^2}\right). \tag{11}$$

Recall $Z_i = \left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G(\mathbf{x}_i\mathbf{x}_i^\top) - \mathbf{\Sigma}\right\rangle$ from (10). Since $\mathbf{\Sigma}$ is block-diagonal on G, we have $P_G(\mathbf{\Sigma}) = \mathbf{\Sigma}$, and therefore

$$\sum_{i=1}^n Z_i = \left\langle M(\mathcal{X}) - \mathbf{\Sigma}, \sum_{i=1}^n (P_G(\mathbf{x}_i\mathbf{x}_i^\top) - \mathbf{\Sigma})\right\rangle = n\left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\rangle.$$

Taking expectations,

$$\mathbb{E}\left[\sum_{i=1}^n Z_i\right] = n\,\mathbb{E}\left[\left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\rangle\right]. \tag{12}$$

By Lemma 6 and Lemma 2, Item 5,

$$\mathbb{E}\left[\left\langle M(\mathcal{X}) - \mathbf{\Sigma}, P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\rangle\right] \ge g_2\frac{dk}{n} - \sqrt{\mathbb{E}\left\|M(\mathcal{X}) - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2 \cdot \mathbb{E}\left\|\mathbb{E}[\mathbf{\Sigma} \mid \mathcal{X}] - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2}.$$

Also, $\|M(\mathcal{X}) - P_G(\widehat{\mathbf{\Sigma}})\|_{\mathrm{F}}^2 \le 2\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2 + 2\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2$, so taking expectations and using $\mathbb{E}\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2 \le \rho^2$ and Lemma 2, Item 5,

$$A := \mathbb{E}\left[\left\|M(\mathcal{X}) - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2\right] \;\le\; 2\rho^2 + 2g_3\frac{dk}{n}. \tag{13}$$

Moreover, by Lemma 3,

$$B := \mathbb{E}\left[\left\|\mathbb{E}[\mathbf{\Sigma} \mid \mathcal{X}] - P_G(\widehat{\mathbf{\Sigma}})\right\|_{\mathrm{F}}^2\right] \;\le\; O\left(\frac{dk^2}{n^2} + \frac{dk^3}{n^3}\right). \tag{14}$$

Assume now from (11), $n \ge C_0 \cdot \frac{dk}{\rho^2}$ for a sufficiently large universal C_0. Then $dk/n \le \rho^2/C_0$, and (13) gives $A \le O(\rho^2)$. Also, $n \gtrsim k$ automatically, so (14) simplifies to $B \le O(dk^2/n^2)$. Hence

$$n\sqrt{A\,B} \;\le\; n \cdot O(\rho) \cdot \sqrt{O\left(\frac{dk^2}{n^2}\right)} \;=\; O(\rho\, k\sqrt{d}).$$

Plugging back into (12) yields $\mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] \geq n \cdot g_2 \frac{dk}{n} - O(\rho k\sqrt{d}) = g_2\, dk - O(\rho k\sqrt{d})$. Since $\rho < c_1\sqrt{d}$, choosing c_1 sufficiently small (absorbing constants) gives

$$\mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] \geq \frac{g_2}{2}\, dk. \tag{15}$$

On the other hand, by Lemma 5,

$$\mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] \leq C\,\rho\,\epsilon\,n. \tag{16}$$

Combining (15) and (16) yields $n \geq \Omega(\frac{dk}{\rho\epsilon})$. Together with (11), this gives the claim with an appropriate constant c_2. □

We are now ready to present the proof of Theorem 2.

Proof of Theorem 2. We note that for the non-private component of the sample complexity, existing lower bounds (Theorem 1 [CZ12]) already demonstrate a lower bound of the form $n = \Omega(\frac{k^2}{\alpha^2})$ for k-RCS covariance matrices satisfying Model 1.

Now, fix any (ϵ, δ)-DP mechanism $M_0 : (\mathbb{R}^d)^n \to \mathbb{S}^{d\times d}$ satisfying the theorem's guarantee: for every covariance matrix $\mathbf{\Sigma}$ satisfying Model 1 with $\gamma = 0$, if $\mathcal{X} \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma})$ then

$$\mathbb{P}\Big(\|M_0(\mathcal{X}) - \mathbf{\Sigma}\|_{\text{op}} \leq \alpha\,\sigma^2\Big) \geq \frac{2}{3}, \qquad \sigma^2 := \|\mathbf{\Sigma}\|_{\text{op}}. \tag{17}$$

Consider $\mathbf{\Sigma}$ drawn from the block prior in Model 3, and then $\mathcal{X} \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma})$. Since (17) holds *for every* valid $\mathbf{\Sigma}$, it also holds after drawing $\mathbf{\Sigma}$ from this prior.

Let $\mathcal{E}_0 := \{\|\mathbf{\Sigma}\|_{\text{op}} \leq 10\}$ as in Lemma 30. By Lemma 30, $\mathbb{P}(\mathcal{E}_0) \geq 1 - \exp(-\Omega(k))$. On $\mathcal{E}_0$, we have $\sigma^2 = \|\mathbf{\Sigma}\|_{\text{op}} \leq 10$. Therefore (17) implies the Frobenius error bound

$$\|M_0(\mathcal{X}) - \mathbf{\Sigma}\|_{\text{F}} \leq \sqrt{d}\,\|M_0(\mathcal{X}) - \mathbf{\Sigma}\|_{\text{op}} \leq \sqrt{d}\,\alpha\,\|\mathbf{\Sigma}\|_{\text{op}} \leq \rho,$$

with probability at least 2/3, where we set

$$\rho := 10\sqrt{d}\,\alpha. \tag{18}$$

Thus, conditional on any fixed $\mathbf{\Sigma}$ with $\|\mathbf{\Sigma}\|_{\text{op}} \leq 10$, the mechanism M_0 satisfies

$$\mathbb{P}\Big(\|M_0(\mathcal{X}) - \mathbf{\Sigma}\|_F \leq \rho\Big) \geq \frac{2}{3}. \tag{19}$$

By Lemma 11 from [Nar24], with the moment bounds in Lemma 2, the tail bound in Lemma 30 for the event $\mathcal{E}_0$, and the lower bound on $d \geq 10$ assumed in the theorem, if for universal constants $b_1, b_2 > 0$,

$$\exp(-b_1 k) \leq \rho \leq b_2\sqrt{d} \tag{20}$$

there exists a mechanism M that is (ϵ, δ)-DP, uses nL samples where $L = \Theta(\log(d/\rho))$, and satisfies

$$\mathbb{E}\left[\|M(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}}^4\right] \;\leq\; O(\rho^4). \tag{21}$$

Applying Lemma 7 to M, this implies

$$nL \;\geq\; c_2 \cdot \max\left\{\frac{dk}{\rho^2},\; \frac{dk}{\rho\epsilon}\right\}, \text{ or equivalently } n \;\geq\; \widetilde{\Omega}\left(\max\left\{\frac{dk}{\rho^2},\; \frac{dk}{\rho\epsilon}\right\}\right)$$

Substituting $\rho = 10\sqrt{d}\,\alpha$ from (18) gives

$$n \;\geq\; \widetilde{\Omega}\left(\max\left\{\frac{dk}{d\alpha^2},\; \frac{dk}{\sqrt{d}\,\alpha\,\epsilon}\right\}\right) = \widetilde{\Omega}\left(\frac{k}{\alpha^2} + \frac{k\sqrt{d}}{\alpha\epsilon}\right). \tag{22}$$

This completes the proof under the parameter regime (20). We now argue for ρ outside the domain of (20) and smaller than $\exp(-b_1 k)$. Indeed if (20) does not hold, then,

$$\frac{1}{10\sqrt{d}\alpha} \geq \frac{1}{\rho} \geq \exp\left(b_1 k\right) \tag{23}$$

Now from existing lower bounds for $d = k = 1$ (e.g., Theorem 5 in [KLSU19]), as highlighted in [Nar24], we already have a lower bound of $\Omega\left(\frac{1}{\alpha^2} + \frac{1}{\alpha\epsilon}\right)$, which is trivially valid for higher dimensions. Using this lower bound for $\alpha \leftarrow \alpha_0 := \exp(-b_1 k)/10\sqrt{d}$, we have

$$n \geq \Omega\left(\frac{d\exp\left(k\right)}{\epsilon}\right) \tag{24}$$

The result then follows by noting that (24) is a valid lower bound for all $\alpha \leq \alpha_0$. □

4 Private Sparse PCA: Algorithms

In this section, we provide algorithms and resulting upper bounds for Problem 2 under Model 2 with positive eigengap γ in Theorem 3.

A key obstacle in designing a differentially private sparse PCA algorithm in this setting is controlling the sensitivity of the output vector. In low dimensions this is often handled via boundedness assumptions of the form $\|\mathbf{x}_i\|_2 = O(1)$, yielding $O(1/n)$ sensitivity for the sample covariance in frobenius norm; in high dimensions, where $\|\mathbf{x}_i\|_2 \asymp \sqrt{d}$, the same reasoning leads to $O(d/n)$ sensitivity. While thresholding the empirical covariance can give $O(k/\sqrt{n})$ operator norm error *with high probability* to the population (see Theorem 1 [BL09], Theorem 6.23 [Wai19]), this does not suffice for privacy, which requires worst-case guarantees. We address this via a new algorithm (Algorithm 2), building on the FriendlyCore primitive [TCK+22], used in recent work [LLA24, KLTY25].

To exploit the sparsity inherent in the model and avoid dimension factors, we approach this problem by trying to control the entrywise sensitivity of the estimator, $\widetilde{\mathbf{\Sigma}}$, between adjacent datasets, while maintaining the k-RCS property to convert entrywise bounds to operator norm bounds. In

Algorithm 2 Sparse PCA for k-RCS covariance matrices : $\mathsf{PrivPCA}(\mathcal{D}, \hat{\lambda}, \gamma, k, \sigma^2, \epsilon, \delta, \tau, \beta, m)$

Require: Samples $\mathbf{x}_1, \dots \mathbf{x}_n \in \mathbb{R}^d$, operator norm estimate $\hat{\lambda}$, gap parameter $\gamma > 0$, sparsity k, sub-Gaussian proxy σ^2, privacy (ϵ, δ), threshold τ, failure β, number of batches m.

1: Partition the samples into m blocks of size $b := n/m$ and compute covariances $\widehat{\mathbf{\Sigma}}_1, \dots, \widehat{\mathbf{\Sigma}}_m$ on disjoint blocks.

2: For $\gamma_i := 1 - \lambda_2(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))/\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))$, define the certified set

$$\mathcal{C} \leftarrow \left\{ i \in [m] : \mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) \text{ is } k\text{-RCS}, \quad \lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)) \geq \left(1 - \frac{\gamma}{2}\right)\hat{\lambda}, \quad \gamma_i \geq \frac{\gamma}{2} \right\}.$$

3: Sample $L \sim \mathsf{BoundedLap}\big(\frac{3}{\epsilon}, \frac{3}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$.

4: **if** $|\mathcal{C}| + L - \frac{3}{\epsilon}\log\big(\frac{12}{\delta}\big) < 0.8\,m$ **then**

5: **return** $\perp$

6: **end if**

7: **for** $i \leftarrow 1 : m$ **do**

8: $f_i \leftarrow \sum_{j\in[m]} \mathcal{I}\Big\{\|\widehat{\mathbf{\Sigma}}_j - \widehat{\mathbf{\Sigma}}_i\|_{\infty,\infty} \leq 4\tau\Big\}$

9: $p_i \leftarrow \min\Big\{\max\Big\{\frac{f_i - m/2}{2m/3 - m/2}, 0\Big\}, 1\Big\}$

10: **end for**

11: $Z \leftarrow \sum_{i\in[m]} p_i$

12: Sample $\xi \sim \mathsf{BoundedLap}\big(\frac{21}{\epsilon}, \frac{21}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$.

13: **if** $Z + \xi - \frac{21}{\epsilon}\log\big(\frac{12}{\delta}\big) \leq 0.8\,m$ **then**

14: **return** $\perp$

15: **end if**

16: $\bar{\mathbf{\Sigma}} \leftarrow \frac{1}{Z}\sum_{i\in[m]} p_i\,\widehat{\mathbf{\Sigma}}_i$

17: $\widehat{\mathbf{v}} \leftarrow \mathbf{v}_1\big(\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})\big), \qquad \widehat{\mathbf{P}} \leftarrow \widehat{\mathbf{v}}\,\widehat{\mathbf{v}}^\top$

18: Set $\Delta_{\mathbf{P}} \leftarrow \frac{68\sqrt{2k}\tau}{\gamma\,\hat{\lambda}}$

19: Set $\sigma_{\text{priv}} \leftarrow \frac{6\Delta_{\mathbf{P}}}{\epsilon}\sqrt{\log\left(\frac{6}{\delta}\right)}$

20: Sample $\mathbf{W}$ with i.i.d. entries $\mathbf{W}_{ij} \sim \mathcal{N}(0, \sigma^2_{\text{priv}})$.

21: $\widetilde{\mathbf{P}} \leftarrow \text{top}_{k^2}\Big(\widehat{\mathbf{P}} + \mathbf{W}\Big), \qquad$ symmetrize $\widetilde{\mathbf{P}} \leftarrow (\widetilde{\mathbf{P}} + \widetilde{\mathbf{P}}^\top)/2$

22: **return** $\widehat{\mathbf{v}}_{\text{priv}} := \mathbf{v}_1(\widetilde{\mathbf{P}})$

particular, classical concentration results (Fact 4) show that for subgaussian distributions, given b samples from a σ-sub-Gaussian distribution (Definition 3), we have with probability atleast $1 - \beta$,

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\infty,\infty} \lesssim \sigma^2\sqrt{\frac{\log(d/\beta)}{b}} =: \tau \tag{25}$$

Since this is true with high probability for all datapoints via a union-bound, it then seems natural to use this fact directly with the FriendlyCore algorithm to weigh each point by the number of their neighbours within an $\|.\|_{\infty,\infty}$ ball (Line 9). To this effect, we first partition the data into m disjoint blocks and form blockwise covariance estimates. We then aggregate the block covariances using weights derived from these agreement counts to obtain a weighted estimator $\bar{\mathbf{\Sigma}}$.

This however poses another challenge in analyzing the final sensitivity. For $\bar{\mathbf{\Sigma}}, \bar{\mathbf{\Sigma}}'$ on adjacent

datasets, we would like to control the sensitivity of their leading eigenvectors after thresholding at $O(\tau)$. However, for analyzing the sensitivity, we cannot assume concentration to the population $\mathbf{\Sigma}$, and must provide a worst case bound. Entrywise concentration as in (25) does not by itself guarantee that each of the blockwise covariances $\{\widehat{\mathbf{\Sigma}}_i\}_{i\in[b]}$ satisfy the k-RCS property after thresholding, but it does guarantee that a large fraction of them would satisfy such a property (see Fact 7).

This motivates us to add this property as check in Line 4, ensuring that if the check passes, then we have a large fraction of $\{\widehat{\mathbf{\Sigma}}_i\}_{i\in[n]}$ satisfying the k-RCS property after thresholding. We note that the certified set, $\mathcal{C}$, in Algorithm 2 includes other checks, which are also true with high probability and required to analyse sensitivity of $\mathbf{v}_1(\mathcal{T}_\tau(\bar{\mathbf{\Sigma}}))$ via an application of Wedin's theorem (Lemma 28). The check in Line 13 is required for similar reasons. The key property of this covariance matrix is its sensitivity is $\widetilde{O}(k\sqrt{\frac{m}{n}})$, which does not depend on d.

Note that the two sparsity assumptions play different roles. The k-RCS assumption is used before eigenvector recovery, to convert entrywise stability into operator-norm control and to keep the sensitivity of the stabilized covariance estimate dimension-independent. The sparsity of v_1 is used only in the final recovery step: under Model 2, the projector $v_1 v_1^\top$ is k^2-sparse, so top-k^2 truncation after the noisy release preserves the signal. If v_1 is dense, this truncation step is no longer aligned with the target projector.

Theorem 3 (Private k-sparse PCA upper bound)**.** *Under Model 2, for Problem 2 there exists an algorithm (Algorithm 2) which satisfies, for*

$$n \geq \widetilde{\Omega}\left(\frac{k^4}{\gamma^2\,\Delta\epsilon^3}\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2}\right)$$

with probability at least $1-\beta$ over the samples and the algorithm, $\sin^2\angle(\widehat{\mathbf{v}}_{\text{priv}}, \mathbf{v}_1(\mathbf{\Sigma})) \leq \Delta$. *Here $\widetilde{\Omega}$ hides polylogarithmic factors in $(d, 1/\epsilon, 1/\delta, 1/\beta)$.*

Given Lemma 11, the sensitivity of the projector, $\widehat{\mathbf{P}}$ then follows by a standard application of Wedin's theorem (Lemma 3) and triangle inequality, using the fact that $\mathbf{\Sigma}^\star$ has a sufficiently large eigengap. This helps us bound $\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}'}\|_{\text{F}}$ (see Lemma 12 for the detailed proof) for projectors on adjacent datasets $\mathcal{D}, \mathcal{D}'$. To formally analyze privacy, Algorithm 2 follows a propose-test-release framework. The two tests (Lines 4-10 and Lines 13-15) are differentially private by the bounded Laplace mechanism (Fact 2) and hence the event of exiting is itself (ϵ, δ)-DP (up to constants). All subsequent releases are performed only after passing these private tests, and on this certified branch, we prove uniform sensitivity bounds (Lemmas 11 and 12), allowing us to invoke the Gaussian mechanism (Fact 1).

We next establish the utility guarantee. Utility is proved by showing that under Model 2 with the stated sample complexity in Theorem 1, the tests in Line 4 and Line 13 pass (Lemmas 14 and 15). This however requires the number of batches $m = \Omega(\frac{1}{\epsilon})$, leading to an extra $\frac{1}{\epsilon}$ factor in our sample complexity. The key difference with the privacy analysis is that now we show $\bar{\mathbf{\Sigma}}$ is in fact close to the population $\mathbf{\Sigma}$, instead of $\mathbf{\Sigma}^\star$. Once this is established, $\widetilde{\mathbf{P}}$ can be shown to be close to the true eigenprojector $\mathbf{P} := \mathbf{v}_1(\mathbf{\Sigma})\mathbf{v}_1(\mathbf{\Sigma})^\top$, which is sparse (see Model 2). We show that the top_{k^2} operation retains the large values of $\mathbf{P}$, which controls the distance from $\mathbf{P}$ in frobenius norm and hence the operator norm, while again using the sparsity to avoid a d dependence.

We set

$$\tau := c_\tau\, \sigma^2 \sqrt{\frac{m \log(d)}{n}} \qquad \Big(\text{equivalently } \tau = c_\tau\, \sigma^2 \sqrt{\frac{\log(d)}{b}}\Big), \tag{26}$$

for a sufficiently large absolute constant $c_\tau > 0$. This choice matches entrywise concentration at scale σ^2 under σ-sub-Gaussianity.

We use PrivNorm, a private operator norm estimator algorithm for k-RCS covariance matrices defined in Algorithm 3 in Appendix C as a subroutine in Algorithm 2. Theorem 6 shows that under the sample complexity bound in Theorem 3 with Algorithm 3 (PrivNorm), which is a slight variation of Algorithm 2, we have with probability $1-\beta$, an (ϵ, δ)-DP estimator $\hat{\lambda}$ satisfying

$$\left|\hat{\lambda} - \lambda_1(\mathbf{\Sigma})\right| \le \frac{\gamma \lambda_1(\mathbf{\Sigma})}{16}, \text{ which provides } \hat{\lambda} \ge 0.9\lambda_1(\mathbf{\Sigma})$$

The design and proof technique of Algorithm 3 largely follows Algorithm 2, differing in the definition of the certified set, $\mathcal{C}$ (Line 4), and the final output.

We also note that the block decomposition is not used as a privacy-amplifying subsampling step. The final private release is applied after the FriendlyCore aggregation, which depends on all block covariances through the agreement scores and private tests. Thus standard amplification-by-subsampling does not directly apply to this analysis. Moreover, the utility proof requires $m = \Omega(1/\varepsilon)$ blocks so that the propose-test-release checks pass with high probability under bounded Laplace noise, while each block must remain large enough for entrywise covariance concentration. This interaction leads to the current ε^{-3} dependence; improving this dependence remains open.

4.1 Privacy analysis

Lemma 8. *Let* $\widetilde{c} := |\mathcal{C}| + L$, $L \sim \mathsf{BoundedLap}\big(\frac{3}{\epsilon}, \frac{3}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$ *and* $\widetilde{z} := Z + \xi$, $\xi \sim \mathsf{BoundedLap}\big(\frac{21}{\epsilon}, \frac{21}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$. *Then,* $\widetilde{c}, \widetilde{z}$ *are* $(\epsilon/3, \delta/3)$*-DP.*

Proof. Consider neighboring datasets $\mathcal{D}' = \mathcal{D} \setminus \{\widehat{\mathbf{\Sigma}}_m\} \cup \{\widehat{\mathbf{\Sigma}}'_m\}$. For $\mathcal{C}$, only one block covariance can change, so $||\mathcal{C}| - |\mathcal{C}'|| \le 1$. For Z, considering the modified datapoint and the rest of the data separately,

$$Z - Z' = \sum_{i \in [m]} p_i - p'_i \le 1 + (m-1) \times \frac{1}{m/6} \le 7. \tag{27}$$

The claims then follows from Fact 2. □

Lemma 9 (Common center for surviving points)**.** *Let* τ *be as in* (26)*. If* $|\mathcal{C}| > 0.6m$*, then there exists a matrix* $\widetilde{\mathbf{\Sigma}}$ *such that*

$$\mathcal{A} := \{\widehat{\mathbf{\Sigma}}_i : \ p_i > 0\} \ \subseteq \ \mathbb{B}_\infty(\widetilde{\mathbf{\Sigma}}, 8\tau), \qquad \widetilde{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}_\ell \text{ for some } \ell \in \mathcal{C},$$

and therefore $\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})$ *is* k*-RCS,* $\lambda_1(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})) \ge \hat{\lambda}$*, and* $\gamma(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})) := 1 - \lambda_2(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}}))/\lambda_1(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})) \ge \gamma/2$.

Proof. Any i with $p_i > 0$ satisfies $|\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau)| > f_i > m/2$. Since $|\mathcal{C}| > 0.6m$, there exists $\ell \in \mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau) \cap \mathcal{C}$; set $\widetilde{\mathbf{\Sigma}} := \widehat{\mathbf{\Sigma}}_\ell$. Then $\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})$ is k-RCS, $\lambda_1(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})) \ge \hat{\lambda}$ and $\gamma(\mathcal{T}_\tau(\widetilde{\mathbf{\Sigma}})) \ge \gamma/2$

by $\ell \in \mathcal{C}$. For any other j with $p_j > 0$, the balls $\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau)$ and $\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_j, 4\tau)$ intersect since $|\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau)|, |\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_j, 4\tau)| > m/2$, hence by triangle inequality $\|\widehat{\mathbf{\Sigma}}_j - \widetilde{\mathbf{\Sigma}}\|_{\infty,\infty} \le 8\tau$. □

Lemma 10. *Consider sets $A, B, C, D \subseteq [m+1]$ each of size αm for $\alpha \ge 0.8$ and $m > 30$. Then $|A \cap B \cap C \cap D| > 2$.*

Proof. Let $t := |A \cap B \cap C \cap D|$. Then,

$$4\alpha m \le |A| + |B| + |C| + |D| \le 4t + 3\,(m + 1 - t) = 3\,(m+1) + t.$$

Therefore, $t \ge m\,(4\alpha - 3) - 3 > 2$. □

Lemma 11 (Sensitivity of the weighted average)**.** *Let $\bar{\mathbf{\Sigma}} := \frac{1}{Z}\sum_{i\in[m]} p_i \widehat{\mathbf{\Sigma}}_i$ and define $\bar{\mathbf{\Sigma}}'$ analogously on a neighboring dataset. Condition on the event that both runs reach Line 15 of Algorithm 2. Then, there exists matrix $\mathbf{\Sigma}^\star \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ such that $\mathcal{T}_\tau(\mathbf{\Sigma}^\star)$ is k-RCS, $\widehat{\gamma}(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)) \ge \gamma/2$ and $\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)) \ge \hat{\lambda}$, and*

$$\left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}') - \mathcal{T}_\tau(\mathbf{\Sigma}^\star)\right\|_{\mathrm{op}},\ \left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_\tau(\mathbf{\Sigma}^\star)\right\|_{\mathrm{op}}\ \le 34k\tau.$$

Proof. Let $\mathcal{D}, \mathcal{D}'$ be neighboring datasets and denote the two induced collections of block covariances by $\{\widehat{\mathbf{\Sigma}}_i\}$ and $\{\widehat{\mathbf{\Sigma}}'_i\}$ which differ in at most one element. Let $\mathcal{D}' = \mathcal{D} \setminus \{\widehat{\mathbf{\Sigma}}_m\} \cup \{\widehat{\mathbf{\Sigma}}'_m\}$. Consider the sets, $\mathcal{A} : \{\widehat{\mathbf{\Sigma}}_i \in \mathcal{D} : p_i > 0\}$ and $\mathcal{A}' : \{\widehat{\mathbf{\Sigma}}'_i \in \mathcal{D}' : p'_i > 0\}$. Given that we reach Line 15, $|\mathcal{A}|, |\mathcal{A}'|, |\mathcal{C}|, |\mathcal{C}'| \ge 0.8m$. Then using Lemma 10, there exist at least three elements in $\mathcal{A}\cap\mathcal{A}'\cap\mathcal{C}\cap\mathcal{C}'$, and consequently have at least one element from $\mathcal{D}\cap\mathcal{D}'$. Denote that element by $\mathbf{\Sigma}^*$.

