Proceedings of Machine Learning Research vol 291:1–49, 2025 38th Annual Conference on Learning Theory

Optimal Differentially Private Sampling of Unbounded Gaussians

Valentio Iverson VIVERSON@UWATERLOO.CA
Cheriton School of Computer Science, University of Waterloo
Gautam Kamath G@CSAIL.MIT.EDU
Cheriton School of Computer Science, University of Waterloo
Argyris Mouzakis AMOUZAKI@UWATERLOO.CA
Cheriton School of Computer Science, University of Waterloo

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We provide the first $\widetilde{\mathcal{O}}(d)$-sample algorithm for sampling from unbounded Gaussian distributions under the constraint of (ε, δ)-differential privacy. This is a quadratic improvement over previous results for the same problem, settling an open question of Ghazi et al. (2023).

Keywords: Differential Privacy, Private Sampling, Gaussian Distribution

1. Introduction

Consider the following problem: given a sample from a probability distribution, output a sample from the same probability distribution. As stated, this problem is trivial: one can simply output the sample received. However, when *privacy* is a concern, this is clearly inappropriate, as the sample may contain sensitive personal information. To address this issue, one may turn to the concept of *Differential Privacy* (DP) (Dwork et al., 2006), a rigorous notion of privacy that ensures that the output of a procedure does not reveal too much information about any individual's data.

One well-studied option is to privately *learn* the underlying distribution. That is, given samples from a distribution, output a (differentially private) estimate of it. After this estimated distribution is released, one may freely output any desired number of samples from the distribution without any further privacy concerns (appealing to the post-processing property of differential privacy).

However, this strategy may be costly. Indeed, in both the private and non-private setting, learning a distribution is clearly a much harder problem than simply outputting a single sample from it. In an influential paper, Raskhodnikova, Sivakumar, Smith, and Swanberg (Raskhodnikova et al., 2021) introduced the problem of privately *sampling* from a distribution. Given a collection of samples from a distribution, privately output a single sample from (approximately) the same underlying distribution. The authors provided upper and lower bounds for private sampling on discrete distributions and binary product distributions, showing that, in various cases, private sampling is either no easier or dramatically easier than private learning.

We note that this problem has immediate implications for machine learning in the context of private generative models. Rather than privately training a generative model (which may be data-intensive), one could instead use private sampling if only a limited number of samples (e.g., prompt completions) are needed.

Subsequently, Ghazi, Hu, Kumar, and Manurangsi Ghazi et al. (2023) studied private sampling of Gaussian distributions. Technically speaking, the most interesting case is when we are sampling from distributions with unknown covariance matrix Σ. The authors provide two algorithms for privately sampling from d-dimensional Gaussians subject to (ε, δ)-DP:

1. If $\mathbb{I} \preceq \Sigma \preceq K\mathbb{I}$ (i.e., the covariance is bounded), they give an algorithm with sample complexity $\widetilde{\mathcal{O}}\left(dK^2/\varepsilon\right)$;
2. Without any assumptions (i.e., a potentially unbounded covariance), they give an algorithm with sample complexity $\widetilde{\mathcal{O}}\left(d^2/\varepsilon\right)$.

They also prove a sample complexity lower bound of $\widetilde{\Omega}(d/\varepsilon)$ (when $\mathbb{I} \preceq \Sigma \preceq 2\mathbb{I}$), thus establishing that, when the covariance is bounded up to constant factors, their algorithm is near-optimal. However, this leaves open the case $K = \omega(1)$. For example, if $K = \Omega\left(\sqrt{d}\right)$, the best known upper bound of $\widetilde{\mathcal{O}}\left(d^2\right)$ is off from the lower bound of $\widetilde{\Omega}\left(d\right)$ by a quadratic factor. Is it always a choice between an algorithm whose sample complexity depends linearly on the dimension d (but also suffers from a dependence on the "range" of the covariance K), and one whose sample complexity is quadratic in d? Or is there an $\widetilde{\mathcal{O}}\left(d\right)$-sample algorithm, regardless of the magnitude of K?

1.1. Results and Techniques

Our main result is an algorithm with $\widetilde{\mathcal{O}}\left(d\right)$ sample complexity, for private sampling from Gaussians with arbitrary (i.e., potentially unbounded) covariance.

Theorem 1 *There exists an (ε, δ)-DP algorithm which, given $\mathcal{O}\left(\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}\left(d + \log\left(\frac{\log\left(\frac{1}{\delta}\right)}{\alpha\varepsilon}\right)\right)\right)$, samples drawn i.i.d. from a distribution $\mathcal{N}\left(\mu, \Sigma\right)$, outputs a sample from a distribution which is α-close in TV-distance to $\mathcal{N}\left(\mu, \Sigma\right)$.*

Observe that, in terms of its dependence on d in the sample complexity, this nearly matches the $\Omega\left(d/(\varepsilon\sqrt{\log d})\right)$ lower bound of Ghazi et al. (2023). Additionally, it is quadratically less than the sample complexity of privately *learning* Gaussians in the same setting (Kamath et al., 2019; Aden-Ali et al., 2021; Kamath et al., 2022b; Ashtiani and Liaw, 2022; Kothari et al., 2022; Kamath et al., 2022a), which is $\widetilde{\Theta}\left(d^2/\alpha^2 + d^2/(\alpha\varepsilon)\right)$.

We give a brief technical overview of our approach. As a starting point, we use the algorithm of Ghazi et al. (2023) which privately samples from a Gaussian with covariance $\mathbb{I} \preceq \Sigma \preceq K\mathbb{I}$ using $\widetilde{\mathcal{O}}\left(dK^2\right)$ samples. Recall the following basic property of Gaussians: given n samples $X_1, \dots, X_n$ from $\mathcal{N}\left(\mu, \Sigma\right)$, their average $1/n \sum_{i\in[n]} X_i$ is distributed as $\mathcal{N}\left(\mu, 1/n\Sigma\right)$. Simply outputting this average is insufficient for two reasons: i) the variance of the resulting distribution is too small by a factor of n, and ii) the algorithm is not differentially private. Ideally, one could solve both problems at once by adding $\mathcal{N}\left(0, \frac{n-1}{n}\Sigma\right)$ noise, but recall that Σ itself is unknown. As a proxy for this, Ghazi et al. (2023) uses the *samples themselves* to sample from this distribution. Another basic property of Gaussians is 2-stability: it implies that, if $z \in \mathbb{R}^n$ is a random unit vector and $X_1, \dots, X_n$ are samples from a Gaussian, then $\sum_{i\in[n]} z_i X_i$ is a sample from the same Gaussian.[1] Thus, letting $Y_i := 1/\sqrt{2}\left(X_{n+i} - X_{n+m+i}\right) \sim \mathcal{N}\left(0, \Sigma\right)$, the random variable $1/n \sum_{i\in[n]} X_i + \sqrt{1 - 1/n}\sum_{i\in[m]} z_i Y_i$ is distributed as $\mathcal{N}\left(\mu, \Sigma\right)$. To complete the privacy analysis, two additional components are needed: a) "clipping" the datapoints to a predetermined range (based on the upper bound $\Sigma \preceq K\mathbb{I}$) and b) a "Propose-Test-Release" (PTR) step (Dwork and Lei, 2009) to ensure that the collection of Y_i's have large (empirical) variance in all directions (which is implied with high probability given the lower bound $\mathbb{I} \preceq \Sigma$). Unfortunately, due to these two steps, the resulting sample complexity is $\Omega\left(dK^2\right)$.

1. See Lemma 35 for a proof of this property.

We modify this framework, turning to recent results of Brown, Hopkins, and Smith Brown et al. (2023).[2] Their main result is an $\widetilde{\mathcal{O}}(d)$ sample algorithm for covariance-aware private mean estimation – that is, mean estimation where the error in the estimate may scale proportional to the underlying distribution's variance in that direction. To do this, they design algorithms STABLEMEAN and STABLECOV. Focusing on STABLEMEAN (STABLECOV is similar), this algorithm takes in a dataset $X_1, \ldots, X_n$, and outputs a vector of weights $v_1, \ldots, v_n$ and a score s.[3] This algorithm has two key properties: a) if the data is truly Gaussian, then the weights will be uniform (i.e., $v_i = 1/n$ for all i) with high probability, and b) if the score s is sufficiently small (which can be verified using a PTR step), then the weighted empirical mean $\sum_{i\in[n]} v_i X_i$ will have bounded sensitivity when modifying one datapoint. We highlight that bounded sensitivity is a property of the overall sum, and individual summands may change dramatically if a single datapoint is changed – this is in contrast to clipping-based strategies where individual summands also have bounded sensitivity. These elegant primitives allow one to perform private mean estimation without any explicit clipping.

We employ these primitives in the framework of Ghazi et al. (2023): we output the quantity $\sum_{i\in[n]} v_i X_i + \sqrt{m\left(1-1/n\right)} \sum_{i\in[m]} z_i \sqrt{w_i} Y_i$, where the v_i's and w_i's are the weights outputtted by STABLEMEAN and STABLECOV, respectively. That is, we incorporate these weights to bound the sensitivity, as an alternative to clipping. We also require a preceding PTR step in order to ensure that the scores output by STABLEMEAN and STABLECOV are both sufficiently small. To perform the utility analysis, it suffices to show that the weights are uniform with high probability, and the analysis is similar to before. The privacy analysis however requires newfound care.

We illustrate in the univariate case (i.e., $Y_i \in \mathbb{R}$), focusing exclusively on the term $\sum_{i\in[m]} z_i U_i$, where we use U_i to denote $\sqrt{w_i} Y_i$. We accordingly define U_i', which is defined with respect to a neighboring dataset $Y_1', \ldots, Y_n'$. Ghazi et al. (2023) previously used a coupling-based argument to reduce to examining only the term involving the Y_i that differs (i.e., the datapoint such that $Y_i \neq Y_i'$). But, as mentioned before, changing a single Y_i may result in changes to *all* U_i's. Since the stability property of STABLECOV tells us that the empirical variance with these weights has bounded sensitivity, all we know is that $\sum_{i\in[m]} U_i^2 \in \left(1 \pm \frac{1}{m}\right) \sum_{i\in[m]} U_i'^2$. Our goal is to argue that, for fixed U and U', $\langle z, U\rangle$ is similar in distribution to $\langle z, U'\rangle$, where all the randomness is in the choice of the random unit vector $z \in \mathbb{R}^m$. Perhaps surprisingly, if U and U' satisfy the aforementioned bounded sensitivity property, then this suffices to prove the desired statement. To this end, it is convenient to think of $U \in \mathbb{R}^m$ as a vector. By rotational symmetry, $\langle z, U\rangle$ is equal in distribution to $\|U\|_2 \langle z, e_1\rangle$, where e_1 is the first standard basis vector, and similarly, $\langle z, U'\rangle$ is equal to $\|U'\|_2 \langle z, e_1\rangle$. It is not too hard to argue that the two transformed distributions are close to each other (as required in Hockey-Stick divergence), as $\langle z, e_1\rangle^2$ can be seen to be a Beta distribution, and the two distributions are scalings of this distribution by similar constants. This exposes the key ideas in the univariate case for the "variance" term of the output, handling the general case requires additional care.

2. A concurrent work to theirs of Kuditipudi, Duchi, and Haque Kuditipudi et al. (2023) obtains a similar result for covariance-aware mean estimation via similar techniques. We focus on Brown et al. (2023) since we directly appeal to their primitives; it is conceivable that methods in Kuditipudi et al. (2023) could have been used instead.

3. In Brown et al. (2023), STABLEMEAN outputs a non-private estimate for the mean of the distribution which is a weighted version of the empirical mean that uses the weight vector v. We have slightly changed the format of the output, since it facilitates aspects of our exposition. For more details on this, we point readers to Section C.

1.2. Related Work

Private sampling is still a relatively nascent topic. In addition to the aforementioned works (Raskhodnikova et al., 2021; Ghazi et al., 2023), Cheu and Nayak Cheu and Nayak (2025) study private *multi-sampling*, which aims to privately draw *multiple independent samples* from a distribution.

The related problem of private distribution learning is a much more mature topic of study. Specifically, a long line of works investigates private learning of Gaussian distributions, e.g., Karwa and Vadhan (2018); Bun et al. (2019); Bun and Steinke (2019); Kamath et al. (2019); Biswas et al. (2020); Du et al. (2020); Cai et al. (2021); Huang et al. (2021); Hopkins et al. (2023); Bie et al. (2022); Ben-David et al. (2023); Aumüller et al. (2023); Alabi et al. (2023); Asi et al. (2023); Liu et al. (2021, 2022); Kamath et al. (2023); Aden-Ali et al. (2021); Kamath et al. (2022b); Ashtiani and Liaw (2022); Kothari et al. (2022); Kamath et al. (2022a); Narayanan (2023); Portella and Harvey (2024). Some works focus on the case where the covariance matrix is potentially unbounded. While packing lower bounds imply that learning is not possible for $(\varepsilon, 0)$-DP (Hardt and Talwar, 2010; Bun et al., 2019; Hopkins et al., 2022), several works successfully show that learning is possible (despite unbounded covariance) under (ε, δ)-DP, including Karwa and Vadhan (2018) in the univariate case, and Aden-Ali et al. (2021); Kamath et al. (2022b); Kothari et al. (2022); Ashtiani and Liaw (2022); Alabi et al. (2023) for the multivariate case. Our work is most related to the aforementioned works on covariance-aware mean estimation (Brown et al., 2023; Kuditipudi et al., 2023), which can accurately estimate the mean with $\widetilde{\mathcal{O}}\left(d\right)$ samples even when the covariance is unknown. That is, we rely heavily on the tools introduced by Brown et al. (2023), hinting at the potential of deeper technical connections between private estimation and private sampling.

As described above, private sampling can be seen as a less cumbersome alternative to private distribution learning, when only a limited number of samples are required. There is an analogous line of work on private prediction (see, e.g., Papernot et al. (2017); Dwork and Feldman (2018); Bassily et al. (2018)), which can be seen as a more lightweight alternative to private PAC-learning, when only a limited number of predictions are required.

A long line of empirical work studies private learning of deep generative models, see, e.g., Xie et al. (2018); Beaulieu-Jones et al. (2019); Cao et al. (2021); Bie et al. (2023); Dockhorn et al. (2023); Harder et al. (2023). Some recent empirical works study private sampling from generative models Lin et al. (2024); Xie et al. (2024), a morally similar problem to ours, though these works generally require having enough data to train an effective non-private model.

An interesting related question in the non-private setting is that of *sample amplification* (Axelrod et al., 2020). Given n samples from a distribution, can one generate $m \gg n$ samples from (approximately) the same distribution? This line of work served as an inspiration for private sampling (Raskhodnikova et al., 2021).

