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

Syamantak Kumar SYAMANTAK@UTEXAS.EDU
Department of Computer Science, UT Austin

Shourya Pandey SHOURYAP@UTEXAS.EDU
Department of Computer Science, UT Austin

Purnamrita Sarkar PURNA.SARKAR@AUSTIN.UTEXAS.EDU
Department of Statistics and Data Science, UT Austin

Kevin Tian KJTIAN@CS.UTEXAS.EDU
Department of Computer Science, UT Austin

Editors: Steve Hanneke and Tor Lattimore

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 (Wang and Xu, 2021) 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 problem 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.

Keywords: Differential privacy, sparse covariance estimation, sparse PCA, minimax lower bounds

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 (Johnstone and Lu, 2009; Baik et al., 2005; Paul, 2007). Thus in high dimensions, meaningful inference requires additional structural assumptions on the covariance

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

matrix, such as row sparsity, sparsity of the principal eigenspace, or a combination thereof (Bickel and Levina, 2009; Cai et al., 2010). 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 (Johnstone and Lu, 2009; Vu and Lei, 2013; Berthet and Rigollet, 2013; Wang et al., 2016; Amini and Wainwright, 2008; Ma, 2013; Deshpande and Montanari, 2016; Kumar and Sarkar, 2024; Qiu et al., 2023).

At the same time, many applications involving sensitive data require rigorous differential privacy (DP) guarantees. Compared to the vast literature on differentially private PCA (Chaudhuri et al., 2013; Dwork et al., 2014; Liu et al., 2022), 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 (Ge et al., 2018; Wang and Xu, 2021; Li and Wang, 2023) 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 (Wang and Xu, 2021) 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 suggests 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 for preliminaries on differential privacy and matrix concentration.

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*

2. Under the scaling discussed near Eq. (2) of (Wang and Xu, 2021), 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 (Bickel and Levina, 2009)	$k^2 + \sqrt{d} \cdot k^{1.5}$ (Thm. 22)	$d \cdot k^2$ (Wang and Xu, 2021)
Lower Bound	k^2 (Cai and Zhou, 2012)	$k^2 + \sqrt{d} \cdot k$ (Thm. 27)	None

a leading eigenvector $\mathbf{v}_1 \in \mathbb{R}^d$, and eigenvalues satisfying $\lambda_2(\mathbf{\Sigma}) \le (1-\gamma)\lambda_1(\mathbf{\Sigma})$. 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 $\text{nnz}(\mathbf{v}_1) \le k$.*

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 $1-\beta$,*

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

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 $\hat{\mathbf{v}} \in \mathbb{R}^d$ satisfying, with probability at least $1-\beta$,*

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

Private sparse covariance estimation. In Appendix D, 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 22) 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 27) shows that this $\sqrt{d}$ dependence is tight, even under approximate DP. Our bound is proven by adapting the fingerprinting techniques of (Narayanan, 2024) for DP covariance estimation, and extending them to sparse models via a graph-based construction (Construction 1). 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.

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 (Bickel and Levina, 2009)[3]	k^4 (Thm. 3)	$d \cdot k^2$ (Wang and Xu, 2021)
Lower Bound (Model 1)	k (Vu and Lei, 2013)	d (Thm. 12)	None

Private sparse PCA. In Sections 3 and 4, we respectively give new upper and lower bounds for Problem 2, both with and without the eigenvector sparsity assumption in Model 2. Our results are summarized in Table 2.

Our main upper bound (Theorem 3) demonstrates that under Model 2, poly$(k, \log d)$ samples suffice to solve Problem 2 subject to approximate DP (omitting dependences on other parameters). Our algorithm uses the FRIENDLYCORE primitive of (Tsfadia et al., 2022) 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 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 12, 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, (Kamath et al., 2020)) 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., $k =$ 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 (Bun et al., 2014; Steinke and Ullman, 2017).[4] However, Theorem 12 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 poly(d) sample complexity is avoidable (when, say, $k = O(\text{polylog}(d))$. Our Theorems 3 and 12 demonstrate that the modeling assumptions can provably shift the sample complexity landscape for DP problems in sparse covariance estimation.

One notable caveat is that the privacy guarantees in Theorems 3 and 12 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 poly$(k, \log d)$ samples.

3. This result holds even under Model 1.
4. We also mention conceptually-related results by (Kasiviswanathan et al., 2011; Cheu and Ullman, 2021; Nissim and Yan, 2022), which showed similar exponential gaps under variants of DP, particularly, the local or shuffle DP models.

Towards investigating this possibility, in Theorem 47 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 $\|\mathbf{\Sigma}\|_{\text{op}}$. We show that $\gtrsim d/\epsilon$ samples are required to solve the problem even under approximate DP, by adapting the private Assouad's method of (Acharya et al., 2021). 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 poly$(k, \log d)$ sample complexities under different models. We leave proving a poly(d) lower bound, or excitingly, a 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 55 demonstrates that $\gtrsim d$ samples are needed under approximate DP, for a much broader parameter range than previously known: e.g., Section 4.2, (Cai et al., 2024) and Theorem 5.4, (Liu et al., 2022) 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 Bickel and Levina (2009); Rothman et al. (2009); Cai and Liu (2011); Deshpande and Montanari (2016). In parallel, sparse PCA has been extensively studied via optimization-based formulations and algorithmic relaxations, including ℓ_1-penalized or regression-style approaches Zou et al. (2006), semidefinite relaxations d'Aspremont et al. (2004); Amini and Wainwright (2008), and iterative schemes such as truncated power methods Yuan and Zhang (2013). A complementary thread establishes statistical limits and computational barriers in sparse PCA, clarifying when polynomial-time methods can (or cannot) attain minimax-optimal rates Berthet and Rigollet (2013); Wang et al. (2016).

Differential privacy for subspace estimation when $d \le n$. Differential privacy (DP) provides a rigorous framework for protecting individuals' contributions in statistical analyses Dwork and Roth (2014). 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 Dwork et al. (2014); Cai et al. (2024) and private iterative methods for dominant subspaces Hardt and Price (2014); Liu et al. (2022); Düngler and Sanyal (2025). More broadly, private low-rank approximation and private linear-algebraic primitives have been developed as building blocks for downstream tasks Kapralov and Talwar (2013). 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 Kamath et al. (2019).

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 Ge et al. (2018); Wang and Xu (2021); Li and Wang (2023) with $O(1)$ norm bounded assumptions, which often does not hold for high-dimensional data. Differentially private top-k and sparse selection mechanisms, often used to locate large coordinates/entries before estimating magnitudes, have been studied extensively Durfee and Rogers (2019); Qiao et al. (2021). 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 Hardt and Talwar (2010), as well as DP analogues of classical minimax techniques (Assouad/Fano/Le Cam) Acharya et al. (2021). Fingerprinting codes and their variants provide sharp lower bounds for answering many queries and for private statistical estimation, highlighting fundamental gaps between non-private and approximate-DP sample complexity Hardt and Talwar (2010); Bun et al. (2014); Steinke and Ullman (2015). Recent work streamlines and strengthens the fingerprinting approach for modern estimation problems Narayanan (2024). 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

General notation. We use $\widetilde{\Theta}(\cdot)$, $\widetilde{O}(\cdot)$, and $\widetilde{\Omega}(\cdot)$ to hide polylogarithmic factors in the ambient problem parameters. We use $\mathrm{poly}(\cdot)$ and $\mathrm{polylog}(\cdot)$ for unspecified polynomial and polylogarithmic dependences. For $n \in \mathbb{N}$ we let $[n] := \{i \in \mathbb{N} : i \le n\}$. For $a, b \in \mathbb{N}$, $a \mid b$ denotes that a divides 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 $\|\mathbf{v}\|_p$ 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 $\mathbb{R}^d$. For equal-length strings or vectors $\mathbf{x}, \mathbf{y}$, we write $d_{\text{ham}}(\mathbf{x}, \mathbf{y}) := \sum_i \mathbb{1}(\mathbf{x}_i \neq \mathbf{y}_i)$ for their Hamming distance.

Probability notation. 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 two probability distributions P, Q on the same measurable space, $\mathrm{TV}(P, Q) := \sup_{\mathcal{A}} |P(\mathcal{A}) - Q(\mathcal{A})|$ denotes total variation distance. We write $X \sim P$ when X has law P, and write $X \overset{\mathrm{d}}{=} Y$ when X

and Y have the same distribution. We write $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{iid}}{\sim} P$ to denote independent samples with common law P, and $X \overset{\text{unif.}}{\sim} S$ to denote that X is uniform on a finite set S.

Notations on matrices. For a symmetric matrix $\mathbf{M}$, we order its eigenvalues as $\lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{M}) \geq \cdots$ and write $\mathbf{v}_1(\mathbf{M})$ for a unit leading eigenvector when it is well-defined. In set-builder notation, we use a colon to separate the ambient set from the defining condition. 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 $\mathbf{A}_{i,:}, \mathbf{A}_{:,i}$ to 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. 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 $\langle \mathbf{M}, \mathbf{N} \rangle = \text{Tr}(\mathbf{M}\mathbf{N})$. 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 $\|\mathbf{M}\|_{\text{op}} := \sqrt{\lambda_1(\mathbf{M}\mathbf{M}^\top)}$, $\|\mathbf{M}\|_{\text{F}} := \sqrt{\sum_{i,j \in [d]} \mathbf{M}_{ij}^2}$, $\|\mathbf{M}\|_{\infty,\infty} := \sup_{i,j \in [d]} |\mathbf{M}_{ij}|$. Denote by $\mathbb{B}_\infty(\mathbf{M}, \tau)$ the set of matrices $\mathbf{M}'$ satisfying $\|\mathbf{M} - \mathbf{M}'\|_{\infty,\infty} \leq \tau$.

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 $i \in [m]$, $\text{nnz}(\mathbf{M}_{i,:}) \leq k$ and for all $j \in [n]$, $\text{nnz}(\mathbf{M}_{:,j}) \leq k$.*

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 $\text{top}_k(\mathbf{v})$ 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:

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

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

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

3. Private Sparse PCA: Upper Bound

In this section, we provide our upper bounds for Problem 2 under Model 2, our strengthening of Model 1 that posits the leading eigenvector is k-sparse. Importantly, our main result (Theorem 3) solves Problem 2 with a sample complexity scaling *polylogarithmically* in d (and polynomially in other problem parameters), bypassing the curse of dimensionality.

Definition 2 (Goodness) *We say that $\mathbf{M} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$ is $(k, \hat{\lambda}, \gamma, \tau)$-good if $\mathcal{T}_\tau(\mathbf{M})$ is k-RCS, and*

$$\lambda_1(\mathcal{T}_\tau(\mathbf{M})) \geq (1 - \gamma)\,\hat{\lambda}, \qquad \lambda_2(\mathcal{T}_\tau(\mathbf{M})) \leq (1 - \gamma)\,\lambda_1(\mathcal{T}_\tau(\mathbf{M})).$$

Algorithm 1: $\mathsf{PrivPCA}(\mathcal{D}, \hat{\lambda}, \gamma, k, \epsilon, \delta, \tau, m)$

Input: Dataset $\mathcal{D} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i \in [n]}$, operator norm estimate $\hat{\lambda} > 0$, gap $\gamma > 0$, sparsity $k \in [d]$, privacy $(\epsilon, \delta) \in (0, 1)^2$, threshold $\tau > 0$, and number of batches m.

Output: Either $\perp$ or a private unit vector $\hat{\mathbf{v}}$.

$b \leftarrow n/m$;
$\widehat{\mathbf{\Sigma}}_i \leftarrow b^{-1} \sum_{j=(i-1)b+1}^{ib} \mathbf{x}_j \mathbf{x}_j^\top$ for all $i \in [m]$;
$\mathcal{C} \leftarrow \{i \in [m] : \widehat{\mathbf{\Sigma}}_i$ is $(k, \hat{\lambda}, \gamma/2, \tau)$-good$\}$;
Sample $L \sim \mathsf{BoundedLap}(\frac{3}{\epsilon}, \frac{3}{\epsilon} \log(\frac{12}{\delta}))$;
if $|\mathcal{C}| + L - \frac{3}{\epsilon} \log(\frac{12}{\delta}) < 0.8m$ **then**
    return $\perp$;
end
for $i \in [m]$ **do**
    $f_i \leftarrow \sum_{j \in [m]} \mathbb{1}\{\|\widehat{\mathbf{\Sigma}}_j - \widehat{\mathbf{\Sigma}}_i\|_{\infty,\infty} \le 2\tau\}$;
    $p_i \leftarrow \min\{\max\{(f_i - m/2)/(m/6), 0\}, 1\}$;
end
$Z \leftarrow \sum_{i \in [m]} p_i$;
Sample $\xi \sim \mathsf{BoundedLap}(\frac{21}{\epsilon}, \frac{21}{\epsilon} \log(\frac{12}{\delta}))$;
if $Z + \xi - \frac{21}{\epsilon} \log(\frac{12}{\delta}) < 0.8m$ **then**
    return $\perp$;
end
$\bar{\mathbf{\Sigma}} \leftarrow Z^{-1} \sum_{i \in [m]} p_i \widehat{\mathbf{\Sigma}}_i$;
$\mathbf{u} \leftarrow \mathbf{v}_1(\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}))$ and $\widehat{\mathbf{P}} \leftarrow \mathbf{u}\mathbf{u}^\top$;
$\sigma_{\text{priv}} \leftarrow \frac{80\sqrt{2}k\tau}{\gamma\hat{\lambda}} \frac{6}{\epsilon} \sqrt{\log(\frac{6}{\delta})}$;
Sample $\mathbf{G} \in \mathbb{R}^{d \times d}$ with $\mathbf{G}_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{priv}}^2)$;
$\widetilde{\mathbf{P}} \leftarrow \mathrm{top}_{k^2}(\widehat{\mathbf{P}} + \mathbf{G})$;
return $\hat{\mathbf{v}} := \mathbf{v}_1(\frac{1}{2}(\widetilde{\mathbf{P}} + \widetilde{\mathbf{P}}^\top))$;

Algorithm 1 proceeds in two stages. The first stage computes an aggregated estimate $\bar{\mathbf{\Sigma}}$ as a weighted average over m batches of sample covariances, after certifying that enough batch covariances are good (Definition 2). Importantly, this guarantees that $\bar{\mathbf{\Sigma}}$ has bounded sensitivity under $\|\cdot\|_{\infty,\infty}$, using an argument inspired by the FriendlyCore framework of Tsfadia et al. (2022) (more specifically, its recent simplified variants in Lowy et al. (2024); Kumar et al. (2025)). The second stage employs a standard clip-and-PCA procedure on the top eigenspace projector of $\bar{\mathbf{\Sigma}}$, after applying entrywise Gaussian noise to ensure privacy.

For a sufficiently large constant $C > 0$, throughout the analysis we set

$$\tau := C\sigma^2 \sqrt{\frac{\log(\frac{d}{\beta})}{b}}, \qquad b := \frac{n}{m}. \tag{3}$$

Theorem 3 (Private k-sparse PCA upper bound) *There is an algorithm (Algorithm 1 with τ set as in (3), and using Proposition 5 to compute $\hat{\lambda}$) that solves Problem 2, for $\mathcal{D} = \{\mathbf{x}_i\}_{i\in[n]}$ drawn from Model 2, with*

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

for a sufficiently large constant.

The main difficulty is that the sensitivity analysis must hold for worst-case adjacent datasets; it cannot assume that the block covariances are close to the population covariance $\mathbf{\Sigma}$. For this reason, Algorithm 1 uses a propose-test-release structure. The first private test checks that many block covariance estimates are certified: after thresholding at level τ, they are k-RCS, have a sufficiently large leading eigenvalue, and have a sufficient eigengap. The second private test checks that the total FriendlyCore weight Z is large. If either test fails, the algorithm outputs $\perp$.

On the certified branch, we prove deterministic sensitivity bounds. Consider two adjacent datasets, producing weighted estimates $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Sigma}}'$. Since the two runs both pass the tests, there are many certified and many surviving block estimates in each run. A counting argument then shows that the two runs share at least one certified surviving block covariance. We denote this common block covariance by $\mathbf{M}$. Importantly, $\mathbf{M}$ is not the population covariance; it is a data-dependent common center used only in the sensitivity proof.

Lemma 4 (Sensitivity of the weighted average) *Let m be sufficiently large, let $\mathcal{D}, \mathcal{D}'$ be neighboring datasets, and condition on two runs of Algorithm 1 on $\mathcal{D}, \mathcal{D}'$ both reaching Line 17. There exists $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ such that $\mathbf{M}$ is $(k, \hat{\lambda}, \gamma/2, \tau)$-good, and, denoting by $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Sigma}}'$ the matrices computed on Line 17 by these two runs,*

$$\max\left\{\left\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_{\tau}(\mathbf{M})\right\|_{\text{op}}, \left\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}') - \mathcal{T}_{\tau}(\mathbf{M})\right\|_{\text{op}}\right\} \le 10k\tau.$$

Conditioned on Line 18 being reached on a run of Algorithm 1, the matrix $\widehat{\mathbf{P}}$ computed on this line is $\Delta_{\mathbf{P}}$-sensitive in $\|\cdot\|_{\text{F}}$, where

$$\Delta_{\mathbf{P}} := \frac{80\sqrt{2}k\tau}{\gamma\hat{\lambda}}.$$

Indeed, fix adjacent $\mathcal{D}$, $\mathcal{D}'$ and let $\mathbf{M}$ be the result of Lemma 4. By goodness of $\mathbf{M}$, $\mathcal{T}_{\tau}(\mathbf{M})$ has a unique leading eigenvector $\mathbf{w}$, with associated projector $\mathbf{W} := \mathbf{w}\mathbf{w}^\top$. By Lemma 4, $\|\mathcal{T}_{\tau}(\mathbf{M}) - \mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})\|_{\text{op}} \le 10k\tau$. Thus, Lemma 14 with $\mathbf{A} \leftarrow \mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})$, $\mathbf{B} \leftarrow \mathcal{T}_{\tau}(\mathbf{M})$, and gap $\leftarrow \frac{\gamma\hat{\lambda}}{2}$, implies

$$\left\|\widehat{\mathbf{P}} - \mathbf{W}\right\|_{\text{op}} \le \frac{40k\tau}{\gamma\hat{\lambda}}.$$

The same bound symmetrically applies for $\|\widehat{\mathbf{P}}' - \mathbf{W}\|_{\text{op}}$, and thus because $\widehat{\mathbf{P}}$, $\widehat{\mathbf{P}}'$ are rank-one,

$$\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\text{F}} \le \sqrt{2}\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\text{op}} \le \sqrt{2}\left(\left\|\widehat{\mathbf{P}} - \mathbf{W}\right\|_{\text{op}} + \left\|\widehat{\mathbf{P}}' - \mathbf{W}\right\|_{\text{op}}\right) \le \frac{80\sqrt{2}k\tau}{\gamma\hat{\lambda}}.$$

The Frobenius sensitivity of the map $\mathcal{D} \mapsto \widehat{\mathbf{P}}(\mathcal{D})$ assuming Line 18 is reached is therefore at most $\Delta_{\mathbf{P}}$. The Gaussian mechanism (Fact 2), applied to the vectorization of $\widehat{\mathbf{P}}$, then yields $(\frac{\epsilon}{3}, \frac{\delta}{3})$-DP for releasing $\widehat{\mathbf{P}} + \mathbf{G}$ for the stated setting of σ_{priv}. All subsequent steps (top_{k^2}, symmetrization, and computing $\mathbf{v}_1(\cdot)$) are postprocessings of this private statistic. The two tests are differentially private by the bounded Laplace mechanism, so basic composition gives (ϵ, δ)-DP for the final output.

Utility is proved by showing that under Model 2, the tests pass with high probability and the aggregated $\bar{\mathbf{\Sigma}}$ is close to the population $\mathbf{\Sigma}$. The k-RCS assumption is used before eigenvector recovery, to convert entrywise stability into operator-norm control. The sparsity of $\mathbf{v}_1$, imposed only in Model 2, is used in the final recovery step: the projector $\mathbf{v}_1\mathbf{v}_1^\top$ is k^2-sparse, so top-k^2 truncation after the noisy release preserves the signal.

We require one additional ingredient: a private eigenvalue estimation procedure under Model 1, whose proof is deferred to Appendix G.

Proposition 5 *Let $(\epsilon, \delta, \beta) \in (0,1)^3$, and let $\mathcal{D}$ be drawn from Model 1 with*

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

for a sufficiently large constant. There is an (ϵ, δ)-DP algorithm, PrivNorm *(Algorithm 3), which returns $\hat{\lambda}$ satisfying*

$$\hat{\lambda} \in \left[(1 - \tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma}), (1 + \tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma})\right]$$

with probability at least $1 - \beta/2$.

4. Private Sparse PCA: Lower Bound

In this section, we prove Theorem 12, our lower bound for PCA for k-RCS covariance matrices (Problem 2 under Model 1) with *pure DP*. Our argument uses the following well-established *packing lower bound* framework with a careful combinatorial construction.

Proposition 6 (Kamath et al. (2020), 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 disjoint $i, j \in [m]$. Suppose $\mathcal{A} : \Gamma \to [m]$ 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 $n = \Omega(\frac{\log m}{\alpha\epsilon})$.

Our construction for instating Proposition 6, combines two well-known combinatorial ingredients. The first is a variant of the Gilbert-Varshamov bound (see e.g., Theorem 7, Graham and Sloane (2003)).

