Proceedings of Machine Learning Research vol 313:1–11, 2026 37th International Conference on Algorithmic Learning Theory

On Purely Private Covariance Estimation

Tommaso d'Orsi TOMMASO.DORSI@UNIBOCCONI.IT
Bocconi University, Italy.

Gleb Novikov GLEB.NOVIKOV@HSLU.CH
Lucerne School of Computer Science and Information Technology, Switzerland.

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

We present a simple perturbation mechanism for the release of d-dimensional covariance matrices Σ under pure differential privacy. For large datasets with at least $n \geqslant d^2/\varepsilon$ elements, our mechanism recovers the provably optimal Frobenius norm error guarantees of Nikolov (2023), while simultaneously achieving best known error for all other p-Schatten norms, with $p \in [1, \infty]$. Our error is information-theoretically optimal for all $p \geqslant 2$, in particular, our mechanism is the first purely private covariance estimator that achieves optimal error in spectral norm.

For small datasets $n < d^2/\varepsilon$, we further show that by projecting the output onto the nuclear norm ball of appropriate radius, our algorithm achieves the optimal Frobenius norm error $O(\sqrt{d\,\mathrm{Tr}(\Sigma)/n})$, improving over the known bounds of $O(\sqrt{d/n})$ of Nikolov (2023) and $O\big(d^{3/4}\sqrt{\mathrm{Tr}(\Sigma)/n}\big)$ of Dong et al. (2022).

1. Introduction

For a set of vectors $X = \{x_1, \ldots, x_n\} \subset \mathcal{B}_d$, where $\mathcal{B}_d$ is a d-dimensional Euclidean unit ball[1], the covariance matrix is defined as[2]

$$\Sigma := \frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top .$$

Because of its widespread use in training prediction models, the task of releasing a private version of Σ has received significant attention (see Amin et al. (2019); Dong et al. (2022); Nikolov (2023); Cohen-Addad et al. (2024) and references therein). In contrast to ad-hoc private training methods, the advantage of working with a synthetic version $\hat{\Sigma}$ of the covariance is that one can replace Σ with $\hat{\Sigma}$ and directly run standard training algorithms. This approach not only fits seamlessly in existing training pipelines, but typically also leads to lower privacy costs, as these are paid only once, upon computing the private estimate.

The de facto standard for privacy is differential privacy (Dwork et al. (2006)). A randomized algorithm $\mathcal{A}$ is said to be (ε, δ)-differentially private if, for any pair of datasets $X, X' \subseteq \mathbb{R}^d$ differing in at most 1 element, and for any measurable event S in the range of $\mathcal{A}$, it holds

$$\mathbb{P}\left(\mathcal{A}(X) \in S\right) \leqslant e^{\varepsilon} \cdot \mathbb{P}\left(\mathcal{A}(X') \in S\right) + \delta.$$

1. $X \subset \mathcal{B}_d$ is a standard assumption in the literature on private covariance estimation.
2. We work with the non-centered covariance. All our results are also valid for estimation of the centered covariance $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top - \left(\frac{1}{n}\sum_i x_i\right)\left(\frac{1}{n}\sum_i x_i\right)^\top$ if the parameters of the algorithms and error bounds are adjusted by a small absolute constant factor.

© 2026 T. d'Orsi & G. Novikov.

Algorithm 1 ε-Differentially Private Covariance Estimation via Perturbations

Input: Σ, ε

1. Return $\Sigma + Z$ with Z drawn from the nuclear-Laplace law $p(Z) \propto \exp\big(-\|Z\|_*/\rho\big)$, where $\rho = 2/(\varepsilon n)$.

The setting $\delta = 0$ is usually refer to as *pure* privacy, while the setting $\delta > 0$ is called approximate differential privacy. Pure privacy is the strictest notion of privacy, it disallows any complete privacy breach and naturally offers privacy protection to groups of elements in the datasets, as opposed to just individuals (see Nikolov (2023) for a more comprehensive discussion).

In the context of private covariance estimation, the difference between pure and approximate differential privacy is surprisingly stark. For $\delta > 0$, efficient algorithms can achieve accuracy $\|\hat{\Sigma} - \Sigma\|_{\mathrm{F}}^2 \leqslant O\left(\sqrt{\log(1/\delta)}/\varepsilon\right) \cdot \min\left\{d/n,\, d^{1/4}\sqrt{\varepsilon/n}\right\}$ and this bound is known to be tight (Nikolov et al. (2013); Dwork et al. (2015); Kasiviswanathan et al. (2010)). In contrast, pure privacy lower bounds for the special case of 2-way marginals already rule out a Frobenius-norm error smaller than $\min\left\{d^{3/2}/\varepsilon n,\, \sqrt{d/n\varepsilon}\right\}$ (Hardt and Talwar (2010); Nikolov (2023)). Here the second term is smaller for small sample sets $n < d^2/\varepsilon$.