For all surviving points $\widehat{\mathbf{\Sigma}}_i \in \mathcal{D}$, we have $|\mathbb{B}(\widehat{\mathbf{\Sigma}}_i, 4\tau)| > m/2$ and $|\mathbb{B}(\mathbf{\Sigma}^*, 4\tau)| > m/2$, which implies by triangle inequality, following Lemma 9, $\mathbb{B}(\mathbf{\Sigma}^*, 8\tau)$ contains all surviving points in $\mathcal{D}$. By expanding the radius to 16τ, we cover all surviving points in $\mathcal{D}'$ as well, proving the first claim about $\mathbf{\Sigma}^\star$.

Write $\mathbf{M}_i := \widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}^\star$ and $\mathbf{M}'_i := \widehat{\mathbf{\Sigma}}'_i - \mathbf{\Sigma}^\star$; then $\|\mathbf{M}_i\|_{\infty,\infty}, \|\mathbf{M}'_i\|_{\infty,\infty} \le 16\tau$ for all indices with nonzero weight in either run. Then, for $\mathbf{M}_i := \widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}^\star$ and $\mathbf{M}'_i := \widehat{\mathbf{\Sigma}}'_i - \mathbf{\Sigma}^\star$, this yields,

$$\left\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star\right\|_{\infty,\infty} = \left\|\frac{1}{Z}\sum_{i\in[m]} p_i \mathbf{M}_i\right\|_{\infty,\infty} \le 16\tau,\ \ \left\|\bar{\mathbf{\Sigma}}' - \mathbf{\Sigma}^\star\right\|_{\infty,\infty} = \left\|\frac{1}{Z}\sum_{i\in[m]} p_i \mathbf{M}'_i\right\|_{\infty,\infty} \le 16\tau.$$

For the second claim, note that since $\mathbf{\Sigma}^\star \in \mathcal{C}$ and $\mathbf{\Sigma}^\star \in \mathcal{C}'$, then $\mathcal{T}_\tau(\mathbf{\Sigma}^\star)$ is k-RCS and

$$\left\|\bar{\mathbf{\Sigma}} - \mathcal{T}_\tau(\mathbf{\Sigma}^\star)\right\|_{\infty,\infty} \le \left\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star\right\|_{\infty,\infty} + \|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathbf{\Sigma}^\star\|_{\infty,\infty} \le 16\tau + \tau \le 17\tau.$$

Similarly, $\left\|\bar{\mathbf{\Sigma}}' - \mathcal{T}_\tau(\mathbf{\Sigma}^\star)\right\|_{\infty,\infty} \le 17\tau$. The proof then follows using Lemma 27. □

Lemma 12 (Sensitivity of the principal eigenprojector)**.** *Condition on the event that both runs reach Line 15 of Algorithm 2. Let $\mathbf{A} := \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})$ and $\mathbf{A}' := \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')$. Let $\widehat{\mathbf{P}} := \mathbf{v}_1(\mathbf{A})\mathbf{v}_1(\mathbf{A})^\top$ and $\widehat{\mathbf{P}}' := \mathbf{v}_1(\mathbf{A}')\mathbf{v}_1(\mathbf{A}')^\top$. Then, if*

$$34k\tau \le \frac{\gamma}{16}\hat{\lambda}\ ,\ \textit{equivalently}\ n \ge \Omega\left(\frac{\sigma^4}{\hat{\lambda}^2}\cdot\frac{k^2 m}{\gamma^2}\log(d)\right), \tag{28}$$

we have,

$$\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{F}} \leq \frac{68\sqrt{2}k\tau}{\gamma\,\hat{\lambda}} =: \Delta_{\mathbf{P}}.$$

Proof. Let $\mathbf{\Sigma}^\star$ satisfy Lemma 11. Since $\gamma(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)) := 1 - \lambda_2(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))/\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)) \geq \gamma/2 > 0$, then $\mathcal{T}_\tau(\mathbf{\Sigma}^\star)$ has a unique leading eigenvector $\mathbf{v}^\star$. Let $\mathbf{P}^\star := \mathbf{v}^\star(\mathbf{v}^\star)^\top$. Furthermore, under the bound on τ in the lemma statement, using the properties of $\mathbf{\Sigma}^\star$ from Lemma 11,

$$\left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')\right\|_{\mathrm{op}},\ \left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})\right\|_{\mathrm{op}} \leq 34k\tau \leq \frac{\gamma}{16}\hat{\lambda} \leq \frac{1}{16}\cdot 2\widehat{\gamma}(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))\cdot\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)).$$

Then, using Lemma 28,

$$\left\|\widehat{\mathbf{P}} - \mathbf{P}^\star\right\|_{\mathrm{op}} \leq \frac{2\left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})\right\|_{\mathrm{op}}}{\gamma(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))\cdot\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))},\ \left\|\widehat{\mathbf{P}}' - \mathbf{P}^\star\right\|_{\mathrm{op}} \leq \frac{2\left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')\right\|_{\mathrm{op}}}{\gamma(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))\cdot\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))}.$$

Then, since $\gamma(\mathcal{T}_\tau(\mathbf{\Sigma}^\star))\cdot\lambda_1(\mathcal{T}_\tau(\mathbf{\Sigma}^\star)) \geq \frac{\gamma\hat{\lambda}}{2}$ and $\left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})\right\|_{\mathrm{op}}, \left\|\mathcal{T}_\tau(\mathbf{\Sigma}^\star) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')\right\|_{\mathrm{op}} \leq 34k\tau$, we have by triangle inequality,

$$\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{op}} \leq \frac{68k\tau}{\gamma\hat{\lambda}}.$$

Finally since $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{P}}'$ are rank one, then $\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{F}} \leq \sqrt{2}\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{op}}$, completing the proof. □

Corollary 2 (Privacy)**.** *For the choice of σ_{priv} in Algorithm 2, the final output $\widehat{\mathbf{v}}_{\mathrm{priv}}$ is (ϵ, δ)-DP.*

Proof. By Lemma 12, the Frobenius sensitivity of the map $\mathcal{D} \mapsto \widehat{\mathbf{P}}(\mathcal{D})$ on the non-abort branch is at most $\Delta_{\mathbf{P}} \leq \frac{68\sqrt{2}k\tau}{\gamma\hat{\lambda}}$. The Gaussian mechanism (Fact 1) applied to the vectorization of $\widehat{\mathbf{P}}$ yields $(\epsilon/3, \delta/3)$-DP for releasing $\widehat{\mathbf{P}} + \mathbf{W}$ when $\sigma_{\mathrm{priv}} \geq \frac{3\Delta_{\mathbf{P}}}{\epsilon}\sqrt{\log(6/\delta)}$. All subsequent steps ($\mathrm{top}_{k^2}$, symmetrization, and computing $\mathbf{v}_1(\cdot)$) are post-processing, so using Lemma 8 along with basic composition, the final output is (ϵ, δ)-DP. □

4.2 Utility analysis

Lemma 13 (Relative gap and operator norm are stable)**.** *Let $\mathbf{\Sigma} \succeq 0$ satisfy $\lambda_2(\mathbf{\Sigma})/\lambda_1(\mathbf{\Sigma}) \leq 1 - \gamma$ for some $\gamma \in (0, 1)$. Let $\mathbf{M} \succeq 0$ satisfy $\|\mathbf{M} - \mathbf{\Sigma}\|_{\mathrm{op}} \leq \eta$ with $\eta \leq \frac{\gamma}{8}\lambda_1(\mathbf{\Sigma})$. Then*

$$\lambda_1(\mathbf{M}) \geq \left(1 - \frac{\gamma}{8}\right)\lambda_1(\mathbf{\Sigma}), \quad \gamma(\mathbf{M}) := 1 - \frac{\lambda_2(\mathbf{M})}{\lambda_1(\mathbf{M})} \geq \gamma/2.$$

Proof. By Weyl's inequality,

$$\lambda_1(\mathbf{M}) \ \geq\ \lambda_1(\mathbf{\Sigma}) - \eta, \qquad \lambda_2(\mathbf{M}) \ \leq\ \lambda_2(\mathbf{\Sigma}) + \eta \ \leq\ (1-\gamma)\lambda_1(\mathbf{\Sigma}) + \eta.$$

Hence

$$\frac{\lambda_2(\mathbf{M})}{\lambda_1(\mathbf{M})} \ \leq\ \frac{(1-\gamma)\lambda_1(\mathbf{\Sigma}) + \eta}{\lambda_1(\mathbf{\Sigma}) - \eta}.$$

With $\eta \leq \frac{\gamma}{8}\lambda_1(\mathbf{\Sigma})$, then $\frac{1-\gamma+\gamma/8}{1-\gamma/8} \leq 1 - \gamma/2$, and therefore $\gamma(\mathbf{M}) = 1 - \lambda_2(\mathbf{M})/\lambda_1(\mathbf{M}) \geq \gamma/2$. □

Lemma 14 (Many blocks are certified)**.** *Under Model 2 and τ set as in (26). Assume*

$$n \geq \Omega\left(\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2} \cdot \frac{k^2 m}{\gamma^2} \log(d)\right), \text{ and } m \geq \frac{60}{\epsilon} \log\left(\frac{12}{\delta\beta}\right). \tag{29}$$

Then with probability at least $1 - \beta/8$ we have $|\mathcal{C}| \geq 0.9m$ and the test in Line 4 passes.

Proof. Fix $i \in [m]$ and consider the block covariance $\widehat{\mathbf{\Sigma}}_i$ based on $b = n/m$ samples. By Fact 7 (applied to block size b and failure $1/100$), with probability at least $99/100$ we have simultaneously:

$$\left\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\right\|_{\infty,\infty} \leq 2\tau, \qquad \mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) \text{ is } k\text{-RCS}, \qquad \left\|\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) - \mathbf{\Sigma}\right\|_{\text{op}} \leq k\tau.$$

On this event, the bound on n in (29) implies $\|\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) - \mathbf{\Sigma}\|_{\text{op}} \leq \frac{\gamma}{16}\lambda_1(\mathbf{\Sigma})$, and therefore Lemma 13 yields $\widehat{\gamma}_i = \gamma(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)) \geq \gamma/2 \geq \gamma/4$ and $\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)) \geq \left(1 - \frac{\gamma}{8}\right)\lambda_1(\mathbf{\Sigma})$. Under the stated bounds on n, m, by Theorem 6, $|\hat{\lambda} - \lambda_1(\mathbf{\Sigma})| \leq \frac{\gamma\lambda_1(\mathbf{\Sigma})}{16}$, and therefore, $\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)) \geq (1 - \gamma/2)\hat{\lambda}$. Thus $\Pr(i \in \mathcal{C}) \geq 0.99$ for each fixed i.

Let $\chi_i := \mathcal{I}\{i \in \mathcal{C}\}$ and note $\mathbb{E}[\chi_i] \geq 0.99$. Then by a standard Chernoff bound, for $\rho \in (0, 1)$, with probability at least $1 - \exp(-\frac{\rho^2}{2+\rho}\sum_{i\in[m]} \mathbb{E}[\chi_i])$,

$$(1 - \rho)\sum_{i\in[m]} \mathbb{E}[\chi_i] \leq \sum_{i\in[m]} \chi_i.$$

Then, using $\mathbb{E}[\chi_i] \geq 0.99$ and $\rho = 0.3$, we have with probability $1 - \exp(-m/60)$,

$$|\mathcal{C}| = \sum_{i\in m} \chi_i \geq 0.9m.$$

Finally, the choice of $m > \log(\frac{1}{\beta})$ leads to a total failure probability of $\beta/8$. Finally, since $|\mathcal{C}| \geq 0.9m$, the bound on m ensures that in the worst case, $L = -\frac{3}{\epsilon}\log\left(\frac{12}{\delta}\right)$, the check passes. □

Lemma 15 (Survivors are close and many)**.** *Let τ be as in (26). Then, for*

$$m \geq \frac{210}{\epsilon}\log\left(\frac{12}{\delta\beta}\right)$$

we have, with probability atleast $1 - \beta/8$, every index i with $p_i > 0$ satisfies $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 6\tau$, $Z \geq 0.9m$ and the test in Line 13 passes.

Proof. Using Fact 7, for any fixed $i \in [m]$, with probability atleast $1 - \frac{1}{d^{10}} \geq 0.99$, $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 2\tau$. Defining indicator random variables $\theta_i := \mathcal{I}(\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 2\tau)$ and using the Chernoff argument as in Lemma 14, with probability atleast $1 - \exp(-m/60)$,

$$\sum_{i\in m} \theta_i \geq 0.9m.$$

Therefore, for 0.9 fraction of $i \in [m]$, $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 2\tau$ with probability atleast $1 - \exp(-m/60)$. We condition on this event for the rest of the proof and denote the set of such points by $\mathcal{G}$.

Consider some index i such that $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} > 6\tau$. Then clearly, the balls $\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau)$ and $\mathbb{B}_\infty(\mathbf{\Sigma}, 2\tau)$ are disjoint. However, by our previous argument $|\mathbb{B}_\infty(\mathbf{\Sigma}, 2\tau)| \geq |\mathcal{G}| \geq 0.9m$. Hence $f_i \leq |\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 4\tau)| \leq 0.1m$. Hence, for these indices $p_i = 0$. This establishes the first conclusion.

Conversely, if $i \in \mathcal{G}$ then for every $j \in \mathcal{G}$, $\|\widehat{\mathbf{\Sigma}}_j - \widehat{\mathbf{\Sigma}}_i\|_{\infty,\infty} \leq \|\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}\|_{\infty,\infty} + \|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 4\tau$, so $f_i \geq 0.9m$ and thus $p_i = 1$. Therefore $Z = \sum_i p_i \geq \sum_{i\in\mathcal{G}} p_i \geq 0.9m$.

Finally, the test in Line 4 passes since the bound on m ensures that in the worst case, for $\xi = -\frac{21}{\epsilon}\log\left(\frac{12}{\delta}\right)$, the test is successful under the event $Z \geq 0.9m$. □

Lemma 16 (Top-k^2 truncation controls operator norm)**.** *Let* $\mathbf{P} := \mathbf{v}_1(\mathbf{\Sigma})\mathbf{v}_1(\mathbf{\Sigma})^\top$ *and note* $\mathbf{P}$ *has at most* k^2 *nonzero entries. Suppose* $\mathbf{X} \in \mathbb{S}^{d\times d}$ *satisfies* $\|\mathbf{X} - \mathbf{P}\|_{\infty,\infty} \leq \eta$ *and define* $\widetilde{\mathbf{P}} := \mathrm{top}_{k^2}(\mathbf{X})$ *(ties broken arbitrarily). Then*

$$\left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\mathrm{F}} \;\leq\; 2\eta\sqrt{2k^2} \;= 2\sqrt{2}k\eta.$$

Proof. Fix any entry (i, j). If $(\mathbf{P})_{ij} = 0$, then $|\mathbf{X}_{ij}| \leq \eta$. Hence any false positive entry kept by top_{k^2} has magnitude at most η. Conversely, if $(\mathbf{P})_{ij} \neq 0$ and $|(\mathbf{P})_{ij}| > 2\eta$, then $|\mathbf{X}_{ij}| \geq |(\mathbf{P})_{ij}| - \eta > \eta$, while every zero entry in $\mathbf{P}$ has magnitude at most η in $\mathbf{X}$; since $\mathbf{P}$ has at most k^2 nonzero entries, top_{k^2} cannot truncate an element larger in magnitude than 2η. Thus any true entry that is dropped must satisfy $|(\mathbf{P})_{ij}| \leq 2\eta$. Therefore, the difference $\widetilde{\mathbf{P}} - \mathbf{P}$ is supported on at most $2k^2$ entries and each such entry has magnitude at most 2η. Hence $\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}} \leq 2\eta\sqrt{2k^2}$, and $\|.\|_{\mathrm{op}} \leq \|.\|_{\mathrm{F}}$. □

Lemma 17 (Entrywise control of the projector)**.** *Let* $\mathbf{A} := \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})$, $\widehat{\mathbf{v}} := \mathbf{v}_1(\mathbf{A})$, $\widehat{\mathbf{P}} := \widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top$ *and* $\mathbf{P} := \mathbf{v}_1(\mathbf{\Sigma})\mathbf{v}_1(\mathbf{\Sigma})^\top$*. Then, for*

$$n \geq \Omega\left(\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2}\cdot\frac{k^2 m}{\gamma^2}\log\left(\frac{dm}{\beta}\right)\right), \text{ and } m \geq \Omega\left(\frac{1}{\epsilon}\log\left(\frac{1}{\beta\delta}\right)\right)$$

we have with probability atleast $1 - \beta/2$*,*

$$\left\|\widehat{\mathbf{P}} - \mathbf{P}\right\|_{\infty,\infty} \;\leq\; \left\|\widehat{\mathbf{P}} - \mathbf{P}\right\|_{\mathrm{F}} \;=\; \sqrt{2}\sin\angle(\widehat{\mathbf{v}}, \mathbf{v}_1(\mathbf{\Sigma})) \;\leq\; \frac{34\sqrt{2}k\tau}{\gamma\,\lambda_1(\mathbf{\Sigma})}.$$

Proof. Let us first condition on the event, $\mathcal{E} := \{\|\mathbf{A} - \mathbf{\Sigma}\|_{\mathrm{op}} \leq 34k\tau \leq \gamma\lambda_1(\mathbf{\Sigma})/2\}$. The identity $\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}} = \sqrt{2}\sin\angle(\widehat{\mathbf{v}}, \mathbf{v}_1)$ follows using (2). Then using Lemma 28 gives $\sin\angle(\widehat{\mathbf{v}}, \mathbf{v}_1) \leq 2\|\mathbf{A} - \mathbf{\Sigma}\|_{\mathrm{op}}/(\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))$. Using $\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}) \geq \gamma\lambda_1(\mathbf{\Sigma})$ and $\|\mathbf{A} - \mathbf{\Sigma}\|_{\mathrm{op}} \leq 34k\tau$ yields the claim. Finally we use $\|\cdot\|_{\infty,\infty} \leq \|\cdot\|_{\mathrm{F}}$.

Now we show that $\mathbb{P}(\mathcal{E}) \geq 1 - \beta/2$. We first note that via a union bound over the events in Lemma 14 and 15, the stated bounds on n and m ensure that the tests in Lines 4 and 13 pass. Lemma 15 then shows that all surviving points with $p_i > 0$ satisfy $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq 6\tau$ and $Z > 0.8m$. Then

$$\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\infty,\infty} = \left\|\frac{1}{Z}\sum_{i\in[m]} p_i(\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma})\right\|_{\infty,\infty} \leq 6\tau.$$

Then since $\mathbf{\Sigma}$ is k-RCS, the conclusion on $\|\mathbf{A} - \mathbf{\Sigma}\|_{\mathrm{op}}$ follows due to Lemma 27. □

We are finally ready to present the proof of Theorem 3.

Proof of Theorem 3. Run Algorithm 2 as

$$\hat{\lambda} \leftarrow \mathsf{PrivNorm}\left(\mathcal{D}_1, \gamma, k, \sigma^2, \frac{\epsilon}{2}, \frac{\delta}{2}, \sqrt{2}\tau, \frac{\beta}{2}\right), \quad \widehat{\mathbf{v}}_{\text{priv}} \leftarrow \mathsf{PrivPCA}\left(\mathcal{D}_2, \hat{\lambda}, \gamma, k, \sigma^2, \frac{\epsilon}{2}, \frac{\delta}{2}, \sqrt{2}\tau, \frac{\beta}{2}\right)$$

with τ set as (26) with $m \geq \Omega(\frac{1}{\epsilon}\log(\frac{1}{\delta\beta}))$ and $\mathcal{D}_1 := \{\mathbf{x}_i\}_{i\in[n/2]}$, $\mathcal{D}_2 := \{\mathbf{x}_i\}_{i\in[n/2+1,n]}$ and $\mathsf{PrivNorm}$ is defined in Algorithm 3.

Define $\mathbf{P} := \mathbf{v}_1(\mathbf{\Sigma})\mathbf{v}_1(\mathbf{\Sigma})^\top$. Lemma 17 gives, with probability atleast $1-\beta/2$,

$$\left\|\widehat{\mathbf{P}} - \mathbf{P}\right\|_{\infty,\infty} \leq \frac{34\sqrt{2}k\tau}{\gamma\lambda_1(\mathbf{\Sigma})}.$$

Next, by a standard Gaussian tail bound and union bound over d^2 entries,

$$\|\mathbf{W}\|_{\infty,\infty} \leq 2\sigma_{\text{priv}}\sqrt{\log(4d/\beta)} \qquad \text{with probability at least } 1-\beta/2.$$

Since $\mathbf{X} := \widehat{\mathbf{P}} + \mathbf{W}$ satisfies $\|\mathbf{X}-\mathbf{P}\|_{\infty,\infty} \leq \left\|\widehat{\mathbf{P}} - \mathbf{P}\right\|_{\infty,\infty} + \|\mathbf{W}\|_{\infty,\infty}$, Lemma 16 yields with probability atleast $1-\beta$,

$$\left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\text{op}} \leq 2\sqrt{2}k\left(\frac{34\sqrt{2}k\tau}{\gamma\lambda_1(\mathbf{\Sigma})} + 2\sigma_{\text{priv}}\sqrt{\log(4d/\beta)}\right). \tag{30}$$

Since $\mathbf{P}$ is symmetric, the same bound holds for the symmetrization $(\widetilde{\mathbf{P}}+\widetilde{\mathbf{P}}^\top)/2$. Under the sample complexity bound on n in the theorem statement, the RHS of (30) is smaller than $1/2$. Finally, $\mathbf{P}$ has eigenvalues $(1,0,\ldots,0)$, so its eigengap is 1; Lemma 28 then implies

$$\sin\angle\left(\mathbf{v}_1(\widetilde{\mathbf{P}}), \mathbf{v}_1(\mathbf{\Sigma})\right) \leq 4\left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\text{op}},$$

and thus $\sin^2\angle(\mathbf{v}_1(\widetilde{\mathbf{P}}), \mathbf{v}_1(\mathbf{\Sigma})) \leq 16\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\text{op}}^2$. Substituting $\sigma_{\text{priv}} = \frac{6\Delta_{\mathbf{P}}}{\epsilon}\sqrt{\log(6/\delta)}$ and $\Delta_{\mathbf{P}} \leq \frac{34\sqrt{2}k\tau}{\gamma\hat{\lambda}}$ from Lemma 12, we have, with probability atleast $1-\beta$, for

$$n \geq \Omega\left(\frac{k^4 m}{\gamma^2\,\Delta}\left(\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2} + \frac{\sigma^4}{\hat{\lambda}^2\epsilon^2}\log\left(\frac{d}{\beta}\right)\right)\log(d)\log\left(\frac{1}{\delta}\right)\right), \text{ and } m \geq \Omega\left(\frac{1}{\epsilon}\log\left(\frac{1}{\delta\beta}\right)\right) \tag{31}$$

$\sin^2\angle(\mathbf{v}_1(\widetilde{\mathbf{P}}), \mathbf{v}_1(\mathbf{\Sigma})) \leq \Delta$ and $\mathbf{v}_1(\widetilde{\mathbf{P}})$ satisfies (ϵ,δ)-DP. Combining the two bounds in (31), we have

$$n \geq \Omega\left(\frac{k^4}{\gamma^2\,\Delta\epsilon^3}\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2}\log\left(\frac{d}{\beta}\right)\log(d)\log\left(\frac{1}{\delta}\right)\log\left(\frac{1}{\delta\beta}\right)\right)$$

Finally, we provide our bound for $\hat{\lambda}$. Theorem 6 shows that under the above bound, a slight variation of Algorithm 2, we have with probability $1-\beta$, an (ϵ,δ)-DP estimator $\hat{\lambda}$ satisfying

$$\left|\hat{\lambda} - \lambda_1(\mathbf{\Sigma})\right| \leq \frac{\gamma\lambda_1(\mathbf{\Sigma})}{16}, \text{ which provides } \hat{\lambda} \geq 0.9\lambda_1(\mathbf{\Sigma}). \tag{32}$$

Then, by choosing ϵ, δ, β smaller by a constant factor, and a union bound, the result follows from (31) and (32). ∎

5 Private Sparse PCA: Lower Bounds

In this section, we provide our lower bounds for Problem 2 under Model 1 with $\gamma > 0$.