2. Preliminaries

2.1. General Notation and Basic Facts

Linear Algebra Preliminaries. We write $[n] := \{1, \ldots, n\}$ and $[\alpha \pm R] := [\alpha - R, \alpha + R]$. Also, given a vector $v \in \mathbb{R}^d$, v_i refers to its i-th component, while $v_{-i} \in \mathbb{R}^{d-1}$ refers to the vector one gets by removing the i-th component, and $v_{\leqslant i} \in \mathbb{R}^i$ ($v_{>i} \in \mathbb{R}^{d-i}$) describes the vector one gets by keeping (removing) the first i components. The set of all d-dimensional unit vectors is denoted by $\mathbb{S}^{d-1}$, while the set of the standard basis vectors of $\mathbb{R}^d$ is denoted by $\{e_i\}_{i \in [d]}$. For a matrix

$M \in \mathbb{R}^{n \times m}$, M_{ij} describes the element at the i-th row and j-th column, and by *singular values* of M we refer to the square roots of the eigenvalues of the matrix $M^\top M \in \mathbb{R}^{m \times m}$. Furthermore, $\|M\|_2$ is the *spectral norm* (the largest singular value of the matrix), $\|M\|_F$ is the *Frobenius norm* (the square root of the sum of the squares of the singular values), and $\|M\|_{\text{tr}}$ is the *trace norm* (the sum of the singular values). Additionally, for $M \in \mathbb{R}^{d \times d}$, given a positive-definite matrix $\Sigma \in \mathbb{R}^{d \times d}$, the *Mahalanobis norm* of M with respect to Σ is $\|M\|_\Sigma := \left\|\Sigma^{-\frac{1}{2}} M \Sigma^{-\frac{1}{2}}\right\|_F$. Given a pair of symmetric matrices $A, B \in \mathbb{R}^{d \times d}$, we write $A \succeq B$ if and only if $x^\top (A - B) x \geqslant 0, \forall x \in \mathbb{R}^d$. Finally, by $\mathbb{I}_{n \times n}$ we will denote the identity matrix of size $n \times n$, and by $0_{n \times m}$ we will denote the all 0s matrix of size $n \times m$, where the subscripts will be dropped if the size is clear from the context. For additional background on linear algebra, we refer readers to Appendix A.

Probability & Statistics Preliminaries. Given a sample space $\mathcal{X}$, we denote the set of all distributions over $\mathcal{X}$ by $\Delta(\mathcal{X})$. By $\text{supp}(\mathcal{D})$ we denote the *support* of $\mathcal{D}$, i.e., the set over which $\mathcal{D}$ assigns positive density. For any distribution $\mathcal{D} \in \Delta(\mathcal{X})$, $\mathcal{D}^{\otimes n}$ denotes the product distribution over $\mathcal{X}^n$, where each marginal distribution is $\mathcal{D}$. Thus, given a dataset $X := (X_1, \dots, X_n)$ drawn i.i.d. from $\mathcal{D}$, we write $X \sim \mathcal{D}^{\otimes n}$. Whenever using the symbol of probability, we may use a subscript what the randomness is over in cases where it might not be clear from the context, e.g., $\underset{X}{\mathbb{P}}[\cdot]$. Some of the distributions that we will encounter are the Gaussian distribution with mean μ and covariance Σ (denoted by $\mathcal{N}(\mu, \Sigma)$), the uniform distribution over a set $\mathcal{X}$ (denoted by $\mathcal{U}(\mathcal{X})$), and, given a vector $v \sim \mathcal{U}\left(\mathbb{S}^{n-1}\right)$, the distribution of its first i components (denoted by $v_{\leqslant i} \sim \mathcal{U}\left(\mathbb{S}^{n-1}\right)_{\leqslant i}$). We will use the either $X \overset{d}{=} Y$ or $X \sim Y$ when the random variables X and Y are *equal in distribution*.

A fundamental concept that will feature heavily in this work is that of f*-divergences*, which are used to quantify the distance between distributions. We now give the formal definition of this notion.

Definition 2 *Let $P, Q \in \Delta(\mathcal{X})$ such that* $\text{supp}(P) \subseteq \text{supp}(Q)$.[4] *Also, let $f \colon \mathbb{R}_{\geqslant 0} \to (-\infty, \infty]$ be a convex function such that $|f(x)| < \infty, \forall x > 0$, $f(1) = 0$, and $f(0) = \lim_{x \to 0^+} f(x)$. Then, the f-divergence from P to Q is defined as $D_f(P \| Q) := \int_{\mathcal{X}} f\left(\frac{dP}{dQ}\right) dQ$. We call f the* generator *of D_f.*

An f-divergence that will be of particular importance to this work is the *Hockey-Stick divergence*. We define it below.

Definition 3 *Let $P, Q \in \Delta(\mathcal{X})$ with density functions p, q, respectively. For $\varepsilon \geqslant 0$, the Hockey-Stick divergence from P to Q is defined as:*

$$D_{e^\varepsilon}(P \| Q) := \max_{S \subseteq \mathcal{X}} \{P(S) - e^\varepsilon Q(S)\} = \frac{1}{2} \int_{\mathcal{X}} |p(x) - e^\varepsilon q(x)| \, dx - \frac{1}{2}(e^\varepsilon - 1).$$

Remark 4 *The two equivalent expressions given above for the Hockey-Stick divergence correspond to different generators f. Specifically, the former definition corresponds to $f(x) := \max\{x - e^\varepsilon, 0\}, \forall x \in \mathbb{R}_{\geqslant 0}$, whereas the latter corresponds to $f(x) := \frac{1}{2}|x - e^\varepsilon| - \frac{1}{2}(e^\varepsilon - 1), \forall x \in \mathbb{R}$. Verifying the*

4. If $\text{supp}(P) \nsubseteq \text{supp}(Q)$ the definition is still valid, but can result in $D_f(P \| Q) = \infty$. However, for the f-divergences considered in present work, this will not be the case, and we will be able to calculate them using simplified versions of the general definition without concern about the distributions' supports.

equivalence of the two definitions is a simple exercise, where it suffices to consider the second definition and partition the domain depending on whether $p(x) > e^{\varepsilon} q(x)$ or otherwise. Additionally, both definitions reduce to TV-distance *when $\varepsilon = 0$.*

For additional facts from probability and statistics, we point readers to Appendix B.

2.2. Privacy Preliminaries

We define differential privacy here and introduce some of its most fundamental properties.

Definition 5 (Approximate Differential Privacy (DP) (Dwork et al., 2006)) *We say that a randomized algorithm $M \colon \mathcal{X}^n \to \mathcal{Y}$ satisfies (ε, δ)-DP if for every pair of neighboring datasets $X, X' \in \mathcal{X}^n$ (i.e., datasets that differ in at most one entry), we have:*

$$\mathbb{P}\left[M(X) \in Y\right] \leqslant e^{\varepsilon} \mathbb{P}\left[M\left(X'\right) \in Y\right] + \delta, \forall Y \subseteq \mathcal{Y}.$$

Remark 6 *We note that the definition of (ε, δ)-DP (Definition 5) can be captured using the Hockey-Stick divergence (Definition 3). Indeed, the definition of (ε, δ)-DP is equivalent to saying that, for every pair of adjacent datasets X, X', we have $D_{e^{\varepsilon}}\left(M(X) \| M(X')\right) \leqslant \delta$.*

Definition 7 (DP under condition (Kothari et al., 2022)) *Let $\Psi \colon \mathcal{X}^n \to \{0, 1\}$ be a predicate. An algorithm $M \colon \mathcal{X}^n \to \mathcal{Y}$ is* (ε, δ)-DP under condition Ψ *for $\varepsilon, \delta > 0$ if, for every neighboring datasets $X, X' \in \mathcal{X}^n$ both satisfying Ψ, we have:*

$$\mathbb{P}\left[M(X) \in Y\right] \leqslant e^{\varepsilon} \mathbb{P}\left[M\left(X'\right) \in Y\right] + \delta, \forall Y \subseteq \mathcal{Y}.$$

Lemma 8 (Composition for Algorithm with Halting (Kothari et al., 2022)) *Assume that we have algorithms $M_1 : \mathcal{X}^n \to \mathcal{Y}_1 \cup \{\texttt{Fail}\}$ and $M_2 : \mathcal{Y}_1 \times \mathcal{X}^n \to \mathcal{Y}_2$. Furthermore, let M denote the following algorithm: let $y_1 := M_1(X)$ and, if $y_1 = \texttt{Fail}$, then halt and output $\texttt{Fail}$ or else, output $y_2 := M_2(y_1, X)$.*

Let Ψ be any condition that, if X does not satisfy Ψ, then $M_1(X)$ always returns $\texttt{Fail}$. Let M_1 be $(\varepsilon_1, \delta_1)$-DP, and M_2 be $(\varepsilon_2, \delta_2)$-DP under condition Ψ. Then, M is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP.

3. Near-Optimal Sampling from Unbounded Gaussian Distributions

3.1. The Sampler and its Guarantees

Here, we give our sampler for Gaussian distributions with unbounded parameters. Our algorithm combines elements of Algorithm 1 of Brown et al. (2023) with Algorithm 2 of Ghazi et al. (2023), as was highlighted in Section 1.1. We start by giving the pseudocode here. Observe that Algorithm 1 makes essential use of StableCov and StableMean. In the following, we will only state the guarantees of these algorithms which are necessary for proving our claims. For a complete presentation of the details of these algorithms, we point readers to Appendix D. Finally, let us note that the format of the output of the algorithms is slightly different compared to their original implementations of Brown et al. (2023). Specifically, the original implementations output the estimates $\widehat{\mu}$ and $\widehat{\Sigma}$ directly. Conversely, the output of our version of StableMean consists of a vector of weights $v \in \mathbb{R}^{n_1}$, while the output of StableCov is a $\mathbb{R}^{d \times n_2}$ matrix W whose columns are of the form $\sqrt{w_i} Y_i$, where $Y_i := 1/\sqrt{2}(X_{n_1+i} - X_{n_1+n_2+i})$.

Algorithm 1: Unbounded Gaussian Sampler

input : Dataset $X := (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n \times d}$; error $\alpha \in [0, 1]$; privacy parameters $\varepsilon, \delta > 0$
require: $n \geqslant n_1 + 2n_2$

`/* Initialize */`

$\lambda_0 \leftarrow 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8 \ln\left(\frac{3n}{\alpha}\right)$;

$n_1 \leftarrow C_1 \frac{\sqrt{\lambda_0} \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for a sufficiently large constant $C_1 \geqslant 1$;

$n_2 \leftarrow C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for a sufficiently large constant $C_2 \geqslant 1$;

$k \leftarrow \left\lceil \frac{6 \log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$; $M \leftarrow 6k + \left\lceil 18 \log\left(\frac{16n}{\delta}\right) \right\rceil$;

`/* Compute stable estimates */`

$(W, \text{Score}_1) \leftarrow \text{StableCov}_{\lambda_0, k}(X_{>n_1})$; $\widehat{\Sigma} \leftarrow WW^\top$;

$R \sim \mathcal{U}(\{R' \subseteq [n] : |R'| = M\})$; $(v, \text{Score}_2) \leftarrow \text{StableMean}_{\widehat{\Sigma}, \lambda_0, k, R}(X_{\leqslant n_1})$;
$\widehat{\mu} \leftarrow \sum_{i \in [n_1]} v_i X_i$;

`/* PTR step */`

if $\mathcal{M}_{\text{PTR}}^{\left(\frac{\varepsilon}{3}, \frac{\delta}{6}\right)}(\max\{\text{Score}_1, \text{Score}_2\}) = \mathtt{Pass}$ **then**
  $z \sim \mathcal{U}\left(\mathbb{S}^{n_2 - 1}\right)$;
  $Z \leftarrow \widehat{\mu} + \sqrt{(1 - 1/n_1)\, n_2} W z$;
  return Z;
else
  return `Fail`
end

We now formally state the guarantees of Algorithm 1.

Theorem 9 *For any $\alpha \in [0, 1]$, $\varepsilon \in [0, 1]$, $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, given $n = \mathcal{O}\left(\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}\left(d + \log\left(\frac{\log\left(\frac{1}{\delta}\right)}{\alpha\varepsilon}\right)\right)\right)$, input samples, Algorithm 1 has the following guarantees:*

- ***Privacy.*** *Algorithm 1 is (ε, δ)-DP.*
- ***Utility.*** *Let $\mathcal{N}(\mu, \Sigma)$ be any Gaussian distribution in d-dimensions with full-rank covariance. Then, given a dataset $X \sim \mathcal{N}(\mu, \Sigma)^{\otimes n}$, Algorithm 1 has runtime polynomial with respect to the size of the input, and produces output $Z \in \mathbb{R}^d \cup \{\mathtt{Fail}\}$ such that $d_{\text{TV}}(Z, \mathcal{N}(\mu, \Sigma)) \leqslant \alpha$.*

Remark 10 *Theorem 9 is stated in terms of Gaussians with full-rank covariances Σ. However, the result holds even for degenerate ones, with a slight modification of the algorithm. Specifically, we would have to add a pre-processing step that employs a subspace-finding algorithm (e.g., Ashtiani and Liaw (2022)) to identify the subspace corresponding to the directions of non-zero variance, and then project everything to that subspace. These extra steps require $\mathcal{O}\left(\frac{d \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$ samples, and run in polynomial time (see Theorem 4.1 of Ashtiani and Liaw (2022)), while all subsequent steps will* *require the same number of samples that Theorem 9 predicts, but d will be replaced by the rank of Σ. Thus, the sample and time complexity guarantees of the resulting algorithm are within the same order of magnitude as the ones given in Theorem 9.*

Remark 11 *We do not perform an explicit analysis of the time complexity of our algorithm, but the claim of polynomial runtime should be clear. Indeed, Algorithm 1 can essentially be viewed as a direct combination of the operations performed by the algorithms given in Brown et al. (2023) and Ghazi et al. (2023), both of which have polynomial runtime.*

The rest of this work will be devoted to proving Theorem 9. The utility guarantee is a direct consequence of the following theorem:

Theorem 12 *Let $X := (X_1, \dots, X_n) \sim \mathcal{N}(\mu, \Sigma)^{\otimes n}$, with $n := n_1 + 2n_2$ and $n_2 \geqslant \mathcal{O}\left(d + \log\left(\frac{1}{\alpha}\right)\right)$. Also, let $\lambda_0 \geqslant 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8\ln\left(\frac{3n}{\alpha}\right)$. Then, $d_{\mathrm{TV}}(Z, \mathcal{N}(\mu, \Sigma)) \leqslant \alpha$.*

The proof is deferred to Appendix E. Indeed, to establish that guarantee, it suffices to argue that, with high probability, there will be no outliers in the dataset due to the strong concentration properties of the Gaussian distribution, which will result in the weights assigned to datapoints by STABLECOV and STABLEMEAN being uniform. Thus, we will focus on demonstrating how the privacy analysis works (Theorem 19).

3.2. Establishing the Privacy Guarantee

Let Z and Z' be the outputs of Algorithm 1 on adjacent input datasets X and X', respectively. In this section, we will establish that, when $n = \mathcal{O}\left(\frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$, we get $D_{e^\varepsilon}(Z \| Z') \leqslant \delta$ (essentially relying on Remark 6 to establish privacy).

We note that the structure of Algorithm 1 falls into the framework of DP under condition (Definition 7) and Lemma 8 due to the use of a PTR argument. Given that the PTR step is the same as in Brown et al. (2023), and its analysis is essentially covered by Lemmas 42 and 43, we will mainly focus on establishing the privacy guarantee of the subsequent steps under the condition that the PTR step has succeeded. Furthermore, we note that Lemma 41 relies on the set R that is sampled in Line 6 being degree-representative (Definition 40). Thus, in addition to the PTR step succeeding, we also need to condition on R being degree-representative.

Taking into account the above discussion, let us fix an input dataset X, and a representation set R. For brevity, we write $\Psi(X, R) = 1$ when both the outcome of the PTR step is `Pass` and R is degree-representative, and $\Psi(X, R) = 0$ if either of the previous fails.[5] Thus, we want to study the quantity $D_{e^\varepsilon}((Z | \Psi(X, R) = 1) \| (Z' | \Psi(X', R) = 1))$. This can be written as:

$$D_{e^\varepsilon}\left(\widehat{\mu} + \sqrt{(1 - 1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1 - 1/n_1)\, n_2} W' z\right).$$

Observe that, in the above, we assumed without loss of generality that the random bits of Algorithm 1 related to the sampling of R and z are coupled in the two executions of the algorithm under input X and X'. This allows us to take $R' = R$ and $z' = z$.