Nikolov (2023) introduced a variant of the projection mechanism, based on the Johnson-Lindestrauss transform (Johnson (1984)) matching these bounds up to constant factors. For datasets with points of small average Euclidean norm, Dong et al. (2022) designed a pure differentially private algorithm achieving a Frobenius norm error of $\tilde{O}(d^{3/4}\sqrt{\operatorname{Tr}\Sigma/n} + \sqrt{d}/n)$, thus going beyond Nikolov (2023) for the special case of matrices satisfying $\operatorname{Tr}\Sigma \ll O(1/\varepsilon) \cdot \min\left\{1/n,\, d^{-1/4}\right\}$.

In this work, we show that particularly simple perturbation and projection mechanisms can be used to tie together these two upper bounds, and improve over them. Our analysis not only tightens the Frobenius norm bound, accounting for the structure of the covariance matrix, but further shows the error can be controlled in a stronger sense: bounding the spectral discrepancy simultaneously in *every* Schatten norm $\|\cdot\|_{\mathrm{S}_p}$ with $p \in [1, \infty]$.

Notation. For $p \in [1, \infty]$ the p-Schatten norm of A is defined as $\|A\|_{S_p} = (\sum_i \sigma_i^p)^{1/p}$, where $\sigma_1 \geqslant \ldots \geqslant \sigma_d \geqslant 0$ are singular values of A. For simplicity we denote the spectral norm ($p = \infty$) with $\|A\|$, the Frobenius norm ($p = 2$) with $\|A\|_{\mathrm{F}}$ and the nuclear norm ($p = 1$) with $\|A\|_*$.

1.1. Results

Thanks to the simplicity of our mechanisms, we can directly introduce them and states their guarantees. Our first result shows that the following additive perturbation mechanism yields tight error bounds for all Schatten norms.

Efficient and streamlined samplers for the nuclear-Laplace law are easy to construct and can be seen as a specific instantiation of the K-norm mechanism of Hardt and Talwar (2010). Therefore the algorithm runs in polynomial time.

Algorithm 2 ε-Differentially Private Covariance Estimation via Projection

Input: Σ, ε

1. Run Algorithm 1 on Σ, $\varepsilon/2$. Let $\hat{\Sigma}$ be its output.
2. Return the orthogonal projection $\tilde{\Sigma}$ of $\hat{\Sigma}$ onto the nuclear norm ball

$$\left\{ Y \in \mathbb{R}^{d\times d} \mid \|Y\|_* \leqslant \max\left(0,\ 2\|\Sigma\|_* + \mathrm{Lap}\left(\tfrac{10}{\varepsilon n}\right)\right) \right\}.$$

Theorem 1 *Let $\varepsilon > 0$. Algorithm 1 is ε-DP, runs in polynomial time, and with high probability its output $\hat{\Sigma}$ satisfies*

$$\left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{S}_p} \leqslant \frac{3 \cdot d^{1+1/p}}{\varepsilon n}$$

for every $p \in [1, \infty]$.

Theorem 1 guarantees bounded error *simultaneously* for all Schatten norms, including the nuclear norm ($p = 1$) and the spectral norm ($p = \infty$), implying spectral discrepancies between the true covariance and the pure differentially private estimate are small under any natural reweighing. For large datasets $n \geqslant d^2/\varepsilon$ –typical of real-world applications– the Theorem recovers the guarantees of Nikolov (2023) for the Frobenius norm error. Furthermore, as for $p \geqslant 2$

$$\left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{S}_p} \geqslant d^{1/p-1/2} \left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{F}}, \tag{1.1}$$

the error of Theorem 1 is *optimal* for $p \geqslant 2$ in the large sample regime $n \geqslant d^2/\varepsilon$. Indeed, the information-theoretic lower bound (Nikolov (2023)) shows that if $n \geqslant d^2/\varepsilon$, no pure DP estimator can achieve error $o(d^{3/2}/\varepsilon n)$ in Frobenius norm, and hence no estimator can achieve error $o(d^{1+1/p}/\varepsilon n)$ in Schatten p-norm for $p \geqslant 2$. In particular, our spectral norm bound $O(d/\varepsilon n)$ is information-theoretically optimal in the large sample regime $n \geqslant d^2/\varepsilon$.

Our second result shows that the Frobenius norm error can be improved, projecting the perturbed covariance obtained from Algorithm 1 back onto the nuclear ball of radius $\|\Sigma\|_*$, up to a tiny perturbation required to maintain privacy.

Theorem 2 *Let $\varepsilon > 0$. Algorithm 2 is ε-DP, runs in polynomial time, and with probability $1-d^{-10}$ its output $\tilde{\Sigma}$ satisfies*

$$\left\|\tilde{\Sigma} - \Sigma\right\|_{\mathrm{F}} \leqslant O(1) \cdot \min\left\{ \frac{d^{3/2}}{\varepsilon n},\ \sqrt{\frac{d \cdot \operatorname{Tr}\Sigma}{\varepsilon n}} + \frac{\sqrt{d \log d}}{\varepsilon n} \right\}.$$

Theorem 2 always recovers the result of Nikolov (2023), but improves over it whenever $\operatorname{Tr}\Sigma < d^2/n$. As values of the trace of the covariance matrix can be significantly smaller than 1, Theorem 2 can lead to important accuracy improvements both in the small sample regime $n \leqslant d^2/\varepsilon$ where the improvement is of the order $O(1/\operatorname{Tr}\Sigma)$, and in the large sample regime. Theorem 2 also strictly improves over the result of Dong et al. (2022) (that also captured the dependence on $\operatorname{Tr}(\Sigma)$) in all regimes by a $d^{1/4}$ factor.