5.1 Pure DP lower bound

We start by proving a $\Omega(d/\epsilon)$ sample complexity lower bound in Theorem 4 for *pure* differential privacy for the PCA task (Problem 2) over a natural family of k-RCS covariance matrices satisfying Model 1. We first have the following DP hypothesis-selection lower bound result due to [KSU20].

Proposition 1 ([KSU20], Lemma 6.2)**.** *Let $\alpha \in (0, 1]$ and $\epsilon \geq 0$, and let $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ be a family of distributions such that $\mathrm{TV}(P_i, P_j) \leq \alpha$ for all $i, j \in [m]$. Suppose $\mathcal{A} : \mathcal{X} \to [m]$ is an ϵ-DP algorithm such that for all $i \in [m]$,*

$$\mathbb{P}_{\mathbf{x}_1,\ldots,\mathbf{x}_n \overset{\text{iid}}{\sim} P_i,\, \mathcal{A}} \left[\mathcal{A}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = i\right] \geq \frac{2}{3}.$$

Then $n = \Omega(\frac{\log m}{\alpha\epsilon})$.

We require some properties of bipartite expander graphs for our lower bound instance.

Definition 4. *A bipartite graph $G = (V, E)$ is said to be (c_1, c_2)-biregular if its vertex set can be partitioned as $V = L \cup R$ such that $\deg(u) = c_1$ for all $u \in L$ and $\deg(v) = c_2$ for all $v \in R$.*

For biregular bipartite graphs, the top eigenpair has an explicit form.

Lemma 18 (Top eigenpair of a biregular bipartite graph)**.** *Let $G = (L \cup R, E)$ be a (c_1, c_2)-biregular bipartite graph, and let $\mathbf{A} \in \mathbb{R}^{|V|\times|V|}$ be its adjacency matrix. Then $\lambda_1(\mathbf{A}) = \sqrt{c_1 c_2}$, and the unit vector*

$$\mathbf{v} := \frac{1}{\sqrt{c_1|L| + c_2|R|}} \begin{pmatrix} \sqrt{c_1}\,\mathbf{1}_L \\ \sqrt{c_2}\,\mathbf{1}_R \end{pmatrix}$$

satisfies $\mathbf{A}\mathbf{v} = \lambda_1(\mathbf{A})\,\mathbf{v}$.

Proof. Write

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{0} \end{pmatrix},$$

where $\mathbf{B} \in \{0, 1\}^{|L|\times|R|}$ is the bipartite incidence matrix. Biregularity implies $\mathbf{B}\mathbf{1}_R = c_1\mathbf{1}_L$ and $\mathbf{B}^\top\mathbf{1}_L = c_2\mathbf{1}_R$. A direct multiplication shows $\mathbf{A}\mathbf{v} = \sqrt{c_1 c_2}\,\mathbf{v}$.

For the top eigenvalue, note that $\mathbf{A}^2 = \mathrm{diag}(\mathbf{B}\mathbf{B}^\top, \mathbf{B}^\top\mathbf{B})$ and $(\mathbf{B}\mathbf{B}^\top)\mathbf{1}_L = \mathbf{B}(\mathbf{B}^\top\mathbf{1}_L) = \mathbf{B}(c_2\mathbf{1}_R) = c_1c_2\,\mathbf{1}_L$, so c_1c_2 is an eigenvalue of $\mathbf{B}\mathbf{B}^\top$. Since $\mathbf{B}\mathbf{B}^\top$ is entrywise nonnegative with every row sum equal to c_1c_2, its spectral radius is exactly c_1c_2, and hence $\lambda_1(\mathbf{A}^2) = c_1c_2$, i.e. $\lambda_1(\mathbf{A}) = \sqrt{c_1c_2}$. □

For our lower bound, we use the following expander existence result of [GM21].

Proposition 2 (Theorem 1.2, [GM21])**.** *For any positive integers g, h, t, there exists a bipartite graph $G_{g,h,t} = (V, E)$ with partite sets L and R such that*

- $|L| = tg$ *and* $|R| = g$,
- $\deg(u) = h$ *for all* $u \in L$ *and* $\deg(v) = th$ *for all* $v \in R$,
- $\lambda_2(\mathbf{A}) \le \sqrt{h-1} + \sqrt{th-1}$, *where* $\mathbf{A}$ *is the adjacency matrix of* $G_{g,h,t}$.

We are now ready to describe the construction of a k-RCS covariance matrix using a bipartite expander graph.

Lemma 19 (k-RCS covariance construction)**.** *Assume* $3 \mid d$, *and let* $k \ge 6$ *be an even integer satisfying* $k \le d/3$. *Then there exists* $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ *such that:*

1. *(Sparsity)* $\mathrm{nnz}(\mathbf{\Sigma}_{i,:}) \le k$ *for all* $i \in [d]$.
2. *(Bounded entries)* $\mathbf{\Sigma}_{ii} = 1$ *for all* i, *and for* $i \ne j$, $\mathbf{\Sigma}_{ij} \in \{0, \sqrt{2}/(k-2)\}$.
3. *(Non-trivial eigengap)* $\lambda_1(\mathbf{\Sigma}) = 2$ *and* $\lambda_2(\mathbf{\Sigma}) \le 1 + \frac{\sqrt{k-4}+\sqrt{2k-6}}{k-2} < 1.97$.
4. *(Top eigenvector) writing* $V = L \cup R$ *with* $|L| = 2d/3$ *and* $|R| = d/3$,

$$\mathbf{v}_1(\mathbf{\Sigma}) = \sqrt{\frac{3}{4d}} \begin{pmatrix} 1 \cdot \mathbf{1}_L \\ \sqrt{2} \cdot \mathbf{1}_R \end{pmatrix}.$$

Proof. Let $g = d/3$, $t = 2$, and $h = (k-2)/2$ in Proposition 2, and let $\mathbf{A}$ be the adjacency matrix of the resulting $(h, 2h)$-biregular bipartite graph on d vertices with bipartition (L, R). Then $\mathbf{A}$ has maximum degree $2h = k - 2$.

Define

$$\mathbf{\Sigma} := \mathbf{I}_d + \frac{\sqrt{2}}{k-2}\mathbf{A}.$$

Then $\mathbf{\Sigma}$ is symmetric with diagonal entries 1 and off-diagonal entries in $\{0, \sqrt{2}/(k-2)\}$. Moreover, each row of $\mathbf{A}$ has at most $k-2$ nonzeros, so each row of $\mathbf{\Sigma}$ has at most $(k-2)+1 = k-1 \le k$ nonzeros, proving the k-RCS claim.

Since G is $(h, 2h)$-biregular, Lemma 18 gives $\lambda_1(\mathbf{A}) = \sqrt{2}\,h = (k-2)/\sqrt{2}$ and

$$\lambda_1(\mathbf{\Sigma}) = 1 + \frac{\sqrt{2}}{k-2}\lambda_1(\mathbf{A}) = 1 + \frac{\sqrt{2}}{k-2} \cdot \frac{k-2}{\sqrt{2}} = 2.$$

Also $\mathbf{A}$ is bipartite so $-\lambda_1(\mathbf{A})$ is an eigenvalue of $\mathbf{A}$, implying $\lambda_{\min}(\mathbf{\Sigma}) = 0$ and hence $\mathbf{\Sigma} \succeq \mathbf{0}$.

For λ_2, Proposition 2 yields $\lambda_2(\mathbf{A}) \le \sqrt{h-1} + \sqrt{2h-1}$ and hence

$$\lambda_2(\mathbf{\Sigma}) \le 1 + \frac{\sqrt{2}}{k-2}\left(\sqrt{h-1} + \sqrt{2h-1}\right) = 1 + \frac{\sqrt{k-4} + \sqrt{2k-6}}{k-2}.$$

The numerical bound < 1.97 holds for all $k \ge 6$.

Finally, Lemma 18 gives the top eigenvector of $\mathbf{A}$ proportional to the vector

$$\begin{pmatrix} \sqrt{h}\mathbf{1}_L \\ \sqrt{2h}\mathbf{1}_R \end{pmatrix},$$

normalizing and using $|L| = 2d/3$, $|R| = d/3$ yields the displayed formula for $\mathbf{v}_1(\mathbf{\Sigma})$. Since $\mathbf{\Sigma} = \mathbf{I} + \frac{\sqrt{2}}{k-2}\mathbf{A}$ shares eigenvectors with $\mathbf{A}$, this is also $\mathbf{v}_1(\mathbf{\Sigma})$. □

We next require a standard result showing the existence of a packing on the hypercube $\{0,1\}^{3m}$ with every pair separated by at least a Hamming distance, $d_{\text{ham}}(\mathbf{x},\mathbf{y}) := \sum_{i\in[3m]} \mathcal{I}(\mathbf{x}_i \neq \mathbf{y}_i)$ of $m/2$. This is classically attributed to Gilbert and Varshamov (see e.g Theorem 7 [GS03]).

Lemma 20. *Let m be a positive integer and let $\mathcal{S}$ be the set of all binary strings of length $3m$ and Hamming weight m. Then, there exists a subset $C_m \subseteq \mathcal{S}$ of size $\exp(\Omega(m))$ such that for any distinct $\mathbf{x},\mathbf{y} \in C$, $d_{\text{ham}}(\mathbf{x},\mathbf{y}) \geq m/2$.*

Proof. For any distinct strings $x, y \in \mathcal{S}$, with $d_{\text{ham}}(x,y) = 2t$, where t is the number of indices i such that $x_i = 0$ and $y_i = 1$, we have for any x, the number of strings $y \neq x$ such that $d_{\text{ham}}(x,y) \leq m/2$ is equal to

$$\sum_{t=1}^{m/4} \binom{m}{t}\binom{2m}{t} \leq \frac{m}{4}\binom{m}{m/4}\binom{2m}{m/4} \leq \frac{m}{4} 2^{mH_2(1/4)+2mH_2(1/8)} \leq \frac{m\cdot 2^{1.9m}}{4}.$$

where for $p \in (0,1)$, $H_2(p) := -p\log(p) - (1-p)\log(1-p)$. Since $|\mathcal{S}| = \binom{3m}{m}$, a subset C of size at least

$$\frac{\binom{3m}{m}}{\frac{m\cdot 2^{1.9m}}{4}} \geq \frac{4}{m\cdot 2^{1.9m}} \cdot \frac{0.4}{\sqrt{m}} \left(\frac{27}{4}\right)^m \geq \frac{6\cdot 4^m}{m^{3/2}} \geq \frac{2\times 1.8^m}{m^{1.5}},$$

with the desired Hamming distance property exists. □

Theorem 4 (Pure-DP PCA lower bound for k-RCS covariance)**.** *Assume $3 \mid d$, and let $k \geq 6$ be an even integer with $k \leq d/3$. Fix any $\epsilon \geq 0$. Suppose there exists an ϵ-DP algorithm $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ such that for every $\mathbf{\Sigma}, \mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ generated from Model 1,*

$$\mathbb{P}\left[\sin^2 \angle(\mathcal{A}(X_1,\ldots,X_n),\ \mathbf{v}_1(\mathbf{\Sigma})) \leq \frac{1}{400}\right] \ \geq\ \frac{2}{3}.$$

Then we must have $n = \Omega(d/\epsilon)$.

Proof. Let $\mathbf{\Sigma}_\star$ be the matrix from Lemma 19. Let $\mathcal{K}$ be the collection of all covariance matrices $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ such that $\text{nnz}(\mathbf{\Sigma}_{i,:}) \leq k$ for all $i \in [d]$ and $\max_{i,j} |\mathbf{\Sigma}_{ij}| \leq 1$. Set $m = d/3$ and take a code $\mathcal{C}_m \subseteq \{0,1\}^d$ from Lemma 20, so $|\mathcal{C}_m| = \exp(\Omega(d))$, every $\mathbf{x} \in \mathcal{C}_m$ has Hamming weight $m = d/3$, and for distinct $\mathbf{x},\mathbf{y} \in \mathcal{C}_m$,

$$d_{\text{ham}}(x,y) \ \geq\ \frac{m}{2} \ =\ \frac{d}{6}.$$

For each $\mathbf{x} \in \mathcal{C}_m$, let $\Pi_\mathbf{x}$ be any permutation matrix that maps the coordinates with $\mathbf{x}_i = 1$ onto the R-block (of size $d/3$) and those with $\mathbf{x}_i = 0$ onto the L-block (of size $2d/3$), relative to the bipartition (L,R) in Lemma 19. Define

$$\mathbf{\Sigma}_\mathbf{x} \ :=\ \Pi_\mathbf{x}\mathbf{\Sigma}_\star\Pi_\mathbf{x}^\top, \qquad P_\mathbf{x} \ :=\ \mathcal{N}(0,\mathbf{\Sigma}_\mathbf{x}).$$

Permutation preserves PSD-ness, entrywise bounds, and row sparsity, so $\mathbf{\Sigma}_\mathbf{x} \in \mathcal{K}$ for all $\mathbf{x} \in \mathcal{C}_m$.

Let $\mathcal{D}_\mathbf{x} := \mathcal{N}(0,\mathbf{\Sigma}_\mathbf{x})$. Then for every $\mathbf{x}$:

1. $\mathbf{\Sigma_x} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ and $\mathrm{nnz}((\mathbf{\Sigma_x})_{i,:}) \le k$ for all $i \in [d]$.
2. $\lambda_1(\mathbf{\Sigma_x}) = 2$ and
$$\frac{\lambda_2(\mathbf{\Sigma_x})}{\lambda_1(\mathbf{\Sigma_x})} \;\le\; \frac{1}{2}\Big(1 + \frac{\sqrt{k-4}+\sqrt{2k-6}}{k-2}\Big) \;\le\; 1-\gamma_0,$$
where
$$\gamma_0 \;:=\; 1 - \frac{1}{2}\Big(1 + \frac{\sqrt{k-4}+\sqrt{2k-6}}{k-2}\Big) \in (0, 1/2).$$
In particular, for all $k \ge 6$ one may take $\gamma = 0.01$.
3. $\mathcal{D}_{\mathbf{x}}$ is σ-sub-Gaussian in the sense of Definition 3 with $\sigma = \sqrt{2}$, and has covariance $\mathbf{\Sigma_x}$.

Consequently, for any sample size n, i.i.d. samples $\mathbf{x}_1, \dots, \mathbf{x}_n \overset{\text{iid}}{\sim} \mathcal{D}_{\mathbf{x}}$ satisfy Model 1 with parameters $(k, d, n, \gamma, \sigma) = (k, d, n, \gamma_0, \sqrt{2})$ (or with any $\gamma \le \gamma_0$, e.g. $\gamma = 0.01$).

Let $\mathbf{v_x} := \mathbf{v}_1(\mathbf{\Sigma_x})$. By Lemma 19, for each coordinate $i \in [d]$,
$$\mathbf{e}_i^\top \mathbf{v_x} = \begin{cases} \sqrt{\frac{3}{4d}}, & \mathbf{x}_i = 0, \\ \sqrt{\frac{3}{2d}}, & \mathbf{x}_i = 1. \end{cases}$$

Therefore, for any distinct $\mathbf{x}, \mathbf{y} \in \mathcal{C}_m$,
$$\|\mathbf{v_x} - \mathbf{v_y}\|_2^2 = d_{\mathrm{ham}}(\mathbf{x}, \mathbf{y}) \cdot \Big(\sqrt{\frac{3}{2d}} - \sqrt{\frac{3}{4d}}\Big)^2 \ge \frac{d}{6} \cdot \frac{3}{4d} \cdot (\sqrt{2}-1)^2 = \frac{(\sqrt{2}-1)^2}{8} > \frac{1}{50}. \tag{33}$$

Define the post-processed output $\widehat{\mathbf{v}}$ by normalizing $\mathcal{A}$:
$$\widehat{\mathbf{v}} \;:=\; \begin{cases} \frac{\mathcal{A}(\mathbf{x}_1,\dots,\mathbf{x}_n)}{\|\mathcal{A}(\mathbf{x}_1,\dots,\mathbf{x}_n)\|_2}, & \mathcal{A}(\mathbf{x}_1,\dots,\mathbf{x}_n) \neq \mathbf{0}_d, \\ \mathbf{e}_1, & \text{otherwise.} \end{cases}$$

This is a deterministic post-processing, hence $\widehat{\mathbf{v}}$ remains ϵ-DP. Moreover, $\sin^2 \angle(\widehat{\mathbf{v}}, \mathbf{v_x}) = \sin^2 \angle(\mathcal{A}(X), \mathbf{v_x})$ by scale-invariance of (2). Define a decoder $\mathcal{B} : (\mathbb{R}^d)^n \to \mathcal{C}_m$ by
$$\mathcal{B}(\mathbf{x}_1, \dots, \mathbf{x}_n) \;:=\; \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{C}_m} \min_{s\in\{\pm 1\}} \|\widehat{\mathbf{v}} - s\mathbf{v_x}\|_2 .$$

Again $\mathcal{B}$ is a post-processing of $\mathcal{A}$, and hence ϵ-DP. Fix any $\mathbf{x} \in \mathcal{C}_m$ and suppose $\sin^2 \angle(\widehat{\mathbf{v}}, \mathbf{v_x}) \le 1/400$. By Corollary 3, this implies
$$\min_{s\in\{\pm1\}} \|\widehat{\mathbf{v}} - s\mathbf{v_x}\|_2^2 \;\le\; 2 \cdot \frac{1}{400} \;=\; \frac{1}{200}, \qquad \text{so} \qquad \min_{s\in\{\pm1\}} \|\widehat{\mathbf{v}} - s\mathbf{v_x}\|_2 \;\le\; \sqrt{\frac{1}{200}}.$$

Let $d_{\mathrm{sign}}(\mathbf{u}, \mathbf{v}) := \min_{s\in\{\pm1\}} \|\mathbf{u} - s\mathbf{v}\|_2$. Then we first note that d_{sign} satisfies the triangle inequality. Indeed choosing $s_2 \in \arg\min_s \|\mathbf{a} - s\mathbf{c}\|_2$ and $s_1 \in \arg\min_s \|\mathbf{c} - s\mathbf{b}\|_2$ gives $d_{\mathrm{sign}}(\mathbf{a}, \mathbf{b}) \le \|\mathbf{a} - s_2 s_1 \mathbf{b}\|_2 \le \|\mathbf{a} - s_2\mathbf{c}\|_2 + \|s_2\mathbf{c} - s_1 s_2 \mathbf{b}\|_2 = d_{\mathrm{sign}}(\mathbf{a}, \mathbf{c}) + d_{\mathrm{sign}}(\mathbf{c}, \mathbf{b})$. Rearranging yields $d_{\mathrm{sign}}(\mathbf{a}, \mathbf{b}) \ge d_{\mathrm{sign}}(\mathbf{c}, \mathbf{b}) - d_{\mathrm{sign}}(\mathbf{a}, \mathbf{c})$.

In our construction $\langle \mathbf{v_x}, \mathbf{v_y} \rangle \geq 0$ so $d_{\text{sign}}(\mathbf{v_x}, \mathbf{v_y}) = \|\mathbf{v_x} - \mathbf{v_y}\|_2$, giving $\min_s \|\hat{\mathbf{v}} - s\mathbf{v_y}\|_2 \geq \|\mathbf{v_y} - \mathbf{v_y}\|_2 - \min_s \|\hat{\mathbf{v}} - s\mathbf{v_x}\|_2$. Then using the separation bound in (33),

$$\min_{s\in\{\pm 1\}} \|\widehat{\mathbf{v}} - s\mathbf{v_y}\|_2 \;\geq\; \|\mathbf{v_x} - \mathbf{v_y}\|_2 - \min_{s\in\{\pm 1\}} \|\widehat{\mathbf{v}} - s\mathbf{v_x}\|_2 \;>\; \sqrt{\frac{1}{50}} - \sqrt{\frac{1}{200}} \;=\; \sqrt{\frac{1}{200}}.$$

Hence $\mathcal{B}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbf{x}$ whenever $\sin^2 \angle(\widehat{\mathbf{v}}, \mathbf{v_x}) \leq 1/400$. By the assumed guarantee on $\mathcal{A}$, we conclude that

$$\mathbb{P}_{\mathbf{x}_1,\ldots,\mathbf{x}_n \overset{\text{iid}}{\sim} P_{\mathbf{x}}} \left[\mathcal{B}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbf{x}\right] \geq \frac{2}{3}.$$

We apply Proposition 1 to the family $\{P_{\mathbf{x}} : \mathbf{x} \in \mathcal{C}_m\}$. We may take $\alpha = 1$ since $\mathrm{TV}(P_{\mathbf{x}}, P_{\mathbf{y}}) \leq 1$ always. With $m = |\mathcal{C}_m| = \exp(\Omega(d))$, Proposition 1 gives

$$n \;=\; \Omega\left(\frac{\log |\mathcal{C}_m|}{\epsilon}\right) \;=\; \Omega\left(\frac{d}{\epsilon}\right).$$

This completes the proof. □

5.2 Approximate DP lower bound

We now provide our arguments for *approximate* differential privacy for the PCA task (Problem 2) over Model 1. We use the following private form of Assouad's Lemma due to [ASZ21], along with a proof deferred to Appendix D.

Lemma 21. *Let $\mathcal{V}$ denote a subset of $\{\pm 1\}^d$ such that each $\mathbf{s} \in \mathcal{V}$ is associated with a distribution $p_{\mathbf{s}} \in \mathcal{P}$ and parameter $\boldsymbol{\theta}(p_{\mathbf{s}})$. Let the loss function ℓ satisfy $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$,*

$$\ell\left(\boldsymbol{\theta}\left(p_{\mathbf{u}}\right), \boldsymbol{\theta}\left(p_{\mathbf{v}}\right)\right) \geq 2\tau . \sum_{i\in[d]} \mathcal{I}\left(\mathbf{u}_i \neq \mathbf{v}_i\right), \;\; \text{and } \ell\left(\boldsymbol{\theta}\left(p_{\mathbf{u}}\right), \boldsymbol{\theta}\left(p_{\mathbf{v}}\right)\right) \leq 2\left(\ell\left(\boldsymbol{\theta}\left(p_{\mathbf{u}}\right), \boldsymbol{\theta}\left(p_{\mathbf{w}}\right)\right) + \ell\left(\boldsymbol{\theta}\left(p_{\mathbf{w}}\right), \boldsymbol{\theta}\left(p_{\mathbf{v}}\right)\right)\right)$$

For each coordinate $i \in [d]$, define the mixture distributions

$$p_{+i} := \frac{1}{|\mathcal{V}_{+i}|} \sum_{\mathbf{s}\in\mathcal{V}, \mathbf{s}_i = +1} p_{\mathbf{s}}, \;\; p_{-i} := \frac{1}{|\mathcal{V}_{-i}|} \sum_{\mathbf{s}\in\mathcal{V}, \mathbf{s}_i = -1} p_{\mathbf{s}}$$

where $\mathcal{V}_{+i} := \{\mathbf{s} \in \mathcal{V}, \mathbf{s}_i = +1\}$ and $\mathcal{V}_{-i} := \{\mathbf{s} \in \mathcal{V}, \mathbf{s}_i = -1\}$. If for all $i \in [d]$, there exists a coupling $(\mathcal{X}, \mathcal{Y})$ such that $\mathcal{X} := \{\mathbf{x}_j\}_{j\in[n]} \sim p_{+i}$ and $\mathcal{Y} := \{\mathbf{y}_j\}_{j\in[n]} \sim p_{-i}$ with $\mathbb{E}\left[\sum_{j\in[n]} \mathcal{I}\left(\mathbf{x}_j \neq \mathbf{y}_j\right)\right] \leq D$, then

$$R\left(\mathcal{P}, \ell, \epsilon, \delta\right) := \min_{\hat{\boldsymbol{\theta}} \text{ is } (\epsilon,\delta)\text{-}DP} \max_{p\in\mathcal{P}} \mathbb{E}_{\mathcal{X}\sim p}\left[\ell\left(\hat{\boldsymbol{\theta}}\left(\mathcal{X}\right), \boldsymbol{\theta}\left(p\right)\right)\right] \geq \frac{\tau d}{2} . \frac{\min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\}}{|\mathcal{V}|} . \left(0.9e^{-10\epsilon D} - 10D\delta\right)$$

We start by a graph based construction of a family of k-RCS covariance matrices.