5. The assumption about R being degree-representative will only matter in Section 3.2.3 really, since that is the only section that relies on Lemma 41.

Observe that, by Lemma 38, we get:

$$\begin{aligned}
&D_{e^{\varepsilon}}\left(\widehat{\mu}+\sqrt{(1-1/n_1)\,n_2}Wz\middle\|\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}W'z\right)\\
&\leqslant D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}+\sqrt{(1-1/n_1)\,n_2}Wz\middle\|\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}Wz\right)\\
&\quad+e^{\frac{\varepsilon}{2}}D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}Wz\middle\|\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}W'z\right)\\
&\leqslant 2\begin{pmatrix} D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}+\sqrt{(1-1/n_1)\,n_2}Wz\middle\|\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}Wz\right)\\ +D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}Wz\middle\|\widehat{\mu}'+\sqrt{(1-1/n_1)\,n_2}W'z\right)\end{pmatrix}, \qquad (1)
\end{aligned}$$

where the last inequality used the assumption that $\varepsilon \leqslant 1$.

We stress that the application of Lemma 38 represents a clear difference from the privacy analysis of Ghazi et al. (2023). The reason for that has to do with the fact that $\widehat{\mu}$ and $\widehat{\mu}'$ are weighted sums whose weights depend on the output of STABLECOV. Thus, although the subset of the dataset that is fed to STABLEMEAN is disjoint from the one that is fed to STABLECOV, changing one of the points that are given to STABLECOV will affect the execution of STABLEMEAN due to $\widehat{\Sigma}$ and $\widehat{\Sigma}'$ being different. Conversely, $\widehat{\mu}$ and $\widehat{\mu}'$ in Ghazi et al. (2023) are just the truncated empirical means.

The analysis will now be done in four steps, with a separate section devoted to each:

1. as a warm-up, we will analyze the second (and more interesting term) of (1) in the univariate setting (Section 3.2.1),
2. we will generalize the previous analysis to the multivariate setting (Section 3.2.2),
3. we will analyze the first term of (1) in the multivariate setting (Section 3.2.3),
4. we will tie everything together to complete the proof of the privacy guarantee (Section 3.2.4).

The second and third steps in particular constitute the bulk of the work required to establish our result. Essentially, they involve reasoning about the effect that small multiplicative and additive shifts have on the distribution of $z_{\leqslant d} \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (see Lemmas 48 and 56, respectively).

3.2.1. WARM-UP: ANALYZING THE SECOND TERM OF (1) FOR UNIVARIATE DATA

We start by analyzing the second term of (1) in the univariate setting. The proof makes crucial use of Lemma 39 to argue that $\frac{\|W'\|_2}{\|W\|_2}$ will be close to 1, which in turn implies that the Hockey-Stick divergence must be small. We will give the proof in the three steps, with one statement devoted to each step of the proof:

1. we will first show that the Hockey-Stick divergence term that we are considering can be equivalently expressed as the divergence between z_1 and $T := \frac{\|W'\|_2}{\|W\|_2} z_1$ (Lemma 13),
2. subsequently, we will identify a sufficient condition for points $t \in \operatorname{supp}(z_1)$ that implies that the log-density ratio $\ln\left(\frac{f_{z_1}(t)}{f_T(t)}\right)$ is at most $\frac{\varepsilon}{2}$ (Lemma 14),
3. finally, we will show that our sample complexity suffices for the second term of (1) to be at most δ (Proposition 15).

We start with the first of the three lemmas. The proof combines the *Data-Processing Inequality* (DPI - Fact 36) with the guarantees implied by Lemma 39.

Lemma 13 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right], \lambda_0 \geqslant 1$. Also, assume that $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for some appropriately large absolute constant $C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^n$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X,R) = \Psi(X',R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(z_1 \| T\right),$$

where $T := \frac{\|W'\|_2}{\|W\|_2} z_1$.

Proof First, by the DPI (Fact 36 - which is satisfied as an equality in this case), we get:

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(W z \middle\| W' z\right). \tag{2}$$

We assume that X and X' differ at one of the points that are fed to STABLECOV, since otherwise (2) is trivially 0. We note that, since we are dealing with univariate data, the shape of the matrix W will be $1 \times n_2$, i.e., it is a row vector whose elements are of the form $\sqrt{w_i} Y_i$ for $Y_i := \frac{1}{\sqrt{2}}\left(X_{n_1+i} - X_{n_1+n_2+i}\right), \forall i \in [n_2]$. By the definition and properties of the SVD (Fact 22 and Remark 23), we get that W and W' can be written in the form $W = \|W\|_2 e_1^\top V^\top$ and $W' = \|W'\|_2 e_1^\top V'^\top$, where $V, V' \in \mathbb{R}^{n_2 \times n_2}$ are rotation matrices. However, rotational invariance for the uniform distribution over the unit sphere (Fact 33) implies that $V^\top z \overset{d}{=} z$, yielding:

$$D_{e^{\frac{\varepsilon}{2}}}\left(W z \middle\| W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(\|W\|_2 e_1^\top z \middle\| \|W'\|_2 e_1^\top z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(\|W\|_2 z_1 \middle\| \|W'\|_2 z_1\right). \tag{3}$$

We note now that $\|W\|_2 = \widehat{\sigma}, \|W'\|_2 = \widehat{\sigma}'$. Given that, in Algorithm 1, k is set to be equal to $\left\lceil \frac{6\log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$, our bound on n_2 implies that the condition $n_2 \geqslant 16e^2\lambda_0 k$ from Lemma 39 is satisfied. Since we assumed that $\Psi(X,R) = \Psi(X',R) = 1$, Lemma 39 implies that $\|W\|_2, \|W'\|_2 > 0$. Thus, applying the DPI (Fact 36 - again as an equality) yields the desired result. ■

We note now that, by the definition of the Hockey-Stick divergence (Definition 3), we have the upper bound:

$$D_{e^{\frac{\varepsilon}{2}}}\left(z_1 \| T\right) \leqslant \underset{t \sim z_1}{\mathbb{P}}\left[f_{z_1}(t) > e^{\frac{\varepsilon}{2}} f_T(t)\right], \tag{4}$$

where by f_{z_1} and f_T we denote the densities of z_1 and T. Thus, to reason about the above, it suffices to identify a set of points t in the support of z_1 can lead to the condition $\ln\left(\frac{f_{z_1}(t)}{f_T(t)}\right) \leqslant \frac{\varepsilon}{2}$ being satisfied for the log-density ratio. Then, the probability of the event $f_{z_1}(t) > e^{\frac{\varepsilon}{2}} f_T(t)$ will be upper-bounded by the probability of the aforementioned condition failing. We identify the desired condition in the following lemma.

Lemma 14 *In the setting of Lemma 13, we have that $|t| \leqslant \sqrt{\frac{2}{3} \cdot \frac{\varepsilon}{\varepsilon + 16e^2\lambda_0}} \implies \ln\left(\frac{f_{z_1}(t)}{f_T(t)}\right) \leqslant \frac{\varepsilon}{2}$.*

Proof First, observe that standard theory about transformations of random vectors (Fact 32) yields $f_T(t) = \frac{\|W\|_2}{\|W'\|_2} f_{z_1}\left(\frac{\|W\|_2}{\|W'\|_2} t\right), \forall t \in \left[\pm \frac{\|W'\|_2}{\|W\|_2}\right]$. This should be contrasted with the support of z_1, which is the interval $[\pm 1]$. By Lemma 39, we get that $\sqrt{1-\gamma} \leqslant \frac{\|W'\|_2}{\|W\|_2} \leqslant \frac{1}{\sqrt{1-\gamma}}$ for $\gamma := \frac{8e^2\lambda_0}{n_2}$. Our choice of n_2 implies that $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4} \implies \sqrt{1-\gamma} \geqslant \sqrt{\frac{3}{4}}$. The range that we consider for t is $|t| \leqslant \sqrt{\frac{2}{3} \cdot \frac{\varepsilon}{\varepsilon + 16e^2\lambda_0}} < \sqrt{\frac{2}{3}} < \sqrt{\frac{3}{4}}$. This implies that the values of t we consider lie in the support of T, so the log-density ratio is finite. With this at hand, we work to establish the desired upper bound on the log-density ratio. To that end, the density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) and the standard inequality $\ln(x) \leqslant x - 1, \forall x > 0$ yield:

$$
\begin{aligned}
\ln\left(\frac{f_{z_1}(t)}{f_T(t)}\right) &= \ln\left(\frac{\|W'\|_2}{\|W\|_2}\right) + \frac{n_2 - 3}{2} \ln\left(\frac{1-t^2}{1-\left(\frac{\|W\|_2}{\|W'\|_2} t\right)^2}\right) \\
&< \frac{1}{2}\ln\left(\frac{1}{1-\gamma}\right) + \frac{n_2}{2}\ln\left(\frac{1-t^2}{1-\left(\frac{1}{\sqrt{1-\gamma}} t\right)^2}\right) \\
&\leqslant \frac{1}{2}\left(\frac{1}{1-\gamma} - 1\right) + \frac{n_2}{2}\left[\frac{1-t^2}{1-\left(\frac{1}{\sqrt{1-\gamma}} t\right)^2} - 1\right] \\
&= \frac{1}{2}\left(\frac{1}{1-\gamma} - 1\right) + \frac{n_2}{2} \cdot \frac{\frac{1}{1-\gamma} - 1}{1 - \frac{t^2}{1-\gamma}} t^2 \\
&= \frac{1}{2}\left(\frac{1}{1-\gamma} - 1\right) + \frac{n_2}{2} \cdot \frac{\frac{\gamma}{1-\gamma}}{1 - \frac{t^2}{1-\gamma}} t^2 \\
&= \frac{1}{2}\left(\frac{1}{1-\gamma} - 1\right) + \frac{1}{2} \cdot \frac{n_2\gamma}{(1-\gamma) - t^2} t^2 \\
&= \frac{1}{2}\left(\frac{1}{1-\gamma} - 1\right) + \frac{4e^2\lambda_0}{\frac{1-\gamma}{t^2} - 1}.
\end{aligned} \tag{5}
$$

By our assumptions on γ and ε, each of the terms of (5) is upper-bounded by $\frac{\varepsilon}{4}$. ∎

We can now show the main result of this section.

Proposition 15 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right], \lambda_0 \geqslant 1$. Also, assume that $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for some appropriately large absolute constant $C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^n$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X, R) = \Psi(X', R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$
D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W' z\right) \leqslant \delta.
$$

Proof By Lemma 13, (3), and Lemma 14, we get that:

$$
D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1-1/n_1)\, n_2} W' z\right) \leqslant \underset{t \sim z_1}{\mathbb{P}}\left[|t| > \sqrt{\frac{2}{3} \cdot \frac{\varepsilon}{\varepsilon + 16e^2\lambda_0}}\right].
$$

By the properties of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_1$ (Fact 30), we have that $z_1^2 \sim \operatorname{Beta}\left(\frac{1}{2}, \frac{n_2-1}{2}\right)$. By our bound on n_2 and Beta concentration (Fact 31), we get the desired result (for C_2 sufficiently large), ■

3.2.2. ANALYZING THE SECOND TERM OF (1) FOR MULTIVARIATE DATA

We now proceed to show the multivariate analogue of Proposition 15. As in the univariate setting, we will break down the proof into a number of lemmas to make the presentation more modular and accessible. Specifically, the main steps are:

1. we first show that the term of interest can be equivalently expressed as the divergence between $z_{\leqslant d}$ and $T := \Lambda^{-\frac{1}{2}} U^\top U' \Lambda'^{\frac{1}{2}} z_{\leqslant d}$, where $\Lambda, \Lambda' \succ 0$ are diagonal matrices, and U, U' are rotation matrices with $\widehat{\Sigma} := W W^\top = U \Lambda U^\top$ and $\widehat{\Sigma}' := W' W'^\top = U \Lambda' U^\top$ (Lemma 48),

2. then, we identify the density of T (Lemma 49),

3. subsequently, we will identify sufficient conditions for points $t \in \operatorname{supp}(z_1)$ that imply that the log-density ratio $\ln\left(\frac{f_{z_1}(t)}{f_T(t)}\right)$ is at most $\frac{\varepsilon}{2}$ (Lemma 51),

4. we will consider the matrix $A := Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) Q - \frac{\varepsilon}{4 n_2} \mathbb{I}_{n_2 \times n_2}$, where $Q \in \mathbb{R}^{d \times n_2}$ denotes the projection matrix that keeps the first d components of a vector in $\mathbb{R}^{n_2}$, and bound its trace, as well as its spectral and Frobenius norms (Lemma 52 and Corollary 53),

5. finally, we show that our sample complexity suffices for the second term of (1) to be at most δ (Proposition 16).

Observe that the fourth step in the above outline is completely missing from the univariate setting. The reason this step is required here is because, as we will see over the course of the proof, Beta concentration (Fact 31) will not be strong enough to allow us to complete the proof with optimal sample complexity. This is in contrast to the univariate setting, and the proof of Proposition 15 specifically. Instead, our proof will involve a reduction to an argument that uses Gaussian random vectors, allowing us us to appeal to the Hanson-Wright Inequality (Fact 28).

We note that the argument given in this section is not our original version of the argument. Specifically, our initial proof did not rely on the aforementioned reduction, instead using Beta concentration (Fact 31), albeit with a slightly sub-optimal sample complexity by a $\operatorname{polylog}(\lambda_0)$-factor. We consider that proof to be of significant technical interest, which is why we have included it in Appendix G. The appendix also includes a detailed discussion about why the univariate approach cannot be directly applied in the multivariate setting.

Having described the structure of the argument, we state the main result of the section, and refer readers to Appendix F for all the technical details.

Proposition 16 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, $\lambda_0 \geqslant d$. Let us assume that $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$, for some appropriately large absolute constant $C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^{n \times d}$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X, R) = \Psi(X', R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{(1 - 1/n_1)\, n_2} W z \middle\| \widehat{\mu}' + \sqrt{(1 - 1/n_1)\, n_2} W' z\right) \leqslant \delta.$$

3.2.3. ANALYZING THE FIRST TERM OF (1)

We now analyze the first term of (1). As in the previous section, the proof will be broken down into a number of intermediate steps, which we outline below:

1. we first show that the term of interest can be equivalently expressed as the divergence between $z_{\leqslant d}$ and $z_{\leqslant d} + \ell$, where $\ell := \sqrt{\frac{n_1}{n_1-1} \cdot \frac{1}{n_2}} \Lambda^{-\frac{1}{2}} U^{\top} \left(\widehat{\mu}' - \widehat{\mu}\right)$[6] and $\|\ell\|_2$ is "small" (Lemma 56),
2. subsequently, we identify sufficient conditions for points $t \in \operatorname{supp}\left(z_{\leqslant d}\right)$ to imply that the log-density ratio $\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_{z_{\leqslant d}+\ell}(t)}\right)$ is at most $\frac{\varepsilon}{2}$ (Lemma 57),
3. finally, we will show that our sample complexity suffices for the first term of (1) to be at most δ (Proposition 17).

The argument is structured similarly to Lemma 4.4 from Ghazi et al. (2023). However, differences arise along the way, due to the fact that the corresponding proof in Ghazi et al. (2023) used that datapoints are truncated, whereas we have to appeal to Lemma 41. Additionally, the original proof of Lemma 4.4 given in Ghazi et al. (2023) contains a bug in the step where it is argued that the sample complexity bound suffices for the log-density ratio to be upper-bounded by ε (the step corresponding to Lemma 57). The bug is fixable (with a minor adjustment in the final sample complexity), and has been acknowledged in private communication with the authors of that work. For all the above reasons, we present the proof in full detail.