2. Techniques

To illustrate our ideas we start by reviewing the algorithm of Nikolov (2023). On a high level, this consists of three steps: (1) project the data onto a low dimensional space, (2) add noise via a K-norm mechanism, and (3) project the result back onto the space of covariance matrices (the intersection of the nuclear unit ball and the positive semidefinite cone). As the cost of traditional perturbation mechanisms tends to be proportional to the ambient dimension, the goal of the first step is to reduce this privacy cost by embedding the dataset into a small subspace, thus paying a cost proportional only to the effective dimension of the data. The result $\tilde{\Sigma}$ of steps (1) and (2) can then be shown to be private and close to (the projection of) Σ up to a Frobenius distance of order $O(d^{3/2}/n\varepsilon)$. The third step is an orthogonal projection onto a convex set containing the ground truth, thus it cannot increase the ℓ_2-distance to Σ, but, remarkably, for small datasets can provably shrink it to $O(\sqrt{d/n\varepsilon})$.

Despite its utility in terms of the Frobenius norm error, this algorithm cannot be expected to be accurate in other Schatten norms, in a strong sense. The dimensionality reduction step rely on a Johnson-Lindestrauss transform, but such mappings are known *not* to exist for spaces that are far from being Euclidean and, more specifically, for p-norm space with $p \in [1, \infty] \setminus \{2\}$ (Brinkman and Charikar (2005); Lee et al. (2005); Johnson and Naor (2010)).

The perturbation mechanism. To bypass this inherent limitation, we do not attempt to explicitly embed the dataset in a low-dimensional space and instead carefully craft the additive noise to inject. Since adjacent covariance matrices Σ, Σ' satisfy $\|\Sigma - \Sigma'\|_* \leqslant 2/n$, the K-norm mechanism (Hardt and Talwar (2010)) with density $p(Z)$ proportional to $\exp(-\varepsilon n\|Z\|_*/2)$ is ε-DP. Note that it is possible to sample from such distribution efficiently, leveraging any efficient weak separation oracle for the nuclear norm[3]. Hence it only remains to bound Schatten norms of Z sampled from this distribution, and in fact, it is enough to bound $\|Z\|$ by $O(\frac{d}{\varepsilon n})$ since $\|Z\|_{\mathrm{S}_p} \leqslant \|Z\| \cdot d^{1/p}$.

Consider a singular value decomposition of $Z = UDV^\top$. By rotational symmetry, U, D, V are mutually independent. For real $d \times d$ matrices, the Lebesgue measure dZ factors under this map

$$\Phi : (U, \sigma, V) \mapsto U \operatorname{diag}(\sigma) V^\top, \qquad \sigma = (\sigma_1, \ldots, \sigma_d) \in \mathbb{R}^d_{\geqslant 0},$$

as[4]

$$dZ = C_d \prod_{1 \leqslant i < j \leqslant d} |\sigma_i^2 - \sigma_j^2| \, d\sigma \;\; d\mu_{\mathrm{Haar}}(U) \;\; d\mu_{\mathrm{Haar}}(V),$$

for a (dimension-dependent) constant $C_d > 0$, where $d\sigma$ is the Lebesgue measure on $\mathbb{R}^d$, and $d\mu_{\mathrm{Haar}}$ denotes the Haar probability measure on the orthogonal group.

Let $R = \sum_i \sigma_i$ be the *radial* component and $w_i = \sigma_i/R$ be *relative weights* of the singular values. Then $d\sigma = R^{d-1}\, dR\, d\lambda_\Delta(w)$, where $d\lambda_\Delta(w)$ is the $(d-1)$-dimensional Lebesgue measure on the simplex $\sum_{i=1}^d w_i = 1\,, w_i > 0$. Substituting $\sigma_i = Rw_i$ yields the joint density

$$p(U, R, w, V) \propto \underbrace{e^{-R/\rho}\, R^{d-1}\, R^{d\cdot(d-1)}}_{=: f(R)} \cdot \underbrace{\prod_{i<j} |w_i^2 - w_j^2|}_{=: g(w)} \cdot d\mu_{\mathrm{Haar}}(U)\, d\mu_{\mathrm{Haar}}(V)\, dR\, d\lambda_\Delta(w)\,.$$

3. See Lemma A.2. in Hardt and Talwar (2010) for more details.

4. This factorization is standard, and can be found, in particular, in Lemma 1.5.3 from Chikuse (2003).

This implies that to show the desired concentration of the spectral norm, it is enough to study the product of two independent random variables: R and $\max_{i\in[d]} w_i$. It is well-known that the distribution of R is[5] $\mathrm{Gamma}(d^2, \rho)$, and it is well-concentrated around ρd^2 (recall that $\rho = 2/(\varepsilon n)$). The result would then follow if we could show that with high probability $\max_{i\in[d]} w_i \leqslant 1/d + \tilde{O}(1/d^{3/2})$ as by union bound $\|Z\| \leqslant O(d^2/\varepsilon \cdot n) \cdot O(1/d) \leqslant O(\frac{d}{\varepsilon n})$.