Lemma 22. *Assume d is even and fix an even integer $k \geq 6$. For each sign vector $\mathbf{s} \in \{\pm 1\}^d$, define the diagonal sign matrix $\mathbf{D_s} := \mathbf{diag}\left(\mathbf{s}\right)$ and the covariance*

$$\boldsymbol{\Sigma_s} \;:=\; \mathbf{I}_d \;+\; \frac{1}{r}\mathbf{D_s}\mathbf{A}\mathbf{D_s}.$$

for $r := k - 1$. Then there exits a matrix $\mathbf{A} \in \{0, 1\}^{d\times d}$ such that the following hold:

1. *(Row-sparsity)* $\mathrm{nnz}((\mathbf{\Sigma_s})_{i,:}) \le k$ *for all* $i \in [d]$, *i.e.* $\mathbf{\Sigma_s}$ *is* k*-RCS.*

2. *(PSD)* $\mathbf{\Sigma_s} \succeq \mathbf{0}$ *and* $\lambda_1(\mathbf{\Sigma_s}) = 2$.

3. *(Nontrivial eigengap) writing* $\mathbf{v_s} := \mathbf{v}_1(\mathbf{\Sigma_s})$,

$$\frac{\lambda_2(\mathbf{\Sigma_s})}{\lambda_1(\mathbf{\Sigma_s})} \;\le\; \frac{1+\frac{2\sqrt{r-1}}{r}}{2} \;=\; 1-\gamma_0, \qquad \gamma_0 := 1-\frac{1+\frac{2\sqrt{k-2}}{k-1}}{2} \in (0,1).$$

   In particular, γ_0 *is an absolute constant for every fixed* $k \ge 6$.

4. *(Top eigenvector)* $\mathbf{v_s} = \mathbf{s}/\sqrt{d}$.

Proof. Let $g = d/2$, $t = 1$, and $h = r$ in Proposition 2, and let $\mathbf{A}$ be the adjacency matrix of the resulting (h,h)-biregular bipartite graph, G, on d vertices with bipartition (L,R) and edge set E. Each row of $\mathbf{A}$ has exactly r nonzeros. Multiplication by $\mathbf{D_s}$ preserves the zero pattern, hence each row of $\mathbf{D_sAD_s}$ has r nonzeros; adding $2\mathbf{I}_d$ adds the diagonal, so each row of $\mathbf{\Sigma_s}$ has at most $r+1 = k$ nonzeros.

Since $\mathbf{D_s}$ is orthogonal, $\mathbf{D_sAD_s}$ is similar to $\mathbf{A}$, hence they share eigenvalues. Because G is r-regular, $\lambda_1(\mathbf{A}) = r$, and because G is bipartite, $-r$ is also an eigenvalue. Thus the eigenvalues of $\mathbf{\Sigma_s}$ are $\{1+\lambda_j(\mathbf{A})/r\}_{j\in[d]}$, so

$$\lambda_{\min}(\mathbf{\Sigma_s}) \;=\; 1+\frac{-r}{r} \;=\; 0, \qquad \lambda_1(\mathbf{\Sigma_s}) \;=\; 1+\frac{r}{r} \;=\; 2,$$

and therefore $\mathbf{\Sigma_s} \succeq \mathbf{0}$. Proposition 2 implies $|\lambda_j(\mathbf{A})| \le 2\sqrt{r-1}$ for every nontrivial j, hence

$$\lambda_2(\mathbf{\Sigma_s}) \;=\; 1+\frac{\lambda_2(\mathbf{A})}{r} \;\le\; 1+\frac{2\sqrt{r-1}}{r}.$$

Dividing by $\lambda_1(\mathbf{\Sigma_s}) = 2$ yields the claimed ratio. Let $\mathbf{1} \in \mathbb{R}^d$ be the all-ones vector. Since $\mathbf{A1} = r\mathbf{1}$, we have

$$(\mathbf{D_sAD_s})\mathbf{s} \;=\; \mathbf{D_sA}(\mathbf{D_s s}) \;=\; \mathbf{D_sA1} \;=\; r\,\mathbf{D_s1} \;=\; r\mathbf{s}.$$

Therefore $\mathbf{s}$ is a top eigenvector of $\mathbf{D_sAD_s}$ with eigenvalue r, hence also a top eigenvector of $\mathbf{\Sigma_s}$, and normalizing gives $\mathbf{v_s} = \mathbf{s}/\sqrt{d}$. □

Definition 5 (Edge-spike distribution)**.** *Fix a graph* $G = (V,E)$ *on* d *vertices that is* r*-regular, and write* $\overrightarrow{E}$ *for its set of directed edges (obtain each undirected edge in both directions), so* $|\overrightarrow{E}| = rd$. *For a sign vector* $\mathbf{s} \in \{\pm 1\}^d$, *define a single-sample distribution* $\mathcal{D}_\mathbf{s}$ *on* $\mathbb{R}^d$ *by:*

1. *draw* $(U,V) \sim \mathrm{Unif}(\overrightarrow{E})$,

2. *draw* $g \sim \mathcal{N}(0,1)$ *independently,*

3. *output*

$$\mathbf{x} \;:=\; c\,g\,\mathbf{D_s}\Big(\mathbf{e}_U + \mathbf{e}_V\Big), \qquad \textit{where } c := \sqrt{d/2}.$$

Let p_e *denote the law of* $\mathbf{X} = (\mathbf{x}_1,\dots,\mathbf{x}_n) \in (\mathbb{R}^d)^n$ *where* $\mathbf{x}_1,\dots,\mathbf{x}_n \overset{\text{iid}}{\sim} \mathcal{D}_\mathbf{s}$.

Lemma 23 (Covariance and sub-Gaussianity)**.** *Let G be r-regular and let $\mathcal{D}_{\mathbf{s}}$ be as in Definition 5. Then:*

1. $\mathbb{E}[\mathbf{x}] = 0$ *and* $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbf{\Sigma}_{\mathbf{s}}$ *where* $\mathbf{\Sigma}_{\mathbf{s}}$ *is as in Lemma 22.*
2. $\mathcal{D}_{\mathbf{s}}$ *is σ-sub-Gaussian in the sense of Definition 3 with* $\sigma^2 = d$

Proof. Conditioning on (U, V) and using $\mathbb{E}[g^2] = 1$, and $\mathbb{E}[g] = 0$,

$$\mathbb{E}[\mathbf{x}\mathbf{x}^\top \mid U, V] = \mathbf{D}_{\mathbf{s}}\Big(c^2(\mathbf{e}_U + \mathbf{e}_V)(\mathbf{e}_U + \mathbf{e}_V)^\top\Big)\mathbf{D}_{\mathbf{s}}.$$

Taking expectation over $(U, V) \sim \mathrm{Unif}(\overrightarrow{E})$ gives

$$\mathbb{E}[(\mathbf{e}_U + \mathbf{e}_V)(\mathbf{e}_U + \mathbf{e}_V)^\top] = \mathbb{E}[\mathbf{e}_U\mathbf{e}_U^\top] + \mathbb{E}[\mathbf{e}_V\mathbf{e}_V^\top] + \mathbb{E}[\mathbf{e}_U\mathbf{e}_V^\top] + \mathbb{E}[\mathbf{e}_V\mathbf{e}_U^\top].$$

Since G is a regular graph, the uniform distribution over edges naturally induces a uniform distribution over the nodes. Therefore $\mathbb{E}[\mathbf{e}_U\mathbf{e}_U^\top] = \mathbf{I}_d/d$; similarly $\mathbb{E}[\mathbf{e}_V\mathbf{e}_V^\top] = \mathbf{I}_d/d$. Also,

$$\mathbb{E}[\mathbf{e}_U\mathbf{e}_V^\top] = \frac{1}{|\overrightarrow{E}|}\sum_{(u,v)\in\overrightarrow{E}} \mathbf{e}_u\mathbf{e}_v^\top = \frac{1}{rd}\,\mathbf{A}, \qquad \mathbb{E}[\mathbf{e}_V\mathbf{e}_U^\top] = \frac{1}{rd}\,\mathbf{A}^\top = \frac{1}{rd}\,\mathbf{A}.$$

Hence

$$\mathbb{E}[(\mathbf{e}_U + \mathbf{e}_V)(\mathbf{e}_U + \mathbf{e}_V)^\top] = \frac{2}{d}\mathbf{I}_d + \frac{2}{rd}\mathbf{A}.$$

With $c^2 = d/2$, we conclude

$$\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbf{D}_{\mathbf{s}}\Big(\frac{d}{2}\Big(\frac{2}{d}\mathbf{I}_d + \frac{2}{rd}\mathbf{A}\Big)\Big)\mathbf{D}_{\mathbf{s}} = \mathbf{D}_{\mathbf{s}}\Big(\mathbf{I}_d + \frac{1}{r}\mathbf{A}\Big)\mathbf{D}_{\mathbf{s}} = \mathbf{I}_d + \frac{1}{r}\mathbf{D}_{\mathbf{s}}\mathbf{A}\mathbf{D}_{\mathbf{s}} = \mathbf{\Sigma}_{\mathbf{s}}.$$

Fix any unit $\mathbf{u} \in \mathbb{R}^d$. Condition on (U, V) and note that $\mathbf{u}^\top\mathbf{D}_{\mathbf{s}}(\mathbf{e}_U + \mathbf{e}_V)$ is deterministic given (U, V). Thus

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(t\,\mathbf{u}^\top\mathbf{x}\Big)\;\Big|\;U, V\Big] &= \mathbb{E}\Big[\exp\Big(t\,c\,g\,\mathbf{u}^\top\mathbf{D}_{\mathbf{s}}(\mathbf{e}_U + \mathbf{e}_V)\Big)\Big] \\ &= \exp\left(\frac{t^2c^2}{2}\,\Big(\mathbf{u}^\top\mathbf{D}_{\mathbf{s}}(\mathbf{e}_U + \mathbf{e}_V)\Big)^2\right).\end{aligned}$$

Moreover, $\|\mathbf{D}_{\mathbf{s}}(\mathbf{e}_U + \mathbf{e}_V)\|_2^2 = \|\mathbf{e}_U + \mathbf{e}_V\|_2^2 = 2$, hence $\big(\mathbf{u}^\top\mathbf{D}_{\mathbf{s}}(\mathbf{e}_U + \mathbf{e}_V)\big)^2 \le \|\mathbf{u}\|_2^2 \cdot 2 = 2$. Therefore

$$\mathbb{E}\Big[\exp\Big(t\,\mathbf{u}^\top\mathbf{x}\Big)\;\Big|\;U, V\Big] \le \exp\left(\frac{t^2c^2}{2}\cdot 2\right) = \exp\left(\frac{t^2d}{2}\right).$$

Taking expectation over (U, V) preserves the bound, proving $\sigma^2 = d$. □

Definition 6. *Let $m := d/2$. For each $\mathbf{s} \in \{\pm 1\}^m$, define its embedding $\bar{\mathbf{s}} \in \{\pm 1\}^d$ by*

$$\bar{\mathbf{s}} \;:=\; (\mathbf{1}_{d-m},\, \mathbf{x}) \in \{\pm 1\}^d, \qquad \textit{and write}\;\; \mathbf{\Sigma}_{\mathbf{s}} := \mathbf{\Sigma}_{\bar{\mathbf{s}}},\;\; \mathbf{v}_{\mathbf{s}} := \mathbf{v}_{\bar{\mathbf{s}}}.$$

Let $\mathcal{V} := \{\pm 1\}^m$ index the resulting family $\{\mathbf{\Sigma}_{\mathbf{s}}\}_{\mathbf{s}\in\mathcal{V}}$.

Lemma 24. *For any $\mathbf{s}, \mathbf{s}' \in \mathcal{V}$, letting $t := d_{\mathrm{ham}}(\mathbf{s}, \mathbf{s}')$, we have*

$$\langle \mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'} \rangle \;=\; 1 - \frac{2t}{d} \qquad \textit{and} \qquad \sin^2 \angle(\mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'}) = 1 - \langle \mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'} \rangle^2 \;\geq\; \frac{2t}{d}.$$

In particular, the loss $\ell(\mathbf{u}, \mathbf{v}) := \sin^2 \angle(\mathbf{u}, \mathbf{v})$ satisfies, for all $\mathbf{s}, \mathbf{s}' \in \mathcal{V}$,

$$\ell(\mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'}) \;\geq\; 2\tau \sum_{i \in [m]} \mathbf{1}\left(\mathbf{s}_i \neq \mathbf{s}'_i\right) \qquad \textit{with} \;\; \tau := \frac{1}{d}.$$

Proof. Since $\mathbf{v}_{\mathbf{s}} = \bar{\mathbf{s}}/\sqrt{d}$, we have

$$\langle \mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'} \rangle = \frac{1}{d} \sum_{j=1}^{d} \bar{\mathbf{s}}_j \bar{\mathbf{s}}'_j = \frac{d-m}{d} + \frac{1}{d} \sum_{i=1}^{m} \mathbf{s}_i \mathbf{s}'_i = \frac{d-m}{d} + \frac{m-2t}{d} = 1 - \frac{2t}{d}.$$

Moreover, $t \leq m = d/2$ implies $\langle \mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'} \rangle \geq 0$, and hence

$$1 - \langle \mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'} \rangle^2 = 1 - \left(1 - \frac{2t}{d}\right)^2 = \frac{4t}{d}\left(1 - \frac{t}{d}\right) \;\geq\; \frac{2t}{d}.$$

The final display is immediate from $t = \sum_{i \in [m]} \mathbf{1}(\mathbf{s}_i \neq \mathbf{s}'_i)$ and $\tau = 1/d$. □

Lemma 25 (Coupling for the coordinate mixtures)**.** *Fix any coordinate $i \in [m]$. Let p_{+i} be the mixture of $\{p_{\mathbf{s}} : \mathbf{x} \in \mathcal{V},\ \mathbf{s}_i = +1\}$ and p_{-i} the mixture of $\{p_{\mathbf{s}} : \mathbf{x} \in \mathcal{V},\ \mathbf{s}_i = -1\}$ (as in Lemma 21) for $\mathcal{V}$ defined in Definition 6. Then there exists a coupling $(\mathcal{X}, \mathcal{Y})$ of p_{+i} and p_{-i} with $\mathcal{X} := \{\mathbf{x}_j\}_{j \in [n]} \sim p_{+i}$ and $\mathcal{Y} := \{\mathbf{y}_j\}_{j \in [n]} \sim p_{-i}$ such that*

$$\mathbb{E}\left[\sum_{j \in [n]} \mathcal{I}\left(\mathbf{x}_j \neq \mathbf{y}_j\right)\right] \;\leq\; \frac{2n}{d}.$$

Proof. Let $\mathcal{V}_{+i} := \{\mathbf{e} \in \mathcal{V}, \mathbf{e}_i = +1\}$ and $\mathcal{V}_{-i} := \{\mathbf{e} \in \mathcal{V}, \mathbf{e}_i = -1\}$. Sample $\mathbf{s} \sim \mathsf{Unif}(\mathcal{V}_{+i})$ and let $\mathbf{s}' \in \mathcal{V}_{-i}$ be obtained by flipping only the i^{th} coordinate of $\mathbf{s}$. Note that since $\mathcal{V} := \{\pm 1\}^m$, then $|\mathcal{V}_{+i}| = |\mathcal{V}_{-i}| = 2^{m-1}$ and each $\mathbf{s}$ is mapped to a unique $\mathbf{s}'$.

Now generate i.i.d. base randomness for $j \in [n]$:

$$(U_j, V_j) \sim \mathrm{Unif}(\overrightarrow{E}), \quad g_j \sim \mathcal{N}(0, 1),$$

all mutually independent. Define

$$\mathbf{x}_j \;:=\; \mathbf{D}_{\bar{\mathbf{s}}}\Big(c\, g_j(\mathbf{e}_{U_j} + \mathbf{e}_{V_j})\Big).$$

To couple p_{-i}, set

$$\mathbf{y}_j \;:=\; \mathbf{D}_{\bar{\mathbf{s}}'}\Big(c\, g_j(\mathbf{e}_{U_j} + \mathbf{e}_{V_j})\Big).$$

If $i \notin \{U_j, V_j\}$, then $\mathbf{D}_{\bar{\mathbf{s}}}(\mathbf{e}_{U_j} + \mathbf{e}_{V_j}) = \mathbf{D}_{\bar{\mathbf{s}}'}(\mathbf{e}_{U_j} + \mathbf{e}_{V_j})$, and hence $\mathbf{x}_j = \mathbf{y}_j$ on this event. Therefore

$$\mathbb{P}\left(\mathbf{x}_j \neq \mathbf{y}_j\right) \;\leq\; \mathbb{P}\left(i \in \{U_j, V_j\}\right) = \mathbb{P}(U_j = i) + \mathbb{P}(V_j = i) = \frac{1}{d} + \frac{1}{d} = \frac{2}{d}.$$

Summing over $j \in [n]$ yields the desired bound on the expected Hamming distance between the datasets $\mathcal{X}$ and $\mathcal{Y}$. □

Lemma 26. *Assume d is even and let $k \geq 6$ be even. Let $r := k-1$ and let G be an r-regular bipartite Ramanujan graph on d vertices. Let $\mathcal{P} := \{p_{\mathbf{s}} : \mathbf{s} \in \mathcal{V}\}$ be the family of n-sample distributions from Definition 5 paired with the padded subcube $\mathcal{V}$ in Definition 6. Consider the PCA loss $\ell(\hat{\mathbf{v}}, \mathbf{v}) := \sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v})$ and define the private minimax risk*

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;:=\; \min_{\hat{\mathbf{v}} \text{ is } (\epsilon,\delta)\text{-DP}} \max_{\mathbf{s}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}} \left[\sin^2 \angle\left(\hat{\mathbf{v}}(\mathcal{X}), \mathbf{v}_{\mathbf{s}}\right)\right], \; \mathcal{X} := \{\mathbf{x}_j\}_{j\in[n]} \sim p_{\mathbf{s}}$$

Then,

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;\geq\; \frac{1}{8}\left(0.9 \exp\left(-\frac{20\epsilon n}{d}\right) \;-\; \frac{20 n\delta}{d}\right),$$

Moreover, every $p_{\mathbf{s}}$ satisfies Model 1 with parameters $(k, d, n, \gamma, \sigma) = (k, d, n, \gamma_0, \sqrt{d})$, where γ_0 is the constant from Lemma 22.

Proof. By Lemma 22 and Lemma 23, under $p_{\mathbf{s}}$ we have i.i.d. samples from a σ-sub-Gaussian distribution with $\sigma = \sqrt{d+1}$, covariance $\mathbf{\Sigma}_{\mathbf{s}}$ that is k-RCS and satisfies the eigengap ratio with $\gamma = \gamma_0$, so Model 1 holds.

Next, by Lemma 24, the loss satisfies the Assouad separation condition with $\tau = 1/d$ over the hypercube $\mathcal{V} = \{\pm 1\}^m$, where $m = d/2$.

Finally, for each coordinate $i \in [m]$, Lemma 25 provides a coupling between the mixtures p_{+i} and p_{-i} with expected sample Hamming distance at most $D \leq 2n/d$.

Applying Lemma 21 with $|\mathcal{V}_{+i}| = |\mathcal{V}_{-i}| = 2^{m-1}$ and $|\mathcal{V}| = 2^m$ gives

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;\geq\; \frac{\tau m}{2} \cdot \frac{1}{2} \cdot \left(0.9 e^{-10\epsilon D} - 10 D\delta\right) = \frac{m}{4d}\left(0.9 e^{-10\epsilon D} - 10 D\delta\right).$$

Substituting $m = d/2$ and $D \leq 2n/d$ yields the claimed bound. □

Theorem 5 (Approx-DP PCA lower bound for k-RCS covariances)**.** *Assume d is even and let $k \geq 6$ be even. Fix privacy parameters $\epsilon \in (0, 1]$ and $\delta \leq \frac{\epsilon}{100}$. Let γ_0 be the constant from Lemma 22. There exists a family of distributions $\{\mathcal{D}_{\mathbf{s}}\}_{\mathbf{s}\in\{\pm1\}^{d/2}}$ over $\mathbb{R}^d$ with covariances $\mathbf{\Sigma}_{\mathbf{s}} := \mathrm{Cov}(\mathcal{D}_{\mathbf{s}})$ such that, for every $\mathbf{s}$, the pair $(\mathbf{\Sigma}_{\mathbf{s}}, \mathcal{D}_{\mathbf{s}})$ satisfies Model 1 with parameters $(k, d, n, \gamma, \sigma) = (k, d, n, \gamma_0, \sqrt{d})$. Moreover, if there exists an (ϵ, δ)-DP algorithm $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ such that for every $\mathbf{s}$ and i.i.d. samples $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}_{\mathbf{s}}$,*

$$\mathbb{E}\Big[\sin^2 \angle(\mathcal{A}(\mathcal{X}),\; \mathbf{v}_1(\mathbf{\Sigma}_{\mathbf{s}}))\Big] \;\leq\; \frac{1}{100},$$

then necessarily $n \geq \frac{d}{100\,\epsilon}$.

Proof. Fix d, k, ϵ, δ as in the theorem statement. Let $r := k - 1$ and let G be an r-regular bipartite Ramanujan graph on d vertices. Consider the associated family $\mathcal{P} = \{p_{\mathbf{s}} : \mathbf{s} \in \mathcal{V}\}$ from Lemma 26, where $p_{\mathbf{s}}$ is the law of n i.i.d. samples from the edge-spike distribution $\mathcal{D}_{\mathbf{s}}$ (Definitions 5 and 6).

By Lemma 26, for every $\mathbf{s} \in \mathcal{V}$ the distribution $\mathcal{D}_{\mathbf{s}}$ is σ-sub-Gaussian with $\sigma = \sqrt{d}$ and covariance $\mathbf{\Sigma}_{\mathbf{s}}$ that is k-RCS with parameter k and satisfies the eigengap ratio $\lambda_2(\mathbf{\Sigma}_{\mathbf{s}})/\lambda_1(\mathbf{\Sigma}_{\mathbf{s}}) \leq 1 - \gamma_0$. Hence $(\mathbf{\Sigma}_{\mathbf{s}}, \mathcal{D}_{\mathbf{s}})$ lies in Model 1 with parameters $(k, d, n, \gamma, \sigma) = (k, d, n, \gamma_0, \sqrt{d})$.

Define the loss $\ell(\hat{\mathbf{v}}, \mathbf{v}) := \sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v})$ and the private minimax risk

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;:=\; \min_{\hat{\mathbf{v}} \text{ is } (\epsilon,\delta)\text{-DP}} \; \max_{\mathbf{s}\in\mathcal{V}} \; \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}} \left[\sin^2 \angle\left(\hat{\mathbf{v}}(\mathcal{X}), \mathbf{v}_{\mathbf{s}}\right)\right].$$

Lemma 26 gives

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;\geq\; \frac{1}{8}\left(0.9 \exp\left(-\frac{20\epsilon n}{d}\right) \;-\; \frac{20n\delta}{d}\right). \tag{34}$$

Suppose for contradiction that $n < d/(100\epsilon)$. Since $\epsilon n/d < 1/100$, we have

$$\exp\left(-\frac{20\epsilon n}{d}\right) \;\geq\; \exp\left(-\frac{1}{5}\right) \;\geq\; 0.8.$$

Moreover, using $\delta \leq \epsilon/100$ and $n < d/(100\epsilon)$,

$$\frac{20n\delta}{d} \;\leq\; 20 \cdot \frac{1}{100\epsilon} \cdot \delta \;\leq\; 20 \cdot \frac{1}{100\epsilon} \cdot \frac{\epsilon}{100} \;=\; \frac{1}{500}.$$

Substituting these into (34) yields

$$R(\mathcal{P}, \ell, \epsilon, \delta) \;\geq\; \frac{1}{8}\left(0.9 \cdot 0.8 - \frac{1}{500}\right) \;\geq\; \frac{1}{8}\left(0.72 - \frac{1}{500}\right) \;>\; \frac{1}{100}.$$

This contradicts the assumed guarantee $\max_{\mathbf{s}\in\mathcal{V}} \mathbb{E}[\sin^2 \angle(\mathcal{A}(\mathcal{X}), \mathbf{v}_1(\mathbf{\Sigma}_{\mathbf{s}}))] \leq 1/100$, since that would imply $R(\mathcal{P}, \ell, \epsilon, \delta) \leq 1/100$ by taking $\hat{\mathbf{v}} = \mathcal{A}$ in the definition of the minimax risk. Therefore, we must have $n \geq d/(100\epsilon)$, completing the proof. □

Remark 1. *For comparison, the construction in Theorem 5 has two natural non-private benchmarks. If one applies the generic Model 1 analysis based only on σ-sub-Gaussianity and k-RCS structure, thresholding gives $\|\Sigma - \Sigma_s\|_{\text{op}} \lesssim k\sigma^2\sqrt{\log d/n}$. Since here $\sigma^2 = d$, $\lambda_1(\Sigma_s) = \Theta(1)$, and the eigengap is constant, this generic route requires $n \gtrsim d^2k^2 \log d$ samples for constant PCA error. This reflects the spikiness of the construction rather than an intrinsic non-private difficulty.*

For the specific edge-spike family in Theorem 5, there is also a simple distribution-specific non-private interpretation. Each sample reveals a sampled edge (U, V) of the underlying graph together with the parity $s_U s_V$, since $\operatorname{sign}(x_U x_V) = s_U s_V$. Thus recovering the leading eigenvector $v_s = s/\sqrt{d}$ reduces non-privately to recovering the vertex signs from sampled edge parities, up to a global sign. This gives the elementary bounds $\Omega(d) \leq n_{\text{nonpriv,family}} \leq O(kd \log d)$ where the upper bound follows by coupon-collecting all $O(kd)$ edges of the base graph. Thus $\widetilde{O}(kd)$ samples suffice non-privately.