Having described the structure of the argument, we state the main result of the section, and refer readers to Appendix H for all the technical details.

Proposition 17 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, $\lambda_0 \geqslant d$. Also, assume that $n_1 \geqslant C_1 \frac{\sqrt{\lambda_0}\log\left(\frac{1}{\delta}\right)}{\varepsilon}$ and $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for appropriately large absolute constants $C_1, C_2 \geqslant 1$. Finally, let $X, X' \in \mathbb{R}^{n \times d}$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X,R) = \Psi(X',R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu} + \sqrt{\left(1 - 1/n_1\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1 - 1/n_1\right) n_2} W z\right) \leqslant \delta.$$

3.2.4. COMPLETING THE PROOF OF THE PRIVACY GUARANTEE

Thanks to the work undertaken in the previous sections, we can now complete the proof of the privacy guarantee. We start with a statement that completes the calculation we initiated at the start of Section 3.2.

Corollary 18 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, $\lambda_0 \geqslant d$. Also, assume that $n_1 \geqslant C_1 \frac{\sqrt{\lambda_0}\log\left(\frac{1}{\delta}\right)}{\varepsilon}$ and $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for appropriately large absolute constants $C_1, C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^{n \times d}$ be adjacent datasets and $R \subseteq [n_1]$ such that $\Psi(X,R) = \Psi(X',R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$D_{e^{\varepsilon}}\left(\widehat{\mu} + \sqrt{\left(1 - 1/n_1\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1 - 1/n_1\right) n_2} W' z\right) \leqslant \delta.$$

6. Λ and U come from the spectral decomposition of $\widehat{\Sigma}$.

Proof We set $\delta \to \frac{\delta}{4}$ in Propositions 16 and 17. Then, the result follows directly from (1). ■

In all of the above, we have been assuming that $\Psi(X, R) = \Psi(X', R) = 1$. Now, we will also take into account the effect that privately checking this has on the privacy guarantees. This leads us to the main theorem of this section, which brings the proof of the privacy guarantee full circle.

Theorem 19 *Let $\varepsilon \in [0, 1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, $\lambda_0 \geqslant d$. Also, assume that $n = \mathcal{O}\left(\frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$. Then, Algorithm 1 is (ε, δ)-DP.*

Proof The proof follows the same structure as the argument in the proof of Lemma 15 in Brown et al. (2023). First, by our choice of M in Line 4, the set R sampled in Line 6 must be degree-representative, except with probability $\frac{\delta}{6}$ (this follows from hypergeometric concentration - Fact 26). Under this condition, Lemmas 42 and 43 imply that $\mathcal{M}_{\text{PTR}}^{\left(\frac{\varepsilon}{3}, \frac{\delta}{6}\right)}$ satisfies $\left(\frac{\varepsilon}{3}, \frac{\delta}{6}\right)$-DP. Finally, under both of the above conditions, setting $\varepsilon \to \frac{2}{3}\varepsilon$ and $\delta \to \frac{2}{3}\delta$ in Corollary 18, we get that the steps in Lines 8-10 of Algorithm 1 satisfy $\left(\frac{2}{3}\varepsilon, \frac{2}{3}\delta\right)$-DP. Then, the desired result follows directly from Lemma 8. ■

3.3. Putting Everything Together

Having established the utility and privacy guarantees in Theorems 12 and 19, respectively, it remains to put the two together, which will allow us to identify the final sample complexity of our algorithm, thus completing the proof of Theorem 9. We do so in this section.

Proof [Proof of Theorem 9]. We assume that we have n samples, of which the first n_1 are fed to StableCov, and the last $2n_2$ are fed to StableMean. By Theorem 19, we get that $n_1 \geqslant \Theta\left(\frac{\sqrt{\lambda_0} \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$ and $n_2 \geqslant \Theta\left(\frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$, i.e., it must be the case that $n \geqslant \Theta\left(\frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$. By Theorem 12, we get that $\lambda_0 \geqslant \Theta\left(d + \sqrt{d \log\left(\frac{3n}{\alpha}\right)} + \log\left(\frac{n}{\alpha}\right)\right)$ and $n_2 \geqslant \Theta\left(d + \log\left(\frac{1}{\alpha}\right)\right)$. The latter condition is trivially satisfied by the bound on n_2 implied by Theorem 19. Thus, it remains now to identify a bound on n such that:

$$n \geqslant \Theta\left(\frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right) \geqslant \Theta\left(\frac{\left(d + \sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + \ln\left(\frac{n}{\alpha}\right)\right) \log\left(\frac{1}{\delta}\right)}{\varepsilon}\right).$$

The above is satisfied when $n = \mathcal{O}\left(\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}\left(d + \log\left(\frac{\log\left(\frac{1}{\delta}\right)}{\alpha\varepsilon}\right)\right)\right)$. ■

Acknowledgments

The authors would like to thank Gavin Brown and the reviewers at COLT for helpful comments and discussions. GK and AM are supported by a Canada CIFAR AI Grant, an NSERC Discovery Grant, and an unrestricted gift from Google. Finally, AM is also supported by the Onassis Foundation (Scholarship ID: F ZT 053-1/2023-2024).

References

Ishaq Aden-Ali, Hassan Ashtiani, and Gautam Kamath. On the sample complexity of privately learning unbounded high-dimensional gaussians. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, ALT '21, pages 185–216. JMLR, Inc., 2021.

Sushant Agarwal, Gautam Kamath, Mahbod Majid, Argyris Mouzakis, Rose Silver, and Jonathan Ullman. Private mean estimation with person-level differential privacy. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SODA '25, pages 2819–2880. SIAM, 2025.

Daniel Alabi, Pravesh K Kothari, Pranay Tankala, Prayaag Venkat, and Fred Zhang. Privately estimating a Gaussian: Efficient, robust and optimal. In *Proceedings of the 55th Annual ACM Symposium on the Theory of Computing*, STOC '23. ACM, 2023.

Hassan Ashtiani and Christopher Liaw. Private and polynomial time algorithms for learning Gaussians and beyond. In *Proceedings of the 35th Annual Conference on Learning Theory*, COLT '22, pages 1075–1076, 2022.

Hilal Asi, Jonathan Ullman, and Lydia Zakynthinou. From robustness to privacy and back. *arXiv preprint arXiv:2302.01855*, 2023.

Martin Aumüller, Christian Janos Lebeda, Boel Nelson, and Rasmus Pagh. Plan: variance-aware private mean estimation. *arXiv preprint arXiv:2306.08745*, 2023.

Brian Axelrod, Shivam Garg, Vatsal Sharan, and Gregory Valiant. Sample amplification: Increasing dataset size even when learning is impossible. In *Proceedings of the 37th International Conference on Machine Learning*, ICML '20, pages 442–451. JMLR, Inc., 2020.

Raef Bassily, Om Thakkar, and Abhradeep Guha Thakurta. Model-agnostic private learning. In *Advances in Neural Information Processing Systems 31*, NeurIPS '18, pages 7102–7112. Curran Associates, Inc., 2018.

Brett K Beaulieu-Jones, Zhiwei Steven Wu, Chris Williams, Ran Lee, Sanjeev P Bhavnani, James Brian Byrd, and Casey S Greene. Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes*, 12(7):e005122, 2019.

Shai Ben-David, Alex Bie, Clément L. Canonne, Gautam Kamath, and Vikrant Singhal. Private distribution learning with public data: The view from sample compression. In *Advances in Neural Information Processing Systems 36*, NeurIPS '23, pages 7184–7215. Curran Associates, Inc., 2023.

Alex Bie, Gautam Kamath, and Vikrant Singhal. Private estimation with public data. In *Advances in Neural Information Processing Systems 35*, NeurIPS '22. Curran Associates, Inc., 2022.

Alex Bie, Gautam Kamath, and Guojun Zhang. Private GANs, revisited. *Transactions on Machine Learning Research*, 2023.

Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. In *Advances in Neural Information Processing Systems 33*, NeurIPS '20, pages 14475–14485. Curran Associates, Inc., 2020.

Gavin Brown, Marco Gaboardi, Adam Smith, Jonathan Ullman, and Lydia Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. In *Advances in Neural Information Processing Systems 34*, NeurIPS '21. Curran Associates, Inc., 2021.

Gavin Brown, Samuel B Hopkins, and Adam Smith. Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. In *Proceedings of the 36th Annual Conference on Learning Theory*, COLT '23, pages 5578–5579, 2023.

Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 181–191. Curran Associates, Inc., 2019.

Mark Bun, Gautam Kamath, Thomas Steinke, and Zhiwei Steven Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 156–167. Curran Associates, Inc., 2019.

T Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

Tianshi Cao, Alex Bie, Arash Vahdat, Sanja Fidler, and Karsten Kreis. Don't generate me: Training differentially private generative models with sinkhorn divergence. In *Advances in Neural Information Processing Systems 34*, NeurIPS '21, pages 12480–12492. Curran Associates, Inc., 2021.

Albert Cheu and Debanuj Nayak. Differentially private multi-sampling from distributions. In *Proceedings of the 36th International Conference on Algorithmic Learning Theory*, ALT '25, pages 289–314. JMLR, Inc., 2025.

Tim Dockhorn, Tianshi Cao, Arash Vahdat, and Karsten Kreis. Differentially private diffusion models. *Transactions on Machine Learning Research*, 2023.

Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.

Cynthia Dwork and Vitaly Feldman. Privacy-preserving prediction. In *Proceedings of the 31st Annual Conference on Learning Theory*, COLT '18, pages 1693–1702, 2018.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the 41st Annual ACM Symposium on the Theory of Computing*, STOC '09, pages 371–380. ACM, 2009.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, Berlin, Heidelberg, 2006. Springer.

Badih Ghazi, Xiao Hu, Ravi Kumar, and Pasin Manurangsi. On differentially private sampling from gaussian and product distributions. In *Advances in Neural Information Processing Systems 36*, NeurIPS '23, pages 77783–77809. Curran Associates, Inc., 2023.

Frederik Harder, Milad Jalali, Danica J Sutherland, and Mijung Park. Pre-trained perceptual features improve differentially private image generation. *Transactions on Machine Learning Research*, 2023.

Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd Annual ACM Symposium on the Theory of Computing*, STOC '10, pages 705–714. ACM, 2010.

Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. In *Proceedings of the 54th Annual ACM Symposium on the Theory of Computing*, STOC '22, pages 1406–1417. ACM, 2022.

Samuel B Hopkins, Gautam Kamath, Mahbod Majid, and Shyam Narayanan. Robustness implies privacy in statistical estimation. In *Proceedings of the 55th Annual ACM Symposium on the Theory of Computing*, STOC '23. ACM, 2023.

Ziyue Huang, Yuting Liang, and Ke Yi. Instance-optimal mean estimation under differential privacy. In *Advances in Neural Information Processing Systems 34*, NeurIPS '21. Curran Associates, Inc., 2021.

Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 1853–1902, 2019.

Gautam Kamath, Argyris Mouzakis, and Vikrant Singhal. New lower bounds for private estimation and a generalized fingerprinting lemma. In *Advances in Neural Information Processing Systems 35*, NeurIPS '22, pages 24405–24418. Curran Associates, Inc., 2022a.

Gautam Kamath, Argyris Mouzakis, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A private and computationally-efficient estimator for unbounded gaussians. In *Proceedings of the 35th Annual Conference on Learning Theory*, COLT '22, pages 544–572, 2022b.

Gautam Kamath, Argyris Mouzakis, Matthew Regehr, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A bias-variance-privacy trilemma for statistical estimation. *arXiv preprint arXiv:2301.13334*, 2023.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science*, ITCS '18, pages 44:1–44:9, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Pravesh K Kothari, Pasin Manurangsi, and Ameya Velingker. Private robust estimation by stabilizing convex relaxations. In *Proceedings of the 35th Annual Conference on Learning Theory*, COLT '22, pages 723–777, 2022.

Rohith Kuditipudi, John Duchi, and Saminul Haque. A pretty fast algorithm for adaptive private mean estimation. In *Proceedings of the 36th Annual Conference on Learning Theory*, COLT '23, pages 2511–2551, 2023.

Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, Harsha Nori, and Sergey Yekhanin. Differentially private synthetic data via foundation model apis 1: Images. In *Proceedings of the 12th International Conference on Learning Representations*, ICLR '24, 2024.

Xiyang Liu, Weihao Kong, Sham Kakade, and Sewoong Oh. Robust and differentially private mean estimation. In *Advances in Neural Information Processing Systems 34*, NeurIPS '21. Curran Associates, Inc., 2021.

Xiyang Liu, Weihao Kong, and Sewoong Oh. Differential privacy and robust statistics in high dimensions. In *Proceedings of the 35th Annual Conference on Learning Theory*, COLT '22, pages 1167–1246, 2022.

Shyam Narayanan. Better and simpler lower bounds for differentially private statistical estimation. *arXiv preprint arXiv:2310.06289*, 2023.

Nicolas Papernot, Martín Abadi, Ulfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *Proceedings of the 5th International Conference on Learning Representations*, ICLR '17, 2017.

Victor S Portella and Nick Harvey. Lower bounds for private estimation of gaussian covariance matrices under all reasonable parameter regimes. *arXiv preprint arXiv:2404.17714*, 2024.

Hedvika Ranosova. Spherically Symmetric Measures. *Bachelor's thesis, Department of Probability and Mathematical Statistics, Charles University*, 2021.

Sofya Raskhodnikova, Satchit Sivakumar, Adam Smith, and Marika Swanberg. Differentially private sampling from distributions. In *Advances in Neural Information Processing Systems 34*, NeurIPS '21, pages 28983–28994. Curran Associates, Inc., 2021.

Igal Sason and Sergio Verdu. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016.

Matthew Skala. Hypergeometric tail inequalities: ending the insanity. *arXiv preprint arXiv:1311.5939*, 2013.

Chulin Xie, Zinan Lin, Arturs Backurs, Sivakanth Gopi, Da Yu, Huseyin A Inan, Harsha Nori, Haotian Jiang, Huishuai Zhang, Yin Tat Lee, Bo Li, and Sergey Yekhanin. Differentially private synthetic data via foundation model apis 2: Text. *arXiv preprint arXiv:2403.01749*, 2024.

Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*, 2018.

Anru Zhang and Yuchen Zhou. On the non-asymptotic and sharp lower tail bounds of random variables. *Stat*, 9(1):e314, 2020.

Appendix A. Facts from Linear Algebra

In this appendix, we state a number of useful facts from linear algebra. We start with a fact about the semi-definite ordering.

Fact 20 *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric positive-definite matrices. Then:*

$$A \preceq B \iff \mathbb{I} \preceq A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \iff B^{-\frac{1}{2}} A B^{-\frac{1}{2}} \preceq \mathbb{I}.$$

Next, we have a fact about matrix norms. Specifically, we recall the property known as *rotational invariance*.

Fact 21 *Let $A \in \mathbb{R}^{n \times m}$. For any rotation matrix $U \in \mathbb{R}^{n \times n}$, we have $\|UA\|_2 = \|A\|_2$, $\|UA\|_F = \|A\|_F$, and $\|UA\|_{\mathrm{tr}} = \|A\|_{\mathrm{tr}}$. The above also holds for AU, for any rotation matrix $U \in \mathbb{R}^{m \times m}$.*

We now introduce basic matrix factorizations, namely the *spectral decomposition* for symmetric matrices, and the *singular value decomposition (SVD)* for general matrices.