Bounding the singular values of Z. The analysis of the maximal entry of w turns out to be fairly delicate. Note that w is a vector with density proportional to $e^{-V(w)} := \prod_{i<j} |w_i^2 - w_j^2|$, restricted to the d-dimensional simplex $\sum_i w_i = 1$, $w_i > 0$. Our approach consists of showing that this distribution is κ-strongly log-concave. This allows us to leverage the entry-wise sub-Gaussian concentration:

$$\mathbb{P}(\,|w_i - \mathbb{E}\, w_i| \geqslant t\,) \;\leqslant\; 2\exp\!\big(-c\kappa\, t^2\big) \tag{2.1}$$

for some absolute constant $c > 0$ and all $t > 0$. The desired concentration bound thus can be achieved if the curvature parameter of $V(w)$ satisfies $\kappa \geqslant \tilde{\Omega}(d^3)$. To this end, we first observe that this distribution is $\Omega(d)$-strongly log-concave. Indeed, since

$$V(w) \;=\; -\sum_{1\leqslant i<j\leqslant d} \log|w_i - w_j| \;-\; \sum_{1\leqslant i<j\leqslant d} \log|w_i + w_j|,$$

we get

$$\nabla^2 V(w) \;=\; \sum_{i<j} \frac{(e_i - e_j)(e_i - e_j)^\top}{(w_i - w_j)^2} \;+\; \sum_{i<j} \frac{(e_i + e_j)(e_i + e_j)^\top}{(w_i + w_j)^2}\,.$$

Now $0 \leqslant w_i + w_j \leqslant 1$ implies $1/(w_i + w_j)^2 \geqslant 1$, and hence

$$\sum_{i<j} \frac{(e_i + e_j)(e_i + e_j)^\top}{(w_i + w_j)^2} \succeq \sum_{i<j} (e_i + e_j)(e_i + e_j)^\top\,.$$

Note that each non-diagonal entry of $\sum_{i<j}(e_i + e_j)(e_i + e_j)^\top$ is equal to 1, while each diagonal entry is equal to $d-1$, therefore

$$\sum_{i<j} (e_i + e_j)(e_i + e_j)^\top = (d-2)\mathrm{I}_d + \mathbf{1}\mathbf{1}^\top\,,$$

where $\mathbf{1}$ denotes the all-ones vector. For $d \geqslant 3$, $(d-2)\mathrm{I}_d \succeq \Omega(d)\cdot \mathrm{I}_d$, so the distribution is strongly log-concave[6]. Since $\sum_{i=1}^d w_i = 1$, by permutation symmetry $\mathbb{E}\, w_i = 1/d$ for each $i \in [d]$, and by union bound we get $\max_{i\in[d]} w_i \leqslant \tilde{O}(1/\sqrt{d})$ with high probability. Unfortunately, this bound only yields $\|Z\| \leqslant \tilde{O}(d^{3/2}/\varepsilon n)$, which is significantly worse than what we aimed for.

To derive a sharper strong convexity bound, we condition w on the high-probability event $\max_{i\in[d]} w_i \leqslant \tilde{O}(1/\sqrt{d})$. Under this conditioning, for all $i, j \in [d]$, $w_i + w_j \leqslant \tilde{O}(1/\sqrt{d})$ and $1/(w_i + w_j)^2 \geqslant \tilde{\Omega}(d)$. Crucially, this implies $\tilde{\Omega}(d^2)$-strong convexity of $V(w)$ on the set $\max_{i\in[d]} w_i \leqslant \tilde{O}(1/\sqrt{d})$, which, in turn, implies that $\max_{i\in[d]} w_i \leqslant \tilde{O}(1/d)$ with high probability.

5. See, for example, Remark 4.2 in Hardt and Talwar (2010).
6. Since we would like to show a high-probability bound (as $d \to \infty$), we can always assume $d \geqslant 3$. However, if $d = 2$, it is also true that $\nabla^2 V(w) \succeq \Omega(1)\cdot \mathrm{I}_2$.

This bound yields $\|Z\| \leqslant \tilde{O}(d/\varepsilon n)$, which is slightly worse than the $O(d/\varepsilon n)$ we need. To get this (optimal) bound, we can repeat the previous argument with the new set: we condition w on the (high-probability) event $\max_{i\in[d]} w_i \leqslant \tilde{O}(1/d)$, and get $\tilde{\Omega}(d^3)$-strong convexity of $V(w)$ on this set, implying now that with high probability $\max_{i\in[d]} w_i \leqslant 1/d + \tilde{O}(1/d^{3/2})$, and leading to the desired bound on $\|Z\|$.

The projection mechanism. The perturbation mechanism in Algorithm 1 yields optimal accuracy for every p-norm, provided the dataset is sufficiently large. Our approach to handle small datasets is to orthogonally project $\Sigma + Z$ back to the nuclear norm ball of radius $r > 0$. Note that, this orthogonal projection may distort p-norms for $p \neq 2$ and thus destroys the guarantees we carefully obtained. Nevertheless, if $r \geqslant \|\Sigma\|_{\mathrm{S}_1} = \operatorname{Tr}\Sigma$, it cannot increase the error in Frobenius norm: $\|\tilde{\Sigma} - \Sigma\|_{\mathrm{F}} \leqslant \|\hat{\Sigma} - \Sigma\|_{\mathrm{F}}$ (since the nuclear norm ball is convex).