Theorem 5 shows that under approximate DP, any algorithm with constant expected $\sin^2$ error requires $n = \Omega(d/\varepsilon)$. Hence the theorem should be viewed as a partial approximate-DP lower bound showing an ambient-dimensional privacy barrier for a spiky k-RCS family.

Remark 2 (From constant-probability accuracy to expected error via boosting)**.** *Theorem 5 is stated for algorithms achieving a small* expected $\sin^2$ *error. A standard boosting argument shows that essentially the same lower bound applies to algorithms that succeed with constant probability, up to an additional logarithmic factor in the sample size. Concretely, suppose $\mathcal{A}_0 : (\mathbb{R}^d)^{n_0} \to \mathbb{R}^d$ is an (ϵ, δ)-DP algorithm such that for every $\mathbf{\Sigma}, \mathcal{X} := \{\mathbf{x}_i\}_{i\in[n_0]} \overset{\text{iid}}{\sim} \mathcal{D}$ generated from Model 1,*

$$\mathbb{P}\left[\sin^2 \angle(\mathcal{A}_0(\mathcal{X}),\; \mathbf{v}_1(\mathbf{\Sigma})) \leq \alpha\right] \;\geq\; \frac{2}{3},$$

for some target accuracy $\alpha \in (0,1)$. Let $T \geq 1$ and set $n := Tn_0$. Given n samples, split them into T disjoint blocks of size n_0 and run $\mathcal{A}_0$ independently on each block to obtain unit vectors $\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_T$. Define a post-processed estimator $\mathcal{A}_{\text{boost}}$ by sign-aligning these vectors and returning a robust aggregate, (see e.g. Lemma 3.10 in [KS24]). By parallel composition and post-processing, $\mathcal{A}_{\text{boost}}$ is still (ϵ, δ)-DP. Consequently, there exist absolute constants $C, c > 0$ such that

$$\mathbb{P}\left[\sin^2 \angle(\mathcal{A}_{\text{boost}}(\mathcal{X}),\ \mathbf{v}_1(\mathbf{\Sigma})) \leq C\alpha\right] \ \geq\ 1 - e^{-cT}.$$

Since $\sin^2 \angle(\cdot,\cdot) \in [0,1]$, this implies the expected-error bound

$$\mathbb{E}\Big[\sin^2 \angle(\mathcal{A}_{\text{boost}}(\mathcal{X}),\ \mathbf{v}_1(\mathbf{\Sigma}))\Big] \ \leq\ C\alpha + e^{-cT}.$$

Choosing $T = \Theta(\log(1/\alpha))$ makes the failure term $e^{-cT} \leq \alpha$, giving $\mathbb{E}[\sin^2] \leq (C+1)\alpha$. Therefore, if an (ϵ, δ)-DP algorithm achieves $\sin^2$ error at most α with probability at least $2/3$ using n_0 samples, then there is an (ϵ, δ)-DP boosted algorithm achieving expected $\sin^2$ error $O(\alpha)$ using $n = O(n_0 \log(1/\alpha))$ samples. Applying Theorem 5 to the boosted algorithm yields the same sample complexity lower bound for constant-probability algorithms, up to this additional $\log(1/\alpha)$ factor. In particular, for any fixed constant accuracy (e.g. $\alpha = 1/100$), the logarithmic factor is an absolute constant, and one still obtains $n_0 = \Omega(d/\epsilon)$.

References

[ALT24] Hilal Asi, Daogao Liu, and Kevin Tian. Private stochastic convex optimization with heavy tails: Near-optimality from simple reductions. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024*, 2024.

[ASZ21] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private Assouad, Fano, and Le Cam. In Vitaly Feldman, Katrina Ligett, and Sivan Sabato, editors, *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 48–78. PMLR, 16–19 Mar 2021.

[AW08] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *2008 IEEE international symposium on information theory*, pages 2454–2458. IEEE, 2008.

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[BL09] Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *arXiv: Statistics Theory*, 2009.

[BR13] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on learning theory*, pages 1046–1066. PMLR, 2013.

[BUV14] Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 1–10, 2014.

[CL11] Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

[CSS13] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *The Journal of Machine Learning Research*, 14(1):2905–2943, 2013.

[CU21] Albert Cheu and Jonathan R. Ullman. The limits of pan privacy and shuffle privacy for learning and estimation. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1081–1094. ACM, 2021.

[CXZ24] T. Tony Cai, Dong Xia, and Mengyue Zha. Optimal differentially private pca and estimation for spiked covariance matrices, 2024.

[CZ12] T. Tony Cai and Harrison H. Zhou. Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, 40(5):2389–2420, 2012.

[CZZ10] T. Tony Cai, Cun-Hui Zhang, and Harrison H. Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

[dGJL04] Alexandre d'Aspremont, Laurent Ghaoui, Michael Jordan, and Gert Lanckriet. A direct formulation for sparse pca using semidefinite programming. *Advances in neural information processing systems*, 17, 2004.

[DM16] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. *Journal of Machine Learning Research*, 17(141):1–41, 2016.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[DR19] 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.

[DS25] Johanna Düngler and Amartya Sanyal. An iterative algorithm for differentially private k-PCA with adaptive noise. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[DTTZ14] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20, 2014.

[GM21] Aurelien Gribinski and Adam W Marcus. Existence and polynomial time construction of biregular, bipartite ramanujan graphs of all degrees. *arXiv preprint arXiv:2108.02534*, 2021.

[GS03] Ron Graham and Neil Sloane. Lower bounds for constant weight codes. *IEEE Transactions on Information Theory*, 26(1):37–43, 2003.

[GWWL18] Jason Ge, Zhaoran Wang, Mengdi Wang, and Han Liu. Minimax-optimal privacy-preserving sparse pca in distributed systems. In *International Conference on Artificial Intelligence and Statistics*, pages 1589–1598. PMLR, 2018.

[HP14] Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. *Advances in neural information processing systems*, 27, 2014.

[HT10] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 705–714, 2010.

[JL09] Iain M. Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009. PMID: 20617121.

[KL17a] Vladimir Koltchinskii and Karim Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, pages 110–133, 2017.

[KL17b] Vladimir Koltchinskii and Karim Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics*, 45(1):121 – 157, 2017.

[KLN+11] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, 2011.

[KLSU19] Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pages 1853–1902. PMLR, 2019.

[KLTY25] Syamantak Kumar, Daogao Liu, Kevin Tian, and Chutong Yang. Private geometric median in nearly-linear time. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[KS24] Syamantak Kumar and Purnamrita Sarkar. Oja's algorithm for streaming sparse pca. *Advances in Neural Information Processing Systems*, 37:74528–74578, 2024.

[KSU20] Gautam Kamath, Vikrant Singhal, and Jonathan Ullman. Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*, pages 2204–2235. PMLR, 2020.

[KT13] Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1395–1414. SIAM, 2013.

[LKJO22] Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. Dp-pca: Statistically optimal and differentially private pca. *Advances in neural information processing systems*, 35:29929–29943, 2022.

[LLA24] Andrew Lowy, Daogao Liu, and Hilal Asi. Faster algorithms for user-level private stochastic convex optimization. *Advances in Neural Information Processing Systems*, 37:96071–96100, 2024.

[LW23] Huimin Li and Jinru Wang. Differentially private sparse covariance matrix estimation under lower-bounded moment assumption. *Mathematics*, 11(17):3670, 2023.

[Ma13] Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772 – 801, 2013.

[Nar24] Shyam Narayanan. Better and simpler lower bounds for differentially private statistical estimation. *IEEE Transactions on Information Theory*, 2024.

[NY22] Kobbi Nissim and Chao Yan. The sample complexity of distribution-free parity learning in the robust shuffle model. *J. Priv. Confidentiality*, 12(2), 2022.

[Pau07] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[QLR23] Yixuan Qiu, Jing Lei, and Kathryn Roeder. Gradient-based sparse principal component analysis with extensions to online learning. *Biometrika*, 110(2):339–360, June 2023.

[QSZ21] Gang Qiao, Weijie Su, and Li Zhang. Oneshot differentially private top-k selection. In *International Conference on Machine Learning*, pages 8672–8681. PMLR, 2021.

[RLZ09] 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.

[SU15] Thomas Steinke and Jonathan Ullman. Interactive fingerprinting codes and the hardness of preventing false discovery. In *Conference on learning theory*, pages 1588–1628. PMLR, 2015.

[SU17] Thomas Steinke and Jonathan Ullman. Tight lower bounds for differentially private selection. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 552–563. IEEE, 2017.

[TCK+22] Eliad Tsfadia, Edith Cohen, Haim Kaplan, Yishay Mansour, and Uri Stemmer. FriendlyCore: Practical differentially private aggregation. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 21828–21863. PMLR, 17–23 Jul 2022.

[Ver10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[VL13] Van Q. Vu and Jing Lei. Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics*, 41(6):2905–2947, 2013.

[Wai19] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

[WBS16] Tengyao Wang, Quentin Berthet, and Richard J. Samworth. Statistical and computational trade-offs in estimation of sparse principal components. *The Annals of Statistics*, 44(5):1896–1930, 2016.

[WX21] Di Wang and Jinhui Xu. Differentially private high dimensional sparse covariance matrix estimation. *Theoretical Computer Science*, 865:119–130, 2021.

[YZ13] Xiao-Tong Yuan and Tong Zhang. Truncated power method for sparse eigenvalue problems. *The Journal of Machine Learning Research*, 14(1):899–925, 2013.

[ZHT06] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.

[ZL16] Zeyuan Allen Zhu and Yuanzhi Li. Even faster SVD decomposition yet without agonizing pain. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 974–982, 2016.

Acknowledgments

SK gratefully acknowledges funding support from the Amazon AI PhD Fellowship.

A Utility Results

Fact 9. *If* $\mathbf{M} \in \mathbb{R}^{d\times d}$ *satisfies* $\mathrm{nnz}(\mathbf{M}_{i,:}) \le k$ *and* $\mathrm{nnz}(\mathbf{M}_{:,i}) \le k$ *for all* $i \in [d]$*, then*

$$\|\mathbf{M}\|_{\mathrm{op}} \;\le\; \sqrt{\|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty} \;\le\; k\,\|\mathbf{M}\|_{\infty,\infty},$$

where $\|\mathbf{M}\|_\infty := \max_i \sum_j |[\mathbf{M}]_{ij}|, \|\mathbf{M}\|_1 := \max_j \sum_i |[\mathbf{M}]_{ij}|$ *and* $\|\mathbf{M}\|_{\infty,\infty} = \max_{i\in[d],j\in[d]} |[\mathbf{M}]_{ij}|$.

Proof. Each row sum is at most $k\,\|\mathbf{M}\|_{\infty,\infty}$ and each column sum is at most $k\,\|\mathbf{M}\|_{\infty,\infty}$, hence $\|\mathbf{M}\|_\infty, \|\mathbf{M}\|_1 \le k\,\|\mathbf{M}\|_{\infty,\infty}$. The claim follows from $\|\mathbf{M}\|_{\mathrm{op}} \le \sqrt{\|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty}$. □

Lemma 27. *Let* $\tau > 0$*,* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d\times d}$ *with* k*-RCS* $\mathbf{A}$ *and* $\|\mathbf{B}-\mathbf{A}\|_{\infty,\infty} \le \tau$*. Then for any* $\rho > \tau$*,*

$$\|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\|_{\mathrm{op}} \le 2k\rho$$

Proof. We divide the analysis into two cases,

1. If $[\mathbf{A}]_{ij} = 0$, then $[\mathbf{B}]_{ij} \le \tau$, and therefore, $[\mathcal{T}_\rho(\mathbf{B})]_{ij} = 0$
2. Else, $\|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\|_{\infty,\infty} \le \|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{B}\|_{\infty,\infty} + \|\mathbf{B}-\mathbf{A}\|_{\infty,\infty} \le \rho + \tau \le 2\rho$

Define the matrices $\mathbf{X} \in \{0,1\}^{d\times d}, [\mathbf{X}]_{ij} = \mathcal{I}([\mathbf{A}]_{ij} \neq 0)$, and $\mathbf{Y} := |\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}|$, where $|.|$ is computed entrywise. Then for all $i, j \in [d]$,

$$[\mathbf{Y}]_{ij} \le |[\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}]_{ij}| \le 2\rho[\mathbf{X}]_{ij}$$

Then, since $\mathbf{Y}$, $2\rho\mathbf{X}$ have non-negative entries, then by Fact 6, $\|\mathbf{Y}\|_{\mathrm{op}} \le \|2\rho\mathbf{X}\|_{\mathrm{op}} \le 2\rho k$ where the last inequality follows from Fact 9. The proof the follows by noting that $\|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\|_{\mathrm{op}} \le \|\mathbf{Y}\|_{\mathrm{op}}$. □

Lemma 28. *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$*. Fix* $p \in [d-1]$ *and let* $\mathbf{V} \in \mathbb{R}^{d\times p}$ *have orthonormal columns spanning the top-*p *eigenspace of* $\mathbf{B}$*, and let* $\widehat{\mathbf{V}} \in \mathbb{R}^{d\times p}$ *have orthonormal columns spanning the top-*p *eigenspace of* $\mathbf{A}$*. Define* $\mathsf{gap} := \lambda_p(\mathbf{B}) - \lambda_{p+1}(\mathbf{B}) > 0$*, and assume that* $\|\mathbf{A}-\mathbf{B}\|_{\mathrm{op}} \le \frac{\mathsf{gap}}{4}$*. Then,*

$$\left\|\mathbf{V}\mathbf{V}^\top - \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top\right\|_{\mathrm{op}} \le \frac{2\,\|\mathbf{A}-\mathbf{B}\|_{\mathrm{op}}}{\mathsf{gap}}.$$

Proof. Let $\widehat{\mathbf{V}}_\perp \in \mathbb{R}^{d\times(d-p)}$ be any orthonormal basis for $\mathrm{span}(\widehat{\mathbf{V}})^\perp$. Let $\mathbf{E} := \mathbf{A}-\mathbf{B}$, and set

$$\mu := \frac{\lambda_p(\mathbf{B}) + \lambda_{p+1}(\mathbf{B})}{2}, \qquad \tau := \frac{\mathsf{gap}}{2}.$$

By Weyl's inequality,

$$\lambda_{p+1}(\mathbf{A}) \le \lambda_{p+1}(\mathbf{B}) + \|\mathbf{E}\|_{\mathrm{op}} = \mu - \tau + \|\mathbf{E}\|_{\mathrm{op}} \le \mu - \frac{\tau}{2} < \mu,$$

and similarly,

$$\lambda_p(\mathbf{A}) \geq \lambda_p(\mathbf{B}) - \|\mathbf{E}\|_{\mathrm{op}} = \mu + \tau - \|\mathbf{E}\|_{\mathrm{op}} \geq \mu + \frac{\tau}{2} > \mu.$$

Hence $\widehat{\mathbf{V}}_\perp$ spans the eigenspace of $\mathbf{A}$ with eigenvalues $\leq \mu$, while $\mathbf{V}$ spans the eigenspace of B with eigenvalues $\geq \mu + \tau$. Applying Wedin's theorem (Fact 3) with $\mathbf{U} = \widehat{\mathbf{V}}_\perp$ yields

$$\left\|\widehat{\mathbf{V}}_\perp^\top \mathbf{V}\right\|_{\mathrm{op}} \leq \frac{\|\mathbf{A} - \mathbf{B}\|_{\mathrm{op}}}{\tau} = \frac{2\,\|\mathbf{A} - \mathbf{B}\|_{\mathrm{op}}}{\mathsf{gap}}. \tag{35}$$

We now relate this to $\|\widehat{\mathbf{V}}_\perp^\top \mathbf{V}\|_{\mathrm{op}}$. Let $\mathbf{P} := \mathbf{V}\mathbf{V}^\top$ and $\widehat{\mathbf{P}} := \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top$. Since $\mathbf{P}^2 = \mathbf{P}$ and $\widehat{\mathbf{P}}^2 = \widehat{\mathbf{P}}$,

$$(\mathbf{P} - \widehat{\mathbf{P}})^2 = \mathbf{P} - \mathbf{P}\widehat{\mathbf{P}} - \mathbf{P}\widehat{\mathbf{P}} + \widehat{\mathbf{P}}.$$

Using $\widehat{\mathbf{P}}_\perp := \mathbf{I} - \widehat{\mathbf{P}}$ and $\mathbf{P}_\perp := \mathbf{I} - \mathbf{P}$, we can rewrite this as

$$(\mathbf{P} - \widehat{\mathbf{P}})^2 = \mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{P} + (\mathbf{I} - \mathbf{P})\widehat{\mathbf{P}}(\mathbf{I} - \mathbf{P}) = \mathbf{P}\,\widehat{\mathbf{P}}_\perp\,\mathbf{P} + \mathbf{P}_\perp\,\widehat{\mathbf{P}}\,\mathbf{P}_\perp.$$

These two terms act on orthogonal subspaces $\mathrm{span}(\mathbf{P})$ and $\mathrm{span}(\mathbf{P}_\perp)$, hence

$$\left\|(\mathbf{P} - \widehat{\mathbf{P}})^2\right\|_{\mathrm{op}} = \max\left\{\left\|\mathbf{P}\,\widehat{\mathbf{P}}_\perp\,\mathbf{P}\right\|_{\mathrm{op}},\ \left\|\mathbf{P}_\perp\,\widehat{\mathbf{P}}\,\mathbf{P}_\perp\right\|_{\mathrm{op}}\right\}.$$

Moreover,

$$\left\|\mathbf{P}\,\widehat{\mathbf{P}}_\perp\,\mathbf{P}\right\|_{\mathrm{op}} = \left\|(\widehat{\mathbf{P}}_\perp\mathbf{P})^\top(\widehat{\mathbf{P}}_\perp\mathbf{P})\right\|_{\mathrm{op}} = \left\|\widehat{\mathbf{P}}_\perp\mathbf{P}\right\|_{\mathrm{op}}^2 = \left\|\widehat{\mathbf{V}}_\perp^\top\mathbf{V}\right\|_{\mathrm{op}}^2,$$

and similarly $\|\mathbf{P}_\perp\,\widehat{\mathbf{P}}\,\mathbf{P}_\perp\|_{\mathrm{op}} = \|\widehat{\mathbf{V}}_\perp^\top\mathbf{V}\|_{\mathrm{op}}^2$. Therefore,

$$\left\|\mathbf{P} - \widehat{\mathbf{P}}\right\|_{\mathrm{op}} = \sqrt{\left\|(\mathbf{P} - \widehat{\mathbf{P}})^2\right\|_{\mathrm{op}}} = \left\|\widehat{\mathbf{V}}_\perp^\top\mathbf{V}\right\|_{\mathrm{op}}.$$

Combining with the Wedin bound in (35),

$$\left\|\mathbf{P} - \widehat{\mathbf{P}}\right\|_{\mathrm{op}} = \left\|\widehat{\mathbf{V}}_\perp^\top\mathbf{V}\right\|_{\mathrm{op}} \leq \frac{\|\mathbf{A} - \mathbf{B}\|_{\mathrm{op}}}{\tau} = \frac{2\,\|\mathbf{A} - \mathbf{B}\|_{\mathrm{op}}}{\mathsf{gap}}.$$

This completes the proof. □

Lemma 29 (Lemma F.4, [LKJO22]). *Let $\mathbf{x}, \mathbf{y}$ be unit vectors. Then*

$$1 - \langle \mathbf{x}, \mathbf{y}\rangle^2 \ \leq\ \|\mathbf{x} - \mathbf{y}\|_2^2. \tag{36}$$

Moreover, if $\|\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{2}$ (equivalently, $\|\mathbf{x} - \mathbf{y}\|_2^2 \leq 2$), then

$$1 - \langle \mathbf{x}, \mathbf{y}\rangle^2 \ \geq\ \frac{1}{2}\,\|\mathbf{x} - \mathbf{y}\|_2^2. \tag{37}$$

Proof. Let $\rho := \langle \mathbf{x}, \mathbf{y}\rangle \in [-1, 1]$. Since $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$,

$$\|\mathbf{x} - \mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\,\langle \mathbf{x}, \mathbf{y}\rangle = 2 - 2\rho = 2(1 - \rho). \tag{38}$$

Also,

$$1-\rho^2 = (1-\rho)(1+\rho). \tag{39}$$

For (36), using $1+\rho \le 2$ (since $\rho \le 1$), we have

$$1-\rho^2 = (1-\rho)(1+\rho) \le 2(1-\rho) = \|\mathbf{x}-\mathbf{y}\|_2^2,$$

where the last equality follows from (38). This yields (36).

For (37), assume $\|\mathbf{x}-\mathbf{y}\|_2 \le \sqrt{2}$, i.e. $\|\mathbf{x}-\mathbf{y}\|_2^2 \le 2$. Then by (38),

$$2(1-\rho) = \|\mathbf{x}-\mathbf{y}\|_2^2 \le 2 \quad \Longrightarrow \quad \rho \ge 0,$$

and hence $1+\rho \ge 1$. Therefore,

$$1-\rho^2 = (1-\rho)(1+\rho) \ge (1-\rho) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2,$$

again using (38). This proves (37). □

Corollary 3 (Sign-invariant conversion between $\sin^2$ and ℓ_2)**.** *Let $\mathbf{x}, \mathbf{y}$ be unit vectors, and let $s \in \{\pm 1\}$ be chosen so that $\langle \mathbf{x}, s\mathbf{y}\rangle \ge 0$ (e.g. take $s := \operatorname{sign}(\langle \mathbf{x}, \mathbf{y}\rangle)$, with the convention $\operatorname{sign}(0) = 1$). Then $\|\mathbf{x} - s\mathbf{y}\|_2 \le \sqrt{2}$ and*

$$\frac{1}{2}\|\mathbf{x}-s\mathbf{y}\|_2^2 \;\le\; 1-\langle \mathbf{x}, \mathbf{y}\rangle^2 \;\le\; \|\mathbf{x}-s\mathbf{y}\|_2^2.$$

In particular, if $1-\langle x, y\rangle^2 \le \Delta$, then $\min_{s\in\{\pm 1\}} \|\mathbf{x}-s\mathbf{y}\|_2^2 \le 2\Delta$.

Proof. Let $\rho := \langle \mathbf{x}, \mathbf{y}\rangle \in [-1, 1]$ and choose s so that $\langle \mathbf{x}, s\mathbf{y}\rangle = |\rho| \ge 0$. Then $\|\mathbf{x}-s\mathbf{y}\|_2^2 = 2(1-|\rho|) \le 2$, hence $\|\mathbf{x}-s\mathbf{y}\|_2 \le \sqrt{2}$. Applying Lemma 29 to $(\mathbf{x}, s\mathbf{y})$ gives

$$1-|\rho|^2 \le \|\mathbf{x}-s\mathbf{y}\|_2^2 \qquad \text{and} \qquad 1-|\rho|^2 \ge \frac{1}{2}\|\mathbf{x}-s\mathbf{y}\|_2^2.$$

Since $|\rho|^2 = \rho^2$, this yields the displayed inequality. The final implication follows immediately. □

B Deferred Proofs from Section 3

Lemma 2 (Properties of $\boldsymbol{\Sigma}$ from Model 3)**.** *Let $\boldsymbol{\Sigma}$ and $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ be generated from Model 3 with associated graph G. Let for all $i \in [n]$, $\mathbf{x}_{i,b} \in \mathbb{R}^{|C_b|}$ denotes the i^{th} vector with coordinates in the block C_b. Let $a := \frac{r-k+1}{2}$ and P_G denote the graph projection operator in Definition 6, then the following properties hold.*

1. $P_G(\widehat{\boldsymbol{\Sigma}}) = \mathbf{diag}\left(\widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_2, \cdots, \widehat{\boldsymbol{\Sigma}}_B\right)$ *for* $\widehat{\boldsymbol{\Sigma}}_b := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_{i,b}\mathbf{x}_{i,b}^\top$.
2. $\forall\, i \in [n],\ b \in [B],\ \mathbf{x}_{i,b} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_b)$, *and* $\forall\, b_1 \ne b_2 \in [B],\ \mathbf{x}_{i,b_1} \perp\!\!\!\perp \mathbf{x}_{i,b_2}$.
3. $\forall t > 0,\ \mathbb{P}(\|\boldsymbol{\Sigma}\|_{\text{op}} > t) \le \frac{d}{k}(\frac{e^2}{t})^a$ *and if* $q \le \frac{a}{2}$, *then* $\mathbb{E}[\|\boldsymbol{\Sigma}\|_{\text{op}}^q] \le 2\exp(4q)$.