Lemma 7 *Let $g \in \mathbb{N}$ and let $\mathcal{S} := \{\mathbf{s} \in \{0,1\}^{3g} : \sum_{i \in [3g]} \mathbf{s}_i = g\}$. There exists a subset $C \subseteq \mathcal{S}$ of size $\exp(\Omega(g))$ such that for any distinct $\mathbf{x}, \mathbf{y} \in C$, $d_{\text{ham}}(\mathbf{x}, \mathbf{y}) \geq \frac{g}{2}$.*

Proof The construction is greedy: repeatedly place any $\mathbf{x} \in \mathcal{S}$ in C, with Hamming distance at least $\frac{g}{2}$ from every element in C, until none exist. For any $\mathbf{x} \in \mathcal{S}$, there are at most

$$1 + \sum_{t \in [\frac{g}{4}]} \binom{g}{t}\binom{2g}{t} \le 1 + \frac{g}{4}\binom{g}{\frac{g}{4}}\binom{2g}{\frac{g}{4}} \le 1 + \frac{g}{4} \cdot 2^{gH_2(\frac{1}{4}) + 2gH_2(\frac{1}{8})} \le \frac{g \cdot 2^{1.9g}}{3}.$$

binary vectors $\mathbf{y} \in \mathcal{S}$ at Hamming distance $\le \frac{g}{2}$ from $\mathbf{x}$, where for $p \in (0,1)$, $H_2(p) := -p\log(p) - (1-p)\log(1-p)$. Thus, a volume argument yields

$$|\mathcal{C}| \ge \frac{\binom{3g}{g}}{\frac{g \cdot 2^{1.9g}}{3}} \ge \frac{3}{g \cdot 2^{1.9g}} \cdot \frac{0.4}{\sqrt{g}} \left(\frac{27}{4}\right)^g \ge \frac{4^g}{g^{\frac{3}{2}}} = \exp\left(\Omega(g)\right).$$

■

The second ingredient is the existence of a bipartite expander graph with appropriate spectral characteristics. Our packing instance eventually permutes the vertices of this graph using permutations given by Lemma 7. We first use the following definition.

Definition 8 *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 9 *Let $G = (L \cup R, E)$ be a (c_1, c_2)-biregular bipartite graph, and let $\mathbf{A} \in \{0,1\}^{V \times V}$ be its adjacency matrix. Then $\lambda_1(\mathbf{A}) = \|\mathbf{A}\|_{\text{op}} = \sqrt{c_1 c_2}$, with corresponding eigenvector*

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

Proof The equality $\mathbf{A}\mathbf{v} = \sqrt{c_1 c_2}\mathbf{v}$ is a direct calculation. To show $\|\mathbf{A}\|_{\text{op}} = \sqrt{c_1 c_2}$, 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. Then $\|\mathbf{A}\|_{\text{op}}$ is the largest singular value of $\mathbf{B}$ (this can be seen by e.g., squaring $\mathbf{A}$). Therefore, the conclusion holds by Cauchy-Schwarz:

$$\sup_{\mathbf{x} \in \mathbb{R}^R : \|\mathbf{x}\|_2 = 1} \|\mathbf{B}\mathbf{x}\|_2^2 = \sup_{\mathbf{x} \in \mathbb{R}^R : \|\mathbf{x}\|_2 = 1} \sum_{\ell \in L} \left(\sum_{\substack{r \in R \\ (\ell, r) \in E}} \mathbf{x}_r \right)^2 \le \sup_{\mathbf{x} \in \mathbb{R}^R : \|\mathbf{x}\|_2 = 1} c_1 \sum_{\ell \in L} \sum_{r \in R} \mathbf{x}_r^2 = c_1 c_2.$$

■

We also require the following expander existence result of Gribinski and Marcus (2021).

Proposition 10 (Theorem 1.2, Gribinski and Marcus (2021)) *For any $(g, h, t) \in \mathbb{N}^3$, there exists a bipartite graph $G_{g,h,t} = (V = L \cup R, E)$ such that the following properties hold.*

- $|L| = tg$, $|R| = g$, $\deg(u) = h$ *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 now describe our k-RCS covariance matrix construction, leveraging Proposition 10.

Corollary 11 *Let $k \ge 6$ be even, let $d = 3g$ for $g \ge k$, and let $R \subseteq [d]$ with $|R| = g$ be arbitrary, with $L := [d] \setminus R$. There exists a k-RCS $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ such that the following properties hold:*

$$\lambda_1(\mathbf{\Sigma}) = 2, \quad \lambda_2(\mathbf{\Sigma}) < 1.97, \quad \mathbf{v}_1(\mathbf{\Sigma}) = \sqrt{\frac{3}{4d}} \begin{pmatrix} \mathbf{1}_L \\ \sqrt{2} \cdot \mathbf{1}_R \end{pmatrix}. \tag{4}$$

Proof Let $t = 2$ and $h = \frac{k-2}{2}$ in Proposition 10, and let $\mathbf{A}$ be the adjacency matrix of the resulting $(h, 2h)$-biregular graph with bipartition (L, R). We let

$$\mathbf{\Sigma} := \mathbf{I}_d + \frac{\sqrt{2}}{k-2}\mathbf{A} = \mathbf{I}_d + \frac{1}{\sqrt{2}h}\mathbf{A}.$$

The fact that $\mathbf{\Sigma}$ is k-RCS is immediate from $2h + 1 \le k$, and the remaining properties follow from combining Lemma 9 and Proposition 10, as well as the numerical bound (for $k \ge 6$):

$$\lambda_2(\mathbf{\Sigma}) \le 1 + \frac{\sqrt{2}}{k-2}\left(\sqrt{h-1} + \sqrt{2h-1}\right) = 1 + \frac{\sqrt{k-4} + \sqrt{2k-6}}{k-2} < 1.97.$$

■

Combining Lemma 7 and Corollary 11 in Proposition 6 now yields our lower bound.

Theorem 12 (Pure-DP PCA lower bound for k-RCS covariance) *Let $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ be an ϵ-DP algorithm that, under Model 1 with $\sigma^2 = 1$ and $\gamma = \frac{1}{100}$, solves Problem 2 with $\beta = \frac{1}{3}$ and $\Delta = \frac{1}{400}$ for sufficiently large d, k. Then $n = \Omega(\frac{d}{\epsilon})$.*

Proof Let $k \ge 6$ be even and $d = 3g$ for $g \ge k$. Let C be the subset of $\{0, 1\}^d$ with size $m := |C| = \exp(\Omega(d))$ guaranteed by Lemma 7. For each $\mathbf{x} \in C$, let $\mathbf{\Sigma}_{\mathbf{x}}$ be the k-RCS matrix given by Corollary 11 with $R \leftarrow \text{supp}(\mathbf{x})$. Using (4), the Gaussian $P_{\mathbf{x}} := \mathcal{N}(\mathbf{0}_d, \frac{1}{2}\mathbf{\Sigma}_{\mathbf{x}})$ satisfies Model 1 with $\sigma^2 = 1$ and $\gamma = \frac{1}{100}$. The rest of the proof establishes that whenever $\mathcal{A}$ returns $\hat{\mathbf{v}}$ with

$$\sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v}_1(\mathbf{\Sigma}_{\mathbf{x}})) \le \frac{1}{400}, \tag{5}$$

there is a deterministic $D : \mathbb{R}^d \to C$ with $D(\hat{\mathbf{v}}) = \mathbf{x}$. Note that $D \circ \mathcal{A}$ is also ϵ-DP by postprocessing, so Proposition 6 with $\alpha = 1$, $\mathcal{A} \leftarrow D \circ \mathcal{A}$, and $\mathcal{P} = \{P_{\mathbf{x}}\}_{\mathbf{x} \in C}$, proves the theorem.

We now describe the decoding algorithm D. For all $\mathbf{x} \in C$, denote by shorthand $\mathbf{v}_{\mathbf{x}} := \mathbf{v}_1(\mathbf{\Sigma}_{\mathbf{x}})$ (i.e., the vector in (4)). The algorithm simply selects $\mathbf{y} \in C$ that minimizes $d_{\text{sign}}(\mathbf{y}, \hat{\mathbf{v}})$, where

$$d_{\text{sign}}(\mathbf{u}, \mathbf{v}) := \min\left\{\|\mathbf{u} - \mathbf{v}\|_2, \|\mathbf{u} + \mathbf{v}\|_2\right\}.$$

Note that d_{sign} satisfies the triangle inequality. Indeed, for any three unit vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$, let $\sigma, \sigma' \in \{\pm 1\}$ be signs for which $d_{\text{sign}}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \sigma\mathbf{b}\|_2$ and $d_{\text{sign}}(\mathbf{b}, \mathbf{c}) = \|\mathbf{b} - \sigma'\mathbf{c}\|_2$. Then,

$$d_{\text{sign}}(\mathbf{a}, \mathbf{c}) \leq \|\mathbf{a} - \sigma\sigma'\mathbf{c}\|_2 \leq \|\mathbf{a} - \sigma\mathbf{b}\|_2 + \|\sigma\mathbf{b} - \sigma\sigma'\mathbf{c}\|_2 = d_{\text{sign}}(\mathbf{a}, \mathbf{b}) + d_{\text{sign}}(\mathbf{b}, \mathbf{c}).$$

By Lemma 19, the true $\mathbf{x}$ that indexes $P_{\mathbf{x}}$ satisfies

$$d_{\text{sign}}(\mathbf{v}_{\mathbf{x}}, \hat{\mathbf{v}}) \leq \sqrt{\frac{1}{200}}.$$

On the other hand, for any $\mathbf{y} \in C$ such that $\mathbf{y} \neq \mathbf{x}$, applying (4), we have

$$\|\mathbf{v}_{\mathbf{x}} - \mathbf{v}_{\mathbf{y}}\|_2 = \sqrt{d_{\text{ham}}(\mathbf{x}, \mathbf{y})}\left(\sqrt{\frac{3}{2d}} - \sqrt{\frac{3}{4d}}\right) \geq \sqrt{\frac{d}{6}}\left(\sqrt{\frac{3}{2d}} - \sqrt{\frac{3}{4d}}\right) > \sqrt{\frac{1}{50}}.$$

Moreover, $d_{\text{sign}}(\mathbf{v}_{\mathbf{x}}, \mathbf{v}_{\mathbf{y}}) = \|\mathbf{v}_{\mathbf{x}} - \mathbf{v}_{\mathbf{y}}\|$ because $\langle \mathbf{v}_{\mathbf{x}}, \mathbf{v}_{\mathbf{y}} \rangle > 0$. Finally, if $d_{\text{sign}}(\mathbf{v}_{\mathbf{y}}, \hat{\mathbf{v}}) \leq d_{\text{sign}}(\mathbf{v}_{\mathbf{x}}, \hat{\mathbf{v}})$ for any $\mathbf{y} \neq \mathbf{x}$, the triangle inequality would yield a contradiction:

$$\sqrt{\frac{1}{50}} < d_{\text{sign}}(\mathbf{v}_{\mathbf{x}}, \mathbf{v}_{\mathbf{y}}) \leq d_{\text{sign}}(\mathbf{v}_{\mathbf{x}}, \hat{\mathbf{v}}) + d_{\text{sign}}(\hat{\mathbf{v}}, \mathbf{v}_{\mathbf{y}}) \leq 2\sqrt{\frac{1}{200}}.$$

Thus, D correctly returns $\mathbf{x}$ whenever (5) holds, concluding the proof. ■

Acknowledgments

The authors thank Eric Price for many beneficial conversations that led to the construction of the pure-DP lower bound. SK gratefully acknowledges funding support from the Amazon AI PhD Fellowship. PS gratefully acknowledges NSF grants 2217069 and CCF-2505865.

References

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. URL https://proceedings.mlr.press/v132/acharya21a.html.

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.

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.

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. doi: 10.1214/009117905000000233.

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

Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *arXiv: Statistics Theory*, 2009. URL https://api.semanticscholar.org/CorpusID:14527560.

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.

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. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/41806542.

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. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/20744485.

T. Tony Cai, Zongming Ma, and Yihong Wu. Sparse PCA: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074–3110, 2013. doi: 10.1214/13-AOS1178.

T. Tony Cai, Dong Xia, and Mengyue Zha. Optimal differentially private pca and estimation for spiked covariance matrices, 2024. URL https://arxiv.org/abs/2401.03820.

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

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.

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.

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.

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

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. URL https://openreview.net/forum?id=SWsDJbIloF.

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

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

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.

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.

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

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.

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

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.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

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. doi: 10.1198/jasa.2009.0121. URL https://doi.org/10.1198/jasa.2009.0121. PMID: 20617121.

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

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

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.

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. doi: 10.1137/090756090. URL https://doi.org/10.1137/090756090.

Vladimir Koltchinskii and Karim Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics*, 45(1):121 – 157, 2017a. doi: 10.1214/16-AOS1437. URL https://doi.org/10.1214/16-AOS1437.

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

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

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. URL https://openreview.net/forum?id=b3tygCf3T7.

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

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.

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.

Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772 – 801, 2013. doi: 10.1214/13-AOS1097. URL https://doi.org/10.1214/13-AOS1097.

Shyam Narayanan. Better and simpler lower bounds for differentially private statistical estimation, 2024.

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

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

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

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. doi: 10.1093/biomet/asac041. URL https://doi.org/10.1093/biomet/asac041.

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.

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.

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.

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. URL https://proceedings.mlr.press/v162/tsfadia22a.html.

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

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

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

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

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

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. ISSN 00905364. URL http://www.jstor.org/stable/43974703.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

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

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

The Appendix is organized as follows:

1. Sections A and B provide technical preliminaries useful for our proofs.
2. Section C provides utility results used in subsequent proofs.
3. Section D and E provide our algorithms, lower bounds, and deferred proofs for private sparse covariance estimation.
4. Sections F and G provide the deferred analysis of our private sparse PCA algorithm and its private operator norm estimator.
5. Section H provides our pure and approximate DP lower bounds for sparse PCA.
6. Section I provides our approximate DP lower bound for standard PCA.
7. Section J provides a sparse PCA algorithm based on the exponential mechanism with improved sparsity dependence.

Appendix A. Technical Preliminaries

Matrix concentration and linear algebra. We begin by defining (multivariate) sub-Gaussianity.

Definition 13 (σ-sub-Gaussianity) *A distribution $\mathcal{D}$ supported on $\mathbb{R}^d$ is said to be σ-sub-Gaussian for proxy $\sigma > 0$ if, for any vector $\mathbf{u} \in \mathbb{R}^d$,*

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

We require the following standard facts to bound the approximation error of random sampling.

Lemma 14 *Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d\times d}$, such that $\mathsf{gap} := \lambda_1(\mathbf{B}) - \lambda_2(\mathbf{B}) > 0$. Let $\mathbf{v} := \mathbf{v}_1(\mathbf{A})$ and $\mathbf{u} := \mathbf{v}_1(\mathbf{B})$. Then,*

$$\|\mathbf{v}\mathbf{v}^\top - \mathbf{u}\mathbf{u}^\top\|_{\text{op}} \le \frac{2\|\mathbf{A}-\mathbf{B}\|_{\text{op}}}{\mathsf{gap}}.$$

Proof Recall the identities

$$\|\mathbf{v}\mathbf{v}^\top - \mathbf{u}\mathbf{u}^\top\|_{\text{op}} = \sin\angle(u, v) = \|(\mathbf{I}_d - \mathbf{u}\mathbf{u}^\top)\,\mathbf{v}\mathbf{v}^\top\|_{\text{op}} \tag{6}$$

for any unit vectors $\mathbf{u}$, $\mathbf{v}$. The proof then follows from Corollary 1 of Yu et al. (2015), applied with $\mathbf{\Sigma} \leftarrow \mathbf{B}$, $\widehat{\mathbf{B}} \leftarrow \mathbf{A}$, and $j = 1$. The denominator is $\min(\lambda_0(\mathbf{B}) - \lambda_1(\mathbf{B}), \lambda_1(\mathbf{B}) - \lambda_2(\mathbf{B})) = \lambda_1(\mathbf{B}) - \lambda_2(\mathbf{B})$, since $\lambda_0(\mathbf{B}) = +\infty$ in their convention. ■

Fact 1 (Lemma 6.26, Wainwright (2019)) *Let $\mathcal{D}$ be σ-sub-Gaussian 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 is a universal constant $C > 0$ such that for all $\delta \in (0,1)$,*

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

Lemma 15 *Let $\mathcal{D}$ be σ-sub-Gaussian with covariance $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, let $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{D}$, let $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$, and let $s \in [d]$. There is 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 Exercise 4.7.3, Vershynin (2018) 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. ■

Finally, we often use the following truncation error bound, whose proof is deferred to Appendix C.

Lemma 16 *Let $\rho \geq \tau \geq 0$, and assume $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d\times d}$ satisfy $\|\mathbf{B} - \mathbf{A}\|_{\infty,\infty} \leq \tau$. If $\mathbf{A}$ is k-RCS,*

$$\|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\|_{\text{op}} \leq 2k\rho,$$

and if $\mathbf{A}$ has $\text{nnz}(\mathbf{A}) \leq s$,

$$\|\text{top}_s(\mathbf{B}) - \mathbf{A}\|_{\text{op}} \leq 2\sqrt{2s}\rho.$$

Appendix B. Differential Privacy Preliminaries

Let Γ be some domain, and let $\mathcal{D} \in \Gamma^n$ be a dataset consisting of n elements. We say that two datasets $\mathcal{D}$, $\mathcal{D}' \in \Gamma^n$ 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 17 (Differential privacy) *Let $(\epsilon, \delta) \in [0,1]^2$.*[5] *We say that a randomized algorithm $\mathcal{A} : \Gamma^n \to \Omega$ satisfies (ϵ,δ)-*differential privacy *(or, is (ϵ,δ)-*DP*) if for all events $\mathcal{E} \subseteq \Omega$, and for all neighboring datasets $\mathcal{D}, \mathcal{D}' \in \Gamma^n$, we have*

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

5. In principle, the privacy parameter ϵ can be larger than 1. However, for any $\epsilon \geq 1$, 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 $\epsilon \in [0,1]$.

Differentially private algorithms obey *basic composition* (Theorem B.1, Dwork and Roth (2014)): if $\mathcal{A}_1 : \Gamma^n \to \Omega_1$ is (ϵ_1, δ_1)-DP and $\mathcal{A}_2 : \Gamma^n \times \Omega_1 \to \Omega_2$ 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 next state the Gaussian mechanism. Recall that if $\mathbf{v} : \Gamma^n \to \mathbb{R}^k$ is a vector-valued function of a dataset, we say $\mathbf{v}$ has sensitivity Δ if for all neighboring $\mathcal{D}, \mathcal{D}' \in \Gamma^n$, we have $\|\mathbf{v}(\mathcal{D}) - \mathbf{v}(\mathcal{D}')\|_2 \le \Delta$.

Fact 2 (Theorem A.1, Dwork and Roth (2014)) *Let $\mathbf{v} : \Gamma^n \to \mathbb{R}^k$ 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 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 also require the bounded Laplace mechanism, which is known to give the following guarantee.

Fact 3 (Lemma 9, Asi et al. (2024)) *Let $s : \Gamma^n \to \mathbb{R}$ 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 3 is established in Asi et al. (2024) via a coupling argument, leveraging that $\mathsf{BoundedLaplace}(\lambda)$ and $\mathsf{Lap}(\lambda)$ produce identical samples except with probability δ.

Appendix C. Utility Results

In this section, we prove several utility lemmas that are used throughout the paper.

Lemma 18 *Let $\rho \ge \tau \ge 0$, and assume $\mathbf{A}, \mathbf{B} \in \mathbb{S}^{d \times d}$ satisfy $\|\mathbf{B} - \mathbf{A}\|_{\infty,\infty} \le \tau$. If $\mathbf{A}$ is k-RCS,*

$$\left\|\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\right\|_{\text{op}} \le 2k\rho,$$

and if $\mathbf{A}$ has $\text{nnz}(\mathbf{A}) \le s$,

$$\left\|\text{top}_s(\mathbf{B}) - \mathbf{A}\right\|_{\text{op}} \le 2\sqrt{2s}\rho.$$

Proof We start with the first claim. Define matrices $\mathbf{X} \in \{0, 1\}^{d \times d}$, $\mathbf{Y} \in \mathbb{R}^{d \times d}_{\ge 0}$, where

$$\mathbf{X}_{ij} = \mathbb{1}\left(\mathbf{A}_{ij} \neq 0\right), \quad \mathbf{Y}_{ij} := \left|\left[\mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}\right]_{ij}\right|, \text{ for all } (i, j) \in [d] \times [d].$$

Because $\rho \ge \tau$, each entry of $\mathbf{Y}$ is nonzero only if the same entry of $\mathbf{X}$ is nonzero, and

$$0 \le \mathbf{Y}_{ij} \le (\rho + \tau)\, \mathbf{X}_{ij} \le 2\rho \mathbf{X}_{ij}, \text{ for all } (i, j) \in [d] \times [d].$$

Thus, defining $\mathbf{Z} := \mathcal{T}_\rho(\mathbf{B}) - \mathbf{A}$ (so that $\mathbf{Y} = |\mathbf{Z}|$ for $|\cdot|$ applied entrywise),

$$\begin{aligned} \|\mathbf{Z}\|_{\text{op}} &= \max_{\substack{\mathbf{x},\mathbf{y}\in\mathbb{R}^d \\ \|\mathbf{x}\|_2=\|\mathbf{y}\|_2=1}} \sum_{(i,j)\in[d]\times[d]} \mathbf{x}_i\mathbf{y}_j\mathbf{Z}_{ij} \le \max_{\substack{\mathbf{x},\mathbf{y}\in\mathbb{R}^d \\ \|\mathbf{x}\|_2=\|\mathbf{y}\|_2=1}} \sum_{(i,j)\in[d]\times[d]} |\mathbf{x}_i||\mathbf{y}_j|\mathbf{Y}_{ij} \\ &\le 2\rho \max_{\substack{\mathbf{x},\mathbf{y}\in\mathbb{R}^d \\ \|\mathbf{x}\|_2=\|\mathbf{y}\|_2=1}} \sum_{(i,j)\in[d]\times[d]} |\mathbf{x}_i||\mathbf{y}_j|\mathbf{X}_{ij} = 2\rho \left\|\mathbf{X}\right\|_{\text{op}}, \end{aligned} \tag{7}$$

where we used the Perron-Frobenius theorem in the last line to conclude that the maximizing $\mathbf{x}, \mathbf{y}$ are entrywise nonnegative. It remains to bound $\|\mathbf{X}\|_{\text{op}}$ when $\mathbf{A}$ is k-RCS:

$$\|\mathbf{X}\|_{\text{op}} = \max_{\mathbf{v}\in\mathbb{R}^d:\|\mathbf{v}\|_2\neq 0} \frac{\|\mathbf{X}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \le \sqrt{\left(\max_{\mathbf{v}\in\mathbb{R}^d:\|\mathbf{v}\|_\infty\neq 0} \frac{\|\mathbf{X}\mathbf{v}\|_\infty}{\|\mathbf{v}\|_\infty}\right)\cdot\left(\max_{\mathbf{v}\in\mathbb{R}^d:\|\mathbf{v}\|_1\neq 0} \frac{\|\mathbf{X}\mathbf{v}\|_1}{\|\mathbf{v}\|_1}\right)} \le k,$$

where the first inequality is Problem 5.6.P21 in Horn and Johnson (2012), and the second uses the k-RCS assumption.