Furthermore, a standard strong convexity argument (e.g. the one used in Lemma 5.2 in Cohen-Addad et al. (2024)) implies that if $r \geqslant \operatorname{Tr}\Sigma$, then the orthogonal projection $\Pi(\Sigma + Z)$ satisfies

$$\begin{aligned}
\|\Sigma - \Pi(\Sigma+Z)\|_{\mathrm{F}}^2 &\leqslant O\left(\langle Z, \Sigma - \Pi(\Sigma+Z)\rangle\right) \\
&\leqslant \|Z\| \cdot O\left(\|\Sigma\|_* + \|\Pi(\Sigma+Z)\|_*\right) \\
&\leqslant O(d/n\varepsilon)\left(\|\Sigma\|_* + \|\Pi(\Sigma+Z)\|_*\right) \\
&\leqslant O(dr/n\varepsilon).
\end{aligned}$$

where in the before-last step we used the fact that the spectral norm of Z concentrates around $\Theta(d/n\varepsilon)$ with high probability, and in the last step the radius of the nuclear norm ball. The naive choice $r = 1$ would recover the result of Nikolov (2023). However, this choice can be wasteful. We instead use an ε-DP Laplace estimator of $\operatorname{Tr}\Sigma$. Note that with probability at least $1-\beta$, $r \leqslant O\left(\operatorname{Tr}(\Sigma) + \log(1/\beta)/(n\varepsilon)\right)$, and we get the bound as in Theorem 2.

Finally, if $r < \operatorname{Tr}(\Sigma)$, it means that the magnitude of Laplace noise is larger than $\operatorname{Tr}(\Sigma)$. With probability $1-\beta$, it happens only if $\log(1/\beta)/(n\varepsilon) > \Omega(\operatorname{Tr}(\Sigma))$. In this case, since $r < \operatorname{Tr}(\Sigma) < O\left(\log(1/\beta)/(n\varepsilon)\right)$, both $\tilde{\Sigma}$ and Σ are in the same nuclear norm ball of radius $O\left(\log(1/\beta)/(n\varepsilon)\right)$, so the nuclear (and, hence, also Frobenius) distance between them is bounded by $O\left(\log(1/\beta)/(n\varepsilon)\right) \leqslant O\left(\sqrt{d}/(n\varepsilon)\right)$ with probability at least $1 - \exp(-\Omega(\sqrt{d}))$, and hence the bound in in Theorem 2 is satisfied also in this case.

3. Future work

It is natural to ask whether a generalization of both Theorem 1 and Theorem 2 is possible. Concretely, it would be interesting to determine if there is an estimator satisfying

$$\left\|\tilde{\Sigma} - \Sigma\right\|_{\mathrm{S}_p} \leqslant \tilde{O}\left(1\right) \cdot \min\left\{\frac{d^{1+1/p}}{\varepsilon n},\ d^{1/p}\sqrt{\frac{\operatorname{Tr}\Sigma}{\varepsilon n}} + \frac{d^{1/p}}{\varepsilon n}\right\}.$$

In addition, it would be interesting to see complementary lower bounds. In particular, currently even in the large sample regime $n \geqslant d^2/\varepsilon$ it is only known that our result is optimal for $p \geqslant 2$, but it is not known for $p < 2$. And in the small-sample regime, the picture is even less clear.

4. Analysis of the additive mechanism

In this section we prove Theorem 1.

Setup. Let $A, A' \in \mathbb{R}^{d\times d}$, $\Delta > 0$. We say that A and A' are nuclear-norm adjacent if

$$\|A - A'\|_* \leqslant \Delta\,.$$

Let

$$\mathcal{M}_\rho(A) = A + Z_\rho, \qquad \mathrm{density}(Z_\rho) \propto \exp\left(-\|Z\|_*/\rho\right).$$

Write $\|\cdot\|_{\mathrm{S}_p}$ for Schatten-p norms and denote the singular values of X by $\sigma(X) \in \mathbb{R}^d_+$.

Theorem 3 *If $\rho = \Delta/\varepsilon$, then $\mathcal{M}_\rho$ is ε-differentially private.*

Proof For any y and adjacent $A \sim A'$,

$$\frac{p_{\mathcal{M}_\rho(A)}(y)}{p_{\mathcal{M}_\rho(A')}(y)} = \exp\left(\frac{\|y - A'\|_* - \|y - A\|_*}{\sigma}\right) \leqslant \exp\left(\frac{\|A - A'\|_*}{\sigma}\right) \leqslant e^{\varepsilon}.$$

■

The next Lemms is a well-known fact about the radial component.