4. $\forall b \in [B],\ \ g_1 \frac{k^2}{n} \le \mathbb{E}\big[\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\|_{\mathrm{F}}^2\big] \le g_2 \frac{k^2}{n}$.

5. $g_2 \cdot dk/n \le \mathbb{E}[\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\|_{\mathrm{F}}^2] \le g_3 \cdot dk/n$.

Here $g_1 > 0$, $0 < g_2 < g_3$ are universal constants.

Proof. Let P_G be the graph projection operator from Definition 6. For the first claim, by linearity of P_G,

$$P_G(\widehat{\mathbf{\Sigma}}) = P_G\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\Big) = \frac{1}{n}\sum_{i=1}^n P_G(\mathbf{x}_i\mathbf{x}_i^\top).$$

Since G is a disjoint union of cliques on $\{C_b\}_{b\in[B]}$, the projection P_G zeroes out all entries across distinct blocks and preserves all within-block entries. Hence, for each i,

$$P_G(\mathbf{x}_i\mathbf{x}_i^\top) = \mathbf{diag}\left(\mathbf{x}_{i,1}\mathbf{x}_{i,1}^\top, \ldots, \mathbf{x}_{i,B}\mathbf{x}_{i,B}^\top\right),$$

and therefore

$$P_G(\widehat{\mathbf{\Sigma}}) = \mathbf{diag}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{i,1}\mathbf{x}_{i,1}^\top, \ldots, \frac{1}{n}\sum_{i=1}^n \mathbf{x}_{i,B}\mathbf{x}_{i,B}^\top\right) = \mathbf{diag}\left(\widehat{\mathbf{\Sigma}}_1, \ldots, \widehat{\mathbf{\Sigma}}_B\right).$$

For the second claim, we note that conditional on $\mathbf{\Sigma}$, we have $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{diag}\left(\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_B\right)$. Thus the marginal on coordinates in block C_b is $\mathbf{x}_{i,b} \sim \mathcal{N}(0, \mathbf{\Sigma}_b)$. Moreover, for $b_1 \neq b_2$, the cross-covariance between $\mathbf{x}_{i,b_1}$ and $\mathbf{x}_{i,b_2}$ is zero because $\mathbf{\Sigma}$ has no off-block entries; since the joint vector is Gaussian, zero cross-covariance implies independence, i.e. $\mathbf{x}_{i,b_1} \perp\!\!\!\perp \mathbf{x}_{i,b_2}$.

Note that for each block $b \in [B]$, we have

$$\mathbf{\Sigma}_b \sim \mathsf{InvWishart}((r-k-1)\mathbf{I}_k, r),\ r := 2k \ge 2\log(d/k).$$

By scaling, define matrices, $\forall b \in [B],\ \ \mathbf{M}_b := \frac{\mathbf{\Sigma}_b}{r-k-1}$. Then,

$$\mathbf{M}_b \sim \mathsf{InvWishart}(\mathbf{I}_k, r),\ \lambda_{\mathsf{min}}(\mathbf{\Sigma}_b) = (r-k-1)\lambda_{\mathsf{min}}(\mathbf{M}_b),\ \lambda_{\mathsf{max}}(\mathbf{\Sigma}_b) = (r-k-1)\lambda_{\mathsf{max}}(\mathbf{M}_b). \quad (40)$$

For the third claim, we start from the proof of [Nar24, Lemma 2] (Appendix A) which yields the following sharper form: for $\mathbf{M} \sim \mathsf{InvWishart}(\mathbf{I}_k, r)$ and all $x > 0$,

$$\mathbb{P}\left(\lambda_{\mathsf{max}}(\mathbf{M}) \ge \frac{x}{r}\right) \ \le\ \left(\frac{e^2}{x}\right)^{(r-k+1)/2},$$

Let $a := (r-k+1)/2$. Since $\mathbf{\Sigma}$ is block diagonal, $\lambda_{\mathsf{max}}(\mathbf{\Sigma}) = \max_{b\in[B]} \lambda_{\mathsf{max}}(\mathbf{\Sigma}_b)$, and therefore using (40) and setting $x := \frac{r}{r-k-1}t$, we have for all $t > 0$,

$$\mathbb{P}\left(\lambda_{\mathsf{max}}(\mathbf{\Sigma}) \ge t\right) \le \sum_{b=1}^B \mathbb{P}\left(\lambda_{\mathsf{max}}(\mathbf{\Sigma}_b) \ge t\right) = \sum_{b=1}^B \mathbb{P}\left(\lambda_{\mathsf{max}}(\mathbf{M}_b) \ge \frac{t}{r-k-1}\right) \le \frac{d}{k}\left(\frac{c_0}{t}\right)^a \quad (41)$$

where $c_0 := (\frac{r-k-1}{r})e^2$. Let $Z := \lambda_{\mathsf{max}}(\mathbf{\Sigma})$. Then (41) implies that for all $t \ge 0$,

$$\mathbb{P}(Z \ge t) \le \min\left\{1,\ \frac{d}{k}\Big(\frac{c_0}{t}\Big)^a\right\}.$$

Using $\mathbb{E}[Z^q] = \int_0^\infty qt^{q-1}\mathbb{P}(Z \geq t)\,dt$ for non-negative random variable Z, and letting $t_0 := c_0\,\left(\frac{d}{k}\right)^{1/a}$ (so that $\frac{d}{k}(c_0/t_0)^a = 1$), we obtain

$$\mathbb{E}[Z^q] \leq \int_0^{t_0} qt^{q-1}\,\mathrm{d}t + \int_{t_0}^\infty qt^{q-1}\,\frac{d}{k}\Big(\frac{c_0}{t}\Big)^a\,\mathrm{d}t = t_0^q + q\frac{d}{k}e^{2a}\int_{t_0}^\infty t^{q-1-a}\,\mathrm{d}t.$$

The second integral converges iff $q < a$, in which case it equals $\frac{t_0^{q-a}}{a-q}$. Substituting $t_0 = c_0\,(\frac{d}{k})^{1/a}$ yields

$$\mathbb{E}[Z^q] \leq \left(c_0\left(\frac{d}{k}\right)^{1/a}\right)^q + \frac{q}{a-q}\left(c_0\left(\frac{d}{k}\right)^{1/a}\right)^q = \left(c_0\left(\frac{d}{k}\right)^{1/a}\right)^q \cdot \frac{a}{a-q}.$$

With $B := d/k$, $r - k + 1 = k + 1 \geq \log(B)$, we have

$$B^{q/a} = \exp\Big(\frac{q}{a}\log B\Big) = \exp\Big(\frac{2q\log B}{r-k+1}\Big) \leq \exp\Big(\frac{2q\log B}{\log B}\Big) = e^{2q},$$

and $\frac{a}{a-q} = \frac{r-k+1}{r-k+1-2q} \leq 2$ for $r - k + 1 \geq 4q$. Thus, substituting c_0,

$$\mathbb{E}[\lambda_{\mathsf{max}}(\mathbf{\Sigma})^{\,q}] \leq 2\left(e^2\frac{r-k-1}{r}\right)^q e^{2q} \leq 2\exp\left(4q\right). \tag{42}$$

For the fourth claim, by Claim (1), both $P_G(\widehat{\mathbf{\Sigma}})$ and $\mathbf{\Sigma}$ are block diagonal with the same blocks, so

$$\mathbb{E}\left[\left\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_{\mathrm{F}}^2\right] = \sum_{b=1}^{B}\mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\right\|_{\mathrm{F}}^2\right]. \tag{43}$$

Then using Lemma 2 from [Nar24], for $r > 2k$, we have for a universal constant $c > 0$ and using (40),

$$\frac{1}{(r-k-1)^2}\mathbb{E}\left[\lambda_{\mathsf{min}}(\mathbf{\Sigma}_b)^2\right] = \mathbb{E}\left[\lambda_{\mathsf{min}}(\mathbf{M}_b)^2\right] \geq \frac{c}{r^2} \tag{44}$$

Since $r \geq 2k > 8$, $(r-k-1)/r \geq 1/4$. Then, using (42) and (44), along with Lemma 1 from [Nar24], there exist universal constants $g_1, g_2 > 0$ such that

$$g_1\frac{k^2}{n} \leq \mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\right\|_{\mathrm{F}}^2\right] \leq g_2\frac{k^2}{n}.$$

The result then follows from (43) and noting that $B = d/k$. □

Lemma 30. *Let $\mathbf{\Sigma}$ be drawn from Model 3. Then,*

$$\mathbb{P}\big(\|\mathbf{\Sigma}\|_{\mathrm{op}} \leq 10\big) \;\geq\; 1 - \exp\left(-\Omega\left(k\right)\right).$$

Futher, on the event $\{\|\mathbf{\Sigma}\|_{\mathrm{op}} \leq 10\}$ the sub-Gaussian parameter $\sigma^2 := \|\mathbf{\Sigma}\|_{\mathrm{op}}$ satisfies $\sigma^2 = O(1)$.

Proof. By Lemma 2, Item 3, for all $t > 0$,

$$\mathbb{P}(\|\mathbf{\Sigma}\|_{\mathrm{op}} > t) \leq \frac{d}{k}\Big(\frac{e^2}{t}\Big)^a, \qquad a = \frac{r-k+1}{2} = \Theta(k),$$

with $r = 2k$. Then using $k \geq \Omega(\log(d/k)$,

$$\mathbb{P}(\|\mathbf{\Sigma}\|_{\mathrm{op}} > 10) \leq 1 - \exp\left(\log\left(\frac{d}{k}\right) - \frac{k+1}{2}\log\left(\frac{10}{e^2}\right)\right) \leq 1 - \exp\left(-\Omega\left(k\right)\right).$$

□

C Private Eigenvalue Estimation for RCS Matrices

In this section, we provide an algorithm for privately estimating the operator norm of k-RCS covariance matrix, $\mathbf{\Sigma}$, i.e., satisfying Models 1 or 2, which feeds as an input to Algorithm 2, and may be of independent interest. Our main result for this section is provided in Theorem 6. We largely follow the same algorithm as Algorithm 2, with the same choice of the threshold τ as in (26), deviating in the definition of the certified set and the final output.

Theorem 6. *Let $\epsilon, \delta, \beta \in (0,1)^3$ and $\mathbf{\Sigma}, \mathcal{D} := \{\mathbf{x}_i\}_{i\in[n]}$ be generated based on Model 1 or 2 with sub-Gaussianity parameter $\sigma > 0$. Then, for*

$$n \gtrsim \frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2}\frac{k^2}{\gamma^2\epsilon^3}\log\left(\frac{1}{\beta}\right)\log(d)\log\left(\frac{1}{\delta}\right)\log\left(\frac{1}{\delta\beta}\right)$$

there exists an (ϵ,δ) differential private algorithm, $\mathcal{A} : \mathsf{PrivNorm}(\mathcal{D}, \hat{\lambda}, \gamma, k, \sigma^2, \epsilon, \delta, \tau, \beta, m)$, which returns estimate $\hat{\lambda}$ satisfying, with probability atleast $1-\beta$,

$$\left|\widehat{\lambda} - \lambda_1(\mathbf{\Sigma})\right| \le \frac{\gamma}{10}\lambda_1(\mathbf{\Sigma})$$

Proof of Theorem 6. The intermediate results for this algorithm follow the same structure as Algorithm 2, and therefore we only sketch the proof here.

Let $\mathcal{D}, \mathcal{D}'$ represent adjacent datasets with corresponding estimates in Line 16 of Algorihtm 3 represented as $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Sigma}}'$ respectively.

From the same argument as Lemma 11, the sensitivity of $\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})$ is bounded via triangle inequality as

$$\begin{aligned}\left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}})\right\|_{\mathrm{op}} - \left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')\right\|_{\mathrm{op}} &\le \left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}')\right\|_{\mathrm{op}} \\ &\le \left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_{\tau}(\mathbf{\Sigma}^\star)\right\|_{\mathrm{op}} + \left\|\mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_{\tau}(\mathbf{\Sigma}^\star)\right\|_{\mathrm{op}} \\ &\le 34k\tau\end{aligned}$$

Note that we add Gaussian noise with standard deviation

$$\sigma_{\mathrm{priv}} = \frac{2\Delta_{\mathbf{P}}}{\epsilon}\sqrt{\log\left(\frac{2}{\delta}\right)} \lesssim \frac{k\tau}{\epsilon}\sqrt{\log\left(\frac{2}{\delta}\right)}$$

Therefore, the privacy follows by the Gaussian mechanism (Fact 1).

For utility, we have, using the same argument as Lemma 17, via a union bound over Lemma 14 Lemma 15 both tests pass with high probability for the choice of n, m provided in those results, and then with probability atleast $1 - \exp(-m/60)$,

$$\left\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\infty,\infty} = \left\|\frac{1}{Z}\sum_{i\in[n]} p_i\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\right\|_{\infty,\infty} = \left\|\frac{1}{Z}\sum_{i\in[n]} p_i(\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma})\right\|_{\infty,\infty} \le 2\tau$$

where we used that $Z \ge 0.6m$ and for every $i \in [m]$ with $p_i > 0$, $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \le 2\tau$. Then since $\mathbf{\Sigma}$ is k-RCS, using Lemma 27, $\left\|\mathcal{T}_{2\tau}(\bar{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_{\mathrm{op}} \le 4k\tau$.

Therefore, with probability atleast $1 - \exp(-m/60)$,

$$\left| \left\| \mathcal{T}_{2\tau}(\bar{\mathbf{\Sigma}}) \right\|_{\mathrm{op}} - \left\| \mathbf{\Sigma} \right\|_{\mathrm{op}} \right| \leq \left\| \mathcal{T}_{2\tau}(\bar{\mathbf{\Sigma}}) - \mathbf{\Sigma} \right\|_{\mathrm{op}} \leq 4k\tau \tag{45}$$

Furthermore, with probability atleast $1 - \beta$, $g \sim \mathcal{N}(0, \sigma^2_{\mathrm{priv}})$ satisfies,

$$|g| \leq \frac{k\tau}{\epsilon} \sqrt{\log\left(\frac{2}{\delta}\right)} \sqrt{\log\left(\frac{1}{\beta}\right)} \tag{46}$$

The result then follows by choosing $m \geq \Omega(\frac{1}{\epsilon}\log(\frac{1}{\delta\beta}))$, setting the RHS of (45), (46) smaller than $\frac{\gamma\lambda_1(\mathbf{\Sigma})}{10}$ and using the definition of τ in (26). □

Algorithm 3 Private Operator norm Estimation for k-RCS matrices: $\mathsf{PrivNorm}(\mathcal{D}, \hat{\lambda}, \gamma, k, \sigma^2, \epsilon, \delta, \tau, \beta, m)$

Require: Samples $\mathbf{x}_1, \ldots \mathbf{x}_n \in \mathbb{R}^d$, gap parameter $\gamma > 0$, sparsity k, sub-Gaussian scale σ^2, privacy (ϵ, δ), threshold τ, failure β, number of batches m.

Partition samples into m blocks of size $b := n/m$ and compute block covariances $\widehat{\mathbf{\Sigma}}_1, \ldots, \widehat{\mathbf{\Sigma}}_m$ on disjoint blocks.

Define the certified set

$$\mathcal{C} \leftarrow \left\{ i \in [m] : \mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) \text{ is } k\text{-RCS} \bigwedge \widehat{\gamma}\Big(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)\Big) := 1 - \frac{\lambda_2(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))}{\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))} \geq \gamma/2 \right\}.$$

Sample $L \sim \mathsf{BoundedLap}\big(\frac{3}{\epsilon}, \frac{3}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$.

if $|\mathcal{C}| + L - \frac{3}{\epsilon}\log\big(\frac{12}{\delta}\big) < 0.8\,m$ **then**

  return $\perp$

end if

for $i \leftarrow 1 : m$ **do**

  $f_i \leftarrow \sum_{j \in [m]} \mathbf{1}\Big\{ \|\widehat{\mathbf{\Sigma}}_j - \widehat{\mathbf{\Sigma}}_i\|_{\infty,\infty} \leq 4\tau \Big\}$

  $p_i \leftarrow \min\Big\{ \max\Big\{ \frac{f_i - m/2}{2m/3 - m/2}, 0 \Big\}, 1 \Big\}$

end for

$Z \leftarrow \sum_{i \in [m]} p_i$

Sample $\xi \sim \mathsf{BoundedLap}\big(\frac{21}{\epsilon}, \frac{21}{\epsilon}\log\big(\frac{12}{\delta}\big)\big)$.

if $Z + \xi - \frac{21}{\epsilon}\log\big(\frac{12}{\delta}\big) \leq 0.8\,m$ **then**

  return $\perp$

end if

$\bar{\mathbf{\Sigma}} \leftarrow \frac{1}{Z}\sum_{i \in [m]} p_i\, \widehat{\mathbf{\Sigma}}_i$

Set $\Delta_{\mathbf{P}} \leftarrow 17 \cdot k\tau$

Set $\sigma_{\mathrm{priv}} \leftarrow \frac{6\Delta_{\mathbf{P}}}{\epsilon}\sqrt{\log\left(\frac{6}{\delta}\right)}$

return $\left\| \mathcal{T}_{17\tau}(\bar{\mathbf{\Sigma}}) \right\|_{\mathrm{op}} + \mathcal{N}(0, \sigma^2_{\mathrm{priv}})$

D Deferred Proofs from Section 5

Lemma 31. *Let $\mathcal{P}_1, \mathcal{P}_2$ denote families of distributions over $\mathcal{X}^n$ where $\mathcal{X}$ denotes a sample space over vectors in $\mathbb{R}^d$. Let $p_1 \in \mathsf{co}(\mathcal{P}_1)$ and $p_2 \in \mathsf{co}(\mathcal{P}_1)$ be distributions over $\mathcal{R}^d$ where $\mathsf{co}(\mathcal{P})$ represents the convex hull of distributions in $\mathcal{P}$. Let $(\mathcal{X}, \mathcal{Y}), \mathcal{X} := \{\mathbf{x}_j\}_{j\in[n]}, \mathcal{Y} := \{\mathbf{y}_j\}_{j\in[n]} \in \mathbb{R}^{n\times d}$ be a coupling between p_1 and p_2 with $D := \mathbb{E}\left[\sum_{i\in[n]} \mathcal{I}(\mathbf{x}_i \neq \mathbf{y}_i)\right]$. Then for $\epsilon, \delta \geq 0$, any (ϵ, δ)-differentially private hypothesis testing algorithm $\hat{\boldsymbol{\theta}}$ must satisfy,*

$$P\left(\hat{\theta}, \mathcal{P}_1, \mathcal{P}_2\right) := \max_{i\in\{1,2\}} \max_{p\in\mathcal{P}_i} \mathbb{P}\left(\hat{\boldsymbol{\theta}}(\mathcal{X}) \neq i | \mathcal{X} \sim p\right) \geq \frac{1}{2}\max\left\{1 - d_{\mathsf{TV}}(p_1, p_2), 0.9e^{-10\epsilon D} - 10D\delta\right\}$$

where $d_{\mathsf{TV}}(p_1, p_2)$ denotes the total variational distance between p_1 and p_2.

Proof. The proof follows by the same steps as the proof of Theorem 1 in [ASZ21], noting that the notion of hamming distance for scalars, can be extended to vectors as $\sum_{i\in[n]} \mathcal{I}(\mathbf{x}_i \neq \mathbf{y}_i)$, used in the theorem statement. □

Lemma 21. *Let $\mathcal{V}$ denote a subset of $\{\pm 1\}^d$ such that each $\mathbf{s} \in \mathcal{V}$ is associated with a distribution $p_{\mathbf{s}} \in \mathcal{P}$ and parameter $\boldsymbol{\theta}(p_{\mathbf{s}})$. Let the loss function ℓ satisfy $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$,*

$$\ell\left(\boldsymbol{\theta}(p_{\mathbf{u}}), \boldsymbol{\theta}(p_{\mathbf{v}})\right) \geq 2\tau. \sum_{i\in[d]} \mathcal{I}(\mathbf{u}_i \neq \mathbf{v}_i), \text{ and } \ell\left(\boldsymbol{\theta}(p_{\mathbf{u}}), \boldsymbol{\theta}(p_{\mathbf{v}})\right) \leq 2\left(\ell\left(\boldsymbol{\theta}(p_{\mathbf{u}}), \boldsymbol{\theta}(p_{\mathbf{w}})\right) + \ell\left(\boldsymbol{\theta}(p_{\mathbf{w}}), \boldsymbol{\theta}(p_{\mathbf{v}})\right)\right)$$

For each coordinate $i \in [d]$, define the mixture distributions

$$p_{+i} := \frac{1}{|\mathcal{V}_{+i}|} \sum_{\mathbf{s}\in\mathcal{V}, \mathbf{s}_i = +1} p_{\mathbf{s}}, \quad p_{-i} := \frac{1}{|\mathcal{V}_{-i}|} \sum_{\mathbf{s}\in\mathcal{V}, \mathbf{s}_i = -1} p_{\mathbf{s}}$$

where $\mathcal{V}_{+i} := \{\mathbf{s} \in \mathcal{V}, \mathbf{s}_i = +1\}$ and $\mathcal{V}_{-i} := \{\mathbf{s} \in \mathcal{V}, \mathbf{s}_i = -1\}$. If for all $i \in [d]$, there exists a coupling $(\mathcal{X}, \mathcal{Y})$ such that $\mathcal{X} := \{\mathbf{x}_j\}_{j\in[n]} \sim p_{+i}$ and $\mathcal{Y} := \{\mathbf{y}_j\}_{j\in[n]} \sim p_{-i}$ with $\mathbb{E}\left[\sum_{j\in[n]} \mathcal{I}(\mathbf{x}_j \neq \mathbf{y}_j)\right] \leq D$, then

$$R(\mathcal{P}, \ell, \epsilon, \delta) := \min_{\hat{\boldsymbol{\theta}} \text{ is } (\epsilon,\delta)\text{-DP}} \max_{p\in\mathcal{P}} \mathbb{E}_{\mathcal{X}\sim p}\left[\ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p)\right)\right] \geq \frac{\tau d}{2}.\frac{\min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\}}{|\mathcal{V}|}.\left(0.9e^{-10\epsilon D} - 10D\delta\right)$$

Proof. Let $\mathbf{s} \in \mathcal{V}$ and $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \sim p_{\mathbf{s}}$ and denote the set of all such distributions as $\mathcal{P}_{\mathcal{V}} := \{p_{\mathbf{s}} : \mathbf{s} \in \mathcal{V}\}$. For an estimator $\hat{\boldsymbol{\theta}}(\mathcal{X})$, consider an estimator $\hat{\mathbf{s}}(\mathcal{X}) = \arg\min_{\mathbf{s}\in\mathcal{V}} \ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p_{\mathbf{s}})\right)$. Then by triangle inequality, for any distribution p,

$$\ell\left(\boldsymbol{\theta}(p_{\hat{\mathbf{s}}}), \boldsymbol{\theta}(p)\right) \leq 2\ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p_{\hat{\mathbf{s}}})\right) + 2\ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p)\right) \leq 4\ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p)\right)$$

Hence,

$$R(\mathcal{P}_{\mathcal{V}}, \ell, \epsilon, \delta) = \min_{\hat{\boldsymbol{\theta}} \text{ is } (\epsilon,\delta)\text{-DP}} \max_{\mathbf{s}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\hat{\boldsymbol{\theta}}(\mathcal{X}), \boldsymbol{\theta}(p_{\mathbf{s}})\right)\right] \geq \frac{1}{4} \min_{\hat{\mathbf{s}} \text{ is } (\epsilon,\delta)\text{-DP}} \max_{\mathbf{e}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\boldsymbol{\theta}\left(p_{\hat{\mathbf{s}}(\mathcal{X})}\right), \boldsymbol{\theta}(p_{\mathbf{s}})\right)\right]$$