We next prove the second claim. Let $S := \text{supp}(\text{top}_s(\mathbf{B}))$ and let $T := \text{supp}(\mathbf{A})$. We first have

$$\left|[\text{top}_s(\mathbf{B}) - \mathbf{A}]_{ij}\right| = \left|[\mathbf{B} - \mathbf{A}]_{ij}\right| \le \rho \text{ for all } (i,j) \in S.$$

Conversely, every $(i,j) \notin S$ has $|\mathbf{B}_{ij}| \le \rho$. This is because there are at most s entries of $\mathbf{B}$ with magnitudes $> \rho$ (since $\text{nnz}(\mathbf{A}) \le s$ and $\|\mathbf{B} - \mathbf{A}\|_{\infty,\infty} \le \rho$), so they are all kept by S. Thus,

$$\left|[\text{top}_s(\mathbf{B}) - \mathbf{A}]_{ij}\right| = |\mathbf{A}_{ij}| \le \left|[\mathbf{B} - \mathbf{A}]_{ij}\right| + |\mathbf{B}_{ij}| \le \tau + \rho \le 2\rho \text{ for all } (i,j) \notin S.$$

We have shown that all of the $\le 2s$ nonzero entries of the matrix $\text{top}_s(\mathbf{B}) - \mathbf{A}$ have magnitude $\le 2\rho$, and the conclusion follows from $\|\text{top}_s(\mathbf{B}) - \mathbf{A}\|_{\text{op}} \le \|\text{top}_s(\mathbf{B}) - \mathbf{A}\|_{\text{F}}$. ■

Lemma 19 *For any unit vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, let $\ell(\mathbf{a},\mathbf{b}) := \sin^2\angle(\mathbf{a},\mathbf{b}) = 1 - \langle\mathbf{a},\mathbf{b}\rangle^2$ and $d_{\text{sign}}(\mathbf{a},\mathbf{b}) := \min\{\|\mathbf{a}-\mathbf{b}\|_2, \|\mathbf{a}+\mathbf{b}\|_2\}$. Then,*

$$d_{\text{sign}}(\mathbf{a},\mathbf{b})^2 \le 2\ell(\mathbf{a},\mathbf{b}).$$

Moreover, for all unit vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$,

$$\ell(\mathbf{x},\mathbf{y}) \le 2\ell(\mathbf{x},\mathbf{z}) + 2\ell(\mathbf{z},\mathbf{y}).$$

Proof Let $\rho := \langle\mathbf{x},\mathbf{y}\rangle \in [-1,1]$ and choose s so that $\langle\mathbf{x}, s\mathbf{y}\rangle = |\rho| \ge 0$. Note that $\|\mathbf{x} - s\mathbf{y}\|_2^2 = 2(1-|\rho|) \le 2(1+|\rho|) = \|\mathbf{x} + s\mathbf{y}\|_2^2$. So, $d_{\text{sign}}(\mathbf{x},\mathbf{y}) = \|\mathbf{x} - s\mathbf{y}\|_2$. Since $1 \le 1 + |\rho| \le 2$,

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

Next, we define the matrices

$$\mathbf{P}_{\mathbf{x}} := \mathbf{x}\mathbf{x}^\top, \qquad \mathbf{P}_{\mathbf{y}} := \mathbf{y}\mathbf{y}^\top, \qquad \mathbf{P}_{\mathbf{z}} := \mathbf{z}\mathbf{z}^\top.$$

Note that for any unit vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\|\mathbf{P_a} - \mathbf{P_b}\|_{\mathrm{F}}^2 = 2 - 2\langle \mathbf{a}, \mathbf{b}\rangle^2 = 2\ell(\mathbf{a}, \mathbf{b}).$$

By the triangle inequality on the Frobenius norm,

$$\begin{aligned}
2\ell(\mathbf{x}, \mathbf{y}) = \|\mathbf{P_x} - \mathbf{P_y}\|_{\mathrm{F}}^2 &\le (\|\mathbf{P_x} - \mathbf{P_z}\|_{\mathrm{F}} + \|\mathbf{P_z} - \mathbf{P_y}\|_{\mathrm{F}})^2 \\
&\le 2\,\|\mathbf{P_x} - \mathbf{P_z}\|_{\mathrm{F}}^2 + 2\,\|\mathbf{P_z} - \mathbf{P_y}\|_{\mathrm{F}}^2 = 4\ell(\mathbf{x}, \mathbf{z}) + 4\ell(\mathbf{z}, \mathbf{y}).
\end{aligned}$$

Dividing by two proves the second inequality. ■

Lemma 20 *Let $\mathbf{A} \in \mathbb{S}_{\succeq \mathbf{0}}^{d\times d}$ have eigenvalues $\lambda_1(\mathbf{A}) > \lambda_2(\mathbf{A})$, and let $\mathbf{v} := \mathbf{v}_1(\mathbf{A})$. Then for every unit vector $\mathbf{u} \in \mathbb{R}^d$,*

$$\sin^2 \angle(\mathbf{u}, \mathbf{v}) \le \frac{\lambda_1(\mathbf{A}) - \mathbf{u}^\top \mathbf{A}\mathbf{u}}{\lambda_1(\mathbf{A}) - \lambda_2(\mathbf{A})}.$$

Proof Write $\mathbf{u} = c\mathbf{v} + s\mathbf{w}$, where $\mathbf{w} \perp \mathbf{v}$ is a unit vector. Then since $\mathbf{w} \perp \mathbf{v}$,

$$\mathbf{u}^\top \mathbf{A}\mathbf{u} = c^2\lambda_1(\mathbf{A}) + s^2\mathbf{w}^\top \mathbf{A}\mathbf{w} \le c^2\lambda_1(\mathbf{A}) + s^2\lambda_2(\mathbf{A}).$$

The claim follows from $s = \sin\angle(\mathbf{u}, \mathbf{v})$, and rearranging

$$\lambda_1(\mathbf{A}) - \mathbf{u}^\top \mathbf{A}\mathbf{u} \ge (\lambda_1(\mathbf{A}) - \lambda_2(\mathbf{A}))\sin^2\angle(\mathbf{u}, \mathbf{v}).$$

■

Lemma 21 *Let P and Q be probability distributions on a measurable space such that, for every measurable set A,*

$$P(A) \le e^\epsilon Q(A) + \delta, \qquad Q(A) \le e^\epsilon P(A) + \delta.$$

Let $\epsilon \le 1$, and let h be a measurable function that is square-integrable under both P and Q. Then

$$|\mathbb{E}_P h - \mathbb{E}_Q h| \le 2\epsilon \mathbb{E}_Q\,|h| + 2\sqrt{\delta}\sqrt{\mathbb{E}_P[h^2] + \mathbb{E}_Q[h^2]}.$$

Proof Let $\mu := P + Q$, and let $p := \mathrm{d}P/\mathrm{d}\mu$ and $q := \mathrm{d}Q/\mathrm{d}\mu$. The two inequalities imply

$$\int (p - e^\epsilon q)_+ \,\mathrm{d}\mu \le \delta, \qquad \int (q - e^\epsilon p)_+ \,\mathrm{d}\mu \le \delta. \tag{8}$$

Indeed, the first integral equals $\sup_A\{P(A) - e^\epsilon Q(A)\}$, with the supremum attained at $A = \{p > e^\epsilon q\}$; the second identity is symmetric. For any scalar c, denote $(c)_+ := \max\{c, 0\}$, and define $r_P := (p - e^\epsilon q)_+$ and $r_Q := (q - e^\epsilon p)_+$. We claim that

$$|p - q| \le (e^\epsilon - 1)q + r_P + r_Q. \tag{9}$$

If $p \geq q$, then $p - q \leq (e^\epsilon - 1)q + r_P$. If $q > p$, then $q - p \leq (e^\epsilon - 1)p + r_Q \leq (e^\epsilon - 1)q + r_Q$. This proves (9). Using (9),

$$|\mathbb{E}_P h - \mathbb{E}_Q h| = \left|\int h(p-q)\,\mathrm{d}\mu\right| \leq (e^\epsilon - 1)\mathbb{E}_Q\,|h| + \int |h|\; r_P\,\mathrm{d}\mu + \int |h|\; r_Q\,\mathrm{d}\mu. \tag{10}$$

Since $r_P \leq p$, $r_Q \leq q$, and both integrate to at most δ by (8), Cauchy–Schwarz gives

$$\int |h|\; r_P\,\mathrm{d}\mu \leq \sqrt{\delta\,\mathbb{E}_P[h^2]}, \qquad \int |h|\; r_Q\,\mathrm{d}\mu \leq \sqrt{\delta\,\mathbb{E}_Q[h^2]}.$$

Using $\sqrt{a} + \sqrt{b} \leq \sqrt{2(a+b)} \leq 2\sqrt{a+b}$ and $e^\epsilon - 1 \leq 2\epsilon$ for $\epsilon \leq 1$ proves the claim. ■

Appendix D. Private Sparse Covariance Estimation

In this section, we provide our main results for solving Problem 1 (private covariance estimation) under Model 1. In Section D.1, we prove Theorem 22, which shows $\approx \sqrt{d}$ samples (suppressing other parameter dependences) suffice for private covariance estimation, improving upon the d dependence from prior work Wang and Xu (2021). After some preprocessing, our algorithm (Algorithm 2) simply applies a private top-k selection step to every row of the empirical covariance, and then releases a noised-and-symmetrized variant of the selected entries. In Section D.2, we prove Theorem 27, a complementary lower bound that shows this $\sqrt{d}$ dependence is necessary up to logarithmic factors.

D.1. Upper bound

Theorem 22 *Algorithm 2 solves Problem 1, for $\mathcal{D} = \{\mathbf{x}_t\}_{t\in[n]}$ drawn from Model 1, with*

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

for a sufficiently large constant.

Proof Throughout, we use the parameter setting

$$R \leftarrow \sigma\sqrt{2\log\left(\frac{6nd}{\beta}\right)}$$

in Algorithm 2, where β is the failure probability from Problem 1, and σ is the sub-Gaussian parameter from Model 1. We handle the privacy and utility proofs separately.

Algorithm 2: $\mathsf{PrivCov}(\mathcal{D}, k, \epsilon, \delta, R)$

Input: Dataset $\mathcal{D} = \{\mathbf{x}_t \in \mathbb{R}^d\}_{t\in[n]}$, sparsity $k \in [d]$, privacy $(\epsilon, \delta) \in (0,1)^2$, truncation level $R > 0$.

Output: A private symmetric covariance estimate $\mathbf{M}$.

1 $\mathbf{y}_t \leftarrow \mathsf{sign}(\mathbf{x}_t) \circ \min\{|\mathbf{x}_t|, R\}$ for all $t \in [n]$;
`//` $\min\{\cdot, R\}$, $|\cdot|$, `and` $\mathsf{sign}(\cdot)$ `are entrywise;` $\circ$ `is entrywise multiplication.`
2 $\widehat{\mathbf{\Sigma}} \leftarrow \frac{1}{n}\sum_{t\in[n]} \mathbf{y}_t\mathbf{y}_t^\top$;
3 $\delta_{\text{row}} \leftarrow \frac{\delta}{2d}$, $\epsilon_{\text{row}} \leftarrow \frac{\epsilon}{4}(2d\log(\frac{2}{\delta}))^{-1/2}$, $\Delta \leftarrow \frac{2R^2}{n}$;
4 $b \leftarrow \frac{2\Delta}{\epsilon_{\text{row}}}\sqrt{k\log(\frac{d}{\delta_{\text{row}}})}$;
5 **for** $i \in [d]$ **do**
6 $\quad S_i \leftarrow \text{supp}(\text{top}_k(\widehat{\mathbf{\Sigma}}_{i,:}^\top + \mathbf{w}_i))$, where $[\mathbf{w}_i]_j \overset{\text{iid}}{\sim} \mathsf{Lap}(b)$;
7 $\quad \widetilde{\mathbf{\Sigma}}_{i,:}^\top \leftarrow \widehat{\mathbf{\Sigma}}_{i,:}^\top + \mathbf{z}_i$, where $[\mathbf{z}_i]_j \overset{\text{iid}}{\sim} \mathsf{Lap}(b)$ independently;
8 **end**
9 $\mathbf{M} \leftarrow \mathbf{0}_{d\times d}$;
10 **for** $i \in [d]$ **do**
11 $\quad \mathbf{M}_{ii} \leftarrow \widetilde{\mathbf{\Sigma}}_{ii}$;
12 **end**
13 **for** $1 \le i < j \le d$ **do**
14 $\quad$ **if** $j \in S_i$ ***and*** $i \in S_j$ **then**
15 $\quad\quad \mathbf{M}_{ij}, \mathbf{M}_{ji} \leftarrow \frac{1}{2}(\widetilde{\mathbf{\Sigma}}_{ij} + \widetilde{\mathbf{\Sigma}}_{ji})$;
16 $\quad$ **end**
17 **end**
18 **return** $\mathbf{M}$;

Privacy. The truncation on Line 1 ensures that for adjacent $(\mathcal{D}, \mathcal{D}')$ differing the t^{th} entry,

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

Hence, entries of $\widehat{\mathbf{\Sigma}}$ are Δ-sensitive. By Theorem 2.2 of Qiao et al. (2021) and the definition of the Laplace parameter b, each release of S_i and the entries of $\widetilde{\mathbf{\Sigma}}_{i,:}$ (restricted to S_i) is $(\epsilon_{\text{row}}, \delta_{\text{row}})$-DP. Applying advanced composition (e.g., Theorem 3.20, Dwork and Roth (2014) with $\delta' \leftarrow \frac{\delta}{2}$) now yields (ϵ, δ)-DP, because

$$\epsilon^\star \le \epsilon_{\text{row}}\sqrt{2d\log\frac{2}{\delta}} + 2d\epsilon_{\text{row}}^2 \le \epsilon, \quad d\delta_{\text{row}} + \frac{\delta}{2} = \delta.$$

The output $\mathbf{M}$ is a postprocessing of $\widetilde{\mathbf{\Sigma}}$ restricted to the selected entries, so Algorithm 2 is (ϵ, δ)-DP.

Utility. We begin by bounding the failure probability of the following events:

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

where

$$\lambda_{\text{samp}} := C\sigma^2 \left(\sqrt{\frac{\log(6d/\beta)}{n}} + \frac{\log(6d/\beta)}{n} \right), \quad \lambda_{\text{priv}} := b \log\left(\frac{12d^2}{\beta} \right), \tag{11}$$

and $C > 0$ is from Fact 1. Let $\mathcal{E} := \mathcal{E}_{\text{clip}} \cap \mathcal{E}_{\text{samp}} \cap \mathcal{E}_{\text{noise}}$, i.e., the event that (i) no truncation occurs, (ii) $\widehat{\boldsymbol{\Sigma}}$ concentrates entrywise, and (iii) all sampled Laplace random variables are bounded.

We claim that each of the three events in the definition of $\mathcal{E}$ has failure probability at most $\frac{\beta}{3}$. For $\mathcal{E}_{\text{clip}}$, this follows from sub-Gaussianity applied to each of nd possible entries of an $\mathbf{x}_t$. For $\mathcal{E}_{\text{samp}}$, this is immediate from Fact 1. Finally, for $\mathcal{E}_{\text{noise}}$, this follows from a union bound over $2d^2$ Laplace random variables. We henceforth condition on $\mathcal{E}$, which gives the failure probability of β.

Under $\mathcal{E}$, we have the following bounds uniformly for all $(i, j) \in [d] \times [d]$

$$\left| \widehat{\boldsymbol{\Sigma}}'_{ij} - \boldsymbol{\Sigma}_{ij} \right| \le \lambda \quad \text{and} \quad \left| \widetilde{\boldsymbol{\Sigma}}_{ij} - \boldsymbol{\Sigma}_{ij} \right| \le \lambda, \tag{12}$$

where $\widehat{\boldsymbol{\Sigma}}'$ denotes the matrix that adds $\mathbf{w}_i$ to each row i of $\widehat{\boldsymbol{\Sigma}}$, and where $\lambda := \lambda_{\text{samp}} + \lambda_{\text{priv}}$. We now claim that we have the entrywise bound

$$|\mathbf{M}_{ij} - \boldsymbol{\Sigma}_{ij}| \le 2\lambda \text{ for all } (i, j) \in [d] \times [d]. \tag{13}$$

To prove (13), first fix a row i. We claim that any

$$j \notin S_i \implies |\boldsymbol{\Sigma}_{ij}| \le 2\lambda. \tag{14}$$

This is because if $|\boldsymbol{\Sigma}_{ij}| \ge 2\lambda$, then by applying (12), the only entries in $\widehat{\boldsymbol{\Sigma}}'_{i,:}$ that could have larger magnitude are the other nonzero entries in $\boldsymbol{\Sigma}_{i,:}$, of which there are at most k.

Now, (13) is immediate for any $(i, j) \in [d] \times [d]$ where $i \notin S_j$ or $j \notin S_i$, because then $\mathbf{M}_{ij} = 0$ and it suffices to apply (14) and symmetry of $\boldsymbol{\Sigma}_{ij}$. Further, if $i \in S_j$ and $j \in S_i$, then using the second bound in (12) (for the tuples (i, j) and (j, i)) and the triangle inequality, (13) again follows.

Hence, $\mathbf{M} - \boldsymbol{\Sigma}$ is entrywise dominated in absolute value by $2\lambda\mathbf{B}$, where $\mathbf{B} \in \{0, 1\}^{d \times d}$ has $(i, j)^{\text{th}}$ entry $\mathbb{1}([\mathbf{M} - \boldsymbol{\Sigma}]_{ij} \neq 0)$. Because $\mathbf{M}$ is $(k + 1)$-RCS and $\boldsymbol{\Sigma}$ is k-RCS, $\mathbf{B}$ is $3k$-RCS, and therefore

$$\|\mathbf{M} - \boldsymbol{\Sigma}\|_{\text{op}} \le 2\lambda \|\mathbf{B}\|_{\text{op}} \le 6k\lambda \le \alpha\sigma^2.$$

The first inequality above used that taking absolute values of a matrix only increases its operator norm (see (7)), and entrywise-larger nonnegative matrices have larger operator norms due to the Perron-Frobenius theorem. The second inequality used that the operator norm of a k-RCS matrix is $\le k$ (Problem 5.6.P21 in Horn and Johnson (2012)), and the third used our choice of n and definitions in (11). ■

D.2. Lower bound

In this section, we prove our lower bound in Theorem 27. Our lower bound is based on a *fingerprinting* strategy, adapted from Narayanan (2024). We specifically instantiate this strategy using a construction based on the *Inverse Wishart* distribution. We recall some properties of this distribution.

Fact 4 *Let $\mathbf{\Sigma} \sim \mathsf{InvWishart}(\mathbf{\Psi}, \nu)$, where $\mathbf{\Psi} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, $\nu > d+1$, and let $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma})$. Then, the density of $\mathbf{\Sigma}$ is $\propto \det(\mathbf{\Sigma})^{-\frac{1}{2}(\nu+d+1)} \exp(-\frac{1}{2}\mathrm{Tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1}))$, and letting $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$,*

$$\mathbb{E}[\mathbf{\Sigma}] = \frac{1}{\nu - d - 1}\mathbf{\Psi}, \quad \mathbb{E}[\mathbf{\Sigma} \mid \mathcal{X}] = w\widehat{\mathbf{\Sigma}} + (1-w)\mathbb{E}[\mathbf{\Sigma}], \text{ where } w := \frac{n}{\nu + n - d - 1}.$$

We also define a *graph projection* operation, used to define scores in the fingerprinting framework. For $G := (V, E)$ an undirected graph with $V \equiv [d]$, we define the operator $P_G := \mathbb{S}^{d\times d} \to \mathbb{S}^{d\times d}$ as

$$[P_G(\mathbf{M})]_{ij} := \begin{cases} \mathbf{M}_{ij} & i = j \text{ or } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

We are now ready to define our lower bound instance.