Fact 22 *Let $A \in \mathbb{R}^{n \times m}$. Then, A can be factorized as follows:*

- *If $n = m$ and $A = A^\top$, we can write $A = S \Lambda S^\top$, where $\Lambda := \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$ and $\{\lambda_i\}_{i \in [n]}$ are the eigenvalues of A, and $SS^\top = \mathbb{I}$. For the special case of PSD matrices, S can be assumed to be a rotation matrix. The previous can equivalently be written as $A = \sum\limits_{i \in [n]} \lambda_i s_i s_i^\top$, where $s_i \in \mathbb{R}^n$ are the vectors representing the columns of S.*
- *If $n \geqslant m$, we can write $A = UDV^\top$, where $D := \left(\operatorname{diag}\{\sigma_1, \ldots, \sigma_m\} \middle| 0_{m \times (n-m)}\right)^\top \in \mathbb{R}^{n \times m}$ and $\{\sigma_i\}_{i \in [m]}$ are the singular values of A, and $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$ are rotation matrices. The previous can equivalently be written as $A = \sum\limits_{i \in [m]} \sigma_i u_i v_i^\top$, where $u_i \in \mathbb{R}^n$ and $v_i \in \mathbb{R}^m$ are the vectors representing the columns of U and V, respectively.*

Remark 23 *The statement given above for the SVD is phrased in terms of matrices whose number of rows is lower-bounded by the number of columns. However, the decomposition is valid even without this assumption and, given a matrix that does not satisfy this, we can obtain its SVD by taking its transpose, and then applying the previous statement.*

We include one standard inequality here that upper-bounds the determinant of a positive-definite matrix in terms of its trace. The inequality follows directly from AM-GM.

Fact 24 *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric PSD matrix. Then, we have $\det(A) \leqslant \left(\frac{\mathrm{tr}(A)}{n}\right)^n$.*

Finally, we have a lemma which establishes that, if a symmetric positive-definite matrix A has small distance from the identity matrix, the same holds for A^{-1} as well.

Lemma 25 *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric PD matrix with $A \succeq \mathbb{I}$. Then, it holds that $\|A - \mathbb{I}\|_{\mathrm{tr}} \leqslant \alpha \implies \left\|A^{-1} - \mathbb{I}\right\|_{\mathrm{tr}} \leqslant \alpha$.*

Proof Since the matrix A is symmetric PD, the spectral factorization (Fact 22) implies that it can be written as $A = U\Lambda U^\top$, where Λ is the eigenvalue matrix of A, and U is a rotation matrix. This implies that we have $A^{-1} = U\Lambda^{-1}U^\top$. Thus, by rotational invariance (Fact 21), we get $\|A - \mathbb{I}\|_{\mathrm{tr}} = \|\Lambda - \mathbb{I}\|_{\mathrm{tr}}$ and $\left\|A^{-1} - \mathbb{I}\right\|_{\mathrm{tr}} = \left\|\Lambda^{-1} - \mathbb{I}\right\|_{\mathrm{tr}}$. Let $\{\lambda_i\}_{i\in[n]}$ be the spectrum of A. By the previous and our assumptions that $A \succeq \mathbb{I}$ and $\|A - \mathbb{I}\|_{\mathrm{tr}} \leqslant \alpha$, we can reduce the problem of upper-bounding $\left\|A^{-1} - \mathbb{I}\right\|_{\mathrm{tr}}$ to the following constrained convex maximization problem:

$$\begin{aligned}
\max_{\lambda_i} \quad & \sum_{i\in[n]} \left(1 - \frac{1}{\lambda_i}\right) \\
\text{s.t.} \quad & \sum_{i\in[n]} (\lambda_i - 1) \leqslant a < \frac{1}{2} \\
& \lambda_i \geqslant 1, \forall i \in [n],
\end{aligned}$$

The above can be solved exactly via an application of the KKT conditions, yielding the bound $\left\|A^{-1} - \mathbb{I}\right\|_F \leqslant \frac{\alpha n}{n+\alpha} \leqslant \alpha$. ■

Appendix B. Facts from Probability & Statistics

In this appendix, we expand upon the background from probability and statistics that was given in Section 2.1. We start with a number of results about the densities and concentration properties of distributions of interest. These results are also used at various points in Brown et al. (2023) and Ghazi et al. (2023).

Fact 26 ((Skala, 2013)) *Suppose an urn contains N balls and exactly k of them are black. Let random variable y be the number of black balls selected when drawing n uniformly at random from the urn without replacement. Then, for all $t \geqslant 0$, we have:*

$$\mathbb{P}\left[\left|\frac{y}{n} - \frac{k}{N}\right| \geqslant t\right] \leqslant 2e^{-2t^2 n}.$$

Fact 27 (Lemma 1 (Laurent and Massart, 2000)) *If $X \sim \chi^2(k)$, and $\beta \in [0,1]$, then:*

$$\mathbb{P}\left[X - k \geqslant 2\sqrt{k \ln\left(\frac{1}{\beta}\right)} + 2\ln\left(\frac{1}{\beta}\right)\right] \leqslant \beta.$$

Equivalently, the above can be written as:

$$\mathbb{P}[X \geqslant t] \leqslant e^{-\frac{(\sqrt{2t-k}-\sqrt{k})^2}{4}}, \forall t \geqslant k.$$

Thus, if $Y \in \mathbb{R}^d$ and $Y \sim \mathcal{N}(0, \mathbb{I})$, then $\mathbb{P}\left[\|Y\|_2^2 \geqslant t\right] \leqslant e^{-\frac{(\sqrt{2t-d}-\sqrt{d})^2}{4}}, \forall t \geqslant d$.

Fact 28 (Hanson-Wright Inequality) *Let $X \in \mathbb{R}^d$ be a random vector with $X \sim \mathcal{N}(0, \mathbb{I})$. There exists an absolute constant $C_3 > 0$ such that, for any non-zero symmetric matrix $A \in \mathbb{R}^{d\times d}$:*

$$\mathbb{P}\left[\left|X^\top AX - \mathbb{E}\left[X^\top AX\right]\right| \geqslant t\right] \leqslant 2\exp\left(-C_3 \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right), \forall t \geqslant 0,$$

where $\mathbb{E}\left[X^\top AX\right] = \mathrm{tr}(A)$.

Fact 29 (Folklore - see Fact 3.4 (Kamath et al., 2019)) *Let $X := (X_1, \ldots, X_n) \sim \mathcal{N}(0, \mathbb{I})^{\otimes n}$ and $\widehat{\Sigma} := \frac{1}{n} \sum_{i \in [n]} X_i X_i^\top$. Then, except with probability β, we have:*

$$\left(1 - \mathcal{O}\left(\sqrt{\frac{d + \log\left(\frac{1}{\beta}\right)}{n}}\right)\right)\mathbb{I} \preceq \widehat{\Sigma} \preceq \left(1 + \mathcal{O}\left(\sqrt{\frac{d + \log\left(\frac{1}{\beta}\right)}{n}}\right)\right)\mathbb{I}.$$

Fact 30 (Theorem 2 (Ranosova, 2021)) *Let $v \sim \mathcal{U}\left(\mathbb{S}^{n-1}\right)$. Then, the density of $v_{\leqslant i} \sim \mathcal{U}\left(\mathbb{S}^{n-1}\right)_{\leqslant i}$ is:*

$$f(x) \propto \begin{cases} \left(1 - \|x\|_2^2\right)^{\frac{n-i}{2} - 1}, & \|x\|_2 \leqslant 1 \\ 0, & \|x\|_2 > 1 \end{cases}.$$

Additionally, $\|v_{\leqslant i}\|_2^2 \sim \mathrm{Beta}\left(\frac{i}{2}, \frac{n-i}{2}\right)$.

Fact 31 (Theorem 8 (Zhang and Zhou, 2020)) *Let $X \sim \mathrm{Beta}(\alpha, \beta)$ with $0 < \alpha < \beta$. Then, there exists an absolute constant $C_4 \in (0, 1)$ such that:*

$$\mathbb{P}\left[X \geqslant \frac{\alpha}{\alpha + \beta} + x\right] \leqslant 2e^{-C_4 \min\left\{\frac{\beta^2 x^2}{\alpha}, \beta x\right\}}, \forall x \geqslant 0.$$

We will also need a standard result about transformations of random vectors.

Fact 32 *Let $X \in \mathbb{R}^d$ be a random vector with density f_X. Also, for an invertible function $g\colon \mathbb{R}^d \to \mathbb{R}^d$, let $Y := g(X)$. Then, we have for the density of Y:*

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\det\left(J_{g^{-1}}\right)\right|,$$

where $J_{g^{-1}}$ denotes the Jacobian matrix of the inverse transform g^{-1}.

As a direct consequence of the above and Fact 21, we get the following result, which establishes that the uniform distribution over the unit sphere is *rotationally invariant*.

Fact 33 *Let $v \sim \mathcal{U}\left(\mathbb{S}^{n-1}\right)$. Then, for any rotation matrix $U \in \mathbb{R}^{n \times n}$, we have $Uv \overset{d}{=} v$.*

We now present a result that describes one method for generating random vectors that are uniformly distributed over the unit sphere.

Fact 34 *Let $X := (X_1, \ldots, X_d) \sim \mathcal{N}(0, 1)^{\otimes d}$. We have $\frac{X}{\|X\|_2} \sim \mathcal{U}\left(\mathbb{S}^{d-1}\right)$.*

We conclude our facts about distributions with a lemma that generalizes the stability property of the Gaussian distribution. Namely, stability says that any sum of independent Gaussians or product of a Gaussian with a number also follows a Gaussian distribution. We extend this to random linear combinations when the weights come from a distribution over the unit sphere.

Lemma 35 *Let $\mathcal{D} \in \Delta\left(\mathbb{S}^{n-1}\right)$, and $a \sim \mathcal{D}$. Then, given $X := (X_1, \ldots, X_n) \sim \mathcal{N}(0, \mathbb{I})^{\otimes n}$ that is independent of a, we have $\sum_{i \in [n]} a_i X_i \overset{d}{=} \mathcal{N}(0, \mathbb{I})$.*

Proof Let $X := \sum_{i \in [n]} a_i \mathcal{N}(0, \mathbb{I})$. Let a_0 be a fixed realization of a. Observe that:

$$(X|a = a_0) = \sum_{i \in [n]} a_{0,i} \mathcal{N}(0, \mathbb{I}) \stackrel{d}{=} \sum_{i \in [n]} \mathcal{N}\left(0, a_{0,i}^2 \mathbb{I}\right) \stackrel{d}{=} \mathcal{N}\left(0, \left(\sum_{i \in [n]} a_{0,i}^2\right) \mathbb{I}\right) \stackrel{d}{=} \mathcal{N}(0, \mathbb{I}).$$

Now, the density of X is a weighted average of the densities of $(X|\alpha)$ for all fixed realizations a_0 of a, where the weighting is performed according to the density of $\mathcal{D}$. Thus, we have:

$$f_X(x) = \int_{\text{supp}(\mathcal{D})} f_{\mathcal{D}}(a_0) f_{(X|a=a_0)}(x)\, da_0 = \int_{\text{supp}(\mathcal{D})} f_{\mathcal{D}}(a_0) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, da_0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \forall x \in \mathbb{R},$$

which yields $X \stackrel{d}{=} \mathcal{N}(0, 1)$, as desired. ■

Having presented all the necessary statements about distributions, we continue with a number of properties of f-divergences, namely the *Data-Processing Inequality (DPI)* and *joint convexity*.

Fact 36 *Let $P, Q \in \Delta(\mathcal{X})$ and $X \sim P, Y \sim Q$, and let f be a function satisfying the conditions of Definition 2. For any function $g \colon \mathcal{X} \to \mathcal{Y}$ (deterministic or randomized), we get:*

$$D_f(g(X) \| g(Y)) \leqslant D_f(X \| Y).$$

If g is invertible, the above holds as an equality.

Fact 37 *Let $P, Q \in \Delta(\mathcal{X})$ and $X \sim P, Y \sim Q$, and let f be a function satisfying the conditions of Definition 2. We assume that there exists a $\lambda \in (0, 1)$ and distributions $P_1, P_2 \in \Delta(\mathcal{X})$ and $Q_1, Q_2 \in \Delta(\mathcal{X})$ such that $P = \lambda P_1 + (1 - \lambda) P_2$ and $Q = \lambda Q_1 + (1 - \lambda) Q_2$, i.e., P and Q are mixtures of (P_1, P_2) and (Q_1, Q_2), respectively, with shared mixing weights $(\lambda, 1 - \lambda)$. Then:*

$$D_f(P \| Q) \leqslant \lambda D_f(P_1 \| Q_1) + (1 - \lambda) D_f(P_2 \| Q_2).$$

We now introduce a property which is specific to the Hockey-Stick divergence. As implied by the definition of the Hockey-Stick divergence (Definition 3), $D_{e^\varepsilon}(\cdot \| \cdot)$ does not satisfy a triangle inequality in the general case (and is thus not a distance metric). However, the following property can be interpreted as a weak form of the triangle inequality. We note that the property has appeared in the literature before (see Inequality (404) (Sason and Verdu, 2016)), but it is stated under the assumption that we have distributions $P, R, Q \in \Delta(\mathcal{X})$ with $\text{supp}(P) \subseteq \text{supp}(R) \subseteq \text{supp}(Q)$ (presumably because Sason and Verdu (2016) only mentions $\max\{x - e^\varepsilon, 0\}$ as a generator of D_{e^ε}). This assumption is not actually necessary, which is why we include a proof of the claim. Our proof is elementary and relies on the equivalent form of the Hockey-Stick divergence that uses $\frac{1}{2}|x - e^\varepsilon| - \frac{1}{2}(e^\varepsilon - 1)$ as a generator.