We now have,

$$\max_{\mathbf{s}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\boldsymbol{\theta}\left(p_{\hat{\mathbf{s}}(\mathcal{X})}\right), \boldsymbol{\theta}\left(p_{\mathbf{s}}\right)\right)\right] \geq \frac{1}{|\mathcal{V}|}\sum_{\mathbf{s}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\boldsymbol{\theta}\left(p_{\hat{\mathbf{s}}(\mathcal{X})}\right), \boldsymbol{\theta}\left(p_{\mathbf{s}}\right)\right)\right]$$
$$\geq \frac{2\tau}{|\mathcal{V}|}\sum_{i\in[d]}\sum_{\mathbf{r}\in\mathcal{V}} \mathbb{P}\left(\hat{\mathbf{s}}_i \neq \mathbf{r}_i | \mathbf{s}=\mathbf{r}\right)$$

For each i, we divide $\mathcal{V}$ into two sets $\mathcal{V}_{+i} := \{\mathbf{r}\in\mathcal{V}, \mathbf{r}_i = +1\}$ and $\mathcal{V}_{-i} := \{\mathbf{r}\in\mathcal{V}, \mathbf{r}_i = -1\}$, based on the value of the i^{th} position,

$$\begin{aligned}
&\max_{\mathbf{s}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\theta\left(p_{\hat{\mathbf{s}}(\mathcal{X})}\right), \theta\left(p_{\mathbf{s}}\right)\right)\right] \\
&\geq \frac{2\tau}{|\mathcal{V}|}\sum_{i\in[d]}\left[\sum_{\mathbf{r}\in\mathcal{V},\mathbf{r}_i=+1} \mathbb{P}\left(\hat{\mathbf{s}}_i \neq \mathbf{r}_i | \mathbf{s}=\mathbf{r}\right) + \sum_{\mathbf{r}\in\mathcal{V},\mathbf{r}_i=-1} \mathbb{P}\left(\hat{\mathbf{s}}_i \neq \mathbf{e}_i | \mathbf{s}=\mathbf{r}\right)\right] \\
&= 2\tau\sum_{i\in[d]}\left(\frac{|\mathcal{V}_{+i}|}{|\mathcal{V}|}\mathbb{P}_{\mathcal{X}\sim p_{+i}}\left(\hat{\mathbf{s}}_i \neq +1\right) + \frac{|\mathcal{V}_{-i}|}{|\mathcal{V}|}\mathbb{P}_{\mathcal{X}\sim p_{-i}}\left(\hat{\mathbf{s}}_i \neq -1\right)\right) \\
&\geq \frac{2\tau\min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\}}{|\mathcal{V}|}\sum_{i\in[d]}\sum_{i\in[d]}\min_{\phi_i:\phi_i \text{ is}(\epsilon,\delta)\text{- DP}}\left(\mathbb{P}_{\mathcal{X}\sim p_{+i}}\left(\phi_i\left(\mathcal{X}\right)\neq 1\right) + \mathbb{P}_{\mathcal{X}\sim p_{-i}}\left(\phi_i\left(\mathcal{X}\right)\neq -1\right)\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
R\left(\mathcal{P}, \ell, \epsilon, \delta\right) &\geq R\left(\mathcal{P}_{\mathcal{V}}, \ell, \epsilon, \delta\right) \\
&\geq \frac{1}{4}\min_{\hat{\mathbf{s}} \text{ is } (\epsilon,\delta)\text{-DP}}\max_{\mathbf{e}\in\mathcal{V}} \mathbb{E}_{\mathcal{X}\sim p_{\mathbf{s}}}\left[\ell\left(\boldsymbol{\theta}\left(p_{\hat{\mathbf{s}}(\mathcal{X})}\right), \boldsymbol{\theta}\left(p_{\mathbf{s}}\right)\right)\right] \\
&\geq \frac{\tau\min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\}}{2|\mathcal{V}|}\sum_{i\in[d]}\min_{\phi_i:\phi_i \text{ is}(\epsilon,\delta)\text{- DP}}\left(\mathbb{P}_{\mathcal{X}\sim p_{+i}}\left(\phi_i\left(\mathcal{X}\right)\neq 1\right) + \mathbb{P}_{\mathcal{X}\sim p_{-i}}\left(\phi_i\left(\mathcal{X}\right)\neq -1\right)\right).
\end{aligned}$$

Therefore, we have for each $i\in[d]$, the summand above is the error probability of a hypothesis test between the mixture distributions p_{+i} and p_{-i}. The result then follows by Lemma 31. □

E Approximate DP Lower Bound for Standard PCA

In this section, we adapt the fingerprinting method of [Nar24] to prove a lower bound for approximate DP in standard (non-sparse) PCA in Theorem 7. We work under Model 1 but do *not* assume any k-RCS or sparsity structure. In particular, we assume 1-sub-Gaussian samples with covariance $\boldsymbol{\Sigma}\in\mathbb{S}^{d\times d}_{\succeq\mathbf{0}}$ and an eigengap $\lambda_2(\boldsymbol{\Sigma})/\lambda_1(\boldsymbol{\Sigma}) \leq 1-\gamma$ where γ is a known parameter.

Model 4 (Spiked Inverse Wishart PCA model)**.** *Fix $\gamma\in(0.1, 0.5)$, $\nu = 1600d/\gamma^2$ and unit vector $\mathbf{v}\in\mathbb{R}^d$. Let*

$$\mathbf{M} \;:=\; (1-\gamma)\mathbf{I}_d + \gamma\mathbf{v}\mathbf{v}^\top.$$

Given $\mathbf{M}$, draw

$$\boldsymbol{\Sigma} \;\sim\; \mathsf{InvWishart}\big((\nu-d-1)\mathbf{M}, \nu\big), \qquad \mathbf{x}_1,\ldots,\mathbf{x}_n \mid \boldsymbol{\Sigma} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Let $\widetilde{\mathbf{v}}(\mathcal{X})$ be any (possibly randomized) estimator with $\|\widetilde{\mathbf{v}}\|_2 = 1$, where $\mathcal{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$.

Under the inverse-Wishart construction in Model 4, there exists

$$\mathbf{A} = \frac{1}{\nu}\mathbf{G}^\top\mathbf{G} \text{ with } \mathbf{G} \in \mathbb{R}^{\nu\times d} \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \text{ and } \mathbf{A} \perp\!\!\!\perp V \text{ such that}$$
$$\mathbf{\Sigma} \;=\; r\,\mathbf{M}^{1/2}\mathbf{A}^{-1}\mathbf{M}^{1/2}, \qquad r := \frac{\nu-d-1}{\nu}, \tag{47}$$

Lemma 32. *Let $\mathbf{G} \in \mathbb{R}^{\nu\times d}, \mathbf{A} \in \mathbb{R}^{d\times d}$ be as defined in (47), $\eta \in (0,1)$ and define the event,*

$$\mathcal{E}_\eta := \{(1-\eta)\mathbf{I}_d \preceq \mathbf{A} \preceq (1+\eta)\mathbf{I}\}\,. \tag{48}$$

Then, for $\nu > 4d$, and $\eta := 6\sqrt{\frac{d}{\nu}}$, $\mathbb{P}(\mathcal{E}_\eta) \geq 1 - 2e^{-d/2}$

Proof. Let $s_{\min}(\mathbf{G})$ and $s_{\max}(\mathbf{G})$ denote the smallest and largest singular values of $\mathbf{G}$. By the Gaussian singular value deviation inequality (e.g. [Ver10, Corollary 5.35]), for every $t \geq 0$, with probability at least $1 - 2e^{-t^2/2}$ we have

$$\sqrt{\nu} - \sqrt{d} - t \;\leq\; s_{\min}(\mathbf{G}) \;\leq\; s_{\max}(\mathbf{G}) \;\leq\; \sqrt{\nu} + \sqrt{d} + t.$$

On this event, divide by $\sqrt{\nu}$ to obtain, for $u := \sqrt{\frac{d}{\nu}} + \frac{t}{\sqrt{\nu}}$,

$$1 - u \;\leq\; s_{\min}\Big(\tfrac{1}{\sqrt{\nu}}\mathbf{G}\Big) \;\leq\; s_{\max}\Big(\tfrac{1}{\sqrt{\nu}}\mathbf{G}\Big) \;\leq\; 1 + u.$$

Setting $t = \sqrt{d}$, we have $u := 2\sqrt{\frac{d}{\nu}} < 1$ for $\nu > 4d$. Note that since,

$$\mathbf{A} \;=\; \Big(\tfrac{1}{\sqrt{\nu}}\mathbf{G}\Big)^\top\Big(\tfrac{1}{\sqrt{\nu}}\mathbf{G}\Big),$$

the eigenvalues of $\mathbf{A}$ are the squares of the singular values of $\frac{1}{\sqrt{\nu}}\mathbf{G}$. Therefore, on the same event,

$$(1-u)^2\,\mathbf{I}_d \;\preceq\; \mathbf{A} \;\preceq\; (1+u)^2\,\mathbf{I}_d.$$

Finally, note that $(1+u)^2 = 1+(2u+u^2) \leq 1+\eta$ and also $(1-u)^2 = 1-2u+u^2 \;\geq\; 1-(2u+u^2) \geq 1-\eta$, so indeed $(1-\eta)\mathbf{I}_d \preceq \mathbf{A} \preceq (1+\eta)\mathbf{I}_d$, i.e. $\mathcal{E}_\eta$ holds. □

Lemma 33. *Under Model 4, for any $k \geq 1$ satisfying $\nu - d + 1 > 2k$, $d > 10k$, there exist constants $0 < g_k < G_k$ satisfying,*

$$(1-\gamma)rg_k \leq \Big[\mathbb{E}\Big[\lambda_{\min}(\mathbf{\Sigma})^k\Big]\Big]^{\frac{1}{k}} \leq \Big[\mathbb{E}\Big[\lambda_{\max}(\mathbf{\Sigma})^k\Big]\Big]^{\frac{1}{k}} \leq rG_k.$$

Proof. From the definition of $\mathbf{M}$ in (4), we have $\lambda_{\max}(\mathbf{M}) = 1$ and $\lambda_{\min}(\mathbf{M}) = 1$. Let $\mathbf{X} := \mathbf{M}^{\frac{1}{2}}\mathbf{A}^{-\frac{1}{2}}$ and $\mathbf{Y} := \mathbf{A}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}$. Then,

$$\mathbf{Y}^{-1}(r\mathbf{YX})\mathbf{Y} = r\mathbf{XY} = \mathbf{\Sigma}$$

and therefore, $r\mathbf{XY}$ and $r\mathbf{YX}$ are similar and share eigenvalues. Let $\widetilde{\mathbf{\Sigma}} := r\mathbf{A}^{-\frac{1}{2}}\mathbf{M}\mathbf{A}^{\frac{1}{2}} = r\mathbf{YX}$. Then,

$$r(1-\gamma)\mathbf{A}^{-1} = \lambda_{\min}(\mathbf{M})\mathbf{A}^{-1} \preceq \widetilde{\mathbf{\Sigma}} \preceq r\lambda_{\max}(\mathbf{M})\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

Therefore,

$$\lambda_{\max}(\widetilde{\mathbf{\Sigma}}) \le r\lambda_{\max}(\mathbf{A}^{-1}) = r\nu\lambda_{\max}((\mathbf{G}^\top\mathbf{G})^{-1}), \text{ and}$$
$$\lambda_{\min}(\widetilde{\mathbf{\Sigma}}) \ge r\nu\,(1-\gamma)\,\lambda_{\min}((\mathbf{G}^\top\mathbf{G})^{-1}) = r\,(1-\gamma)\,\lambda_{\min}(\mathbf{A}^{-1}).$$

Substituting r from (47), we have

$$(1-\gamma)\,(\nu-d-1)\,\lambda_{\min}((\mathbf{G}^\top\mathbf{G})^{-1}) \le \lambda_{\min}(\widetilde{\mathbf{\Sigma}}) \le \lambda_{\max}(\widetilde{\mathbf{\Sigma}}) \le (\nu-d-1)\lambda_{\max}((\mathbf{G}^\top\mathbf{G})^{-1}). \quad (49)$$

Now, note that $(\mathbf{G}^\top\mathbf{G})^{-1} \overset{d}{=} \mathsf{InvWishart}(\mathbf{I}_d, \nu)$. Then, using Lemma 2 from [Nar24], for any $k \ge 1$, there exist constants $0 < g_k < G_k$ satisfying,

$$\frac{g_k}{\nu^k} \le \mathbb{E}\left[\lambda_{\min}((\mathbf{G}^\top\mathbf{G})^{-1})^k\right] \le \mathbb{E}\left[\lambda_{\max}((\mathbf{G}^\top\mathbf{G})^{-1})^k\right] \le \frac{G_k}{\nu^k}. \quad (50)$$

The result then follows by combining (49) and (50). □

Corollary 4. *Under the setting of Lemma 33, there exists a universal constant $U > 0$ such that,*

$$\mathbb{E}\left[\left\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\right\|_{\text{op}}^2\right] \le U\left(\frac{d}{n} + \left(\frac{d}{n}\right)^2\right)$$

Proof. The inequality follows from Corollary 2 of [KL17a] with $p = 2$, and the observation $r(\mathbf{\Sigma}) := \frac{\mathsf{Tr}(\mathbf{\Sigma})}{\lambda_1(\mathbf{\Sigma})} \le d$ along with using Lemma 33 to bound $\mathbb{E}[\|\mathbf{\Sigma}\|_{\text{op}}^2]$. □

We now establish a posterior concentration result, following the proof of Lemma 6 in [Nar24].

Lemma 34. *Under Model 4, let $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\widehat{\mathbf{\Sigma}} := \sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top/n$. Then,*

$$\mathbb{E}\left[\|\mathbb{E}\left[\mathbf{\Sigma}|X\right] - \widehat{\mathbf{\Sigma}}\|_{\text{op}}^2\right] \le O\left(\left(\frac{\nu}{n}\right)^2\left(1 + \left(\frac{d}{n}\right) + \left(\frac{d}{n}\right)^2\right)\right).$$

Proof. Note that $\mathbf{\Sigma} \sim \mathsf{InvWishart}\left((\nu-d-1)\mathbf{M}, \nu\right)$, $\mathcal{X} \overset{\mathsf{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma})$. Then,

$$\begin{aligned}
\mathbb{E}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}|\mathcal{X}\right] - \widehat{\mathbf{\Sigma}}\right\|_{\mathrm{op}}^2\right] &= \mathbb{E}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}|\mathcal{X}\right] - \widehat{\mathbf{\Sigma}}\right\|_{\mathrm{op}}^2\right] \\
&\overset{(i)}{=} (1-w_n)^2 \mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbb{E}\left[\mathbf{\Sigma}\right]\right\|_{\mathrm{op}}^2\right], \quad \text{for } w_n := \frac{n}{\nu+n-d-1} \\
&\leq 3(1-w_n)^2 \mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathrm{op}}^2 + \|\mathbf{\Sigma}\|_{\mathrm{op}}^2 + \|\mathbb{E}\left[\mathbf{\Sigma}\right]\|_{\mathrm{op}}^2\right] \\
&\overset{(ii)}{\leq} 3(1-w_n)^2 O\left(1 + \frac{d}{n} + \left(\frac{d}{n}\right)^2\right), \\
&\leq \left(\frac{\nu-d-1}{n+\nu-d-1}\right)^2 O\left(1 + \frac{d}{n} + \left(\frac{d}{n}\right)^2\right) \\
&\leq O\left(\left(\frac{\nu}{n}\right)^2 \left(1 + \left(\frac{d}{n}\right) + \left(\frac{d}{n}\right)^2\right)\right).
\end{aligned}$$

where (i) used Fact 8, and (ii) used the definition of $\mathbf{M}$ to bound $\mathbb{E}[\|\mathbf{M}\|_{\mathrm{op}}^2]$, Lemma 33 to bound $\mathbb{E}[\|\mathbf{\Sigma}\|_{\mathrm{op}}^2]$ and Corollary 4 to bound $\mathbb{E}[\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathrm{op}}^2]$. □

Now let us consider Model 4. Let $\mathbf{v}_1 := \mathbf{v}_1(\mathbf{\Sigma})$ be the top eigenvector of $\mathbf{\Sigma}$, and define the rank-one projector

$$\mathbf{P} := \mathbf{v}_1 \mathbf{v}_1^\top.$$

Let $\mathcal{M}$ be an (ϵ, δ)-DP algorithm which outputs a unit vector

$$\tilde{\mathbf{v}} := \mathcal{M}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^d, \qquad \widetilde{\mathbf{P}} := \tilde{\mathbf{v}}\tilde{\mathbf{v}}^\top.$$

Let $\mathbf{x}_i'$ be an independent copy of $\mathbf{x}_i$, and define the neighboring dataset

$$\mathcal{X}^{\sim i} := (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i', \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$$

and the corresponding output

$$\tilde{\mathbf{v}}^{\sim i} := \mathcal{M}(\mathcal{X}^{\sim i}), \qquad \widetilde{\mathbf{P}}^{\sim i} := \tilde{\mathbf{v}}^{\sim i}(\tilde{\mathbf{v}}^{\sim i})^\top.$$

Define the fingerprinting statistics

$$Z_i := \left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma} \right\rangle, \qquad Z_i' := \left\langle \widetilde{\mathbf{P}}^{\sim i} - \mathbf{P}, \mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma} \right\rangle.$$

Note that conditioned on $\mathbf{\Sigma}$, $\widetilde{\mathbf{P}}^{\sim i}$ is independent of $\mathbf{x}_i$, hence

$$\mathbb{E}[Z_i' \mid \mathbf{\Sigma}] = 0 \qquad \Longrightarrow \qquad \mathbb{E}[Z_i'] = 0. \tag{51}$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Z_i] = \mathbb{E}\left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle. \tag{52}$$

We will use the expectation-comparison inequality of [Nar24, Proposition 5]: for neighboring datasets (here $\mathcal{X}$ and $\mathcal{X}^{\sim i}$), one has

$$\left|\mathbb{E}[Z_i - Z_i']\right| \;\le\; 2\epsilon\,\mathbb{E}\left|Z_i'\right| \;+\; 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]}. \tag{53}$$

Combining (51) and (53) gives an upper bound on $|\mathbb{E}Z_i|$.

Lemma 35 (DP controls the fingerprinting statistic)**.** *Assume* $\mathbf{x}_i \mid \mathbf{\Sigma} \sim \mathcal{N}(0, \mathbf{\Sigma})$ *and* $\|\mathbf{\Sigma}\|_{\mathrm{op}} = \lambda_1(\mathbf{\Sigma})$*. Let* $\rho^4 := \mathbb{E}[\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^4]$*. Then for every* $i \in [n]$*,*

$$|\mathbb{E}[Z_i]| \;\lesssim\; \epsilon\,\rho \;+\; \sqrt{\delta}\,(\rho\, d + \rho)\,.$$

In particular, if $\delta \le \epsilon^2/d^2$*, then*

$$|\mathbb{E}[Z_i]| \;\lesssim\; \epsilon\,\rho.$$

Proof. By (51), $\mathbb{E}Z_i = \mathbb{E}(Z_i - Z_i')$, so (53) yields

$$|\mathbb{E}Z_i| \le 2\epsilon\,\mathbb{E}|Z_i'| + 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]}.$$

We first bound $\mathbb{E}|Z_i'|$ and $\mathbb{E}[(Z_i')^2]$ using the from bound [Nar24, Proposition 3], which states that for fixed $\mathbf{A}$ and $\mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma})$,

$$\mathbb{E}\left[\left\langle \mathbf{A}, \mathbf{x}\mathbf{x}^\top - \mathbf{\Sigma}\right\rangle^2 \,\middle|\, \mathbf{\Sigma}\right] \le 2\,\|\mathbf{\Sigma}\|_{\mathrm{op}}^2\,\|\mathbf{A}\|_{\mathrm{F}}^2\,.$$

Conditioned on $\mathbf{\Sigma}$ and on $\widetilde{\mathbf{P}}^{\sim i}$, apply this with $\mathbf{A} = \widetilde{\mathbf{P}}^{\sim i} - \mathbf{P}$ to obtain

$$\mathbb{E}[(Z_i')^2 \mid \mathbf{\Sigma}, \widetilde{\mathbf{P}}^{\sim i}] \le 2\,\lambda_1(\mathbf{\Sigma})^2\,\left\|\widetilde{\mathbf{P}}^{\sim i} - \mathbf{P}\right\|_{\mathrm{F}}^2.$$

Taking expectations and using Jensen's inequality with Lemma 33 gives

$$\mathbb{E}|Z_i'| \le \sqrt{\mathbb{E}[(Z_i')^2]} \le \sqrt{2}\,\sqrt{\mathbb{E}\,[\lambda_1(\mathbf{\Sigma})^2]}\,\sqrt{\mathbb{E}\left\|\widetilde{\mathbf{P}}^{\sim i} - \mathbf{P}\right\|_{\mathrm{F}}^2} \lesssim \rho,$$

and similarly $\mathbb{E}[(Z_i')^2] \lesssim \rho^2$. For Z_i^2, we use $|Z_i| \le \left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\mathrm{F}} \cdot \left\|\mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma}\right\|_{\mathrm{F}}$ and Cauchy–Schwarz:

$$\mathbb{E}[Z_i^2] \le \sqrt{\mathbb{E}\left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\mathrm{F}}^4 \cdot \mathbb{E}\left\|\mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma}\right\|_{\mathrm{F}}^4} \lesssim \rho^2\, d^2,$$

where the last step uses the standard Gaussian fourth-moment scaling $\mathbb{E}\|\mathbf{x}\mathbf{x}^\top - \mathbf{\Sigma}\|_{\mathrm{F}}^4 \lesssim \mathbb{E}\left[\lambda_1(\mathbf{\Sigma})^4\right] d^4$ and Lemma 33. Plugging these bounds into (53) gives the claim, and when $\delta \le \epsilon^2/d^2$ the $\sqrt{\delta}$-term is absorbed into the ϵ-term. □

Combining Lemma 35 with (52) yields

$$\mathbb{E}\left[\left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\rangle\right] \;=\; \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Z_i] \;\lesssim\; \epsilon\,\rho. \tag{54}$$

Let $\hat{\mathbf{v}} := \mathbf{v}_1(\widehat{\mathbf{\Sigma}})$ be the top eigenvector of $\widehat{\mathbf{\Sigma}}$ and define $\widehat{\mathbf{P}} := \hat{\mathbf{v}}\hat{\mathbf{v}}^\top$. The following deterministic inequality lower bounds the correlation term by the sample misalignment.

Lemma 36 (Rayleigh-gap lower bound)**.** *For any $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ with eigenvalues $\lambda_1 > \lambda_2$ and any sample covariance $\widehat{\mathbf{\Sigma}}$,*

$$\left\langle \widehat{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle \;\geq\; (\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))\, \sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v}_1).$$

Proof. Using $\langle \mathbf{u}\mathbf{u}^\top, \mathbf{A} \rangle = \mathbf{u}^\top \mathbf{A} \mathbf{u}$,

$$\left\langle \widehat{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle = \underbrace{\hat{\mathbf{v}}^\top \widehat{\mathbf{\Sigma}} \hat{\mathbf{v}} - \mathbf{v}_1^\top \widehat{\mathbf{\Sigma}} \mathbf{v}_1}_{\geq 0} + \left(\lambda_1(\mathbf{\Sigma}) - \hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} \right).$$

The first bracket is nonnegative since $\hat{\mathbf{v}}$ maximizes the Rayleigh quotient of $\widehat{\mathbf{\Sigma}}$. Write $\hat{\mathbf{v}} = c\mathbf{v}_1 + \mathbf{r}$ with $\mathbf{r} \perp \mathbf{v}_1$ and $c^2 = 1 - \|\mathbf{r}\|_2^2$. Then $\hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} \leq c^2 \lambda_1 + \lambda_2 \|\mathbf{r}\|_2^2$ and hence

$$\lambda_1(\mathbf{\Sigma}) - \hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} \geq \lambda_1(\mathbf{\Sigma}) - \left(c^2 \lambda_1(\mathbf{\Sigma}) + \lambda_2(\mathbf{\Sigma}) \|\mathbf{r}\|_2^2 \right) = (\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma})) \|\mathbf{r}\|_2^2 = (\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma})) \sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v}_1).$$

□

Theorem 7. *Let $0 < t_1 < 1 < t_2, t_3$ be universal constants. Under Model 4, let $\epsilon, \delta \in (0,1)$ with $\delta < \frac{\epsilon^2}{d^2}$, $e^{-d} < \rho < t_1$ and $n \geq t_2 d$, $d \geq t_3$ for sufficiently small t_1 and sufficiently large t_2, t_3. Then, any (ϵ, δ)-DP algorithm $M : \mathcal{X} \to \widetilde{\mathbf{v}}$ which achieves*

$$\mathbb{E}\left[\left\| \widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^\top - \mathbf{v}_1(\mathbf{\Sigma}) \mathbf{v}_1(\mathbf{\Sigma})^\top \right\|_{\mathrm{F}}^4 \right] \leq \rho^4$$

must necessarily require $n = \Omega\left(\frac{d}{\rho\epsilon}\right)$.