Construction 1 *Let $(k, d) \in \mathbb{N}^2$ with $k \mid d$ and $k \geq 2\log(\frac{d}{k})$, and let $B := \frac{d}{k}$. Let $G = (V, E)$ with $V \equiv [d]$ be a union of B cliques (with no edges between cliques), so that the b^{th} clique is on vertices j with $k(b-1)+1 \leq j \leq kb$. Draw independent $\{\mathbf{\Sigma}_b \in \mathbb{S}^{k\times k}_{\succeq \mathbf{0}}\}_{b\in[B]}$ via*

$$\mathbf{\Sigma}_b \overset{\text{iid}}{\sim} \mathsf{InvWishart}\left((k-1)\mathbf{I}_k, 2k\right), \tag{16}$$

and let $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ be block diagonal with $\{\mathbf{\Sigma}_b\}_{b\in[B]}$ along the diagonal, and $\{\mathbf{x}_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_d, \mathbf{\Sigma})$.

Note that for $\mathbf{\Sigma}$ arising from Construction 1, drawing samples from $\mathcal{N}(\mathbf{0}_d, \mathbf{\Sigma})$ is an instance of Model 1 with $\sigma^2 := \|\mathbf{\Sigma}\|_{\text{op}}$, because $\mathbf{\Sigma}$ is always k-RCS. We next provide some additional useful properties of $\mathbf{\Sigma}$ drawn from Construction 1. As several of these properties follow from small modifications to arguments in Narayanan (2024), we defer a proof to Appendix E.

Lemma 23 *Let $\mathbf{\Sigma}$ be generated as in Construction 1 with associated graph G. For all $i \in [n]$ and $b \in [B]$, let $\mathbf{x}_i^b \in \mathbb{R}^k$ denote the coordinates $j \in [d]$ of $\mathbf{x}_i$ with $k(b-1)+1 \leq j \leq kb$. Then the following hold for universal positive constants $g_1 < g_2$, and*

$$\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top, \quad \widehat{\mathbf{\Sigma}}_b := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i^b(\mathbf{x}_i^b)^\top, \text{ for all } b \in [B].$$

1. *$P_G(\widehat{\mathbf{\Sigma}})$ is block diagonal with $\{\widehat{\mathbf{\Sigma}}_b\}_{b\in[B]}$ along the diagonal.*

2. *For all $i \in [n]$ and $b \in [B]$, $\mathbf{x}_i^b \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma}_b)$, and $\mathbf{x}_i^b \perp\!\!\!\perp \mathbf{x}_i^{b'}$ for all $b \neq b'$, conditional on $\boldsymbol{\Sigma}$.*
3. *For all $t > 0$, $\mathbb{P}[\|\boldsymbol{\Sigma}\|_{\text{op}} > t] \le \frac{d}{k}(\frac{3e^2}{t})^a$ where $a := \frac{k}{2}$, and if $q \le \frac{a}{2}$, then $\mathbb{E}[\|\boldsymbol{\Sigma}\|_{\text{op}}^q] \le 2\exp(6q)$.*
4. *For all $b \in [B]$, $g_1 \frac{k^2}{n} \le \mathbb{E}\big[\|\widehat{\boldsymbol{\Sigma}}_b - \boldsymbol{\Sigma}_b\|_{\text{F}}^2\big] \le g_2 \frac{k^2}{n}$, and $g_1 \frac{dk}{n} \le \mathbb{E}[\|P_G(\widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma}\|_{\text{F}}^2] \le g_2 \frac{dk}{n}$.*

Our proof adapts the fingerprinting-based strategy proposed in Narayanan (2024) to k-RCS matrices. To ease our exposition, we isolate the Bayesian fingerprinting argument of Narayanan (2024) into a self-contained result in Lemma 24, applied in this section and in Appendix I. This result operates under the following *Bayesian replacement model*, instantiated appropriately in applications of Lemma 24.

Model 3 (Bayesian replacement model) *Let Θ and Γ be measurable spaces, let Π be a* prior *distribution on Θ, and let $\{P_\theta : \theta \in \Theta\}$ be* sample *distributions on Γ. Draw $\theta \sim \Pi$ and, conditional on θ, draw $\{X_i\}_{i\in[n]}, \{X_i'\}_{i\in[n]} \overset{\text{iid}}{\sim} P_\theta$. Set $\mathcal{X} := \{X_i\}_{i\in[n]}$, and for all $i \in [n]$, set*

$$\mathcal{X}^{\sim i} := \left\{X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n\right\}.$$

We next state our general framework for establishing approximate DP lower bounds, Lemma 24, which applies whenever a problem-specific score has a large aggregate signal but DP limits the contribution of each sample. The score Z_i measures the correlation contributed by the observed sample X_i, whereas its mean-zero counterpart Z_i' evaluates X_i against an output that did not observe X_i; by DP, Z_i' and Z_i cannot differ by too much. The lower bound then follows by comparing with a lower bound on the total score. We defer the proof to Appendix E.

Lemma 24 *In Model 3, let $p \in \mathbb{N}$, let $g : \Theta \to \mathbb{R}^p$ and $\psi : \Theta \times \Gamma \to \mathbb{R}^p$ be measurable, and let $\mathcal{A} : \Gamma^n \to \mathbb{R}^p$ be (ϵ, δ)-DP for $(\delta, \epsilon) \in (0,1)^2$. For all $i \in [n]$, define*

$$Z_i := \langle \mathcal{A}(\mathcal{X}) - g(\theta), \psi(\theta, X_i)\rangle, \quad Z_i' := \left\langle \mathcal{A}(\mathcal{X}^{\sim i}) - g(\theta), \psi(\theta, X_i)\right\rangle.$$

Suppose that for all $i \in [n]$, $\mathbb{E}[\psi(\theta, X_i) \mid \theta] = 0$ and, for some $L, U > 0$, that

$$\mathbb{E}\left[\sum_{i\in[n]} Z_i\right] \ge L, \quad 2\epsilon\mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]} \le U.$$

Then $n \ge \frac{L}{U}$.

Before instating Lemma 24, we require two additional ingredients. The first is a posterior concentration result, adapted from Lemma 4.7 in Narayanan (2024), which is used to obtain our lower bound L for Lemma 24. We defer its proof to Appendix E.

Lemma 25 *Let $\boldsymbol{\Sigma}$, $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]}$ be generated from Construction 1 and $\widehat{\boldsymbol{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$. Then,*

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

The second is a procedure that takes a covariance estimation algorithm with an expected squared Frobenius norm guarantee for Construction 1 under a good event, and boosts it to have a stronger error guarantee. This procedure simply takes the coordinatewise median of logarithmically-many fresh calls to the base estimation algorithm, and applies the tail behavior of medians.

Lemma 26 (Lemma A.2, Narayanan (2024)) *Suppose there is an (ϵ, δ)-DP algorithm $\mathcal{A}$ that uses n samples $\mathcal{X}$ from Model 1 with $\mathbf{\Sigma}$ drawn from Construction 1 for sufficiently large d, and satisfies*[6]

$$\Pr\left[\|\mathcal{A}(\mathcal{X}) - \mathbf{\Sigma}\|_{\mathrm{F}} \le \rho\right] \ge \frac{2}{3} \text{ if } \|\mathbf{\Sigma}\|_{\mathrm{op}} \le 30,$$

where for a universal constant c, $\exp(-ck) \le \rho \le c\sqrt{d}$. Then there is an (ϵ, δ)-DP algorithm $\mathcal{A}'$ that uses $O(n \log \frac{d}{\rho})$ samples from Model 1 with $\mathbf{\Sigma}$ drawn from Construction 1, and satisfies

$$\mathbb{E}\left[\left\|\mathcal{A}'(\mathcal{X}) - \mathbf{\Sigma}\right\|_{\mathrm{F}}^4\right] = O(\rho^4).$$

We are now ready to conclude our lower bound proof.

Theorem 27 *Let α_0 be a sufficiently small universal constant, and let $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ be an (ϵ, δ)-DP algorithm that, under Model 1 with $\gamma = 0$, $k \ge \frac{1}{\alpha_0} \log(d)$, and sufficiently large d, solves Problem 1 with $\alpha \le \alpha_0$ and $\beta = \frac{1}{3}$. Then if $\delta \le \frac{\epsilon^2}{d^2}$,*

$$n = \Omega\left(\frac{k^2}{\alpha^2} + \min\left\{\frac{k\sqrt{d}}{\alpha\epsilon \log \frac{d}{\alpha}}, \frac{\sqrt{d}\exp(\alpha_0 k)}{\epsilon}\right\}\right).$$

Proof The non-private component, $n = \Omega(\frac{k^2}{\alpha^2})$, is a consequence of Theorem 1, Cai and Zhou (2012).

For the other component, let $\rho := 30\alpha\sqrt{d}$, and suppose that α_0 is chosen small enough such that $\rho \le c\sqrt{d}$ for the constant c from Lemma 26. We split into two regimes. First, suppose that $\rho \ge \exp(-ck)$. Then by applying the bound from Lemma 26, there is an (ϵ, δ)-DP algorithm $\mathcal{A}'$ that uses $N := O(n \log \frac{d}{\alpha})$ samples $\mathcal{X}$ and satisfies

$$\mathbb{E}\left[\left\|\mathcal{A}'(\mathcal{X}) - \mathbf{\Sigma}\right\|_{\mathrm{F}}^4\right] = O(\alpha^4 d^2). \tag{17}$$

We now use this assumption to instate Lemma 24. Concretely, in Lemma 24, take $n \leftarrow N$, $\theta \leftarrow \mathbf{\Sigma}$, $\Pi \leftarrow$ the prior distribution from Construction 1, $P_\theta \leftarrow \mathcal{N}(\mathbf{0}_d, \mathbf{\Sigma})$, $g(\mathbf{\Sigma}) = \mathbf{\Sigma}$, and $\psi(\mathbf{\Sigma}, \mathbf{x}) := P_G(\mathbf{x}\mathbf{x}^\top) - \mathbf{\Sigma}$. We begin with the lower bound L in Lemma 24:

$$\begin{aligned}\sum_{i \in [N]} Z_i &= \sum_{i \in [N]} \left\langle \mathcal{A}'(\mathcal{X}) - \mathbf{\Sigma}, P_G(\mathbf{x}_i\mathbf{x}_i^\top) - \mathbf{\Sigma} \right\rangle \\ &= N \left\langle \mathcal{A}'(\mathcal{X}) - \mathbf{\Sigma}, \bar{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\rangle, \text{ where } \bar{\mathbf{\Sigma}} := P_G\left(\widehat{\mathbf{\Sigma}}\right), \ \widehat{\mathbf{\Sigma}} := \frac{1}{N} \sum_{i \in [n]} \mathbf{x}_i\mathbf{x}_i^\top.\end{aligned}$$

6. The statement in Narayanan (2024) has a constant of 10, but the proof holds for any larger constant upon examination.

Moreover,

$$
\begin{aligned}
\mathbb{E}\left[\langle \mathcal{A}'(\mathcal{X}) - \boldsymbol{\Sigma}, \bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\rangle\right] &= \mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\langle \mathcal{A}'(\mathcal{X}) - \bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\rangle\right] \\
&= \mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\mathbb{E}\left[\langle \mathcal{A}'(\mathcal{X}) - \bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\rangle \mid \mathcal{X}\right]\right] \\
&= \mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\langle \mathbb{E}[\mathcal{A}'(\mathcal{X}) \mid \mathcal{X}] - \bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{\Sigma}} - \mathbb{E}[\boldsymbol{\Sigma} \mid \mathcal{X}]\rangle\right] \\
&\geq \mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] - \mathbb{E}\left[\left\|\mathbb{E}[\mathcal{A}'(\mathcal{X}) \mid \mathcal{X}] - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}} \left\|\bar{\boldsymbol{\Sigma}} - \mathbb{E}[\boldsymbol{\Sigma} \mid \mathcal{X}]\right\|_{\mathrm{F}}\right] \\
&\geq \mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] - \sqrt{\mathbb{E}\left\|\mathcal{A}'(\mathcal{X}) - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}^2 \mathbb{E}\left\|\mathbb{E}\left[\boldsymbol{\Sigma} \mid \mathcal{X}\right] - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}^2}.
\end{aligned}
$$

The third line above used $\mathcal{A}'(\mathcal{X}) \perp\!\!\!\perp \boldsymbol{\Sigma} \mid \mathcal{X}$, the fourth line used Cauchy-Schwarz after conditioning on $\mathcal{X}$, and the last line used Cauchy-Schwarz after taking expectation over $\mathcal{X}$. Finally,

$$
\begin{aligned}
\mathbb{E}\left[\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] &\geq g_1 \cdot \frac{dk}{N}, \\
\mathbb{E}\left\|\mathcal{A}'(\mathcal{X}) - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}^2 \mathbb{E}\left\|\mathbb{E}\left[\boldsymbol{\Sigma} \mid \mathcal{X}\right] - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}^2 &\leq 2\left(\mathbb{E}\left\|\mathcal{A}'(\mathcal{X}) - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2 + \mathbb{E}\left\|\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right) \\
&\quad \cdot \mathbb{E}\left\|\mathbb{E}\left[\boldsymbol{\Sigma} \mid \mathcal{X}\right] - \bar{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}^2 \\
&= O\left(\alpha^2 d + \frac{dk}{N}\right) \cdot O\left(\frac{dk^2}{N^2} + \frac{dk^3}{N^3}\right) \\
&= O\left(\alpha^2 d\right) \cdot O\left(\frac{dk^2}{N^2}\right) \leq \frac{g_1^2}{4} \cdot \frac{d^2k^2}{N^2}.
\end{aligned}
$$

The first line applied Item 4, Lemma 23, the second line used $(a+b)^2 \leq 2(a^2+b^2)$, the fourth line plugged in our bounds from (17), Item 4, Lemma 23 and Lemma 25, and the last line simplified terms using our assumption that $N \geq n = \Omega(\frac{k^2}{\alpha^2})$, by taking α_0 small enough. Combining the above three displays shows that we may take $L = \frac{g_1}{2} \cdot dk$ in our Lemma 24 application.

Next, for the upper bound U in Lemma 24, we first have

$$
\begin{aligned}
\mathbb{E}\left[\left(Z_i'\right)^2\right] &= \mathbb{E}\left[\left\langle \mathcal{A}'(\mathcal{X}^{\sim i}) - \boldsymbol{\Sigma}, P_G(\mathbf{x}_i\mathbf{x}_i^\top) - \boldsymbol{\Sigma}\right\rangle^2\right] \\
&\leq 2\mathbb{E}\left[\|\boldsymbol{\Sigma}\|_{\mathrm{op}}^2 \left\|\mathcal{A}'(\mathcal{X}^{\sim i}) - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^2\right] \\
&\leq 2\sqrt{\mathbb{E}\left[\|\boldsymbol{\Sigma}\|_{\mathrm{op}}^4\right] \mathbb{E}\left[\|\mathcal{A}'(\mathcal{X}^{\sim i}) - \boldsymbol{\Sigma}\|_{\mathrm{F}}^4\right]} = O\left(\alpha^2 d\right),
\end{aligned}
$$

where the second line used Proposition 3.8, Narayanan (2024), and the third line applied our bounds from (17) and Item 3, Lemma 23. Applying Jensen's inequality then gives $\mathbb{E}|Z_i'| = O(\alpha\sqrt{d})$. Finally,

$$
\begin{aligned}
\mathbb{E}\left[Z_i^2\right] &\leq \sqrt{\mathbb{E}\left[\|\mathcal{A}'(\mathcal{X}) - \boldsymbol{\Sigma}\|_{\mathrm{F}}^4\right] \mathbb{E}\left[\left\|P_G(\mathbf{x}_i\mathbf{x}_i^\top) - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^4\right]} \\
&= O(\alpha^2 d) \cdot \sqrt{\mathbb{E}\left[\left\|\mathbf{x}_i\mathbf{x}_i^\top - \boldsymbol{\Sigma}\right\|_{\mathrm{F}}^4\right]} = O(\alpha^2 d^3),
\end{aligned}
$$

where we applied (17) and the bounds in Proposition 3.10 and Eq. (10) of Narayanan (2024). In sum,

$$U = 2\epsilon\mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\sqrt{\mathbb{E}\left[Z_i^2\right] + \mathbb{E}\left[(Z_i')^2\right]} = O\left(\epsilon\alpha\sqrt{d} + \sqrt{\delta}\alpha d^{1.5}\right) = O(\epsilon\alpha\sqrt{d}),$$

for our assumed range of δ. Plugging both bounds into Lemma 24, and adjusting by an $\frac{N}{n} = O(\log\frac{d}{\alpha})$ factor, gives the first term in the $\min$ in our claimed sample complexity, when $\rho \geq \exp(-cd)$.

Finally, we handle the case when $\rho \leq \exp(-ck)$. In this case, it is enough to provide a lower bound when $\rho = 30\alpha\sqrt{d} = \exp(-ck)$ exactly, by monotonicity of the error guarantee in α. For this particular α, our earlier bound simplifies for sufficiently small α_0:

$$\frac{k\sqrt{d}}{\alpha\epsilon\log\frac{d}{\alpha}} = \Omega\left(\frac{\sqrt{d}\exp(\alpha_0 k)}{\epsilon}\right).$$

■

Appendix E. Deferred Proofs from Section D

Lemma 28 *Let $\mathbf{\Sigma}$ be generated as in Construction 1 with associated graph G. For all $i \in [n]$ and $b \in [B]$, let $\mathbf{x}_i^b \in \mathbb{R}^k$ denote the coordinates $j \in [d]$ of $\mathbf{x}_i$ with $k(b-1)+1 \leq j \leq kb$. Then the following hold for universal positive constants $g_1 < g_2$, and*

$$\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]}\mathbf{x}_i\mathbf{x}_i^\top, \quad \widehat{\mathbf{\Sigma}}_b := \frac{1}{n}\sum_{i\in[n]}\mathbf{x}_i^b(\mathbf{x}_i^b)^\top, \text{ for all } b \in [B].$$

1. *$P_G(\widehat{\mathbf{\Sigma}})$ is block diagonal with $\{\widehat{\mathbf{\Sigma}}_b\}_{b\in[B]}$ along the diagonal.*
2. *For all $i \in [n]$ and $b \in [B]$, $\mathbf{x}_i^b \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{\Sigma}_b)$, and $\mathbf{x}_i^b \perp\!\!\!\perp \mathbf{x}_i^{b'}$ for all $b \neq b'$, conditional on $\mathbf{\Sigma}$.*
3. *For all $t > 0$, $\mathbb{P}[\|\mathbf{\Sigma}\|_{\text{op}} > t] \leq \frac{d}{k}(\frac{3e^2}{t})^a$ where $a := \frac{k}{2}$, and if $q \leq \frac{a}{2}$, then $\mathbb{E}[\|\mathbf{\Sigma}\|_{\text{op}}^q] \leq 2\exp(6q)$.*
4. *For all $b \in [B]$, $g_1\frac{k^2}{n} \leq \mathbb{E}\big[\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\|_{\text{F}}^2\big] \leq g_2\frac{k^2}{n}$, and $g_1\frac{dk}{n} \leq \mathbb{E}[\|P_G(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\|_{\text{F}}^2] \leq g_2\frac{dk}{n}$.*

Proof The first two claims are immediate from Construction 1. To prove Item 3, we start from Lemma 3.7, Narayanan (2024), which implies that for $\mathbf{M} \sim \mathsf{InvWishart}(\mathbf{I}_k, 2k)$ and all $x > 0$,

$$\mathbb{P}\left[\lambda_1(\mathbf{M}) \geq \frac{x}{2k}\right] \leq \left(\frac{e^2}{x}\right)^a.$$

Moreover, the law of $(k-1)\mathbf{M}$ for $\mathbf{M} \sim \mathsf{InvWishart}(\mathbf{I}_k, 2k)$ is $\mathsf{InvWishart}((k-1)\mathbf{I}_k, 2k)$, i.e., the law of each $\mathbf{\Sigma}_b$. Thus, setting $\frac{x}{2k} \leftarrow \frac{t}{k-1}$ and using a union bound over all $B = \frac{d}{k}$ blocks,

$$\Pr\left[\max_{b\in[B]}\|\mathbf{\Sigma}_b\|_{\text{op}} > t\right] = \Pr\left[\|\mathbf{\Sigma}\|_{\text{op}} > t\right] \leq \frac{d}{k}\left(\frac{3e^2}{t}\right)^a.$$

To simplify notation, let $A := 3e^2$ and $t_0 := A(\frac{d}{k})^{1/a}$ henceforth. Using $\mathbb{E}[Z^q] = \int_0^\infty qt^{q-1}\mathbb{P}[Z \geq t]\,\mathrm{d}t$ for any nonnegative random variable Z, we obtain

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{\Sigma}\|_{\text{op}}^q\right] &\leq \int_0^{t_0} qt^{q-1}\,\mathrm{d}t + \int_{t_0}^\infty qt^{q-1}\,\frac{d}{k}\left(\frac{A}{t}\right)^a \mathrm{d}t \\
&= t_0^q + \frac{qdA^a}{k}\int_{t_0}^\infty t^{q-1-a}\,\mathrm{d}t \\
&= t_0^q + \frac{qdA^a}{k}\cdot\frac{t_0^{q-a}}{a-q} = t_0^q\cdot\frac{a}{a-q} \\
&\leq 2t_0^q \leq 2\exp(4q)\left(\frac{d}{k}\right)^{\frac{q}{a}} \leq 2\exp(6q)\,.
\end{aligned}$$

The first line above split the integral and used that probabilities are always ≤ 1, the next two lines evaluated the integrals using our assumption $q < a$, and the last line used $A \leq e^4$ and $(\frac{d}{k})^{1/a} \leq e^2$ by our lower bound on k in Construction 1.