Lemma 4 (Nuclear-radial factorization) *(Remark 4.2 in Hardt and Talwar (2010)) Let Z have density $p_\rho(Z) \propto \exp\left(-\|Z\|_*/\rho\right)$. Then*

$$Z = R\,\Theta, \qquad R := \|Z\|_*, \quad \Theta := Z/\|Z\|_* \in \mathbb{S}_* := \{A : \|A\|_* = 1\},$$

with $R \perp \Theta$ and $R \sim \mathrm{Gamma}\left(\textit{shape} = d^2,\ \textit{scale} = \sigma\right)$. In particular

$$\mathbb{E}\,R = d^2\sigma, \qquad \mathbb{E}\,R^2 = d^2\left(d^2+1\right)\sigma^2.$$

Now let us study the distribution of the singular spectrum of Z.

Lemma 5 *Let $Z \in \mathbb{R}^{d\times d}$ be distributed with density*

$$p_\rho(Z) \ \propto\ \exp\left(-\|Z\|_*/\rho\right), \qquad \sigma > 0,$$

and write an SVD $Z = U\,\mathrm{diag}(\sigma_1,\dots,\sigma_d)\,V^\top$ with $\sigma_i \geqslant 0$, $U, V \in \mathrm{O}(d)$. Define the nuclear radius $R := \sum_{i=1}^d \sigma_i = \|Z\|_*$ *and the* normalized singular values

$$w_i \ :=\ \sigma_i/R, \qquad i = 1,\dots,d,$$

so that $w := (w_1,\dots,w_d) \in \Delta_{d-1} := \{w \in \mathbb{R}^d_{\geqslant 0} :\ \sum_i w_i = 1\}$, and set

$$\Theta \ :=\ \frac{Z}{\|Z\|_*} \ =\ U\,\mathrm{diag}(w_1,\dots,w_d)\,V^\top.$$

Then:

(i) R is independent of (U, V, w), and $R \sim \mathrm{Gamma}\left(\textit{shape} = d^2,\ \textit{scale} = \rho\right)$.

(ii) Conditionally on $(R = r)$, the pair (U, V) is independent of w and is Haar distributed on $\mathrm{O}(d) \times \mathrm{O}(d)$.

(iii) The conditional law of w given $R = r$ is independent of r *and has density on Δ_{d-1} (with respect to the $(d-1)$–dimensional Lebesgue measure on the simplex) proportional to*

$$f_\Delta(w) \;\propto\; \prod_{1 \leqslant i < j \leqslant d} |w_i^2 - w_j^2| \,.$$

Consequently, the unordered singular values of Θ have joint density proportional to $f_\Delta(w)$ on Δ_{d-1}.

Proof For real $d \times d$ matrices, Lebesgue measure dZ factors under the SVD map

$$\Phi:\; (U, \sigma, V) \;\mapsto\; U\,\mathrm{diag}(\sigma)\,V^\top, \qquad \sigma = (\sigma_1, \dots, \sigma_d) \in \mathbb{R}^d_{\geqslant 0},$$

as

$$dZ \;=\; C_d \underbrace{\prod_{1 \leqslant i < j \leqslant d} |\sigma_i^2 - \sigma_j^2|}_{\text{Vandermonde in } \sigma^2} \; d\sigma \;\; d\mu_{\mathrm{Haar}}(U) \;\; d\mu_{\mathrm{Haar}}(V),$$

for a constant $C_d > 0$, where $d\sigma$ is Lebesgue measure on $\mathbb{R}^d$ restricted to $\sigma_i \geqslant 0$ and $d\mu_{\mathrm{Haar}}$ denotes Haar probability measure on $\mathrm{O}(d)$ (see Lemma 1.5.3 from Chikuse (2003)).

With $dZ \propto \exp\left(-\sum_i \sigma_i/\rho\right)$, the joint density of (U, σ, V) is therefore

$$\tilde{p}(U, \sigma, V) \;\propto\; \exp\left(-\tfrac{1}{\rho}\sum_{i=1}^{d} \sigma_i\right) \prod_{i<j} |\sigma_i^2 - \sigma_j^2|.$$

Let $R = \sum_i \sigma_i$ and $w_i = \sigma_i/R$; then $w \in \Delta_{d-1}$ and the Jacobian is $d\sigma = R^{d-1}\,dR\,d\lambda_\Delta(w)$. Moreover,

$$\prod_{i<j} |\sigma_i^2 - \sigma_j^2| = \prod_{i<j} |(Rw_i)^2 - (Rw_j)^2| = R^{d\cdot(d-1)} \prod_{i<j} |w_i^2 - w_j^2|.$$

Substituting $\sigma_i = Rw_i$ yields the joint density

$$p(U, R, w, V) \;\propto\; \underbrace{e^{-R/\rho}\,R^{d-1}\,R^{d\cdot(d-1)}}_{=e^{-R/\rho}\,R^{d^2-1}} \cdot \underbrace{\prod_{i<j} |w_i^2 - w_j^2|}_{=:F(w)} \cdot d\mu_{\mathrm{Haar}}(U)\,d\mu_{\mathrm{Haar}}(V)\,dR\,d\lambda_\Delta(w).$$

This factorizes into a function of R times a function of (U, w, V), so:

$$R \;\perp\; (U, V, w), \qquad f_R(R) \;\propto\; e^{-R/\rho} R^{d^2-1},$$

i.e. $R \sim \mathrm{Gamma}(d^2, \rho)$, and given R the pair (U, V) is Haar and independent of w, while w has density proportional to $F(w)$ on Δ_{d-1}, independent of R.