Proof. We connect $\widetilde{\mathbf{P}} := \widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^\top$ to the leading projector $\widehat{\mathbf{P}} := \widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top$ where $\widehat{\mathbf{v}}$ denotes the leading eigenvector of $\widehat{\mathbf{\Sigma}}$, to provide a lower bound on the average score. Using Lemma 36,

$$\begin{aligned} \left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle &= \left\langle \widehat{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle + \left\langle \widetilde{\mathbf{P}} - \widehat{\mathbf{P}}, \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \right\rangle. \\ &\geq \frac{(\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))}{2} \left\| \widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}_1 \mathbf{v}_1^\top \right\|_{\mathrm{F}}^2 - \left\langle \widetilde{\mathbf{P}} - \widehat{\mathbf{P}}, \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \right\rangle \end{aligned} \tag{55}$$

For the second term above, conditioning on $\mathcal{X}$ and using Jensen's inequality, we have

$$\begin{aligned} \mathbb{E}\left[\left\langle \widetilde{\mathbf{P}} - \widehat{\mathbf{P}}, \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \right\rangle | \mathcal{X} \right] &= \left\langle \mathbb{E}\left[\widetilde{\mathbf{P}} | \mathcal{X} \right] - \widehat{\mathbf{P}}, \mathbb{E}[\mathbf{\Sigma} | \mathcal{X}] - \widehat{\mathbf{\Sigma}} \right\rangle \\ &\leq \left\| \mathbb{E}\left[\widetilde{\mathbf{P}} | \mathcal{X} \right] - \widehat{\mathbf{P}} \right\|_{\mathrm{tr}} \left\| \mathbb{E}[\mathbf{\Sigma} | \mathcal{X}] - \widehat{\mathbf{\Sigma}} \right\|_{\mathrm{op}} \\ &\leq \sqrt{\mathbb{E}_M \left[\left\| \widetilde{\mathbf{P}} - \widehat{\mathbf{P}} \right\|_{\mathrm{tr}}^2 \right]} \left\| \mathbb{E}[\mathbf{\Sigma} | \mathcal{X}] - \widehat{\mathbf{\Sigma}} \right\|_{\mathrm{op}} \end{aligned}$$

Therefore, taking expectations

$$\begin{aligned}
\mathbb{E}\left[\left\langle \widetilde{\mathbf{P}}-\widehat{\mathbf{P}}, \boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right\rangle\right] &\leq \mathbb{E}_{\mathcal{X}}\left[\sqrt{\mathbb{E}_M\left[\left\|\widetilde{\mathbf{P}}-\widehat{\mathbf{P}}\right\|_{\mathrm{tr}}^2\right]}\left\|\mathbb{E}[\boldsymbol{\Sigma}|\mathcal{X}]-\widehat{\boldsymbol{\Sigma}}\right\|_{\mathrm{op}}\right] \\
&\leq \sqrt{\mathbb{E}_{M,\mathcal{X}}\left[\left\|\widetilde{\mathbf{P}}-\widehat{\mathbf{P}}\right\|_{\mathrm{tr}}^2\right]\mathbb{E}_{\mathcal{X}}\left[\left\|\mathbb{E}[\boldsymbol{\Sigma}|\mathcal{X}]-\widehat{\boldsymbol{\Sigma}}\right\|_{\mathrm{op}}^2\right]} \\
&\leq \sqrt{2\left(\mathbb{E}_{M,\mathcal{X}}\left[\left\|\widetilde{\mathbf{P}}-\mathbf{P}\right\|_{\mathrm{tr}}^2\right]+\mathbb{E}_{\mathcal{X}}\left[\left\|\widehat{\mathbf{P}}-\mathbf{P}\right\|_{\mathrm{tr}}^2\right]\right)\mathbb{E}_{\mathcal{X}}\left[\left\|\mathbb{E}[\boldsymbol{\Sigma}|\mathcal{X}]-\widehat{\boldsymbol{\Sigma}}\right\|_{\mathrm{op}}^2\right]} \\
&\leq \sqrt{2\left(\rho^2+\mathbb{E}\left[\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top-\mathbf{v}\mathbf{v}^\top\|_{\mathrm{F}}^2\right]\right)\mathbb{E}_{\mathcal{X}}\left[\left\|\mathbb{E}[\boldsymbol{\Sigma}|\mathcal{X}]-\widehat{\boldsymbol{\Sigma}}\right\|_{\mathrm{op}}^2\right]} \qquad (56)
\end{aligned}$$

Combining together with (55),

$$\begin{aligned}
\mathbb{E}\left[\left\langle \widetilde{\mathbf{P}}-\mathbf{P}, \widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\rangle\right] &\geq \mathbb{E}\left[\frac{(\lambda_1(\boldsymbol{\Sigma})-\lambda_2(\boldsymbol{\Sigma}))}{2}\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top-\mathbf{v}_1\mathbf{v}_1^\top\right\|_{\mathrm{F}}^2\right] \\
&\quad -\sqrt{2}\sqrt{\left(\rho^2+\mathbb{E}\left[\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top-\mathbf{v}\mathbf{v}^\top\|_{\mathrm{F}}^2\right]\right)\mathbb{E}_{\mathcal{X}}\left[\left\|\mathbb{E}[\boldsymbol{\Sigma}|\mathcal{X}]-\widehat{\boldsymbol{\Sigma}}\right\|_{\mathrm{op}}^2\right]} \qquad (57)
\end{aligned}$$

Recall $\boldsymbol{\Sigma} := r\mathbf{M}^{\frac{1}{2}}\mathbf{A}^{-1}\mathbf{M}^{\frac{1}{2}}$ where $\mathbf{A}$ is defined in (47). Then,

$$\boldsymbol{\Sigma}-r\mathbf{M} = r\mathbf{M}^{\frac{1}{2}}\left(\mathbf{A}^{-1}-\mathbf{I}_d\right)\mathbf{M}^{\frac{1}{2}} \qquad (58)$$

Let $\mathcal{E}_\eta$ be as defined in Lemma 32 with $\mathbb{P}\left(\mathcal{E}_\eta\right) \geq 1-2e^{-\frac{d}{2}}$. Then, under $\mathcal{E}_\eta$, for $\eta := 2\sqrt{\frac{d}{\nu}}$ with $\nu = 1600d/\gamma^2$, we have

$$-\frac{\gamma}{10} \leq \lambda_{\min}((\mathbf{A})^{-1}-\mathbf{I}_d) \leq \lambda_{\max}((\mathbf{A})^{-1}-\mathbf{I}_d) \leq \frac{\gamma}{10} \qquad (59)$$

Therefore, under $\mathcal{E}_\eta$, using Weyl's inequality, using (58) and (59),

$$|\lambda_1(\boldsymbol{\Sigma})-r| \leq \frac{r\gamma}{10} \text{ and }, \forall i>1 \ |\lambda_i(\boldsymbol{\Sigma})-r(1-\gamma)| \leq \frac{r\gamma}{10} \qquad (60)$$

Therefore, under $\mathcal{E}$, using $\frac{1}{10}<\gamma<\frac{1}{2}$,

$$\begin{aligned}
&1-\frac{3\gamma}{2} \leq \frac{1-\frac{11}{10}\gamma}{1+\frac{\gamma}{10}} \leq \frac{\lambda_2(\boldsymbol{\Sigma})}{\lambda_1(\boldsymbol{\Sigma})} \leq \frac{1-\frac{9\gamma}{10}}{1+\frac{\gamma}{10}} \leq 1-\frac{\gamma}{2}, \text{ and} \\
&d \geq \frac{\mathrm{Tr}\left(\boldsymbol{\Sigma}\right)}{\lambda_1(\boldsymbol{\Sigma})} \geq \frac{\left(1-\frac{\gamma}{10}\right)+(d-1)\left(1-\frac{11}{10}\gamma\right)}{1+\frac{\gamma}{10}} \geq \left(1-\frac{3\gamma}{2}\right)(d-1) \geq \frac{3d}{5} \qquad (61)
\end{aligned}$$

Further, since $\frac{\lambda_2(\boldsymbol{\Sigma})}{\lambda_1(\boldsymbol{\Sigma})}<1$, then under $\mathcal{E}_\eta$, $\boldsymbol{\Sigma}$ posses a unique eigenvector.

Using Theorem-3 Item (3) [KL17b], for any fixed $\boldsymbol{\Sigma}$ satisfying (61), if $\frac{\mathrm{Tr}(\boldsymbol{\Sigma})}{\lambda_1(\boldsymbol{\Sigma})} < cn$ for a universal constant $c<1$, then

$$\mathbb{E}\left[\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top-\mathbf{v}\mathbf{v}^\top\right\|_{\mathrm{F}}^2\right] = \frac{A_1(\boldsymbol{\Sigma})}{n}+O\left(\left(\frac{\mathrm{Tr}(\boldsymbol{\Sigma})}{n\lambda_1(\boldsymbol{\Sigma})}\right)^{\frac{3}{2}}\right), \quad A_1(\boldsymbol{\Sigma}) := 2\sum_{s>1}\frac{\lambda_1(\boldsymbol{\Sigma})\lambda_s(\boldsymbol{\Sigma})}{(\lambda_1(\boldsymbol{\Sigma})-\lambda_s(\boldsymbol{\Sigma}))^2}$$

Then, using $\gamma \in (0.1, 0.5)$,

$$\frac{d-1}{4} \leq 2(d-1)\frac{1-\frac{3\gamma}{2}}{1-\left(1-\frac{3\gamma}{2}\right)^2} \leq A_1(\mathbf{\Sigma}) \leq 2(d-1)\frac{1-\frac{\gamma}{2}}{1-\left(1-\frac{\gamma}{2}\right)^2} \leq 10(d-1)$$

Then, since we are operating under the domain $n \geq t_2 d$ for t_2 sufficiently large, then using the bounds on $\frac{\mathrm{Tr}(\mathbf{\Sigma})}{\lambda_1(\mathbf{\Sigma})}$ in (61), we have $\frac{\mathrm{Tr}(\mathbf{\Sigma})}{n\lambda_1(\mathbf{\Sigma})} \ll 1$, and therefore there exist universal constants $0 < c_1 < c_2$ such that under the event $\mathcal{E}_n$,

$$c_1 \frac{d}{n} \leq \mathbb{E}\left[\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}\mathbf{v}^\top\right\|_{\mathrm{F}}^2 | \mathcal{E}_n\right] \leq c_2 \frac{d}{n}$$

Therefore,

$$\mathbb{E}\left[\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}\mathbf{v}^\top\right\|_{\mathrm{F}}^2\right] \leq \mathbb{E}\left[\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}\mathbf{v}^\top\right\|_{\mathrm{F}}^2 | \mathcal{E}_n\right] + 2\mathbb{P}\left(\mathcal{E}_n\right) \leq c_2 \frac{d}{n} + 4e^{-d}$$

Similarly, using (61),

$$\begin{aligned}\mathbb{E}\left[\frac{(\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))}{2}\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}_1\mathbf{v}_1^\top\right\|_{\mathrm{F}}^2\right] &\geq \mathbb{P}\left(\mathcal{E}_n\right)\mathbb{E}\left[\frac{(\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))}{2}\left\|\widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top - \mathbf{v}_1\mathbf{v}_1^\top\right\|_{\mathrm{F}}^2 \middle| \mathcal{E}_n\right] \\ &\geq c_1 r\left(1 - \frac{12}{10}\gamma\right)\frac{d}{n}\end{aligned}$$

Using Lemma 34 along with $\nu := 1600d/\gamma^2$ and $\gamma \in (0.1, 0.5)$,

$$\mathbb{E}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}|X\right] - \widehat{\mathbf{\Sigma}}\right\|_{\mathrm{op}}^2\right] \leq O\left(\left(\frac{\nu}{n}\right)^2\left(1 + \left(\frac{d}{n}\right) + \left(\frac{d}{n}\right)^2\right)\right) \leq O\left(\left(\frac{d}{n}\right)^2 + \left(\frac{d}{n}\right)^3 + \left(\frac{d}{n}\right)^4\right) \quad (62)$$

Substituting in (57), and using the domain $n \geq t_2 d$ and $\rho^2 > e^{-d}$, we have for a universal constant $c_3 > 0$,

$$\begin{aligned}\mathbb{E}\left[\left\langle\widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\rangle\right] &\geq c_1 r\left(1 - \frac{12}{10}\gamma\right)\frac{d}{n} - c_3\sqrt{\left(\rho^2 + c_2\frac{d}{n} + 4e^{-d}\right)\left(\frac{d}{n}\right)^2} \\ &\geq \frac{2c_1}{5}\frac{d}{n} - c_3\frac{d}{n}\sqrt{\left(\rho^2 + c_2\frac{d}{n}\right)}\end{aligned}$$

Then, for $\rho < t_1$ and $n \geq t_2 d$ for sufficiently small $t_1 < 1$, and sufficiently large $t_2 > 1$,

$$\mathbb{E}\left[\left\langle\widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\rangle\right] \geq \frac{c_1}{5}\frac{d}{n}$$

The claim then follows by comparing with the upper bound in (54). □

Remark 3. *We note that Theorem 7 provides a lower bound on the sample complexity of any algorithm achieving small error in the $\mathbb{E}[\|\cdot\|_{\mathrm{F}}^4]$ metric, via standard reduction argument such as that in Lemma 11 of [Nar24], we can convert this to a lower bound on any (ϵ, δ)-DP algorithm achieving small $\sin^2$ error with probability atleast $2/3$ at the cost of additional logarithmic factors. Further, note that the assumption $n \geq t_2 d$ follows from non-private PCA lower bounds (see e.g Example 15.19 [Wai19]) and can therefore be assumed without loss of generality to prove lower bounds for private estimators.*

F Exponential-mechanism for Sparse PCA

In this section, we provide a simple information-theoretic private algorithm for Problem 2 that only assumes sub-Gaussian samples, a positive eigengap, and sparsity of the leading eigenvector. In particular, unlike Model 2 as defined earlier, this section does not assume that $\mathbf{\Sigma}$ is k-RCS. The algorithm (Algorithm 4) is computationally inefficient, as it ranges over all supports of size k, but it will be useful as a clean baseline.

Algorithm 4 Exponential-mechanism support selection for sparse PCA

Require: Samples $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$, sparsity $k \in [d]$, privacy parameter $\epsilon > 0$, clipping level $\tau > 0$
Ensure: A private unit vector $\hat{\mathbf{v}} \in \mathbb{R}^d$

Define $\mathcal{S}_k \leftarrow \{S \subseteq [d] : |S| = k\}$
for $i = 1, \dots, n$ **do**
  for $j = 1, \dots, d$ **do**
    $(\mathbf{z}_i)_j \leftarrow \mathsf{sign}\left((\mathbf{x}_i)_j\right) \min\left\{\left|(\mathbf{x}_i)_j\right|, \tau\right\}$
  end for
end for
for each $S \in \mathcal{S}_k$ **do**
  Compute

$$\widehat{\mathbf{\Sigma}}_S^\tau \leftarrow \frac{1}{n} \sum_{i=1}^n \mathbf{z}_{i,S} \mathbf{z}_{i,S}^\top$$

  Compute $q_D(S) \leftarrow \lambda_1\left(\widehat{\mathbf{\Sigma}}_S^\tau\right)$
end for
Sample $\hat{S} \in \mathcal{S}_k$ from the distribution

$$\mathbb{P}_D(S) = \frac{\exp\left(\frac{\epsilon n}{4k\tau^2} q_D(S)\right)}{\sum_{T \in \mathcal{S}_k} \exp\left(\frac{\epsilon n}{4k\tau^2} q_D(T)\right)}$$

Let $\hat{\mathbf{v}}_{\hat{S}}$ be a top eigenvector of $\widehat{\mathbf{\Sigma}}_{\hat{S}}^\tau$
Set $\hat{\mathbf{v}}_j \leftarrow (\hat{\mathbf{v}}_{\hat{S}})_j$ for $j \in \hat{S}$, and $\hat{\mathbf{v}}_j \leftarrow 0$ for $j \notin \hat{S}$
return $\hat{\mathbf{v}}$

The following result summarize standard properties of the exponential mechanism; see [DR14] [Definition 3.4, Theorem 3.10, Theorem 3.11, Corollary 3.12].

Lemma 37 (Exponential mechanism)**.** *Let $\mathcal{R}$ be a finite set, and let $q : \mathcal{D}^n \times \mathcal{R} \to \mathbb{R}$ be a score function with global sensitivity $\Delta_q := \sup_{r \in \mathcal{R}} \sup_{D \sim D'} |q(D, r) - q(D', r)|$. Then, the mechanism $\mathcal{M}$ by $\mathbb{P}\left(\mathcal{M}(D) = r\right) \propto \exp\left(\frac{\epsilon}{2\Delta_q} q(D, r)\right)$ is ϵ-DP. Moreover, for every $\beta \in (0, 1)$, with probability at least $1 - \beta$,*

$$q\left(D, \mathcal{M}(D)\right) \geq \max_{r \in \mathcal{R}} q(D, r) - \frac{2\Delta_q}{\epsilon}\left(\log|\mathcal{R}| + \log\frac{1}{\beta}\right).$$

Let

$$\mathcal{S}_k := \{S \subseteq [d] : |S| = k\}.$$

For $\tau > 0$, define clipped samples $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{R}^d$ by

$$(\mathbf{z}_i)_j := \mathsf{sign}\left((\mathbf{x}_i)_j\right) \min\left\{\left|(\mathbf{x}_i)_j\right|, \tau\right\}, \qquad i \in [n],\ j \in [d].$$

For each $S \in \mathcal{S}_k$, define the clipped empirical covariance

$$\widehat{\mathbf{\Sigma}}_S^\tau := \frac{1}{n} \sum_{i=1}^n \mathbf{z}_{i,S} \mathbf{z}_{i,S}^\top,$$

and the score

$$q_D(S) := \lambda_1\left(\widehat{\mathbf{\Sigma}}_S^\tau\right).$$

We then consider the distribution on $\mathcal{S}_k$ given by

$$\mathbb{P}_D(S) \propto \exp\left(\frac{\epsilon n}{4k\tau^2} q_D(S)\right).$$

The first lemma gives privacy.

Lemma 38 (Privacy)**.** *For every $S \in \mathcal{S}_k$, the score $q_D(S)$ has sensitivity at most*

$$\Delta_q \le \frac{2k\tau^2}{n}$$

Consequently, if Algorithm 4 samples exactly from the distribution $\mathbb{P}_D$, then it is ϵ-DP.

Proof. Fix $S \in \mathcal{S}_k$, and let D, D' be adjacent datasets. Then

$$\widehat{\mathbf{\Sigma}}_S^\tau(D) - \widehat{\mathbf{\Sigma}}_S^\tau(D') = \frac{1}{n}\left(\mathbf{z}\mathbf{z}^\top - \mathbf{z}'(\mathbf{z}')^\top\right),$$

where $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^k$ are the clipped restrictions of the differing samples to S. Since $\|\mathbf{z}\|_2^2, \|\mathbf{z}'\|_2^2 \le k\tau^2$,

$$\left\|\widehat{\mathbf{\Sigma}}_S^\tau(D) - \widehat{\mathbf{\Sigma}}_S^\tau(D')\right\|_{\mathrm{op}} \le \frac{\|\mathbf{z}\|_2^2 + \|\mathbf{z}'\|_2^2}{n} \le \frac{2k\tau^2}{n}.$$

Since $\lambda_1(\cdot)$ is 1-Lipschitz with respect to operator norm on symmetric matrices,

$$|q_D(S) - q_{D'}(S)| \le \left\|\widehat{\mathbf{\Sigma}}_S^\tau(D) - \widehat{\mathbf{\Sigma}}_S^\tau(D')\right\|_{\mathrm{op}} \le \frac{2k\tau^2}{n}.$$

The claim follows from the standard privacy guarantee of the exponential mechanism (Lemma 37). □

We next record the utility guarantee. Let $\mathbf{v}_1 := \mathbf{v}_1(\mathbf{\Sigma})$.

Lemma 39 (Utility)**.** *Assume $\mathbf{x}_1, \dots, \mathbf{x}_n \overset{\mathsf{iid}}{\sim} \mathcal{D}$ are mean-zero with covariance $\mathbf{\Sigma} \in \mathbb{S}_{\succeq \mathbf{0}}$, and $\lambda_2(\mathbf{\Sigma}) \le (1-\gamma)\lambda_1(\mathbf{\Sigma})$ for some $\gamma \in (0, 1]$. Assume further that $\mathrm{nnz}(\mathbf{v}_1) \le k$. Let $\hat{\mathbf{v}}$ be the output of Algorithm 4 with $\tau = \Theta\left(\sigma\sqrt{\log(nd/\beta)}\right)$. Then with probability at least $1-\beta$ over the exponential mechanism and the samples $\{\mathbf{x}_i\}_{i\in[n]}$,*

$$\sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v}_1) \lesssim \frac{1}{\gamma} \cdot \frac{\sigma^2}{\lambda_1(\mathbf{\Sigma})} \cdot \left(\sqrt{\frac{k}{n}\log\left(\frac{d}{k\beta}\right)} + \frac{k^2}{\epsilon n}\log\left(\frac{d}{k\beta}\right)\log\left(\frac{nd}{\beta}\right)\right).$$

Proof. For $\eta > 0$, define the event

$$\mathcal{E}_\eta := \left\{ \sup_{S \in \mathcal{S}_k} \left\| \widehat{\mathbf{\Sigma}}_S^\tau - \mathbf{\Sigma}_{S \times S} \right\|_{\mathrm{op}} \leq \eta \right\}.$$

Let $S_\star \in \mathcal{S}_k$ contain $\mathrm{supp}(\mathbf{v}_1)$. Since $\mathbf{v}_1$ is supported on $S_\star$, $\lambda_1\left(\mathbf{\Sigma}_{S_\star \times S_\star}\right) = \lambda_1(\mathbf{\Sigma})$.

On the event $\mathcal{E}_\eta$,

$$q_D(S_\star) = \lambda_1\left(\widehat{\mathbf{\Sigma}}_{S_\star}^\tau\right) \geq \lambda_1(\mathbf{\Sigma}) - \eta.$$

By the utility guarantee of the exponential mechanism (Lemma 37), with probability at least $1 - \beta$,

$$q_D(\hat{S}) \geq q_D(S_\star) - \frac{2\Delta_q}{\epsilon}\left(\log|\mathcal{S}_k| + \log\frac{1}{\beta}\right) \geq \lambda_1(\mathbf{\Sigma}) - \eta - \frac{4k\tau^2}{\epsilon n}\left(\log\binom{d}{k} + \log\frac{1}{\beta}\right).$$

Since $\hat{\mathbf{v}}$ is supported on $\hat{S}$ and is a top eigenvector of $\widehat{\mathbf{\Sigma}}_{\hat{S}}^\tau$,

$$\hat{\mathbf{v}}^\top \widehat{\mathbf{\Sigma}}_{\hat{S}}^\tau \hat{\mathbf{v}} = q_D(\hat{S}).$$

Using again the event $\mathcal{E}_\eta$,

$$\hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} = \hat{\mathbf{v}}_{\hat{S}}^\top \mathbf{\Sigma}_{\hat{S} \times \hat{S}} \hat{\mathbf{v}}_{\hat{S}} \geq \hat{\mathbf{v}}_{\hat{S}}^\top \widehat{\mathbf{\Sigma}}_{\hat{S}}^\tau \hat{\mathbf{v}}_{\hat{S}} - \eta = q_D(\hat{S}) - \eta.$$

Therefore,

$$\hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} \geq \lambda_1(\mathbf{\Sigma}) - 2\eta - \frac{4k\tau^2}{\epsilon n}\left(\log\binom{d}{k} + \log\frac{1}{\beta}\right).$$

Finally,

$$\lambda_1(\mathbf{\Sigma}) - \hat{\mathbf{v}}^\top \mathbf{\Sigma} \hat{\mathbf{v}} \geq (\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))\sin^2\angle(\hat{\mathbf{v}}, \mathbf{v}_1) \geq \gamma\lambda_1(\mathbf{\Sigma})\sin^2\angle(\hat{\mathbf{v}}, \mathbf{v}_1),$$

Under sub-Gaussianity, from Fact 4 and Corollary 1, taking $\tau = \Theta\left(\sigma\sqrt{\log(nd/\beta)}\right)$ ensures that clipping is inactive with high probability. On this event, $\widehat{\mathbf{\Sigma}}_S^\tau$ coincides with the usual empirical covariance restricted to S, and a union bound over $\binom{d}{k}$ supports yields, with probability atleast $1 - \beta$,

$$\eta \lesssim \sigma^2\left(\sqrt{\frac{k\log(ed/k) + \log(1/\beta)}{n}} + \frac{k\log(ed/k) + \log(1/\beta)}{n}\right).$$

which proves the claim. □