For the first part of Item 4, Lemma 3.7 from Narayanan (2024) implies that $\mathbb{E}[\lambda_k(\mathbf{\Sigma}_b)^2] \geq c$ for a universal constant c, and we have obtained a matching upper bound (up to a constant factor) by using Item 3 with $q = 2$. Therefore, by Lemma 3.1 from Narayanan (2024), for universal constants $0 < g_1 < g_2$,

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

The second part of Item 4 follows by summing the above display over all of the $B = \frac{d}{k}$ blocks. ■

Lemma 29 *In Model 3, let $p \in \mathbb{N}$, let $g : \Theta \to \mathbb{R}^p$ and $\psi : \Theta \times \Gamma \to \mathbb{R}^p$ be measurable, and let $\mathcal{A} : \Gamma^n \to \mathbb{R}^p$ be (ϵ, δ)-DP for $(\delta, \epsilon) \in (0,1)^2$. For all $i \in [n]$, define*

$$Z_i := \langle \mathcal{A}(\mathcal{X}) - g(\theta), \psi(\theta, X_i)\rangle\,, \quad Z_i' := \left\langle \mathcal{A}(\mathcal{X}^{\sim i}) - g(\theta), \psi(\theta, X_i)\right\rangle.$$

Suppose that for all $i \in [n]$, $\mathbb{E}[\psi(\theta, X_i) \mid \theta] = 0$ and, for some $L, U > 0$, that

$$\mathbb{E}\left[\sum_{i\in[n]} Z_i\right] \geq L, \quad 2\epsilon\mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]} \leq U.$$

Then $n \geq \frac{L}{U}$.

Proof All expectations below include the internal randomness of $\mathcal{A}$, which is independent of the prior and sample distributions. We first show that the replacement statistic Z_i' is mean-zero. Write $Y_i' := \mathcal{A}(\mathcal{X}^{\sim i})$. Conditional on $(\theta, \mathcal{X}^{\sim i}, Y_i')$, X_i remains distributed as P_θ, since $\mathcal{X}^{\sim i} \perp\!\!\!\perp X_i$, and the internal randomness producing Y_i' from $\mathcal{X}^{\sim i}$ is independent of all samples. Therefore,

$$\begin{aligned}
\mathbb{E}[Z_i' \mid \theta, \mathcal{X}^{\sim i}, Y_i'] &= \left\langle Y_i' - g(\theta), \mathbb{E}[\psi(\theta, X_i) \mid \theta, \mathcal{X}^{\sim i}, Y_i']\right\rangle \\
&= \left\langle Y_i' - g(\theta), \mathbb{E}[\psi(\theta, X_i) \mid \theta]\right\rangle = 0.
\end{aligned} \tag{18}$$

Taking expectations gives $\mathbb{E}[Z_i'] = 0$. Next, for all $i \in [n]$, let

$$W_i := (\theta, X_1, \ldots, X_n, X_i').$$

Conditional on W_i, the datasets $\mathcal{X}$ and $\mathcal{X}^{\sim i}$ are fixed and neighboring. Let P_{W_i} and Q_{W_i} denote the respective output distributions of $\mathcal{A}(\mathcal{X})$ and $\mathcal{A}(\mathcal{X}^{\sim i})$. By Lemma 21 applied with

$$P \leftarrow P_{W_i}, \quad Q \leftarrow Q_{W_i}, \quad h \leftarrow h_{W_i} \text{ where } h_{W_i}(y) := \langle y - g(\theta), \psi(\theta, X_i) \rangle ,$$

we have for $Z_i := h_{W_i}(\mathcal{A}(\mathcal{X}))$, $Z_i' := h_{W_i}(\mathcal{A}(\mathcal{X}^{\sim i}))$, that

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

Taking expectations over W_i, and applying $\mathbb{E}[Z_i'] = 0$ gives

$$\begin{aligned} |\mathbb{E} Z_i| = \left|\mathbb{E}[Z_i - Z_i']\right| &\le \mathbb{E}\left|\mathbb{E}[Z_i - Z_i' \mid W_i]\right| \\ &\le 2\epsilon \mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\,\mathbb{E}\sqrt{\mathbb{E}[Z_i^2 \mid W_i] + \mathbb{E}[(Z_i')^2 \mid W_i]} \\ &\le 2\epsilon \mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]} \le U. \end{aligned} \tag{19}$$

Here, the first line used convexity of $|\cdot|$, the second used our previous upper bound, and the third applied concavity of $\sqrt{\cdot}$ as well as the definition of U. Finally, the claim follows by applying the assumed lower bound and rearranging:

$$L \le \mathbb{E}\left[\sum_{i\in[n]} Z_i\right] \le \sum_{i\in[n]} |\mathbb{E} Z_i| \le nU.$$

■

Lemma 30 *Let* $\mathbf{\Sigma}$, $\mathcal{X} := \{\mathbf{x}_i\}_{i\in[n]}$ *be generated from Construction 1 and* $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$. *Then,*

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

Proof The prior and observation distributions both factor across blocks $b \in [B]$, so $\mathbb{E}[\mathbf{\Sigma}_b \mid \mathcal{X}] = \mathbb{E}[\mathbf{\Sigma}_b \mid \mathcal{X}_b]$, where $\mathcal{X}_b := \{\mathbf{x}_i^b\}_{i\in[n]}$. Fact 4 then gives

$$\mathbb{E}[\mathbf{\Sigma}_b \mid \mathcal{X}_b] - \widehat{\mathbf{\Sigma}}_b = (1-w)(\mathbb{E}[\mathbf{\Sigma}_b] - \widehat{\mathbf{\Sigma}}_b), \quad 1 - w = \frac{k-1}{n+k-1}.$$

We then have

$$\begin{aligned} \mathbb{E}_{\mathcal{X}_b}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}_b \mid \mathcal{X}\right] - \widehat{\mathbf{\Sigma}}_b\right\|_{\mathrm{F}}^2\right] &= \mathbb{E}_{\mathcal{X}_b}\left[\left\|\mathbb{E}\left[\mathbf{\Sigma}_b \mid \mathcal{X}_b\right] - \widehat{\mathbf{\Sigma}}_b\right\|_{\mathrm{F}}^2\right] \\ &= (1-w)^2 \mathbb{E}_{\mathcal{X}_b}\left[\left\|\widehat{\mathbf{\Sigma}}_b - \mathbb{E}\left[\mathbf{\Sigma}_b\right]\right\|_{\mathrm{F}}^2\right] \\ &\le 3(1-w)^2 \mathbb{E}_{\mathcal{X}_b}\left[\|\mathbf{\Sigma}_b\|_{\mathrm{F}}^2 + \left\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\right\|_{\mathrm{F}}^2 + \|\mathbb{E}\left[\mathbf{\Sigma}_b\right]\|_{\mathrm{F}}^2\right] \\ &\le \left(\frac{k-1}{n+k-1}\right)^2 \cdot O\left(k + \frac{k^2}{n} + k\right) = O\left(\frac{k^3}{n^2} + \frac{k^4}{n^3}\right). \end{aligned}$$

The third line above used the triangle inequality with $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, and the fourth line used Fact 4, and Items 4 and 3 of Lemma 23, to bound the three resulting terms (in that order). Summing over the $\frac{d}{k}$ independent blocks proves the claim. ■

Appendix F. Private Sparse PCA: Deferred Proofs

Theorem 3 (Private k-sparse PCA upper bound) *There is an algorithm (Algorithm 1 with τ set as in (3), and using Proposition 5 to compute $\hat{\lambda}$) that solves Problem 2, for $\mathcal{D} = \{\mathbf{x}_i\}_{i\in[n]}$ drawn from Model 2, with*

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

for a sufficiently large constant.

F.1. Privacy analysis

In this section, we provide a privacy analysis of Algorithm 1. Our analysis has two parts: bounding the privacy of the decision to continue executing after running the checks in Lines 2 to 16, and bounding the privacy of the remaining Lines 17 to 22. We begin with the first (simpler) part.

Lemma 31 *The quantities $|\mathcal{C}|+L$ and $Z+\xi$, on Lines 6 and 15 of Algorithm 1, are each $(\frac{\epsilon}{3}, \frac{\delta}{3})$-DP.*

Proof Each sample $\mathbf{x}_i$ only affects one block covariance $j \in [m]$, so $|\mathcal{C}|$ is 1-sensitive (and similarly each $f_{j'}$ is 1-sensitive for all $j' \neq j$). We can thus bound the sensitivity of Z by

$$1 + (m-1) \cdot \frac{6}{m} \le 7. \tag{20}$$

The claims then follow from Fact 3. ■

Our next goal is to establish a sensitivity bound on the projector $\widehat{\mathbf{P}}$ computed on Line 18, so that the Gaussian mechanism on Line 20 ensures privacy. To do so, we first give a helper lemma on the stability of the weighted average $\bar{\mathbf{\Sigma}}$ on Line 17, that is used to compute $\widehat{\mathbf{P}}$.

Lemma 32 (Sensitivity of the weighted average) *Let m be sufficiently large, let $\mathcal{D}, \mathcal{D}'$ be neighboring datasets, and condition on two runs of Algorithm 1 on $\mathcal{D}, \mathcal{D}'$ both reaching Line 17. There exists $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ such that $\mathbf{M}$ is $(k, \hat{\lambda}, \gamma/2, \tau)$-good, and, denoting by $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Sigma}}'$ the matrices computed on Line 17 by these two runs,*

$$\max\left\{\left\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}) - \mathcal{T}_\tau(\mathbf{M})\right\|_{\text{op}}, \left\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}') - \mathcal{T}_\tau(\mathbf{M})\right\|_{\text{op}}\right\} \le 10k\tau.$$

Proof Let $\mathcal{C}$ and $\mathcal{C}'$ denote the sets computed on Line 3 when Algorithm 1 is run with $\mathcal{D}$ and $\mathcal{D}'$ respectively. Similarly, let $\mathcal{S}$ and $\mathcal{S}'$ denote the sets of batch indices $i \in [m]$ that have $p_i > 0$ in Line 10, on these two runs of Algorithm 1. Because the two runs both passed the checks on Lines 6 and 15, we deterministically have that

$$\min\left\{|\mathcal{C}|, |\mathcal{C}'|, |\mathcal{S}|, |\mathcal{S}'|\right\} \geq 0.8m.$$

Next, assume without loss of generality that $\mathcal{D}$, $\mathcal{D}'$ differ in the first batch, and view $\mathcal{C}$, $\mathcal{S}$ as subsets of $[m]$ and $\mathcal{C}'$, $\mathcal{S}'$ as subsets of $[m+1] \setminus \{1\}$ (i.e., we swap out the first batch in $\mathcal{D}$ with a new batch with index $m+1$, equated with the first batch in $\mathcal{D}'$). We claim that $|\mathcal{C} \cap \mathcal{C}' \cap \mathcal{S} \cap \mathcal{S}'|$ is nonempty. To see this, since all of $\mathcal{C}, \mathcal{C}', \mathcal{S}, \mathcal{S}'$ are subsets of $\mathcal{U} := [m+1]$,

$$\begin{aligned}|\mathcal{C} \cap \mathcal{C}' \cap \mathcal{S} \cap \mathcal{S}'| &= |\mathcal{U}| - \left|(\mathcal{U} \setminus \mathcal{C}) \cup (\mathcal{U} \setminus \mathcal{C}') \cup (\mathcal{U} \setminus \mathcal{S}) \cup (\mathcal{U} \setminus \mathcal{S}')\right| \\ &\geq m + 1 - 4(0.2m + 1) \geq 0.2m - 3 \geq 1,\end{aligned}$$

for sufficiently large m. Let $\mathbf{M} = \widehat{\mathbf{\Sigma}}_i$ for the index $i \in \mathcal{C} \cap \mathcal{C}' \cap \mathcal{S} \cap \mathcal{S}'$. By the definitions of $\mathcal{C}$, $\mathcal{C}'$, $\mathbf{M}$ is $(k, \hat{\lambda}, \frac{\gamma}{2}, \tau)$-good. Moreover, every batch $i \in \mathcal{S}$ has that $\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 2\tau)$ covers at least $\frac{m}{2}$ elements of $\{\widehat{\mathbf{\Sigma}}_i\}_{i \in [m]}$, so that $\mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 2\tau) \cap \mathbb{B}_\infty(\mathbf{M}, 2\tau)$ is nonempty. By the triangle inequality, $\mathbb{B}_\infty(\mathbf{M}, 4\tau)$ covers all of the $\{\widehat{\mathbf{\Sigma}}_i\}_{i \in \mathcal{S}}$, and a similar argument applies for $\mathcal{S}'$. Because $\mathbb{B}_\infty(\mathbf{M}, 4\tau)$ is convex, we also have $\bar{\mathbf{\Sigma}} \in \mathbb{B}_\infty(\mathbf{M}, 4\tau)$ and $\bar{\mathbf{\Sigma}}' \in \mathbb{B}_\infty(\mathbf{M}, 4\tau)$. The conclusion for $\bar{\mathbf{\Sigma}}$ follows from

$$\left\|\bar{\mathbf{\Sigma}} - \mathcal{T}_\tau(\mathbf{M})\right\|_{\infty,\infty} \leq \left\|\bar{\mathbf{\Sigma}} - \mathbf{M}\right\|_{\infty,\infty} + \|\mathbf{M} - \mathcal{T}_\tau(\mathbf{M})\|_{\infty,\infty} \leq 4\tau + \tau,$$

and then applying Lemma 16 with $\rho \leftarrow 5\tau$. The conclusion for $\bar{\mathbf{\Sigma}}'$ is symmetric. ∎

Lemma 33 *Conditioned on Line 18 being reached on a run of Algorithm 1, the matrix $\widehat{\mathbf{P}}$ computed on this line is Δ-sensitive in $\|\cdot\|_{\mathrm{F}}$, where*

$$\Delta := \frac{80\sqrt{2}k\tau}{\gamma\hat{\lambda}}. \tag{21}$$

Proof Fix adjacent $\mathcal{D}$, $\mathcal{D}'$ and let $\mathbf{M}$ be the result of Lemma 4. By goodness of $\mathbf{M}$, $\mathcal{T}_\tau(\mathbf{M})$ has a unique leading eigenvector $\mathbf{w}$, with associated projector $\mathbf{W} := \mathbf{w}\mathbf{w}^\top$. By Lemma 4, $\|\mathcal{T}_\tau(\mathbf{M}) - \mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})\|_{\mathrm{op}} \leq 10k\tau$. Thus, Lemma 14 with $\mathbf{A} \leftarrow \mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})$, $\mathbf{B} \leftarrow \mathcal{T}_\tau(\mathbf{M})$, and $\mathsf{gap} \leftarrow \frac{\gamma\hat{\lambda}}{2}$, implies

$$\left\|\widehat{\mathbf{P}} - \mathbf{W}\right\|_{\mathrm{op}} \leq \frac{40k\tau}{\gamma\hat{\lambda}}.$$

The same bound symmetrically applies for $\|\widehat{\mathbf{P}}' - \mathbf{W}\|_{\mathrm{op}}$, and thus because $\widehat{\mathbf{P}}$, $\widehat{\mathbf{P}}'$ are rank-one,

$$\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{F}} \leq \sqrt{2}\left\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}'\right\|_{\mathrm{op}} \leq \sqrt{2}\left(\left\|\widehat{\mathbf{P}} - \mathbf{W}\right\|_{\mathrm{op}} + \left\|\widehat{\mathbf{P}}' - \mathbf{W}\right\|_{\mathrm{op}}\right) \leq \frac{80\sqrt{2}k\tau}{\gamma\hat{\lambda}}.$$

∎

Corollary 34 *Algorithm 1 is (ϵ, δ)-DP.*

Proof By Lemma 33, the Frobenius sensitivity of the map $\mathcal{D} \mapsto \widehat{\mathbf{P}}(\mathcal{D})$ assuming Line 18 is reached is at most Δ, as defined in (21). The Gaussian mechanism (Fact 2), applied to the vectorization of $\widehat{\mathbf{P}}$, then yields $(\frac{\epsilon}{3}, \frac{\delta}{3})$-DP for releasing $\widehat{\mathbf{P}} + \mathbf{G}$ for the stated setting of σ_{priv}. All subsequent steps (top_{k^2}, symmetrization, and computing $\mathbf{v}_1(\cdot)$) are postprocessings of this private statistic, so by using Lemma 31 along with basic composition, the final output is (ϵ, δ)-DP. ■

F.2. Utility analysis

We now analyze the utility of Algorithm 1, and conclude our proof of Theorem 3. To begin, we show that when the data is drawn from Model 1 for sufficiently large n, the algorithm passes the tests in Lines 6 and 15, and the aggregated $\bar{\mathbf{\Sigma}}$ is close to the population covariance $\mathbf{\Sigma}$.

Lemma 35 *Let τ be set as in (3), let $\hat{\lambda} \in [(1 - \frac{\gamma}{10})\lambda_1(\mathbf{\Sigma}), (1 + \frac{\gamma}{10})\lambda_1(\mathbf{\Sigma})]$, and assume that*

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

for sufficiently large constants. Then with probability at least $1 - \frac{\beta}{4}$, if $\mathcal{D}$ is drawn from Model 1, the tests in Lines 6 and 15 pass, $\min\{|\mathcal{C}|, Z\} \geq 0.9m$, and every $i \in [m]$ with $p_i > 0$ satisfies

$$\left\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\right\|_{\infty,\infty} \leq 3\tau.$$

Proof We first prove that Line 6 passes and $|\mathcal{C}| \geq 0.9m$, with probability $\geq 1 - \frac{\beta}{4}$. We claim that for each batch $i \in [m]$, $\widehat{\mathbf{\Sigma}}_i$ is $(k, \hat{\lambda}, \frac{\gamma}{2}, \tau)$-good with probability $\geq \frac{19}{20}$. If so, a standard Chernoff bound (for a large enough constant in m) proves that at least $\frac{9}{10}$ of the $\widehat{\mathbf{\Sigma}}_i$ are good, except with probability $\frac{\beta}{4}$. In this case, taking m so that $|L| \leq \frac{m}{20}$ deterministically, Line 6 always passes.

It remains to prove that each $\widehat{\mathbf{\Sigma}}_i$ is good with probability $\geq \frac{19}{20}$. By Fact 1, taking τ as in (3), with probability $\geq \frac{19}{20}$ we have $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq \tau$, and hence no zero entry of $\mathbf{\Sigma}$ is nonzero in $\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)$. Under this event, Lemma 16 gives

$$\left\|\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i) - \mathbf{\Sigma}\right\|_{\text{op}} \leq 2k\tau \leq \frac{\gamma\lambda_1(\mathbf{\Sigma})}{10},$$

where the last inequality holds for a large enough constant in n. Under the stated assumption on $\hat{\lambda}$, the remaining conditions in Definition 2 now hold by Weyl's inequality.

Henceforth, condition on the earlier event that at least $\frac{9}{10}$ of the $i \in [m]$ have $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} \leq \tau$; call these indices $G \subseteq [m]$. We next prove that if $|G| \geq 0.9m$, $Z \geq 0.9m$, which implies that Line 15 passes. Indeed, every $i \in G$ has $f_i \geq 0.9m$ by the triangle inequality, and therefore

$$Z = \sum_{i \in [m]} p_i \geq |G| = 0.9m.$$

To obtain the last claim, take some $i \in [m]$ with $\|\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}\|_{\infty,\infty} > 3\tau$. By the triangle inequality, for every $j \in G$, $\widehat{\mathbf{\Sigma}}_j \notin \mathbb{B}_\infty(\widehat{\mathbf{\Sigma}}_i, 2\tau)$, and hence $f_i \le 0.1m$. Thus, $p_i = 0$ for any such i. ■

We next prove that under the closeness guarantee afforded by Lemma 35, the remaining steps in Algorithm 1 yield a sufficiently good solution to PCA (Problem 2). Notably, this is the step that leverages the stronger eigenvector sparsity assumption from Model 2.

Lemma 36 *In the setting of Lemma 35, assume that the conclusion of Lemma 35 holds, and that $\mathcal{D}$ is drawn from Model 2. Then, with probability $\ge 1 - \frac{\beta}{4}$, denoting $\mathbf{P} := \mathbf{v}\mathbf{v}^\top$ where $\mathbf{v} = \mathbf{v}_1(\mathbf{\Sigma})$, and $\widetilde{\mathbf{P}}$ as in Line 21 of Algorithm 1,*

$$\left\|\widetilde{\mathbf{P}} - \mathbf{P}\right\|_{\text{op}} \le 2\sqrt{2}k\left(\frac{20k\tau}{\gamma\lambda_1(\mathbf{\Sigma})} + 2\sigma_{\text{priv}}\sqrt{\log\left(\frac{4d}{\beta}\right)}\right).$$

Proof Under Lemma 35's conclusion, convexity implies $\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\infty,\infty} \le 3\tau$, so that Lemma 16 gives

$$\left\|\mathcal{T}_{5\tau}\left(\bar{\mathbf{\Sigma}}\right) - \mathbf{\Sigma}\right\|_{\text{op}} \le 10k\tau \le \frac{\gamma\lambda_1(\mathbf{\Sigma})}{10},$$

under the sparsity assumption in Model 2 and our setting of n. Then, Lemma 14 with $\mathsf{gap} \leftarrow \gamma\lambda_1(\mathbf{\Sigma})$, $\mathbf{A} \leftarrow \mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})$, and $\mathbf{B} \leftarrow \mathbf{\Sigma}$, implies

$$\left\|\widehat{\mathbf{P}} - \mathbf{P}\right\|_{\text{op}} \le \frac{20k\tau}{\gamma\lambda_1(\mathbf{\Sigma})}.$$

Next, applying $\|\cdot\|_{\infty,\infty} \le \|\cdot\|_{\text{op}}$ and a standard tail bound on the maximum of d^2 Gaussian random variables (e.g., Mill's inequality) gives that with probability $\ge 1 - \frac{\beta}{4}$,

$$\left\|\left(\widehat{\mathbf{P}} + \mathbf{G}\right) - \mathbf{P}\right\|_{\infty,\infty} \le \frac{20k\tau}{\gamma\lambda_1(\mathbf{\Sigma})} + 2\sigma_{\text{priv}}\sqrt{\log\left(\frac{8d}{\beta}\right)}.$$

The conclusion follows from Lemma 16, because $\mathbf{P}$ has at most $s = k^2$ nonzero entries. ■

We require one additional ingredient: a private eigenvalue estimation procedure under Model 1, whose proof is deferred to Appendix G.