Lemma 38 *Let $P, R, Q \in \Delta(\mathcal{X})$ with respective probability density functions p, r, q. Then, for any $\varepsilon_1, \varepsilon_2 \geqslant 0$, we have:*

$$D_{e^{\varepsilon_1 + \varepsilon_2}}(P \| Q) \leqslant D_{e^{\varepsilon_1}}(P \| R) + e^{\varepsilon_1} D_{e^{\varepsilon_2}}(R \| Q).$$

Proof By the definition of the Hockey-Stick divergence (Definition 3) and the triangle inequality, we get:

$$\begin{aligned}
D_{e^{\varepsilon_1+\varepsilon_2}}(P\|Q) &= \frac{1}{2}\int_{\mathcal{X}} \left|p(x) - e^{\varepsilon_1+\varepsilon_2} q(x)\right| dx - \frac{1}{2}\left(e^{\varepsilon_1+\varepsilon_2}-1\right) \\
&\leqslant \frac{1}{2}\int_{\mathcal{X}} \left(\left|p(x) - e^{\varepsilon_1} r(x)\right| + \left|e^{\varepsilon_1} r(x) - e^{\varepsilon_1+\varepsilon_2} q(x)\right|\right) dx - \frac{1}{2}\left(e^{\varepsilon_1+\varepsilon_2}-1\right) \\
&= \left[\frac{1}{2}\int_{\mathcal{X}} \left|p(x) - e^{\varepsilon_1} r(x)\right| dx - \frac{1}{2}\left(e^{\varepsilon_1}-1\right)\right] + e^{\varepsilon_1}\int_{\mathcal{X}} \left|r(x) - e^{\varepsilon_2} q(x)\right| dx \\
&\quad + \frac{1}{2}\left(e^{\varepsilon_1}-1\right) - \frac{1}{2}\left(e^{\varepsilon_1+\varepsilon_2}-1\right) \\
&= D_{e^{\varepsilon_1}}(P\|R) + e^{\varepsilon_1}\left[\int_{\mathcal{X}} \left|r(x) - e^{\varepsilon_2} q(x)\right| dx - \frac{1}{2}\left(e^{\varepsilon_2}-1\right)\right] \\
&= D_{e^{\varepsilon_1}}(P\|R) + e^{\varepsilon_1} D_{e^{\varepsilon_2}}(R\|Q)\,.
\end{aligned}$$

■

Appendix C. The Main Algorithm of Brown et al. (2023) and its Properties

We now present the main algorithm given in Brown et al. (2023) for covariance-aware mean estimation. The algorithm's goal is to privately obtain an estimate $\widetilde{\mu}$ for the mean of a Gaussian (or sub-Gaussian) distribution with mean vector μ and covariance matrix Σ that satisfies $\|\widetilde{\mu} - \mu\|_{\Sigma} \leqslant \alpha$ with high probability. Roughly speaking, the algorithm involves producing a pair of non-private estimates $\left(\widehat{\mu}, \widehat{\Sigma}\right)$ for the mean and the covariance of the distribution, respectively. These estimates come from *stable* estimators. This is a class of algorithms which, given well-behaved input datasets (i.e., datasets that do not contain significant outliers), produce outputs which are resilient to small perturbations of the input dataset, despite having no formal robustness and/or privacy guarantees. A PTR step ensues, which privately determines whether the results produced by the algorithms correspond to a "good" dataset. If the estimates fail the test, the algorithm aborts. Conversely, if the test is successful, a point is sampled from the distribution $\mathcal{N}\left(\widehat{\mu}, c^2\widehat{\Sigma}\right)$,[7] which comes with the desired utility and privacy guarantees when the dataset is appropriately large.[8]

Algorithm 2: Covariance-Aware Private Mean Estimator

input : Dataset $X := (X_1, \dots, X_n)^\top \in \mathbb{R}^{n \times d}$; privacy parameters $\varepsilon, \delta > 0$; outlier threshold $\lambda_0 \geqslant 1$

```
/* Initialize                                                    */
```
$k \leftarrow \left\lceil \frac{6 \log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$; $M \leftarrow 6k + \left\lceil 18 \log\left(\frac{16n}{\delta}\right)\right\rceil$; $c^2 \leftarrow \frac{720 e^2 \lambda_0 \log\left(\frac{12}{\delta}\right)}{\varepsilon^2 n^2}$;

```
/* Compute stable estimates (see Appendix D for details)         */
```
$(W, \mathrm{Score}_1) \leftarrow \textsc{StableCov}_{\lambda_0, k}(X)$; $\widehat{\Sigma} \leftarrow WW^\top$;

$R \sim \mathcal{U}(\{R' \subseteq [n] : |R'| = M\})$; $(v, \mathrm{Score}_2) \leftarrow \textsc{StableMean}_{\widehat{\Sigma}, \lambda_0, k, R}(X)$; $\widehat{\mu} \leftarrow \sum\limits_{i \in [n]} v_i X_i$;

```
/* PTR step                                                      */
```
if $\mathcal{M}_{\mathrm{PTR}}^{\left(\frac{\varepsilon}{3}, \frac{\delta}{6}\right)}(\max\{\mathrm{Score}_1, \mathrm{Score}_2\}) = \mathtt{Pass}$ **then**

  $\widetilde{\mu} \sim \mathcal{N}\left(\widehat{\mu}, c^2\widehat{\Sigma}\right)$;

  return $\widetilde{\mu}$;

else

  return `Fail`

end

Algorithm 2 uses a pair of subroutines (StableCov and StableMean), which correspond to the stable estimators mentioned earlier. These algorithms function by assigning weights to points depending on how far they are from other points in the dataset (essentially performing a *soft outlier removal* operation). For brevity of exposition, we will not give the pseudocode of these algorithms here, instead deferring the full presentation to Appendix D. However, we will note that we have made a slight modification to the format of the output of these algorithms. Rather than outputting the weighted sums that correspond to the stable estimates, StableCov outputs a matrix $W \in \mathbb{R}^{d \times n}$ whose columns are of the form $\sqrt{w_i} Y_i$ where $Y_i := \frac{1}{\sqrt{2}}\left(X_i - X_{i + \lfloor \frac{n}{2} \rfloor}\right)$, while StableMean outputs a vector of weights, and the estimates $\left(\widehat{\mu}, \widehat{\Sigma}\right)$ are constructed separately in a post-processing step. The goal of this modification is to facilitate the use of these algorithms in subsequent sections.

For the rest of this section, we will focus on some lemmas which will prove to be particularly important for our analysis in Section 3. These are related to the *stability* guarantees of StableCov and StableMean, as well as the way they interact with the PTR step. They are crucial to the analysis of Brown et al. (2023) for establishing the privacy guarantees in their main result. We note that lemmas related to the accuracy guarantees given in Brown et al. (2023) will not be discussed at

7. c is a multiplicative factor that exists so that the magnitude of the noise is large enough to ensure privacy.

8. Sampling a point from $\mathcal{N}\left(\widehat{\mu}, c^2\widehat{\Sigma}\right)$ essentially amounts to an implementation of the Gaussian mechanism with noise which, instead of being spherical, depends on the shape of $\widehat{\Sigma}$, i.e., $\mathcal{N}\left(\widehat{\mu}, c^2\widehat{\Sigma}\right) = \widehat{\mu} + \widehat{\Sigma}^{\frac{1}{2}}\mathcal{N}\left(0, c^2\mathbb{I}\right)$. This first appeared in Brown et al. (2021) as a means of performing covariance-aware mean estimation (termed the *Emprically Rescaled Gaussian Mechanism*), and then in Alabi et al. (2023) for covariance estimation as part of the *Gaussian Sampling Mechanism*.

any point in this manuscript, the reason being that the utility guarantee in the present work (closeness of output distribution to input distribution in TV-distance) differs significantly from the target guarantee in Brown et al. (2023), and we will not be applying their results.

We stress that the versions of the statements we will give here are slightly different from the original ones. Indeed, the original formulations assume that StableMean and StableCov are fed the same dataset. However, in Section 3, this will not be the case. For that reason, we have made the smallest possible number of changes to the statements to ensure that they are compatible with how they are used in subsequent sections.

We first state a result of Brown et al. (2023) that captures the effect of changing one input datapoint has on the output of StableCov, under the assumption that both the original dataset and the new one are well-concentrated (which is quantified by the score outputted by StableCov).

Lemma 39 (Lemma 11 (Brown et al., 2023)) *Fix dataset size $2n_2$, dimension d, outlier threshold $\lambda_0 \geqslant 1$, and privacy parameters $\varepsilon, \delta \in [0,1]$. Set the discretization parameter k to be $\left\lceil \frac{6\log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$ (as in Algorithm 2). Let X, X' be adjacent d-dimensional datasets of size $2n_2$. Assume $\gamma := \frac{8e^2\lambda_0}{n_2} \leqslant \frac{1}{2k}$ (i.e., $n_2 \geqslant 16e^2\lambda_0 k$). Let:*

$$\begin{aligned}(W_1, \mathrm{Score}) &:= \textsc{StableCov}\,(X, \lambda_0, k)\,,\\ \left(W_2, \mathrm{Score}'\right) &:= \textsc{StableCov}\,\left(X', \lambda_0, k\right).\end{aligned}$$

Assume $\mathrm{Score}, \mathrm{Score}' < k$. *Then $\Sigma_1 := W_1 W_1^\top, \Sigma_2 := W_2 W_2^\top \succ 0$ and $(1-\gamma)\,\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$. Additionally, we have:*

$$\left\|\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}} - \mathbb{I}\right\|_{\mathrm{tr}}, \left\|\Sigma_2^{-\frac{1}{2}}\Sigma_1\Sigma_2^{-\frac{1}{2}} - \mathbb{I}\right\|_{\mathrm{tr}} \leqslant (1+2\gamma)\,\gamma.$$

The previous is complemented by an analogous guarantee for StableMean. However, before stating the guarantee for StableMean, we first need to introduce the notion of a *degree-representative* subset of a dataset that was given in Brown et al. (2023). The motivation behind this definition has to do with the fact that Algorithm 2 uses a *reference set* R in StableMean. This is a subset of the dataset that is used to efficiently implement the soft outlier removal process described earlier.

Definition 40 (Definition 32 (Brown et al., 2023)) *Fix d-dimensional dataset X of size n_1, covariance $\Sigma \succ 0$, outlier threshold $\lambda_0 \geqslant 1$, and reference set $R \subseteq [n_1]$. For all i, let:*

$$\begin{aligned}N_i &:= \left\{j \in [n_1] : \; \|X_i - X_j\|_\Sigma^2 \leqslant e^2\lambda_0\right\},\\ \widetilde{N}_i &:= \left\{j \in R\colon \; \|X_i - X_j\|_\Sigma^2 \leqslant e^2\lambda_0\right\}.\end{aligned}$$

Let $z_i := |N_i|$ and $\widetilde{z}_i := \left|\widetilde{N}_i\right|$ be the sets' sizes. We say that R is degree-representative for X *if for every index i we have $\left|\frac{1}{|R|}\widetilde{z}_i - \frac{1}{n_1}z_i\right| \leqslant \frac{1}{6}$.*

Having given the above definition, we now proceed to state the lemma capturing the stability guarantees of StableMean.

Lemma 41 (Lemma 13 (Brown et al., 2023)) *Fix dataset sizes n_1 and $2n_2$, dimension d, outlier threshold $\lambda_0 \geqslant 1$, and privacy parameters $\varepsilon, \delta \in [0,1]$. Set the discretization parameter k to be $\left\lceil \frac{6\log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$ (as in Algorithm 2). Use reference set $R \subseteq [n_1]$ with $|R| > 6k$. Let X and X' be adjacent d-dimensional datasets of size n_1. Let $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d\times d}$ be positive definite matrices satisfying $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$ for $\gamma := \frac{8e^2\lambda_0}{n_2}$. Assume $n_1 \geqslant 32e^2k$ and $\gamma \leqslant \frac{1}{2k}$ (i.e., $n_2 \geqslant 16e^2\lambda_0 k$). Let:*

$$\begin{aligned}(v, \mathrm{Score}) &:= \textsc{StableMean}\left(X, \Sigma_1, \lambda_0, k, R\right), \\ \left(v', \mathrm{Score}'\right) &:= \textsc{StableMean}\left(X', \Sigma_2, \lambda_0, k, R\right).\end{aligned}$$

If $\mathrm{Score}, \mathrm{Score}' < k$ and R is degree-representative for both X and X', then:

$$\left\|\widehat{\mu} - \widehat{\mu}'\right\|_{\Sigma_1}^2 \leqslant \frac{(1+2\gamma)\, 38e^2\lambda_0}{n_1^2},$$

where $\widehat{\mu} := \sum_{i\in[n]} v_i X_i$ and $\widehat{\mu}' := \sum_{i\in[n]} v_i' X_i'$.

We conclude this section with two results that capture the function of the PTR step of Algorithm 2. The first result describes the conditions under which the PTR step works.

Lemma 42 (Claim 10 (Brown et al., 2023)) *Fix $0 < \varepsilon \leqslant 1$ and $0 < \delta \leqslant \frac{\varepsilon}{10}$. There is an algorithm $\mathcal{M}_{\mathrm{PTR}}^{(\varepsilon,\delta)} \colon \mathbb{R} \to \{\texttt{Pass}, \texttt{Fail}\}$ that satisfies the following conditions:*

1. *Let $\mathcal{U}$ be a set and $g \colon \mathcal{U}^n \to \mathbb{R}_{\geqslant 0}$ a function. If, for all $X, X' \in \mathcal{U}^n$ that differ in one entry, $|g(X) - g(X')| \leqslant 2$, then $\mathcal{M}_{\mathrm{PTR}}^{(\varepsilon,\delta)}(g(\cdot))$ is (ε,δ)-DP.*
2. *$\mathcal{M}_{\mathrm{PTR}}^{(\varepsilon,\delta)}(0) = \texttt{Pass}$.*
3. *For all $z \geqslant \frac{2\log\left(\frac{1}{\delta}\right)}{\varepsilon} + 4$, $\mathcal{M}_{\mathrm{PTR}}^{(\varepsilon,\delta)}(z) = \texttt{Fail}$.*

The second result describes how the guarantees of Lemmas 39 and 41 suffice to satisfy the conditions of Lemma 42.

Lemma 43 (Lemma 9 (Brown et al., 2023)) *Fix dataset size n, dimension d, outlier threshold $\lambda_0 \geqslant 1$, reference set $R \subseteq [n_1]$, and privacy parameters $\varepsilon, \delta \in [0,1]$. Set $k := \left\lceil \frac{6\log\left(\frac{6}{\delta}\right)}{\varepsilon} \right\rceil + 4$ (as in Algorithm 2). Assume $n := n_1 + 2n_2$ with $n_1 \geqslant 32e^2k$ and $n_2 \geqslant 16e^2\lambda_0 k$. Let X and X' be adjacent d-dimensional datasets of size n. Let:*

$$\begin{aligned}(W, \mathrm{Score}_1) &:= \textsc{StableCov}\left(X_{>n_1}, \lambda_0, k\right), \\ \left(W', \mathrm{Score}_1'\right) &:= \textsc{StableCov}\left(X'_{>n_1}, \lambda_0, k\right), \\ (v, \mathrm{Score}_2) &:= \textsc{StableMean}\left(X_{\leqslant n_1}, WW^\top, \lambda_0, k, R\right), \\ \left(v', \mathrm{Score}_2'\right) &:= \textsc{StableMean}\left(X'_{\leqslant n_1}, W'W'^\top, \lambda_0, k, R\right).\end{aligned}$$

Then $\left|\max\left\{\mathrm{Score}_1, \mathrm{Score}_2\right\} - \max\left\{\mathrm{Score}_1', \mathrm{Score}_2'\right\}\right| \leqslant 2$.

Appendix D. Subroutines of Algorithm 2

In this appendix, we include all the details of the subroutines of Algorithm 2, complementing Appendix C.

We start by giving the pseudocode for STABLECOV (Algorithm 3). As we have discussed in Section C, the algorithm's function is to perform a soft outlier removal process, which will result in a sequence of weights. These can be used to construct a weighted version of the empirical covariance which, in addition to being accurate, also has stability guarantees which prove to be useful in guaranteeing privacy. The algorithm makes crucial of use of the subroutine LARGESTGOODSUBSET (Algorithm 4), which is given separately.

Algorithm 3: STABLECOV (X, λ_0, k), for nonprivate covariance estimation

input: Dataset $X := (X_1, \ldots, X_{2m})^\top \in \mathbb{R}^{2m \times d}$, outlier threshold λ_0, discretization parameter $k \in \mathbb{N}$

$\forall i \in [m], Y_i \leftarrow \frac{1}{\sqrt{2}}(X_i - X_{i+m})$; `/* pair and rescale */`

for $\ell = 0, 1, \ldots, 2k$ **do**

  $S_\ell \leftarrow \text{LARGESTGOODSUBSET}_{e^{\frac{\ell}{k}}\lambda_0}(Y)$

end

$\text{Score} \leftarrow \min\left\{k, \min\limits_{0 \leqslant \ell \leqslant k}\{m - |S_\ell| + \ell\}\right\}$;

for $i = 1, \ldots, m$ **do**

  $w_i \leftarrow \frac{1}{km} \sum\limits_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$;

end

$W \leftarrow \left(\sqrt{w_1}Y_1, \ldots, \sqrt{w_m}Y_m\right)^\top$;

return W, Score;

We now give the pseudocode for LARGESTGOODSUBSET. The algorithm starts by constructing the standard empirical covariance of the whole input dataset Y. Then, the algorithm calculates the Mahalanobis norm of all the points Y_i measured with respect to the empirical covariance. Due to the fact that Algorithm 3 starts by performing a centering operation to the dataset, the points that have large Mahalanobis norm are the outliers. These are removed, the empirical covariance of the remaining points is calculated, and the previous check is repeated with those points. The process is repeated iteratively until a fixpoint is reached, i.e., all outliers have been removed.