Finally, $\Theta = Z/\|Z\|_* = U\,\mathrm{diag}(w)\,V^\top$, so the singular values of Θ equal w. ■

Now we are ready to restate and prove Theorem 1. Note that it is enough to prove it for spectral norm, for other norm the result follows from the general ℓ_p-ℓ_∞-norm inequality.

Theorem 6 *Let $Z \in \mathbb{R}^{d\times d}$ have density proportional to $\exp(-\|Z\|_*/\rho)$. Write $R := \|Z\|_*$, $\Theta := Z/\|Z\|_*$ so that $\|\Theta\|_* = 1$, and let $w = (w_1, \dots, w_d)$ be the (unordered) singular values of Θ normalized so that $\sum_i w_i = 1$. There exist an absolute constant $C > 0$ such that with probability at least $1 - 2\exp(-d^{1/3}/C)$,*

$$\|Z\| \leqslant 1.1 \cdot \rho d\,.$$

Proof The density of w is proportional to $e^{-V(w)}$ with

$$V(w) \;=\; -\sum_{1\leqslant i<j\leqslant d} \log|w_i - w_j| \;-\; \sum_{1\leqslant i<j\leqslant d} \log|w_i + w_j|,$$

restricted to the simplex $\sum_i w_i = 1$, $w_i > 0$. On the tangent subspace $\{u : \sum_i u_i = 0\}$,

$$\nabla^2 V(w) \;=\; \sum_{i<j} \frac{(e_i - e_j)(e_i - e_j)^\top}{(w_i - w_j)^2} \;+\; \sum_{i<j} \frac{(e_i + e_j)(e_i + e_j)^\top}{(w_i + w_j)^2} \;\succeq\; (d-2)\,\mathrm{I}_d.$$

Since for $d = 1$ and $d = 2$ the statement of the theorem is always true (for large enough constant C), further we consider the case $d \geqslant 3$. In this case $(d-2)\,\mathrm{I}_d \succeq 0.1d$.

If $\max_i w_i \leqslant \alpha$ then $w_i + w_j \leqslant 2\alpha$ for all $i \neq j$, hence

$$\nabla^2 V(w)\big|_{\{\sum_i u_i = 0\}} \;\succeq\; \frac{0.1d}{(2\alpha)^2}\,\mathrm{I}_d \;\text{ whenever }\; \max_i w_i \leqslant \alpha. \tag{4.1}$$

Since w is strongly log-concave, for each L–Lipschitz F and each $t > 0$,

$$\mathbb{P}\Bigg(|F(w) - \mathbb{E}\,F(w)| \geqslant t \,\Bigg|\, \max_i w_i \leqslant \alpha \Bigg) \;\leqslant\; 2\exp\Bigg(-c\,\frac{d\,t^2}{(2\alpha)^2\,L^2} \Bigg). \tag{4.2}$$

Step 0 (Bound via global curvature). On the full simplex $\{w_i \geqslant 0,\ \sum_i w_i = 1\}$ the Hessian on the tangent subspace $\mathsf{T} := \{u : \sum_i u_i = 0\}$ satisfies

$$\nabla^2 V(w)\big|_{\mathsf{T}} \;\succeq\; 0.1d \cdot \mathrm{I}_d \qquad \text{for all } w.$$

Let $F(w) := w_i$. Then F is 1-Lipschitz on T, and by strong log–concavity with curvature $0.1d$ we have, for some absolute constant c_1 and all $t \geqslant 0$,

$$\mathbb{P}\big(|F(w) - \mathbb{E}\,F(w)| \geqslant t\big) \;\leqslant\; 2\exp\Big(-c_1 d\,t^2 \Big).$$

Note that by permutational symmetry and $\sum_i w_i = 1$, $\mathbb{E}\,w_i = 1/d$. Therefore, by union bound, for each $\beta \in (0,1)$, with probability at least $1-\beta$,

$$\max_i w_i \leqslant \;\frac{1}{d}\; + \sqrt{\frac{\log(d/\beta)}{c_1 d}}\,.$$

Define the set

$$\mathcal{S}_1 := \Big\{ w :\ \max_i w_i \leqslant \alpha_1 \Big\}, \qquad \alpha_1 := C_1 \sqrt{\frac{\log(1/\beta)}{d}}$$

for some large enough absolute constant $C_1 > 0$. Then $\mathbb{P}(w \in \mathcal{S}_1) \geqslant 1 - \beta/4$.