Proposition 37 *Let $(\epsilon, \delta, \beta) \in (0,1)^3$, and let $\mathcal{D}$ be drawn from Model 1 with*

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

for a sufficiently large constant. There is an (ϵ, δ)-DP algorithm, $\mathsf{PrivNorm}$ *(Algorithm 3), which returns $\hat{\lambda}$ satisfying*

$$\hat{\lambda} \in \left[(1 - \tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma}), (1 + \tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma})\right]$$

with probability at least $1 - \beta/2$.

Finally, we combine the pieces to prove Theorem 3.

Proof Throughout the proof, we take m sufficiently large for Lemma 35 to hold, and set τ as in (3). Under these settings, condition on the results of Lemma 35, Lemma 36, and Proposition 5 all holding, which occurs with probability $\geq 1-\beta$. Then by convexity of $\|\cdot\|_{\text{op}}$,

$$\begin{aligned}\left\|\frac{1}{2}\left(\widetilde{\mathbf{P}}+\widetilde{\mathbf{P}}^\top\right)-\mathbf{P}\right\|_{\text{op}} &\leq 2\sqrt{2}k\left(\frac{20k\tau}{\gamma\lambda_1(\mathbf{\Sigma})}+2\sigma_{\text{priv}}\sqrt{\log\left(\frac{4d}{\beta}\right)}\right)\\ &\leq \frac{2\sqrt{2}k^2\tau}{\gamma\lambda_1(\mathbf{\Sigma})}\left(20+\frac{3600\sqrt{2}}{\epsilon}\log\left(\frac{6d}{\beta\delta}\right)\right)\leq\frac{\sqrt{\Delta}}{2},\end{aligned}$$

where we plugged in our choice of σ_{priv} from Line 19, and our choice of τ from (3), by taking

$$b=\Omega\left(\frac{\sigma^4}{\lambda_1(\mathbf{\Sigma})^2}\cdot\frac{k^4\log^3(\frac{d}{\beta\delta})}{\gamma^2\Delta\epsilon^2}\right),\quad m=\Omega\left(\frac{1}{\epsilon}\log\left(\frac{1}{\delta\beta}\right)\right).$$

Our bound on n follows by using the definition $n=mb$. The privacy claim in Problem 2 is immediate from Corollary 34. For the utility claim, it follows from Lemma 14 with $\mathbf{A}\leftarrow\frac{1}{2}(\widetilde{\mathbf{P}}+\widetilde{\mathbf{P}}^\top)$, $\mathbf{B}\leftarrow\mathbf{P}$, and gap $\leftarrow 1$ that $\|\hat{\mathbf{v}}\hat{\mathbf{v}}^\top-\mathbf{P}\|_{\text{op}}\leq\sqrt{\Delta}$, and the sine-squared error follows from the equivalence (6). ■

Appendix G. Private Operator Norm Estimation under Model 1

In this section, we provide an algorithm for privately estimating the operator norm of a k-RCS covariance matrix under Model 1, by proving Proposition 5, which is used in Theorem 3 and may be of independent interest. Our Algorithm 3 is patterned off of Algorithm 1, with the same choice of τ as in (3), deviating in the definition of the certified set and the final output.

Proposition 38 *Let $(\epsilon,\delta,\beta)\in(0,1)^3$, and let $\mathcal{D}$ be drawn from Model 1 with*

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

for a sufficiently large constant. There is an (ϵ,δ)-DP algorithm, PrivNorm *(Algorithm 3), which returns $\hat{\lambda}$ satisfying*

$$\hat{\lambda}\in\left[(1-\tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma}),(1+\tfrac{\gamma}{10})\lambda_1(\mathbf{\Sigma})\right]$$

with probability at least $1-\beta/2$.

We split the proof of Proposition 5 into two parts: Lemma 39 (which handles the privacy), and Lemma 40 (which handles the utility). As much of the proofs are identical to the components used to prove Theorem 3, we primarily highlight the differences.

Lemma 39 *Algorithm 3 is (ϵ,δ)-DP.*

Algorithm 3: $\mathsf{PrivNorm}(\mathcal{D}, \gamma, k, \epsilon, \delta, \tau, m)$

Input: Dataset $\mathcal{D} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i \in [n]}$, gap $\gamma > 0$, sparsity $k \in [d]$, privacy (ϵ, δ), threshold τ, and number of batches $m \in \mathbb{N}$.

Output: Either $\perp$ or a private estimate of $\|\mathbf{\Sigma}\|_{\mathrm{op}}$.

1 $\widehat{\mathbf{\Sigma}}_i \leftarrow b^{-1} \sum_{j=(i-1)b+1}^{ib} \mathbf{x}_j \mathbf{x}_j^\top$ for all $i \in [m]$, where $b := n/m$;

2 $\mathcal{C} \leftarrow \{i \in [m] : \mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)$ is k-RCS, $\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i)) > 0,\ \frac{\lambda_2(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))}{\lambda_1(\mathcal{T}_\tau(\widehat{\mathbf{\Sigma}}_i))} \leq 1 - \frac{\gamma}{2}\}$;

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

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

5   **return** $\perp$;

6 **end**

7 **for** $i \in [m]$ **do**

8     $f_i \leftarrow \sum_{j \in [m]} \mathbb{1}\{\|\widehat{\mathbf{\Sigma}}_j - \widehat{\mathbf{\Sigma}}_i\|_{\infty,\infty} \leq 2\tau\}$;

9     $p_i \leftarrow \min\{\max\{(f_i - m/2)/(m/6), 0\}, 1\}$;

10 **end**

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

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

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

14   **return** $\perp$;

15 **end**

16 $\bar{\mathbf{\Sigma}} \leftarrow Z^{-1} \sum_{i \in [m]} p_i \widehat{\mathbf{\Sigma}}_i$;

17 $\sigma_{\mathrm{priv}} \leftarrow 20k\tau\frac{6}{\epsilon}\sqrt{\log(\frac{6}{\delta})}$;

18 Sample $g \sim \mathcal{N}(0, \sigma_{\mathrm{priv}}^2)$;

19 **return** $\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})\|_{\mathrm{op}} + g$;

Proof Algorithm 3 is identical to Algorithm 1 through Line 16, except the definition of the set $\mathcal{C}$ excludes the $\hat{\lambda}$-dependent part of goodness (Definition 2). Because the proof of Lemma 4 did not use the $\hat{\lambda}$-dependent part of Definition 2, its conclusion still holds, and therefore conditioned on Line 16 being reached, the scalar statistic $\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}})\|_{\mathrm{op}}$ is $20k\tau$-sensitive. The conclusion follows from combining the Gaussian mechanism (Fact 2) with the privacy loss of reaching Line 16 (Lemma 31). ■

Lemma 40 *In the setting of Proposition 5 with n taken as stated, τ set as in (3), and $m = \Theta(\frac{1}{\epsilon}\log\frac{1}{\delta\beta})$ for an appropriate constant, Algorithm 3 returns $\hat{\lambda}$ satisfying $\hat{\lambda} \in [(1-\frac{\gamma}{10})\lambda_1(\mathbf{\Sigma}), (1+\frac{\gamma}{10})\lambda_1(\mathbf{\Sigma})]$, with probability $\geq 1 - \frac{\beta}{2}$.*

Proof First, with probability $\geq 1 - \frac{\beta}{4}$, we have that

$$|g| = O\left(\frac{k\tau}{\epsilon}\sqrt{\log\frac{1}{\delta}} \cdot \sqrt{\log\frac{1}{\beta}}\right) \leq \frac{\gamma}{20}\lambda_1(\mathbf{\Sigma}),$$

by plugging our lower bound on $b = \frac{n}{m}$ in Proposition 5 into the definition of τ in (3).

Next, the proof of Lemma 35 applies up to the point where $\bar{\mathbf{\Sigma}}$ is defined, so applying its conclusion, with probability $\geq 1 - \frac{\beta}{4}$, we have $\|\bar{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\infty,\infty} \leq 3\tau$. Lemma 16 then gives

$$\left\|\mathcal{T}_{5\tau}(\bar{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_{\text{op}} \leq 10k\tau \leq \frac{\gamma}{20}\lambda_1(\mathbf{\Sigma}),$$

again by using our choice of τ. The claim follows from the above displays and Weyl's inequality. ■

Appendix H. Private Sparse PCA: Lower Bounds

In this section, we give dimension-dependent lower bounds for Problem 2 under Model 1, complementing our upper bound, Theorem 3, which uses $\text{poly}(k, \log(d))$ samples under the stronger Model 2. We provide two incomparable lower bounds: the first (Section H.1) holds under pure DP, with the same parameterization as used in Theorem 3, whereas the second (Section H.2) holds under approximate DP, with a different parameterization commented on in Remark 48.

H.1. Pure DP lower bound

We start by proving a $\Omega(d/\epsilon)$ sample complexity lower bound in Theorem 12 for *pure* differential privacy for the PCA task (Problem 2) over a natural family of k-RCS covariance matrices satisfying Model 1. We use the DP hypothesis-selection in Proposition 6 lower bound result due to Kamath et al. (2020).

We use Definition 8 and Lemma 9 from the main text.

For our lower bound, we use the following expander existence result of Gribinski and Marcus (2021).

Proposition 41 (Theorem 1.2, Gribinski and Marcus (2021)) *For any $(g, h, t) \in \mathbb{N}^3$, there exists a bipartite graph $G_{g,h,t} = (V = L \cup R, E)$ such that the following properties hold.*

- $|L| = tg$, $|R| = g$, $\deg(u) = h$ *for all* $u \in L$, *and* $\deg(v) = th$ *for all* $v \in R$.
- $\lambda_2(\mathbf{A}) \leq \sqrt{h-1} + \sqrt{th-1}$, *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.

Corollary 42 *Let $k \geq 6$ be even, let $d = 3g$ for $g \geq k$, and let $R \subseteq [d]$ with $|R| = g$ be arbitrary, with $L := [d] \setminus R$. There exists a k-RCS $\mathbf{\Sigma} \in \mathbb{S}_{\succeq \mathbf{0}}^{d\times d}$ such that the following properties hold:*

$$\lambda_1(\mathbf{\Sigma}) = 2, \quad \lambda_2(\mathbf{\Sigma}) < 1.97, \quad \mathbf{v}_1(\mathbf{\Sigma}) = \sqrt{\frac{3}{4d}}\begin{pmatrix}\mathbf{1}_L \\ \sqrt{2}\cdot\mathbf{1}_R\end{pmatrix}. \tag{4}$$

Proof Let $g = d/3$, $t = 2$, and $h = (k-2)/2$ in Proposition 10, 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 9 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 10 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 9 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, $d_{\text{ham}}(\mathbf{x}, \mathbf{y}) := \sum_{i \in [3m]} \mathbb{1}(\mathbf{x}_i \neq \mathbf{y}_i)$ of $m/2$. This is classically attributed to Gilbert and Varshamov.

Lemma 43 *Let $g \in \mathbb{N}$ and let $\mathcal{S} := \{\mathbf{s} \in \{0,1\}^{3g} : \sum_{i \in [3g]} \mathbf{s}_i = g\}$. There exists a subset $C \subseteq \mathcal{S}$ of size $\exp(\Omega(g))$ such that for any distinct $\mathbf{x}, \mathbf{y} \in C$, $d_{\text{ham}}(\mathbf{x}, \mathbf{y}) \ge \frac{g}{2}$.*

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 $d_{\text{ham}}(x, y) \le m/2$ is equal to

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

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 12 (Pure-DP PCA lower bound for k-RCS covariance) *Let $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ be an ϵ-DP algorithm that, under Model 1 with $\sigma^2 = 1$ and $\gamma = \frac{1}{100}$, solves Problem 2 with $\beta = \frac{1}{3}$ and $\Delta = \frac{1}{400}$ for sufficiently large d, k. Then $n = \Omega(\frac{d}{\epsilon})$.*

Proof Let $k \geq 6$ be even and $d = 3g$ for $g \geq k$. Let C be the subset of $\{0,1\}^d$ with size $m := |C| = \exp(\Omega(d))$ guaranteed by Lemma 7. For each $\mathbf{x} \in C$, let $\mathbf{\Sigma_x}$ be the k-RCS matrix given by Corollary 11 with $R \leftarrow \text{supp}(\mathbf{x})$. Using (4), the Gaussian $P_{\mathbf{x}} := \mathcal{N}(\mathbf{0}_d, \frac{1}{2}\mathbf{\Sigma_x})$ satisfies Model 1 with $\sigma^2 = 1$ and $\gamma = \frac{1}{100}$. The rest of the proof establishes that whenever $\mathcal{A}$ returns $\hat{\mathbf{v}}$ with

$$\sin^2 \angle(\hat{\mathbf{v}}, \mathbf{v}_1(\mathbf{\Sigma_x})) \leq \frac{1}{400}, \tag{22}$$

there is a deterministic $D : \mathbb{R}^d \to C$ with $D(\hat{\mathbf{v}}) = \mathbf{x}$. Note that $D \circ \mathcal{A}$ is also ϵ-DP by postprocessing, so Proposition 6 with $\alpha = 1$, $\mathcal{A} \leftarrow D \circ \mathcal{A}$, and $\mathcal{P} = \{P_{\mathbf{x}}\}_{\mathbf{x} \in C}$, proves the theorem.

We now describe the decoding algorithm D. For all $\mathbf{x} \in C$, denote by shorthand $\mathbf{v_x} := \mathbf{v}_1(\mathbf{\Sigma_x})$ (i.e., the vector in (4)). The algorithm simply selects $\mathbf{y} \in C$ that minimizes $d_{\text{sign}}(\mathbf{y}, \hat{\mathbf{v}})$, where

$$d_{\text{sign}}(\mathbf{u}, \mathbf{v}) := \min\left\{\|\mathbf{u} - \mathbf{v}\|_2, \|\mathbf{u} + \mathbf{v}\|_2\right\}.$$

Note that d_{sign} satisfies the triangle inequality. Indeed, for any three unit vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$, let $\sigma, \sigma' \in \{\pm 1\}$ be signs for which $d_{\text{sign}}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \sigma\mathbf{b}\|_2$ and $d_{\text{sign}}(\mathbf{b}, \mathbf{c}) = \|\mathbf{b} - \sigma'\mathbf{c}\|_2$. Then,

$$d_{\text{sign}}(\mathbf{a}, \mathbf{c}) \leq \|\mathbf{a} - \sigma\sigma'\mathbf{c}\|_2 \leq \|\mathbf{a} - \sigma\mathbf{b}\|_2 + \|\sigma\mathbf{b} - \sigma\sigma'\mathbf{c}\|_2 = d_{\text{sign}}(\mathbf{a}, \mathbf{b}) + d_{\text{sign}}(\mathbf{b}, \mathbf{c}).$$

By Lemma 19, the true $\mathbf{x}$ that indexes $P_{\mathbf{x}}$ satisfies

$$d_{\text{sign}}(\mathbf{v_x}, \hat{\mathbf{v}}) \leq \sqrt{\frac{1}{200}}.$$

On the other hand, for any $\mathbf{y} \in C$ such that $\mathbf{y} \neq \mathbf{x}$, (4) implies that

$$\|\mathbf{v_x} - \mathbf{v_y}\|_2 = \sqrt{d_{\text{ham}}(\mathbf{x}, \mathbf{y})}\left(\sqrt{\frac{3}{2d}} - \sqrt{\frac{3}{4d}}\right) \geq \sqrt{\frac{d}{6}}\left(\sqrt{\frac{3}{2d}} - \sqrt{\frac{3}{4d}}\right) > \sqrt{\frac{1}{50}}.$$

Moreover, $d_{\text{sign}}(\mathbf{v_x}, \mathbf{v_y}) = \|\mathbf{v_x} - \mathbf{v_y}\|$ because $\langle \mathbf{v_x}, \mathbf{v_y} \rangle > 0$. Finally, $d_{\text{sign}}(\mathbf{v_y}, \hat{\mathbf{v}}) > d_{\text{sign}}(\mathbf{v_x}, \hat{\mathbf{v}})$ for any $\mathbf{y} \neq \mathbf{x}$, because triangle inequality yields the following contradiction otherwise:

$$\sqrt{\frac{1}{50}} < d_{\text{sign}}(\mathbf{v_x}, \mathbf{v_y}) \leq d_{\text{sign}}(\mathbf{v_x}, \hat{\mathbf{v}}) + d_{\text{sign}}(\hat{\mathbf{v}}, \mathbf{v_y}) \leq 2d_{\text{sign}}(\mathbf{v_x}, \hat{\mathbf{v}}) \leq 2\sqrt{\frac{1}{200}}.$$

Thus, D correctly returns $\mathbf{x}$ whenever (22) holds. ∎

H.2. Approximate DP lower bound

In this section, we prove Theorem 47, our lower bound for PCA for k-RCS covariance matrices with *approximate DP*. We note that our hard instances follow a slightly different problem parameterization than used by Model 1, which makes it not fully compatible with our corresponding upper bounds in Section 3; we provide additional commentary on this discrepancy in Remark 48.

Our lower bound arguments apply the following variant of the *DP Assouad's lemma* from Acharya et al. (2021); for completeness, we defer a proof to Appendix H.

Proposition 44 *Let $\mathcal{V} \subseteq \{\pm 1\}^d$ and associate each $\mathbf{s} \in \mathcal{V}$ with a distribution $P_{\mathbf{s}}$ over Ω^n for some sample space Ω, and a parameter $\boldsymbol{\theta}_{\mathbf{s}} \in \Theta$. Let symmetric $\ell : \Theta \times \Theta$ satisfy*

$$\begin{aligned} \ell\left(\boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\theta}_{\mathbf{v}}\right) &\geq 2\tau \sum_{i \in [d]} \mathbb{1}\left(\mathbf{u}_i \neq \mathbf{v}_i\right) \text{ for all } (\mathbf{u}, \mathbf{v}) \in \mathcal{V}^2, \\ \ell\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) &\leq 2\left(\ell\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3\right) + \ell\left(\boldsymbol{\theta}_3, \boldsymbol{\theta}_2\right)\right) \text{ for all } (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \in \Theta^3. \end{aligned} \tag{23}$$

For each $i \in [d]$, let $\mathcal{V}_{+i} := \{\mathbf{s} \in \mathcal{V} : \mathbf{s}_i = +1\}$, $\mathcal{V}_{-i} := \{\mathbf{s} \in \mathcal{V} : \mathbf{s}_i = -1\}$, and define

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

If for all $i \in [d]$ with $\min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\} > 0$, there is a coupling Γ of $\mathcal{X} := \{\mathbf{x}_j\}_{j \in [n]} \sim P_{+i}$ and $\mathcal{Y} := \{\mathbf{y}_j\}_{j \in [n]} \sim P_{-i}$ with

$$\mathbb{E}_{(\mathcal{X}, \mathcal{Y}) \sim \Gamma}\left[\sum_{j \in [n]} \mathbb{1}(\mathbf{x}_j \neq \mathbf{y}_j)\right] \leq D, \tag{24}$$

then defining $R(\mathcal{V}, \ell, \epsilon, \delta) := \min_{\mathcal{A} \text{ is } (\epsilon,\delta)\text{-DP}} \max_{\mathbf{s} \in \mathcal{V}} \mathbb{E}_{\mathcal{X} \sim P_{\mathbf{s}}^{\otimes n}}[\ell(\mathcal{A}(\mathcal{X}), \boldsymbol{\theta}_{\mathbf{s}})]$,

$$R(\mathcal{V}, \ell, \epsilon, \delta) \geq \frac{\tau}{2} \cdot \frac{\sum_{i \in [d]} \min\{|\mathcal{V}_{-i}|, |\mathcal{V}_{+i}|\}}{|\mathcal{V}|} \cdot \left(0.9e^{-10\epsilon D} - 10D\delta\right).$$

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

Lemma 45 *Let k, d be sufficiently large, and for each sign vector $\mathbf{s} \in \{\pm 1\}^d$, define the diagonal sign matrix $\mathbf{D}_{\mathbf{s}} := \mathbf{diag}(\mathbf{s})$. There exists a matrix $\mathbf{A} \in \{0, 1\}^{d \times d}$ such that if we let*

$$\boldsymbol{\Sigma}_{\mathbf{s}} := \mathbf{I}_d + \frac{1}{k-1} \mathbf{D}_{\mathbf{s}} \mathbf{A} \mathbf{D}_{\mathbf{s}},$$

the following properties hold: $\boldsymbol{\Sigma}_{\mathbf{s}} \in \mathbb{S}_{\succeq \mathbf{0}}^{d \times d}$ is k-RCS,

$$\lambda_1(\boldsymbol{\Sigma}_{\mathbf{s}}) = 2, \quad \lambda_2(\boldsymbol{\Sigma}_{\mathbf{s}}) \leq \frac{3}{2}, \quad \mathbf{v}_{\mathbf{s}} := \mathbf{v}_1(\boldsymbol{\Sigma}_{\mathbf{s}}) = \frac{1}{\sqrt{d}} \mathbf{s}.$$

Proof Taking $g \leftarrow \frac{d}{2}$, $t = 1$, and $h \leftarrow k - 1$ in Proposition 10, let $\mathbf{A}$ be the adjacency matrix of the resulting (h, h)-biregular bipartite graph, G. Multiplication by $\mathbf{D}_{\mathbf{s}}$ preserves the zero pattern, hence each row of $\mathbf{D}_{\mathbf{s}}\mathbf{A}\mathbf{D}_{\mathbf{s}}$ has h nonzeros, so $\mathbf{\Sigma}_{\mathbf{s}}$ is k-RCS. Since $\mathbf{D}_{\mathbf{s}}$ is orthogonal, $\mathbf{D}_{\mathbf{s}}\mathbf{A}\mathbf{D}_{\mathbf{s}}$ is similar to $\mathbf{A}$, hence they share eigenvalues. Because G is h-regular and bipartite, $\lambda_1(\mathbf{A}) = \|\mathbf{A}\|_{\text{op}} = h$ (Lemma 9). Since the eigenvalues of $\mathbf{\Sigma}_{\mathbf{s}}$ are $\{1 + \frac{1}{h}\lambda_j(\mathbf{A})\}_{j\in[d]}$, $\mathbf{\Sigma}_{\mathbf{s}} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, and

$$\lambda_1(\mathbf{\Sigma}_{\mathbf{s}}) = 1 + \frac{h}{h} = 2, \quad \lambda_2(\mathbf{\Sigma}_{\mathbf{s}}) = 1 + \frac{2\sqrt{h-1}}{h} \le \frac{3}{2},$$

and the statement about $\mathbf{v}_{\mathbf{s}}$ follows from Lemma 9 after conjugation by $\mathbf{D}_{\mathbf{s}}$. ∎

We are now ready to give our family of distributions for use in Proposition 44.