Algorithm 4: `LargestGoodSubset` (Y, λ), subroutine for covariance estimation

input: Dataset $Y := (Y_1, \ldots, Y_m) \in \mathbb{R}^{m \times d}$, outlier threshold λ_0

`/* For vector` v `and singular matrix` A`, define` $\|A^{-1}v\|_2 = +\infty$ `*/`

$S \leftarrow [m]$;
repeat
  $\mathtt{OUT} \leftarrow \left\{ i \in S\colon \left\| \left(\frac{1}{m} \sum_{j \in S} Y_j Y_j^\top \right)^{-\frac{1}{2}} Y_i \right\|_2^2 > \lambda \right\}$;
  $S \leftarrow S \backslash \mathtt{OUT}$
until $\mathtt{OUT} = \varnothing$;
return S;

We now give the pseudocode for STABLEMEAN (Algorithm 5). The algorithm works similarly to STABLECOV, but has some minor differences which we highlight here. First, the algorithm receives the output of STABLECOV as input, which it uses to rescale the points and measure their pair-wise distances accordingly. The goal of this is to result in a dataset that comes from a distribution that is approximately isotropic. Second, the algorithm also receives a reference set R. This is a subset of the dataset that the algorithm uses to identify outliers. Instead of computing all pair-wise distances between points in the dataset, the algorithm will identify outliers by only considering distances of elements of the dataset from points in the set reference set. The purpose of this aspect of the algorithm is to ensure computational efficiency. Similarly to STABLECOV, this process leverages a subroutine which is given separately (LARGESTCORE - Algorithm 6).

Algorithm 5: STABLEMEAN $\left(X, \widehat{\Sigma}, \lambda_0, k, R\right)$ for nonprivate mean estimation

input: Dataset $X := (X_1, \ldots, X_n) \in \mathbb{R}^{n \times d}$, covariance $\widehat{\Sigma}$, outlier threshold λ_0, discretization parameter $k \in \mathbb{N}$, reference set $R \subseteq [n]$

for $\ell = 0, 1, \ldots, 2k$ **do**
  $S_\ell \leftarrow \mathtt{LargestCore}\left(X, \widehat{\Sigma}, e^{\frac{\ell}{k}} \lambda_0, |R| - \ell, R\right)$;
end
$\text{Score} \leftarrow \min \left\{ k, \min\limits_{0 \leqslant \ell \leqslant k} \{n - |S_\ell| + \ell\} \right\}$;
for $i = 1, \ldots, n$ **do**
  $c_i \leftarrow \sum\limits_{\ell = k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$;
end
$Z \leftarrow \sum\limits_{i \in [n]} c_i$;
$\forall i \in [n], v_i \leftarrow \frac{c_i}{Z}$;   `/* set` $w_i \leftarrow 0$ `if` $Z = 0$ `*/`
return v, Score;

We conclude by giving the pseudocode for LARGESTCORE. The algorithm identifies the points in the dataset that have "small" Mahalanobis distance (measured with respect to the output $\widehat{\Sigma}$ of STABLECOV) from at least τ points in the reference set R.

Algorithm 6: `LargestCore` $\left(X, \widehat{\Sigma}, \lambda, \tau, R\right)$, subroutine for mean estimation

input: dataset $X := (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n \times d}$, covariance $\widehat{\Sigma}$, outlier threshold λ, degree threshold τ, reference set $R \subseteq [n]$

for $i \in [n]$ **do**
  $N_i \leftarrow \left\{j \in R : \|X_i - X_j\|_{\widehat{\Sigma}}^2 \leqslant \lambda\right\}$; /* Nearby points in R */
end
return $\{i \in [n] : |N_i| \geqslant \tau\}$;

Appendix E. Establishing the Utility Guarantee

In this section, we will show that, if the input dataset is drawn from $\mathcal{N}(\mu, \Sigma)$, and $\lambda_0 \geqslant 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8 \ln\left(\frac{3n}{\alpha}\right)$, the output Z of Algorithm 1 will satisfy $d_{\mathrm{TV}}(Z, \mathcal{N}(\mu, \Sigma)) \leqslant \alpha$. Showing this will require us to argue that, with high probability, there will be no significant outliers in the dataset. This follows from the strong concentration properties of the Gaussian distribution. In that case, the weights assigned to datapoints by STABLECOV and STABLEMEAN will be uniform. Intuitively, this yields:

$$\widehat{\mu} = \frac{1}{n_1} \sum_{i \in [n_1]} X_i \sim \mathcal{N}\left(\mu, \frac{1}{n_1}\Sigma\right) \text{ and } Wz = \frac{1}{\sqrt{n_2}} \sum_{i \in [n_2]} z_i Y_i \sim \mathcal{N}\left(0, \frac{1}{n_2}\Sigma\right) \implies Z \sim \mathcal{N}(\mu, \Sigma).$$

We stress that the above argument is not formal by any means. Indeed, as mentioned above, for the weights to be uniform, we conditioned on the dataset being well-concentrated. This conditioning affects the input distribution, making it a *truncated Gaussian* instead of a regular Gaussian. Despite this, as we will see, the intuition is correct, and will lead us to a formal result that goes through an application of the *joint convexity* property of f-divergences (Fact 37).

We start working now towards making the above intuition more formal. Let $\bar{\mu}$ and $\bar{\Sigma}$ denote the standard empirical mean and empirical covariance of X and Y respectively, i.e., $\bar{\mu} := \frac{1}{n_1} \sum\limits_{i \in [n_1]} X_i$ and $\bar{\Sigma} := \frac{1}{n_2} \sum\limits_{i \in [n_2]} Y_i Y_i^\top$. In a series of lemmas, we will identify sufficient conditions which establish that, with probability at least $1 - \alpha$, we have $\widehat{\mu} = \bar{\mu}$ and $\widehat{\Sigma} = \bar{\Sigma}$, i.e., $v_i = \frac{1}{n_1}, \forall i \in [n_1]$ and $w_i = \frac{1}{n_2}, \forall i \in [n_2]$, and the PTR step of Line 6 succeeds. To help us in the proof, we define the following events of interest:

$$\begin{aligned}
E_1 &:= \left\{\|X_i - \mu\|_\Sigma^2 \leqslant \frac{\lambda_0}{4}, \forall i \in [n_1]\right\}, \\
E_2 &:= \left\{\|Y_i\|_\Sigma^2 \leqslant \frac{\lambda_0}{4}, \forall i \in [n_2]\right\}, \\
E_3 &:= \left\{\left\|\Sigma^{\frac{1}{2}} \bar{\Sigma}^{-1} \Sigma^{\frac{1}{2}} - \mathbb{I}\right\|_2 \leqslant 2\right\}, \\
E &:= E_1 \cap E_2 \cap E_3.
\end{aligned}$$

Our goal is to show that E is a sufficient condition for the weights outputted by both STABLECOV and STABLEMEAN to be uniform, and to have $\mathrm{Score}_1 = \mathrm{Score}_2 = 0$. However, before doing so, we identify sufficient conditions for each of the events E_1, E_2, and E_3 to hold with probability at least $1 - \frac{\alpha}{3}$. A union bound then implies that E must hold with probability at least $1 - \alpha$.

We first identify a bound on the outlier threshold λ_0 that is used by STABLECOV and STABLEMEAN that ensures that each of the events E_1 and E_2 occurs with probability at least $1 - \frac{\alpha}{3}$.

Lemma 44 *Let $X := (X_1, \dots, X_{n_1}) \sim \mathcal{N}(\mu, \Sigma)^{\otimes n_1}$ and $Y := (Y_1, \dots, Y_{n_2}) \sim \mathcal{N}(0, \Sigma)^{\otimes n_2}$. Then, for $n := n_1 + 2n_2$, given $\lambda_0 \geqslant 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8\ln\left(\frac{3n}{\alpha}\right)$, we have $\mathbb{P}[E_1] \geqslant 1 - \frac{\alpha}{3}$ and $\mathbb{P}[E_2] \geqslant 1 - \frac{\alpha}{3}$.*

Proof We will show the result for E_1, and the proof for E_2 is entirely analogous. We note that, for all $i \in [n_1]$, we have $X_i - \mu \sim \mathcal{N}(0, \Sigma) \implies \Sigma^{-\frac{1}{2}}(X_i - \mu) \sim \mathcal{N}(0, \mathbb{I})$. By a union bound and χ^2 concentration (Fact 27), we get:

$$\mathbb{P}[E_1^c] \leqslant n_1 \underset{Z \sim \mathcal{N}(0,\mathbb{I})}{\mathbb{P}}\left[\|Z\|_2^2 \geqslant \frac{\lambda_0}{4}\right] \leqslant n_1 e^{-\frac{\left(\sqrt{\frac{\lambda_0}{2} - d} - \sqrt{d}\right)^2}{4}} \leqslant \frac{\alpha}{3},$$

where the last inequality follows from our bound on λ_0. ∎

We now proceed to identify a bound on n_2 such that E_3 holds.

Lemma 45 *Let $Y := (Y_1, \dots, Y_{n_2}) \sim \mathcal{N}(0, \mathbb{I})^{\otimes n_2}$. For $n_2 \geqslant \mathcal{O}\left(d + \log\left(\frac{1}{\alpha}\right)\right)$, we have $\mathbb{P}[E_3] \geqslant 1 - \frac{\alpha}{3}$.*

Proof We note that $\Sigma^{-\frac{1}{2}}\bar{\Sigma}\Sigma^{-\frac{1}{2}} = \frac{1}{n_2}\sum\limits_{i \in [n_2]}\left(\Sigma^{-\frac{1}{2}}X_i\right)\left(\Sigma^{-\frac{1}{2}}X_i\right)^\top$, where $\Sigma^{-\frac{1}{2}}X_i \sim \mathcal{N}(0, \mathbb{I})$. By the concentration properties of the empirical covariance of Gaussians (Fact 29) and our sample complexity bound, we have with probability at least $1 - \frac{\alpha}{3}$:

$$\begin{aligned}\frac{1}{3}\mathbb{I} \preceq \Sigma^{-\frac{1}{2}}\bar{\Sigma}\Sigma^{-\frac{1}{2}} \preceq \frac{5}{3}\mathbb{I} \iff \frac{3}{5}\mathbb{I} \preceq \Sigma^{\frac{1}{2}}\bar{\Sigma}^{-1}\Sigma^{\frac{1}{2}} \preceq 3\mathbb{I} &\iff -\frac{2}{5}\mathbb{I} \preceq \Sigma^{\frac{1}{2}}\bar{\Sigma}^{-1}\Sigma^{\frac{1}{2}} - \mathbb{I} \preceq 2\mathbb{I} \\ &\implies \left\|\Sigma^{\frac{1}{2}}\bar{\Sigma}^{-1}\Sigma^{\frac{1}{2}} - \mathbb{I}\right\|_2 \leqslant 2,\end{aligned}$$

yielding the desired result. ∎

We now leverage the previous two lemmas to establish that, with high probability, the output of STABLECOV will coincide with the standard empirical covariance $\bar{\Sigma}$ and Score_1 will be 0.

Lemma 46 *Let $X := (X_1, \dots, X_n) \sim \mathcal{N}(\mu, \Sigma)^{\otimes n}$, where $n := n_1 + 2n_2$ and $n_2 \geqslant \mathcal{O}\left(d + \log\left(\frac{1}{\alpha}\right)\right)$. Also, let $\lambda_0 \geqslant 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8\ln\left(\frac{3n}{\alpha}\right)$. Then, except with probability $\frac{2\alpha}{3}$, $E_2 \cap E_3$ occurs, and the weights w_i assigned to datapoints by $\mathrm{STABLECOV}_{\lambda_0,k}(X_{>n_1})$ will satisfy $w_i = \frac{1}{n_2}, \forall i \in [n_2]$, and we will have $\bar{\Sigma} = \widehat{\Sigma}$, and $\mathrm{Score}_1 = 0$.*

Proof The assumptions of Lemmas 44 and 45 are satisfied, thus a union bound implies that $E_2 \cap E_3$ is realized, except with probability $\frac{2\alpha}{3}$. For the following, we condition over this event. Under this assumption, it suffices to prove that LARGESTGOODSUBSET (Algorithm 4 in Appendix D) will return the entire set $[n_2]$ every time it is called by STABLECOV (Algorithm 3 in Appendix D). For this to happen, it suffices to have $\left\|\bar{\Sigma}^{-\frac{1}{2}} Y_i\right\|_2^2 \leqslant \lambda_0, \forall i \in [n_2]$, where $Y_i := \frac{1}{\sqrt{2}}\left(X_{n_1+i} - X_{n_1+n_2+i}\right) \sim \mathcal{N}(0, \Sigma)$. We note that the definition of the spectral norm implies the upper bound:

$$\left\|\bar{\Sigma}^{-\frac{1}{2}} Y_i\right\|_2^2 = \left\|\left(\bar{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right)\left(\Sigma^{-\frac{1}{2}} Y_i\right)\right\|_2^2 \leqslant \left\|\bar{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 \left\|\Sigma^{-\frac{1}{2}} Y_i\right\|_2^2 = \left\|\bar{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 \|Y_i\|_{\Sigma}^2 .$$

Thus, to establish the desired result, we upper-bound each term in the above separately. The fact that E_2 is realized implies that the second term is upper-bounded by $\frac{\lambda_0}{4}$. As for the first term, we have again by the definition of the spectral norm:

$$\left\|\bar{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 = \left\|\Sigma^{\frac{1}{2}} \bar{\Sigma}^{-1} \Sigma^{\frac{1}{2}}\right\|_2 \leqslant \left\|\Sigma^{\frac{1}{2}} \bar{\Sigma}^{-1} \Sigma^{\frac{1}{2}} - \mathbb{I}\right\|_2 + \|\mathbb{I}\|_2 \leqslant 3, \tag{6}$$

where we appealed to the triangle inequality and the fact that E_3 is assumed to occur.

The desired results follow directly by combining the previous upper bounds. ■

The previous is complemented by a lemma which establishes that, with high probability, the output of STABLEMEAN will coincide with the standard empirical mean $\bar{\mu}$ and Score_2 will be 0.