Step 1 (Boost curvature to $\tilde{\Theta}(d^2)$ and push $\|w\|_\infty$ to $\tilde{O}(1/d)$). On $\mathcal{S}_1$, (4.1) gives curvature $\kappa_1 \geqslant 0.1d/(2\alpha_1)^2 \geqslant \Omega(d^2/\log(1/\beta))$. Apply (4.2) to the 1–Lipschitz linear functionals $w \mapsto w_i$ and union–bound over $i = 1, \ldots, d$:

$$\mathbb{P}\left(\max_{1 \leqslant i \leqslant d} (w_i - \mathbb{E}\, w_i) \;\geqslant\; t \;\middle|\; w \in \mathcal{S}_1 \right) \;\leqslant\; 2d\, \exp(-c_2\, d^2 t^2 / \log(1/\beta)).$$

Choose $t := C_2 \log(d/\beta)/d$ with C_2 large enough absolute constant, so that the right-hand side is $\leqslant \beta/4$. Thus, with probability at least $1 - \beta/2$,

$$\max_i w_i \;\leqslant\; \alpha_2 \;:=\; \frac{1}{d} \;+\; \frac{C_2 \log(d/\beta)}{d} \;\leqslant\; \frac{C_2' \log(d/\beta)}{d}\,. \tag{4.3}$$

Step 2 (Boost curvature to $\tilde{\Theta}(d^3)$ and squeeze $\|w\|_\infty$ to $1/d$ up to $(\log d)/d^{3/2}$). Define the tighter cap $\mathcal{S}_2 := \{w : \max_i w_i \leqslant \alpha_2\}$ with α_2 from (4.3). On $\mathcal{S}_2$, (4.1) yields curvature

$$\kappa_2 \;\geqslant\; \Omega\left(\frac{d^3}{\log^2(d/\beta)} \right).$$

Applying (4.2) as before and union–bounding over i gives, for all $t > 0$,

$$\mathbb{P}\left(\max_{1 \leqslant i \leqslant d} (w_i - 1/d) \;\geqslant\; t \;\middle|\; w \in \mathcal{S}_2 \right) \;\leqslant\; 2d\, \exp\left(-c\, \frac{d^3}{\log^2(d/\beta)}\, t^2 \right).$$

Set $t := C_3 \left(\frac{\log(d/\beta)}{d} \right)^{3/2}$; then $(d^3/\log(d/\beta))\, t^2 \;=\; C_3^2 \log(d/\beta)$. Choosing C_3 large enough so that $2d\, e^{-cC_3^2 \log(d/\beta)} \leqslant \beta/4$, we get with probability at least $1 - 3\beta/4$,

$$\max_i w_i \;\leqslant\; \frac{1}{d} \;+ C_3 \left(\frac{\log(d/\beta)}{d} \right)^{3/2}. \tag{4.4}$$

Radial concentration. Write

$$Z = R\,\Theta, \qquad R := \|Z\|_*, \qquad \Theta := Z/\|Z\|_* \in \mathbb{S}_* := \{A : \|A\|_* = 1\}.$$

As shown in Lemma 4, $R \perp \Theta$ and $R \sim \mathrm{Gamma}\left(d^2, \rho\right)$, i.e. $R = \rho \sum_{i=1}^{d^2} E_i$ with i.i.d. $E_i \sim \mathrm{Exp}(1)$. By Bernstein inequality for the sum of iid exponential distributions, for all $\beta \in (0, 1/e)$, with probability at least $1 - \beta$,

$$d^2\rho - C_4 d\rho\sqrt{\log(1/\beta)} - C_4\rho\log(1/\beta) \leqslant R \leqslant d^2\rho + C_4 d\rho\sqrt{\log(1/\beta)} + C_4\rho\log(1/\beta)\,,$$

where $C_4 > 0$ is an absolute constant.

Putting everything together. (4.4) and (4) imply that with probability at least $1 - \beta$,

$$\|Z\| \;\leqslant\; \left(d^2\rho + C_4\,\rho(d\sqrt{\log(1/\beta)} + \log(1/\beta)) \right) \left(\frac{1}{d} + \frac{C_3\, \log^{3/2}(d/\beta)}{d^{3/2}} \right).$$

For $\beta = \exp(-d^{1/3}/C)$ where $C > 0$ is large enough absolute constant, we get the desired bound

$$\|Z\| \leqslant 1.1 \cdot \rho d\,.$$

■

References

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

Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in l1. *Journal of the ACM (JACM)*, 52 (5):766–788, 2005.

Yasuko Chikuse. *Statistics on special manifolds*, volume 174. Springer Science & Business Media, 2003.

Vincent Cohen-Addad, Tommaso d'Orsi, Alessandro Epasto, Vahab Mirrokni, and Peilin Zhong. Perturb-and-project: differentially private similarities and marginals. In *Proceedings of the 41st International Conference on Machine Learning*, pages 9161–9179, 2024.

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

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53:650–673, 2015.

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.

William B Johnson. Extensions of lipschitz mappings into a hilbert space. *Contemp. Math.*, 26:189–206, 1984.

William B Johnson and Assaf Naor. The johnson–lindenstrauss lemma almost characterizes hilbert space, but not quite. *Discrete & Computational Geometry*, 43(3):542–553, 2010.

Shiva Prasad Kasiviswanathan, Mark Rudelson, Adam Smith, and Jonathan Ullman. The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 775–784, 2010.

James R Lee, Manor Mendel, and Assaf Naor. Metric structures in l1: dimension, snowflakes, and average distortion. *European Journal of Combinatorics*, 26(8):1180–1190, 2005.

Aleksandar Nikolov. Private query release via the johnson-lindenstrauss transform. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4982–5002. SIAM, 2023.

Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 351–360, 2013.