Lemma 46 *Let $G = (V, E)$ be r-regular and let $\mathbf{s} \in \{\pm 1\}^d$. Let $\overrightarrow{E}$ be made by taking each undirected $e \in E$ and adding both corresponding directed edges to $\overrightarrow{E}$, so that $|\overrightarrow{E}| = rd$. Define $\mathcal{D}_{\mathbf{s},G}$ to be the following distribution. Independently draw $g \overset{\text{unif.}}{\sim} \{\pm 1\}$ and $(u, v) \overset{\text{unif.}}{\sim} \overrightarrow{E}$, and output*

$$\mathbf{x} := \sqrt{\frac{d}{2}} g \cdot \mathbf{D}_{\mathbf{s}} \left(\mathbf{e}_u + \mathbf{e}_v\right).$$

Then $\mathcal{D}_{\mathbf{s},G}$ is sub-Gaussian (Definition 13) with $\sigma^2 = d$, and following Lemma 45,

$$\mathbb{E}_{\mathcal{D}_{\mathbf{s},G}}[\mathbf{x}] = \mathbf{0}_d, \quad \mathbb{E}_{\mathcal{D}_{\mathbf{s},G}}[\mathbf{x}\mathbf{x}^\top] = \mathbf{\Sigma}_{\mathbf{s}}.$$

Proof We first verify the moment calculations: $\mathbb{E}_{\mathcal{D}_{\mathbf{s},G}}[\mathbf{x}] = \mathbf{0}_d$ follows since $\mathbb{E}[g] = 0$, and

$$\mathbb{E}_{\mathcal{D}_{\mathbf{s},G}}\left[\mathbf{x}\mathbf{x}^\top\right] = \mathbb{E}_{(u,v)\overset{\text{unif.}}{\sim}\overrightarrow{E}}\left[\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top \mid (u,v)\right]\right] = \frac{d}{2}\mathbf{D}_{\mathbf{s}}\mathbb{E}_{(u,v)\overset{\text{unif.}}{\sim}\overrightarrow{E}}\left[\left(\mathbf{e}_u + \mathbf{e}_v\right)\left(\mathbf{e}_u + \mathbf{e}_v\right)^\top\right]\mathbf{D}_{\mathbf{s}}.$$

Next, it is a straightforward calculation that

$$\mathbb{E}_{(u,v)\overset{\text{unif.}}{\sim}\overrightarrow{E}}\left[\left(\mathbf{e}_u + \mathbf{e}_v\right)\left(\mathbf{e}_u + \mathbf{e}_v\right)^\top\right] = \frac{2}{d}\mathbf{I}_d + \frac{2}{|\overrightarrow{E}|}\sum_{(u,v)\in\overrightarrow{E}} \mathbf{e}_u\mathbf{e}_v^\top = \frac{2}{d}\mathbf{I}_d + \frac{2}{rd}\mathbf{A}.$$

The covariance follows by combining the above two displays with the definition of $\mathbf{\Sigma}_{\mathbf{s}}$ in Lemma 45. Finally, the sub-Gaussianity claim follows because for any $\mathbf{u} \in \mathbb{R}^d$, and any $(u, v) \in \overrightarrow{E}$,

$$\begin{aligned} &\mathbf{u}^\top\mathbf{D}_{\mathbf{s}}(\mathbf{e}_u + \mathbf{e}_v) \le \sqrt{2}\,\|\mathbf{u}\|_2 \\ \implies &\mathbb{E}\left[\exp\left(\mathbf{u}^\top\mathbf{x}\right) \mid (u,v)\right] \le \exp\left(\frac{d}{4}\cdot 2\,\|\mathbf{u}\|_2^2\right) \le \exp\left(\frac{d}{2}\,\|\mathbf{u}\|_2^2\right), \end{aligned}$$

by using 1-sub-Gaussianity of g. Taking expectation over $(u, v) \overset{\text{unif.}}{\sim} \overrightarrow{E}$ completes the proof. ∎

At this point, we have assembled all the pieces required for our lower bound.

Theorem 47 *Let $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ be an (ϵ, δ)-DP algorithm that, under Model 1 with $\sigma^2 = d$ and $\gamma = \frac{1}{4}$, solves Problem 2 with $\beta = \Delta \le \frac{1}{25}$ for sufficiently large d, k, and $\delta \le \frac{\epsilon}{100}$. Then $n = \Omega(\frac{d}{\epsilon})$.*

Proof Let $m := \frac{d}{2}$ and let G be the $(k-1, k-1)$-biregular bipartite graph G from Lemma 45. For each $\mathbf{x} \in \{\pm 1\}^m$, let $\mathbf{s}(\mathbf{x}) \in \{\pm 1\}^d$ concatenate $\mathbf{1}_m$. Let $\mathcal{V} := \{\mathbf{s} \in \{\pm 1\}^d : \mathbf{s} = \mathbf{s}(\mathbf{x})$ for some $\mathbf{x} \in \{\pm 1\}^m\}$. Then for all $\mathbf{s}, \mathbf{s}' \in \mathcal{V}$ with $d_{\text{ham}}(\mathbf{s}, \mathbf{s}') = t$,

$$\langle \mathbf{s}, \mathbf{s}' \rangle = 1 - \frac{2t}{d} \geq 0 \implies \sin^2 \angle(\mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'}) \geq \frac{2t}{d}.$$

Thus, taking Θ to be the unit sphere and $\boldsymbol{\theta} \equiv \mathbf{v}$, the function $\ell(\mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'}) := \sin^2 \angle(\mathbf{v}_{\mathbf{s}}, \mathbf{v}_{\mathbf{s}'})$ satisfies (23) with $\tau = \frac{1}{d}$. The other property in (23) follows from Lemma 19.

Next, to each $\mathbf{s} \in \mathcal{V}$ we associate a distribution $P_{\mathbf{s}}$ given by n independent draws from $D_{\mathbf{s},G}$ (as defined in Lemma 46). We claim that for all pairs of $\mathbf{s} \in \mathcal{V}_{+i}$, $\mathbf{s}' \in \mathcal{V}_{-i}$ differing only in the i^{th} coordinate, we have $\mathrm{TV}(D_{\mathbf{s},G}, D_{\mathbf{s}',G}) \leq \frac{2}{d}$. To see this, upon coupling the random sign $g \in \{\pm 1\}$ and random edge $(u, v) \in \overrightarrow{E}$ used in constructing $D_{\mathbf{s},G}, D_{\mathbf{s}',G}$, the samples differ iff $i \in \{u, v\}$. This occurs with probability $\leq \frac{2}{d}$. Thus, letting Γ in (24) be the product coupling on coordinates, where each coordinate coupling is the TV coupling, shows we may take $D = \frac{2n}{d}$ in (24).

Now, suppose that for some $n = o(\frac{d}{\epsilon})$ there exists an (ϵ, δ)-DP algorithm $\mathcal{A}$ that solves Problem 2 with the stated parameters. Applying Proposition 44 shows that

$$R(\mathcal{V}, \ell, \epsilon, \delta) \geq \frac{\tau}{2} \cdot \frac{d}{4} \cdot \left(0.9e^{-\frac{20\epsilon n}{d}} - \frac{20n\delta}{d}\right) = \frac{1}{8}\left(0.9e^{-\frac{20\epsilon n}{d}} - \frac{20n\delta}{d}\right) > \frac{1}{10},$$

for small enough n. On the other hand, since $\mathcal{A}$ achieves $\beta = \Delta \leq \frac{1}{20}$ in Problem 2, it certifies that $R(\mathcal{V}, \ell, \epsilon, \delta) \leq \frac{1}{10}$, because the worst-case $\sin^2$ error is ≤ 1. This is a contradiction. ■

Remark 48 *For comparison, the construction in Theorem 47 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 47, 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 $\mathrm{sign}(x_U x_V) = s_U s_V$. 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 47 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.

Appendix I. Approximate DP Lower Bound for Standard PCA

In this section, we adapt the fingerprinting method of Narayanan (2024) to prove Theorem 55, a lower bound for approximate-DP PCA in the standard dense setting. The lower-bound instance below uses Gaussian samples with a spiked inverse-Wishart prior and imposes no k-RCS or sparsity structure on the covariance. Its covariance has a constant-order eigengap on the high-probability event used in the proof. Throughout this section, unless explicitly conditioned otherwise, expectations are over the prior, the samples, and the internal randomness of the algorithm. The proof applies Lemma 24 to PCA. As in Section D.2, the proof reduces to upper and lower bounds on a suitable score.

Construction 2 (Spiked inverse-Wishart PCA instance) *Fix $\gamma \in (0.1, 0.5)$, $\nu = \lceil 1600d/\gamma^2 \rceil$, and a 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

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

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

$$\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i \in [n]} \mathbf{x}_i \mathbf{x}_i^\top, \qquad \mathbf{P} := \mathbf{v}_1(\mathbf{\Sigma})\mathbf{v}_1(\mathbf{\Sigma})^\top, \qquad \widehat{\mathbf{P}} := \mathbf{v}_1(\widehat{\mathbf{\Sigma}})\mathbf{v}_1(\widehat{\mathbf{\Sigma}})^\top.$$

We will use the following standard representation of the inverse-Wishart draw. There is a matrix $\mathbf{G} \in \mathbb{R}^{\nu \times d}$ with independent $\mathcal{N}(0,1)$ entries such that, for

$$\mathbf{A} := \frac{1}{\nu}\mathbf{G}^\top\mathbf{G}, \qquad r := \frac{\nu - d - 1}{\nu},$$

we have

$$\mathbf{\Sigma} = r\mathbf{M}^{1/2}\mathbf{A}^{-1}\mathbf{M}^{1/2}. \tag{25}$$

We require the following helpful properties of the spiked inverse-Wishart PCA construction.

Lemma 49 (Properties of the spiked inverse-Wishart PCA construction) *Under Construction 2, the following hold.*

1. *For $\eta := 6\sqrt{\frac{d}{\nu}}$, let $\mathcal{E}_\eta := \{(1-\eta)\mathbf{I}_d \preceq \mathbf{A} \preceq (1+\eta)\mathbf{I}_d\}$. Then $\mathbb{P}(\mathcal{E}_\eta) \geq 1 - 2e^{-d/2}$, and, on $\mathcal{E}_\eta$,*
$$\frac{1}{5}\mathbf{I}_d \preceq \mathbf{\Sigma} \preceq \frac{11}{10}\mathbf{I}_d, \qquad \lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}) \geq \frac{3}{100}.$$
2. *For every fixed integer $q \geq 1$, there are constants $0 < g_q < G_q < \infty$ such that, for all sufficiently large d,*
$$(1-\gamma) r g_q \leq \mathbb{E}[\lambda_{\min}(\mathbf{\Sigma})^q]^{1/q} \leq \mathbb{E}[\lambda_{\max}(\mathbf{\Sigma})^q]^{1/q} \leq r G_q.$$
3. $$\mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathrm{op}}^2\right] \leq O\left(\frac{d}{n} + \left(\frac{d}{n}\right)^2\right).$$

Proof For Item 1, let $u := 2\sqrt{d/\nu}$. By the Gaussian singular-value bound (Vershynin, 2010, Corollary 5.35), the singular values of $\mathbf{G}/\sqrt{\nu}$ lie in $[1-u, 1+u]$ except with probability $2e^{-d/2}$. Since $\mathbf{A} = (\mathbf{G}/\sqrt{\nu})^\top(\mathbf{G}/\sqrt{\nu})$ and $2u + u^2 \leq 3u = \eta$, this is exactly the event $\mathcal{E}_\eta$. On this event, $\left\|\mathbf{A}^{-1} - \mathbf{I}_d\right\|_{\text{op}} \leq \eta/(1-\eta) \leq \gamma/5$, where the last inequality uses $\nu \geq 1600d/\gamma^2$. Thus $\mathbf{\Sigma} = r\mathbf{M}^{1/2}\mathbf{A}^{-1}\mathbf{M}^{1/2}$ is an $r\gamma/5$ operator-norm perturbation of $r\mathbf{M}$. The matrix $r\mathbf{M}$ has top eigenvalue r and remaining eigenvalues $r(1-\gamma)$, so Weyl's inequality gives the stated constant spectral bounds and eigengap, using $r \geq 1/2$ and $\gamma \in (0.1, 0.5)$.

For Item 2, $\mathbf{\Sigma}$ has the same eigenvalues as $r\mathbf{A}^{-1/2}\mathbf{M}\mathbf{A}^{-1/2}$. The deterministic spike satisfies $(1-\gamma)\mathbf{I}_d \preceq \mathbf{M} \preceq \mathbf{I}_d$, so the eigenvalues of $\mathbf{\Sigma}$ are sandwiched between those of $r(1-\gamma)\mathbf{A}^{-1}$ and $r\mathbf{A}^{-1}$. Since $\mathbf{A}^{-1} = \nu(\mathbf{G}^\top\mathbf{G})^{-1}$, (Narayanan, 2024, Lemma 3.7) supplies constant moments for the extreme eigenvalues of $\mathbf{A}^{-1}$; absorbing constants gives the claim.

For Item 3, condition on $\mathbf{\Sigma}$ and apply (Koltchinskii and Lounici, 2017b, Corollary 2). Its right-hand side is $O(\|\mathbf{\Sigma}\|_{\text{op}}^2\{d/n + (d/n)^2\})$, because $\text{Tr}(\mathbf{\Sigma})/\lambda_1(\mathbf{\Sigma}) \leq d$. Taking expectations and using Item 2 with $q = 2$ proves the bound. ■

Lemma 50 (Empirical projector error) *Fix constants $0 < \underline{\lambda} \leq \overline{\lambda}$ and $g > 0$. There are constants $c, C, C_0 > 0$ such that, if $\underline{\lambda}\mathbf{I}_d \preceq \mathbf{\Sigma} \preceq \overline{\lambda}\mathbf{I}_d$, $\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}) \geq g$, $n \geq C_0 d$, then*

$$c\frac{d}{n} \leq \mathbb{E}\left[\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\text{F}}^2 \mid \mathbf{\Sigma}\right] \leq C\frac{d}{n}.$$

Proof Condition on $\mathbf{\Sigma}$. In the notation of (Koltchinskii and Lounici, 2017a, Theorem 3), the leading eigenspace is simple and all factors depending on $\|\mathbf{\Sigma}\|_{\text{op}}/(\lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma}))$ are constants depending only on the assumed spectral bounds. The leading variance coefficient is

$$A_1(\mathbf{\Sigma}) = 2\sum_{j>1}\frac{\lambda_1(\mathbf{\Sigma})\lambda_j(\mathbf{\Sigma})}{(\lambda_1(\mathbf{\Sigma}) - \lambda_j(\mathbf{\Sigma}))^2}.$$

Each summand is bounded above and below by positive constants, so $ad \leq A_1(\mathbf{\Sigma}) \leq bd$ for constants $a, b > 0$ depending only on the spectral bounds and eigengap.

Let $r_{\text{eff}}(\mathbf{\Sigma}) = \text{Tr}(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|_{\text{op}} \leq d$. Items (1) and (2) of (Koltchinskii and Lounici, 2017a, Theorem 3), applied with $m_1 = 1$, give a constant K such that the centered error has expansion

$$\mathbb{E}\left[\|\widehat{\mathbf{P}} - \mathbb{E}[\widehat{\mathbf{P}} \mid \mathbf{\Sigma}]\|_{\text{F}}^2 \mid \mathbf{\Sigma}\right] = \frac{A_1(\mathbf{\Sigma})}{n} + \xi_n, \qquad |\xi_n| \leq K\left(\left(\frac{d}{n}\right)^{3/2} \vee \left(\frac{d}{n}\right)^4\right).$$

The same theorem bounds the bias by

$$\|\mathbb{E}[\widehat{\mathbf{P}} \mid \mathbf{\Sigma}] - \mathbf{P}\|_{\text{F}} \leq K\left(\frac{d}{n} \vee \left(\frac{d}{n}\right)^2\right).$$

The conditional bias-variance identity then writes the desired error as the centered term plus the squared bias. With $x = d/n$, choosing C_0 sufficiently large makes $x \leq 1$ and absorbs $K(x^{3/2} \vee x^4)$

into the leading $A_1(\mathbf{\Sigma})/n$ term. The centered term is therefore between constant multiples of d/n, and the squared bias is only $O((d/n)^2)$. This proves both bounds. ■

We next record the posterior stability estimate needed by the fingerprinting argument.

Lemma 51 (Posterior stability for the PCA prior) *Under Construction 2,*

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

Proof The prior has inverse-Wishart parameters $\mathbf{\Psi} = (\nu - d - 1)\mathbf{M}$ and ν. By Fact 4,

$$\mathbb{E}[\mathbf{\Sigma}] = \mathbf{M}, \qquad \mathbb{E}[\mathbf{\Sigma} \mid \mathcal{X}] = (1 - w_n)\mathbb{E}[\mathbf{\Sigma}] + w_n\widehat{\mathbf{\Sigma}}, \qquad w_n := \frac{n}{\nu + n - d - 1}.$$

Thus $\mathbb{E}[\mathbf{\Sigma} \mid \mathcal{X}] - \widehat{\mathbf{\Sigma}} = (1 - w_n)(\mathbb{E}[\mathbf{\Sigma}] - \widehat{\mathbf{\Sigma}})$ and $1 - w_n = (\nu - d - 1)/(n + \nu - d - 1) \le \nu/n$. It remains to bound $\mathbb{E}\left\|\widehat{\mathbf{\Sigma}} - \mathbb{E}[\mathbf{\Sigma}]\right\|_{\text{op}}^2$. By the triangle inequality and $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$, this expectation is at most

$$3\mathbb{E}\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\text{op}}^2 + 3\mathbb{E}\left\|\mathbf{\Sigma}\right\|_{\text{op}}^2 + 3\left\|\mathbb{E}[\mathbf{\Sigma}]\right\|_{\text{op}}^2.$$

The first term is Lemma 49, Item 3; the second is Item 2 with $q = 2$; and the third is bounded by Jensen's inequality. This gives $O(1 + d/n + (d/n)^2)$ before multiplying by $(1 - w_n)^2$, which proves the claim. ■

Under Construction 2, instantiate Model 3 with $\theta = \mathbf{\Sigma}$, $P_\theta = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, $g(\mathbf{\Sigma}) := \mathbf{P}$, and $\psi(\mathbf{\Sigma}, \mathbf{x}) := \mathbf{x}\mathbf{x}^\top - \mathbf{\Sigma}$. Let $\mathcal{M} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ output a unit vector $\tilde{\mathbf{v}}$, and set $M(\mathcal{X}) := \widetilde{\mathbf{P}} := \tilde{\mathbf{v}}\tilde{\mathbf{v}}^\top$. Let Z_i, Z_i' be the corresponding scores in Lemma 24.

Lemma 52 (PCA score bounds) *Under Construction 2, there are universal constants $0 < c < 1 < C_0, C_1$ such that, if $\mathbb{E}\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\text{F}}^4 \le \rho^4$, $\rho < c$, $n \ge C_0 d$, and $d \ge C_1$, then*

$$\mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] \ge \Omega(d)$$

and, for every $i \in [n]$,

$$2\epsilon\mathbb{E}\left|Z_i'\right| + 2\sqrt{\delta}\sqrt{\mathbb{E}[Z_i^2] + \mathbb{E}[(Z_i')^2]} \le O\left(\rho(\epsilon + d\sqrt{\delta})\right).$$

Proof *Signal.* Let $\text{gap}(\mathbf{\Sigma}) = \lambda_1(\mathbf{\Sigma}) - \lambda_2(\mathbf{\Sigma})$. The empirical projector $\widehat{\mathbf{P}}$ is useful because it simultaneously maximizes the empirical Rayleigh quotient and is close to the population projector. More precisely, $\left\langle \widehat{\mathbf{P}} - \mathbf{P}, \widehat{\mathbf{\Sigma}} \right\rangle \ge 0$ by empirical optimality. Lemma 20 turns the population Rayleigh deficit of $\widehat{\mathbf{P}}$ into projector error: since $\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\text{F}}^2 = 2\sin^2\angle(\mathbf{v}_1(\widehat{\mathbf{\Sigma}}), \mathbf{v}_1(\mathbf{\Sigma}))$, it gives

$\left\langle \mathbf{P} - \widehat{\mathbf{P}}, \boldsymbol{\Sigma} \right\rangle \geq \frac{\mathsf{gap}(\boldsymbol{\Sigma})}{2} \|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2$. Combining the empirical and population inequalities shows that the empirical fluctuation $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ has positive correlation with the empirical projector error.

It remains to replace $\widehat{\mathbf{P}}$ by the algorithm's output $\widetilde{\mathbf{P}}$. Let R be the empirical projector error $\mathbb{E}\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2$, and let H be the posterior-stability quantity controlled by Lemma 51. Conditional on $\mathcal{X}$ and the algorithmic randomness, $\widetilde{\mathbf{P}}$ is fixed and independent of the posterior draw $\boldsymbol{\Sigma}$. Therefore the replacement error is $\mathbb{E}\left\langle \widehat{\mathbf{P}} - \widetilde{\mathbf{P}}, \widehat{\boldsymbol{\Sigma}} - \mathbb{E}[\boldsymbol{\Sigma} \mid \mathcal{X}] \right\rangle$. For each realization, $\widehat{\mathbf{P}} - \widetilde{\mathbf{P}}$ is a difference of rank-one projectors, so its nuclear norm is bounded by a universal constant times its Frobenius norm. Trace-operator norm duality and Cauchy-Schwarz bound the replacement error by $O(\sqrt{(\rho^2 + R)H})$. Here we used $\mathbb{E}\|\widehat{\mathbf{P}} - \widetilde{\mathbf{P}}\|_{\mathrm{F}}^2 \leq 2R + 2\mathbb{E}\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2$ and $\mathbb{E}\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2 \leq \rho^2$. Hence

$$\mathbb{E}\left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\rangle \geq \mathbb{E}\left[\frac{\lambda_1(\boldsymbol{\Sigma}) - \lambda_2(\boldsymbol{\Sigma})}{2} \|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2\right] - O(\sqrt{(\rho^2 + R)H}). \tag{26}$$

On the good event from Lemma 49, the covariance has a constant eigengap and bounded spectrum, so Lemma 50 gives an empirical projector error of order d/n. The complement has probability $2e^{-d/2}$, while $\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^2 \leq 2$. Thus the main term in (26) is $\Omega(d/n)$ and $R = O(d/n + e^{-d/2})$. Lemma 51 gives $H = O((d/n)^2)$ because $\nu = \Theta(d)$ in the construction. For ρ small and $n \geq C_0 d$, the replacement cost is lower order, so $\mathbb{E}\left\langle \widetilde{\mathbf{P}} - \mathbf{P}, \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\rangle \geq \Omega(d/n)$. Multiplying by n gives $\mathbb{E}[\sum_i Z_i] \geq \Omega(d)$.

Moments. For Z_i', the output $M(\mathcal{X}^{\sim i})$ is conditionally independent of $\mathbf{x}_i$ given $\boldsymbol{\Sigma}$. (Narayanan, 2024, Proposition 3.8) gives the conditional bound $\mathbb{E}[(Z_i')^2 \mid \boldsymbol{\Sigma}, \mathcal{X}^{\sim i}] \leq 2\|\boldsymbol{\Sigma}\|_{\mathrm{op}}^2 \|M(\mathcal{X}^{\sim i}) - \mathbf{P}\|_{\mathrm{F}}^2$. Cauchy-Schwarz, the fourth-moment assumption, and Lemma 49, Item 2 with $q = 4$, give $\mathbb{E}[(Z_i')^2] = O(\rho^2)$, hence $\mathbb{E}|Z_i'| = O(\rho)$.

For Z_i, Cauchy-Schwarz gives $\mathbb{E}[Z_i^2] \leq \sqrt{\mathbb{E}\|\widetilde{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^4 \, \mathbb{E}\|\mathbf{x}_i\mathbf{x}_i^\top - \boldsymbol{\Sigma}\|_{\mathrm{F}}^4}$. The Gaussian norm moment bound (Narayanan, 2024, Proposition 3.10), together with Item 2, bounds the second factor by $O(d^4)$. Thus $\mathbb{E}[Z_i^2] = O(\rho^2 d^2)$, which is the claimed DP comparison bound. ■

Proposition 53 (Moment lower bound for dense PCA) *There are universal constants $0 < c < 1 < C_0, C_1$ such that the following holds under Construction 2. Let $\epsilon, \delta \in (0,1)$ satisfy $\delta \leq \epsilon^2/d^2$, let $\rho < c$, $n \geq C_0 d$, and $d \geq C_1$. Every (ϵ, δ)-DP algorithm $\mathcal{M} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ whose unit-vector output $\tilde{\mathbf{v}}$ satisfies*

$$\mathbb{E}\|\tilde{\mathbf{v}}\tilde{\mathbf{v}}^\top - \mathbf{P}\|_{\mathrm{F}}^4 \leq \rho^4$$

must have

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

Proof The score is centered because, conditional on $\boldsymbol{\Sigma}$, $\mathbb{E}[\psi(\boldsymbol{\Sigma}, \mathbf{x}_i) \mid \boldsymbol{\Sigma}] = \mathbb{E}[\mathbf{x}_i\mathbf{x}_i^\top \mid \boldsymbol{\Sigma}] - \boldsymbol{\Sigma} = \mathbf{0}$. Lemma 52 supplies the two certificates in Lemma 24 with $L = \Omega(d)$ and $U = O\left(\rho(\epsilon + d\sqrt{\delta})\right)$. The assumption $\delta \leq \epsilon^2/d^2$ implies $d\sqrt{\delta} \leq \epsilon$, so $U = O(\rho\epsilon)$. Lemma 24 therefore gives

$$n \geq \frac{L}{U} = \Omega\left(\frac{d}{\rho\epsilon}\right).$$

■

Lemma 54 (PCA moment amplification) *Let $\rho_0 \in (0,1)$ be a sufficiently small universal constant. Suppose an (ϵ,δ)-DP algorithm $\mathcal{A}$ uses n samples and solves Problem 2 under the Gaussian instances in Construction 2 with $\Delta = \rho_0^2/100$ and $\beta = 1/3$. Then there is an (ϵ,δ)-DP algorithm $\mathcal{A}'$ using $O(n)$ samples whose unit-vector output $\tilde{\mathbf{v}}$ satisfies*

$$\mathbb{E}\|\tilde{\mathbf{v}}\tilde{\mathbf{v}}^\top - \mathbf{P}\|_{\mathrm{F}}^4 \le \rho_0^4.$$

Proof Set $B = \lceil 18\log(8/\rho_0^4)\rceil$, a universal constant once ρ_0 is fixed. Run $\mathcal{A}$ on B disjoint blocks, and let $\mathbf{P}_j$ be the returned projector on block j. Output a unit vector whose projector $\widehat{\mathbf{P}}$ minimizes the median of $\{\|\mathbf{P}_j - \mathbf{P}_\ell\|_{\mathrm{F}} : \ell \in [B]\}$.

Since Problem 2 is a pointwise guarantee for each Gaussian instance, each block is successful with probability at least $2/3$, conditional on $\mathbf{\Sigma}$. On a successful block, $\|\mathbf{P}_j - \mathbf{P}\|_{\mathrm{F}}^2 = 2\sin^2\angle(\tilde{\mathbf{v}}_j, \mathbf{v}_1(\mathbf{\Sigma})) \le \rho_0^2/50$, so $\|\mathbf{P}_j - \mathbf{P}\|_{\mathrm{F}} \le \rho_0/5$. Hoeffding's inequality gives probability at most $\exp(-B/18) \le \rho_0^4/8$ that at most half the blocks are successful.

If more than half the blocks are successful, any successful $\mathbf{P}_j$ has median distance at most $2\rho_0/5$ from the other projectors. The minimizing $\widehat{\mathbf{P}}$ therefore has median distance at most $2\rho_0/5$, so it is within $2\rho_0/5$ of at least one successful block. Hence $\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}} \le 3\rho_0/5$. Since the Frobenius distance between rank-one projectors is always at most $\sqrt{2}$,

$$\mathbb{E}\|\widehat{\mathbf{P}} - \mathbf{P}\|_{\mathrm{F}}^4 \le \left(\frac{3\rho_0}{5}\right)^4 + 4\cdot\frac{\rho_0^4}{8} \le \rho_0^4.$$

Parallel composition preserves (ϵ,δ)-DP. ■

Theorem 55 *There are universal constants $\Delta_0, C_1 > 0$ such that the following holds. Let $\epsilon \le 1$, $\delta \le \epsilon^2/d^2$, and $d \ge C_1$. If an (ϵ,δ)-DP algorithm $\mathcal{A} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ solves Problem 2 with $\Delta = \Delta_0$ and $\beta = 1/3$ under the Gaussian instances in Construction 2, then*

$$n = \Omega\left(\frac{d}{\epsilon}\right).$$

Proof Apply Lemma 54 with $\Delta_0 = \rho_0^2/100$ for a sufficiently small constant ρ_0. This gives an (ϵ,δ)-DP algorithm $\mathcal{A}'$ using $N = O(n)$ samples and satisfying the hypothesis of Proposition 53. We may assume $n = \Omega(d)$: the lower bound of (Cai et al., 2013, Theorem 3), with ambient dimension and sparsity both equal to d, gives constant projector error unless $n = \Omega(d)$ for constant signal-to-noise ratio and eigengap; the same dense specialization is also consistent with the row-sparse $q = 0$ lower bound of (Vu and Lei, 2013, Theorem 3.1). Thus a sufficiently small constant-error guarantee already forces $n = \Omega(d)$. Increasing the constant number of amplification blocks if necessary ensures $N \ge C_0 d$. Therefore Proposition 53 gives $N = \Omega(d/\epsilon)$, and $N = O(n)$ completes the proof. ■

Algorithm 4: $\mathsf{ExpMechPCA}(\mathcal{D}, k, \epsilon, \delta, \tau)$

Input: Dataset $\mathcal{D} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i\in[n]}$, sparsity parameter $k \in [d]$, privacy parameters $(\epsilon, \delta) \in (0, 1)^2$, threshold $\tau > 0$.

Output: A private k-sparse PCA vector $\hat{\mathbf{v}} \in \mathbb{R}^d$.

$\mathcal{S}_k \leftarrow \{S \subseteq [d] : |S| = k\}$;
for $i \in [n]$ **do**
  $\mathbf{z}_i \leftarrow \mathsf{sign}(\mathbf{x}_i) \circ \min\{|\mathbf{x}_i|, \tau\}$;
  // $\min\{\cdot, \tau\}$, $|\cdot|$, and $\mathsf{sign}(\cdot)$ are entrywise; $\circ$ is entrywise multiplication.
end
$\widehat{\mathbf{\Sigma}} \leftarrow \frac{1}{n}\sum_{i\in[n]} \mathbf{z}_i\mathbf{z}_i^\top$;
for $S \in \mathcal{S}_k$ **do**
  $q(\mathcal{D}, S) \leftarrow \lambda_1(\widehat{\mathbf{\Sigma}}_{S\times S})$;
end
$\pi_{\mathcal{D}} \leftarrow$ distribution over $\mathcal{S}_k$ with $\pi_{\mathcal{D}}(S) \propto \exp\left(\frac{\epsilon n}{8k\tau^2} q(\mathcal{D}, S)\right)$;
$\widehat{S} \sim \pi_{\mathcal{D}}$;
$\sigma_{\text{priv}} \leftarrow \frac{8k\tau^2}{\epsilon n}\sqrt{\log(\frac{4}{\delta})}$;
$\mathbf{G} \leftarrow d \times d$ matrix with $[\mathbf{G}]_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{priv}}^2)$;
$\widetilde{\mathbf{\Sigma}} \leftarrow \widehat{\mathbf{\Sigma}} + \frac{1}{2}(\mathbf{G} + \mathbf{G}^\top)$;
$\hat{\mathbf{v}} \leftarrow \mathbf{0}_d$;
$\hat{\mathbf{v}}_{\widehat{S}} \leftarrow \mathbf{v}_1(\widetilde{\mathbf{\Sigma}}_{\widehat{S}\times\widehat{S}})$;
return $\hat{\mathbf{v}}$;

Appendix J. Exponential Mechanism for Sparse PCA

This appendix gives a simple algorithm for Problem 2 under a generalization of Model 2, where we only assume $\mathbf{\Sigma}$ has a unique, k-sparse leading eigenvector, with no k-RCS assumption imposed.

Our algorithm is based on the exponential mechanism, and uses a brute-force enumeration of all candidate supports (size-k subsets of $[d]$). Notably, after dropping privacy terms and logarithmic factors, Theorem 58 has leading statistical error $\widetilde{O}(\sqrt{k/n})$, matching the information-theoretic non-private lower bound Vu and Lei (2013); Cai et al. (2013) rather than the $\widetilde{O}(k/\sqrt{n})$ rate attained by known polynomial-time approaches Berthet and Rigollet (2013). This improvement is enabled by an exhaustive search over all k-supports via the exponential mechanism.

Algorithm 4 is an instance of the *exponential mechanism*, whose guarantees we now recall.

Lemma 56 (Exponential mechanism, Theorem 3.10 and Corollary 3.12, Dwork and Roth (2014)) *Let $\mathcal{R}$ be a finite set and $q : \Gamma^n \times \mathcal{R} \to \mathbb{R}$ be a score function with global sensitivity*

$$\Delta_q := \sup_{r\in\mathcal{R}} \sup_{\text{neighboring } \mathcal{D},\mathcal{D}'\in\Gamma^n} |q(\mathcal{D}, r) - q(\mathcal{D}', r)|.$$

Then, the mechanism $\mathcal{M}$ defined by $\mathbb{P}(\mathcal{M}(\mathcal{D}) = r) \propto \exp(\frac{\epsilon}{2\Delta_q} q(\mathcal{D}, r))$ is ϵ-DP. Moreover, for every $\beta \in (0,1)$, with probability at least $1-\beta$,

$$q(\mathcal{D}, \mathcal{M}(\mathcal{D})) \geq \max_{r \in \mathcal{R}} q(\mathcal{D}, r) - \frac{2\Delta_q}{\epsilon} \left(\log |\mathcal{R}| + \log \frac{1}{\beta} \right).$$

More concretely, Algorithm 4 assigns to each $S \in \mathcal{S}_k$ a score $q(\mathcal{D}, S)$ given by the largest eigenvalue of $\widehat{\mathbf{\Sigma}}^{\tau}_{S}$, the $S \times S$ submatrix of the clipped empirical covariance. It then applies the exponential mechanism (Lemma 56) to privately select a set $\widehat{S}$, and the remaining steps proceed similarly to Algorithm 1. We begin by proving the privacy of Algorithm 4.

Lemma 57 (Privacy) *Algorithm 4 is (ϵ, δ)-DP.*

Proof Fix $S \in \mathcal{S}_k$, and let $\mathcal{D}, \mathcal{D}'$ be adjacent datasets. Then

$$\widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}) - \widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}') = \frac{1}{n} \left(\mathbf{z}\mathbf{z}^{\top} - \mathbf{z}'(\mathbf{z}')^{\top} \right),$$

where $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^k$ are the clipped restrictions of the differing samples to S. Since $\|\mathbf{z}\|_2^2, \|\mathbf{z}'\|_2^2 \leq k\tau^2$,

$$\|\widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}) - \widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}')\|_{\text{op}} \leq \frac{\|\mathbf{z}\|_2^2 + \|\mathbf{z}'\|_2^2}{n} \leq \frac{2k\tau^2}{n}.$$

By Weyl's inequality,

$$\left| q(\mathcal{D}, S) - q(\mathcal{D}', S) \right| \leq \|\widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}) - \widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D}')\|_{\text{op}} \leq \frac{2k\tau^2}{n}.$$

This proves that the statistic $q(\mathcal{D}, S)$ has global sensitivity $\frac{2k\tau^2}{n}$ for every S. Thus, the support selection step is $\frac{\epsilon}{2}$-DP by Lemma 56. Also, the same calculations as above with the Frobenius norm in place of the operator norm gives the same bound of $\frac{2k\tau^2}{n}$ on the Frobenius norm sensitivity of the map $\mathcal{D} \to \widehat{\mathbf{\Sigma}}^{\tau}_{S}(\mathcal{D})$. For every fixed transcript $\widehat{S} = S$, releasing $\widehat{\mathbf{\Sigma}}^{\tau}_{S} + \mathbf{G}_S$ with independent entrywise Gaussian noise of variance σ^2_{priv} is then $(\frac{\epsilon}{2}, \delta)$-DP by Fact 2. Symmetrization and eigenvector computation are postprocessings, and basic composition gives (ϵ, δ)-DP. ■

We now conclude our analysis of Algorithm 4 by proving a utility bound.

Theorem 58 *Let $(\epsilon, \delta, \Delta, \beta) \in (0,1)^4$, $k \in [d]$, and $\gamma \in (0,1)$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be drawn i.i.d. from a mean-zero σ-sub-Gaussian distribution with covariance $\mathbf{\Sigma} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$. Assume $\lambda_1(\mathbf{\Sigma}) > 0$, $\lambda_2(\mathbf{\Sigma}) \leq (1-\gamma)\lambda_1(\mathbf{\Sigma})$, and $\mathbf{v}_1(\mathbf{\Sigma})$ is k-sparse. Algorithm 4, with $\tau \leftarrow 2\sigma\sqrt{\log \frac{4nd}{\beta}}$, solves Problem 2 with*

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

for a sufficiently large constant.

Proof The privacy claim follows from Lemma 57.

Let $\mathbf{v} := \mathbf{v}_1(\boldsymbol{\Sigma})$ and let $S^\star \in \mathcal{S}_k$ contain supp($\mathbf{v}$). We condition on four events, each with failure probability at most $\beta/4$. First, clipping is inactive for every sample; this follows from the coordinatewise sub-Gaussian tail bound and the choice of τ. Second, the empirical covariance is accurate on every candidate support:

$$\left\|\widehat{\boldsymbol{\Sigma}}_{S\times S} - \boldsymbol{\Sigma}_{S\times S}\right\|_{\text{op}} \leq \frac{\gamma\Delta\lambda_1(\boldsymbol{\Sigma})}{5} \quad \text{for all } S \in \mathcal{S}_k. \tag{27}$$

This is Lemma 15 with the stated sample size. Third, the Gaussian perturbation on the selected support has operator norm at most $\gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$:

$$\left\|\left[\frac{1}{2}\left(\mathbf{G} + \mathbf{G}^\top\right)\right]_{\widehat{S}\times\widehat{S}}\right\|_{\text{op}} \leq \frac{\gamma\Delta\lambda_1(\boldsymbol{\Sigma})}{5}. \tag{28}$$

This follows from the standard Gaussian operator-norm bound (Vershynin, 2018, Theorem 4.4.3) and the choice of n. Fourth, the exponential mechanism loses at most $\gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$ in score; here $\log|\mathcal{S}_k| \leq k\log(ed/k)$ and the score sensitivity is $2k\tau^2/n$.

These events imply that the selected support has nearly optimal population Rayleigh value. Indeed, $S^\star$ contains the true support, so $\lambda_1(\boldsymbol{\Sigma}_{S^\star\times S^\star}) = \lambda_1(\boldsymbol{\Sigma})$. The sample event shows that $S^\star$ has empirical score at least $\lambda_1(\boldsymbol{\Sigma}) - \gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$; the exponential mechanism loses another $\gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$; and the Gaussian perturbation loses another $\gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$. Thus $\lambda_1(\widetilde{\boldsymbol{\Sigma}}_{\widehat{S}\times\widehat{S}}) \geq \lambda_1(\boldsymbol{\Sigma}) - 3\gamma\Delta\lambda_1(\boldsymbol{\Sigma})/5$.

The output $\hat{\mathbf{v}}$ is the top eigenvector of $\widetilde{\boldsymbol{\Sigma}}_{\widehat{S}\times\widehat{S}}$, embedded into $\mathbb{R}^d$. Since $\hat{\mathbf{v}}$ is supported on $\widehat{S}$, the sample and noise events convert the preceding lower bound into the population Rayleigh bound $\hat{\mathbf{v}}^\top\boldsymbol{\Sigma}\hat{\mathbf{v}} \geq \lambda_1(\boldsymbol{\Sigma}) - \gamma\Delta\lambda_1(\boldsymbol{\Sigma})$. Now Lemma 20 gives the conclusion directly: it says that the squared sine error is at most the Rayleigh deficit divided by the eigengap. Let $\hat{\mathbf{v}} = c\mathbf{v} + \mathbf{r}$ where $\mathbf{r}$ is orthogonal to $\mathbf{v}$, $\|\mathbf{r}\|^2 = \sin^2\angle(\hat{\mathbf{v}}, \mathbf{v})$, and $c^2 = 1 - \|\mathbf{r}\|^2$. Then the population Rayleigh lower bound and the eigengap imply $\hat{\mathbf{v}}^\top\boldsymbol{\Sigma}\hat{\mathbf{v}} \leq \lambda_1(\boldsymbol{\Sigma}) - (\lambda_1(\boldsymbol{\Sigma}) - \lambda_2(\boldsymbol{\Sigma}))\sin^2\angle(\hat{\mathbf{v}}, \mathbf{v})$. Here the deficit is at most $\gamma\Delta\lambda_1(\boldsymbol{\Sigma})$, while the eigengap is at least $\gamma\lambda_1(\boldsymbol{\Sigma})$. Therefore $\sin^2\angle(\hat{\mathbf{v}}, \mathbf{v}) \leq \Delta$. ■