Lemma 47 *Let $X := (X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \Sigma)^{\otimes n}$, where $n := n_1 + 2n_2$ and $n_2 \geqslant \mathcal{O}\left(d + \log\left(\frac{1}{\alpha}\right)\right)$. Also, let $\lambda_0 \geqslant 4d + 8\sqrt{d \ln\left(\frac{3n}{\alpha}\right)} + 8\ln\left(\frac{3n}{\alpha}\right)$. Then, except with probability α, E occurs, implying that the weight vector v outputted by $\mathrm{STABLEMEAN}_{\hat{\Sigma}, \lambda_0, k, R}(X_{\leqslant n_1})$ will satisfy $v_i = \frac{1}{n_1}, \forall i \in [n_1]$, and we will have $\bar{\mu} = \hat{\mu}$, and $\mathrm{Score}_2 = 0$.*

Proof The assumptions of Lemmas 44 and 45 are satisfied, thus a union bound implies that E is realized, except with probability α. For the following, we condition over this event. It suffices to prove that LARGESTCORE (Algorithm 6 Appendix D) will return the entire set $[n_1]$ every time it is called by STABLEMEAN (Algorithm 5 in Appendix D). For this to hold, it suffices to have $\|X_i - \mu\|_{\hat{\Sigma}}^2 \leqslant \frac{\lambda_0}{4}, \forall i \in [n_1]$. If the previous is satisfied, the choice of the subset R chosen for the execution of STABLEMEAN in Line 6 will not matter as all the points are close to one another. Indeed, for any $i \in [n_1]$ and $j \in R$, the triangle inequality yields:

$$\|X_i - X_j\|_{\hat{\Sigma}}^2 \leqslant \left(\|X_i - \mu\|_{\hat{\Sigma}} + \|X_j - \mu\|_{\hat{\Sigma}}\right)^2 \leqslant \lambda_0,$$

and therefore $N_i = R$ for all $i \in [n_1]$, and thus LARGESTCORE will always return $[n_1]$, and STABLEMEAN will assign uniform weights to every X_i. Thus, the rest of the proof is devoted to showing that E implies $\|X_i - \mu\|_{\hat{\Sigma}}^2 \leqslant \frac{\lambda_0}{4}, \forall i \in [n_1]$. By the definition of the spectral norm, we have:

$$\|X_i - \mu\|_{\hat{\Sigma}}^2 = \left\|\left(\hat{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right)\left[\Sigma^{-\frac{1}{2}}(X_i - \mu)\right]\right\|_2^2 \leqslant \left\|\hat{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 \left\|\Sigma^{-\frac{1}{2}}(X_i - \mu)\right\|_2^2 = \left\|\hat{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 \|X_i - \mu\|_{\Sigma}^2 .$$

Since we have assumed that E_1 is realized, we immediately get that the second term is upper-bounded by $\frac{\lambda_0}{4}$. For the first term, since $E_2 \cap E_3$ is also assumed to occur, Lemma 46 implies that $\hat{\Sigma} = \bar{\Sigma}$. This allows us to work as in Lemma 46 (namely (6)), and get $\left\|\hat{\Sigma}^{-\frac{1}{2}} \Sigma^{\frac{1}{2}}\right\|_2^2 \leqslant 3$. Then, the desired results follow directly. ■

We can now complete the proof of the utility guarantee. We will proceed by writing the output Z of Algorithm 1 as a mixture of two components, with one corresponding to when E is realized, and the other corresponding to when E^c is realized. Then, the desired result will follow from joint convexity (Fact 37).

Proof [Proof of Theorem 12]. First, we note that Lemmas 46 and 47 imply:

$$(Z|E) = \left(\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i \middle| E\right) \text{ and } \mathbb{P}\left[E^c\right] \leqslant \alpha.$$

Additionally, observe that Lemma 35 implies that:

$$\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i \overset{d}{=} \mathcal{N}\left(\mu, \Sigma\right).$$

Combining the previous two remarks and applying joint convexity (Fact 37), we get:

$$\begin{aligned}
d_{\mathrm{TV}}\left(Z, \mathcal{N}\left(\mu,\Sigma\right)\right) &= d_{\mathrm{TV}}\left(Z, \frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i\right)\\
&= d_{\mathrm{TV}}\left(\mathbb{1}\left\{E\right\}Z + \mathbb{1}\left\{E^c\right\}Z, \begin{array}{c}\mathbb{1}\left\{E\right\}\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i\\ +\mathbb{1}\left\{E^c\right\}\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i\end{array}\right)\\
&\leqslant \cancelto{0}{\mathbb{P}\left[E\right]d_{\mathrm{TV}}\left(\left(Z|E\right), \left(\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i \middle| E\right)\right)}\\
&\quad + \mathbb{P}\left[E^c\right]d_{\mathrm{TV}}\left(\left(Z|E^c\right), \left(\frac{1}{n_1}\sum_{i\in[n_1]} X_i + \sqrt{\left(1-\frac{1}{n_1}\right)n_2}\frac{1}{\sqrt{n_2}}\sum_{i\in[n_2]} z_i Y_i \middle| E^c\right)\right)\\
&\leqslant \mathbb{P}\left[E^c\right]\\
&\leqslant \alpha.
\end{aligned}$$

■

Appendix F. Omitted Proofs from Section 3.2.2

We give here the first lemma from the sequence described in Section 3.2.2. Its proof is analogous that of Lemma 13, but requires some extra care due to the fact that we are dealing with multivariate data.

Lemma 48 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right], \lambda_0 \geqslant d$. Let us assume that $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$, for some appropriately large absolute constant $C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^{n \times d}$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X,R) = \Psi(X',R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d} \| T\right),$$

where $T := \Lambda^{-\frac{1}{2}} U^\top U' \Lambda'^{\frac{1}{2}} z_{\leqslant d}$, while $\Lambda, \Lambda' \succ 0$ are diagonal matrices, and U, U' are rotation matrices such that we have $\widehat{\Sigma} := WW^\top = U\Lambda U^\top$ and $\widehat{\Sigma}' := W'W'^\top = U\Lambda' U^\top$.

Proof As in Lemma 13, the equality case of the DPI (Fact 36) yields:

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(W z \| W' z\right). \tag{7}$$

We assume that X and X' differ at one of the points that are fed to STABLECOV, since otherwise (7) is trivially 0. For multivariate data, the shape of the matrix W will be $d \times n_2$. By properties of the SVD (Fact 22 and Remark 23), we get that W and W' can be written in the form $W = UDV^\top$ and $W' = U'D'V'^\top$, where $U, U' \in \mathbb{R}^{d \times d}$ and $V, V' \in \mathbb{R}^{n_2 \times n_2}$ are rotation matrices, while $D, D' \in \mathbb{R}^{d \times n_2}$ are matrices of the form $D = \left(\operatorname{diag}\{\sigma_1, \ldots, \sigma_d\} \middle| 0_{d \times (n_2 - d)}\right)$ and $D' = \left(\operatorname{diag}\{\sigma'_1, \ldots, \sigma'_d\} \middle| 0_{d \times (n_2 - d)}\right)$. However, rotational invariance for the uniform distribution over the unit sphere (Fact 33) implies that $V^\top z \overset{d}{=} z$, yielding:

$$D_{e^{\frac{\varepsilon}{2}}}\left(W z \| W' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(U D z \| U' D' z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(U \Lambda^{\frac{1}{2}} z_{\leqslant d} \middle\| U' \Lambda'^{\frac{1}{2}} z_{\leqslant d}\right), \tag{8}$$

where $\Lambda := DD^\top = \operatorname{diag}\{\sigma_1^2, \ldots, \sigma_d^2\}$ and $\Lambda' := D'D'^\top = \operatorname{diag}\{\sigma_1'^2, \ldots, \sigma_d'^2\}$.

We note now that:

$$\widehat{\Sigma} = WW^\top = \left(UDV^\top\right)\left(UDV^\top\right)^\top = UD\left(V^\top V\right) D^\top U^\top = UDD^\top U^\top = U\Lambda U^\top,$$

and, analogously, $\widehat{\Sigma}' = U'\Lambda' U'^\top$. Our choice of n_2 implies that the condition $n_2 \geqslant 16 e^2 \lambda_0 k$ of Lemma 39 is satisfied, and, since $\Psi(X,R) = \Psi(X',R) = 1$, we have that $\Lambda, \Lambda' \succ 0 \implies \exists \Lambda^{-1}, \Lambda'^{-1}$. Thus, applying the DPI (Fact 36 - again as an equality) yields the desired result. ■

Observe now that the multivariate analogue of (4) implies that:

$$D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d} \| T\right) \leqslant \underset{t \sim z_{\leqslant d}}{\mathbb{P}}\left[f_{z_{\leqslant d}}(t) > e^{\frac{\varepsilon}{2}} f_T(t)\right]. \tag{9}$$

Thus, we need to identify the density of T. We do so in the following lemma.

Lemma 49 *In the setting of Lemma 48, it holds that:*

$$f_T(t) \propto \sqrt{\det\left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}\right)} \left(1 - s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s\right)^{\frac{n_2 - d}{2} - 1} \mathbb{1}\left\{s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s \leqslant 1\right\},$$

where $s := s(t) := Ut$ with $\|s\|_2 = \|t\|_2$.

Proof The density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) and standard theory about transformations of random vectors (Fact 32) yield that:

$$\begin{aligned} f_T(t) &= \det\left(\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}}\right) f_{z_{\leqslant d}}\left(\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}} t\right) \\ &\propto \det\left(\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}}\right)\left(1 - \left\|\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}} t\right\|_2^2\right)^{\frac{n_2-d}{2}-1}, \end{aligned} \tag{10}$$

for all $t \in \mathbb{R}^d$ satisfying $\left\|\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}} t\right\|_2^2 \leqslant 1$.

We work towards simplifying the above. First, observe that:

$$\left\|\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}} t\right\|_2^2 = t^\top \Lambda^{\frac{1}{2}} U^\top U' \Lambda'^{-1} U'^\top U \Lambda^{\frac{1}{2}} t = (Ut)^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} (Ut) = s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s, \tag{11}$$

where we appealed to the guarantees of Lemma 48 to write $\widehat{\Sigma} = U \Lambda U^\top, \widehat{\Sigma}' = U' \Lambda' U'^\top$ with $\Lambda, \Lambda' \succ 0$ (which ensures that the inverses and square roots of $\widehat{\Sigma}$ and $\widehat{\Sigma}'$ are well-defined).

Second, we have:

$$\begin{aligned} \det\left(\Lambda'^{-\frac{1}{2}} U'^\top U \Lambda^{\frac{1}{2}}\right) &\overset{(a)}{=} \det\left(\Lambda'^{-\frac{1}{2}}\right) \det\left(\Lambda^{\frac{1}{2}}\right) \overset{(b)}{=} \sqrt{\det\left(\Lambda'^{-1}\right) \det(\Lambda)} \\ &\overset{(c)}{=} \sqrt{\det\left(\Lambda^{\frac{1}{2}}\right) \det\left(\Lambda'^{-1}\right) \det\left(\Lambda^{\frac{1}{2}}\right)} \\ &\overset{(d)}{=} \sqrt{\det\left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}\right)}, \end{aligned} \tag{12}$$

where (a) used that the determinant is multiplicative when considering products of square matrices and that rotation matrices have determinant 1, (b) used that the determinant of the square root of a diagonal matrix is the square root of the determinant, (c) used again the multiplicative property, and (d) used the same properties as (a).

Substituting to (10) based on (11) and (12) yields the desired result about the density. It remains to argue that $\|s\|_2 = \|t\|_2$. This follows directly from rotational invariance (Fact 21). ∎

Remark 50 *Lemma 49 establishes that, although $z_{\leqslant d}$ is supported over the origin-centered unit ball, T is supported over an origin-centered ellipsoid, where the orientation and the length of the axes are determined by the eigendecomposition of $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}$. Lemma 39 implies that the shape of this ellipsoid will not be too different from that of the unit ball, and this is something that will be leveraged significantly in Lemma 52.*

Having identified the density of T, and working as we did in Section 3.2.1, we proceed to identify sufficient conditions for the log-density ratio $\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_T(t)}\right)$ to be upper-bounded by $\frac{\varepsilon}{2}$.

Lemma 51 *In the setting of Lemmas 48 and 49, we have that:*

$$s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s \leqslant \frac{1}{2} \text{ and } s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) s \leqslant \frac{\varepsilon}{4 n_2} \implies \ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_T(t)}\right) \leqslant \frac{\varepsilon}{2}.$$

Proof The first of our two conditions ensures that we are considering points in the support of T, and thus the log-density ratio will be bounded. Under that assumption, we work with the density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) and Lemma 49, and obtain:

$$\begin{aligned}
\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_T(t)}\right) &= \ln\left(\frac{1}{\sqrt{\det\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}\right)}}\right) + \frac{n_2-d-2}{2}\ln\left(\frac{1-\|t\|_2^2}{1-s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s}\right) \\
&= \frac{1}{2}\ln\left(\det\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right)\right) + \frac{n_2-d-2}{2}\ln\left(\frac{1-\|s\|_2^2}{1-s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s}\right). \qquad (13)
\end{aligned}$$

We proceed to upper-bound each term of (13) separately. We will show that, as a consequence of our assumptions, both terms are upper-bounded by $\frac{\varepsilon}{4}$. For the first term, we have by the AM-GM inequality for the trace of symmetric PSD matrices (Fact 24):

$$\frac{1}{2}\ln\left(\det\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right)\right) \leqslant \frac{d}{2}\ln\left(\frac{\operatorname{tr}\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right)}{d}\right). \qquad (14)$$

By triangle inequality and Lemma 39, we get:

$$\begin{aligned}
&\left|\operatorname{tr}\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right) - d\right| \leqslant \left\|\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}} - \mathbb{I}\right\|_{\operatorname{tr}} \leqslant (1+2\gamma)\gamma \\
\Longrightarrow\ &\operatorname{tr}\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right) \leqslant d + (1+2\gamma)\gamma \leqslant d + \frac{3}{2}\gamma, \qquad (15)
\end{aligned}$$

where the last inequality used the fact that n_2 is large enough for us to have $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$.

Upper-bounding (14) using (15) and the inequality $\ln(x) \leqslant x-1, \forall x > 0$, we get:

$$\frac{1}{2}\ln\left(\det\left(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}}\right)\right) \leqslant \frac{d}{2}\ln\left(1+\frac{3\gamma}{2d}\right) \leqslant \frac{3}{4}\gamma, \qquad (16)$$

We want the upper bound of (16) to be at most $\frac{\varepsilon}{4}$, which is equivalent to $\gamma \leqslant \frac{\varepsilon}{3}$. However, our assumptions on n_2 and γ directly imply that, completing the calculation for the first term of (13).

We now turn to the second term of (13). Again using the inequality $\ln(x) \leqslant x-1, \forall x > 0$, we get:

$$\begin{aligned}
\frac{n_2-d-2}{2}\ln\left(\frac{1-\|s\|_2^2}{1-s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s}\right) &< \frac{n_2}{2}\left(\frac{1-\|s\|_2^2}{1-s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s} - 1\right) \\
&= \frac{n_2}{2}\cdot\frac{s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)s}{1-s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s} \\
&\leqslant n_2 s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)s \\
&\leqslant \frac{\varepsilon}{4},
\end{aligned}$$

where the last two inequalities follow from applications of the assumptions $s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s \leqslant \frac{1}{2}$ and $s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)s \leqslant \frac{\varepsilon}{4n_2}$, respectively. ■

The next steps focus on upper-bounding the probability of either of the two conditions of Lemma 51 failing. For the first of the two conditions, it is possible to derive the bound using Beta concentration (Fact 31), as we did in the univariate case. However, the second condition cannot be handled that way, if we want to get a result with optimal sample complexity (see Appendix G for a detailed discuss as to why this is the case). For that reason, our argument will have to go through the Hanson-Wright inequality (Fact 28). To facilitate the implementation of the argument, we give the following auxiliary lemma.

Lemma 52 *Assume we are in the setting of Lemma 48. Let $Q \in \mathbb{R}^{d \times n_2}$ be the projection matrix which, when acting on a vector in $\mathbb{R}^{n_2}$, keeps its first d components. Finally, let $A := Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) Q - \frac{\varepsilon}{4n_2} \mathbb{I}_{n_2 \times n_2}$, where we assume that $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} \succeq \mathbb{I}_{d \times d}$. Then, we have:*

- $\operatorname{tr}(A) \leqslant \frac{3}{2}\gamma - \frac{\varepsilon}{4} < 0$,
- $\|A\|_2 \leqslant \frac{\varepsilon}{4n_2} \left(\frac{128e^2\lambda_0}{3} - 1\right)$,
- $\|A\|_F^2 \leqslant \frac{d\varepsilon^2}{16n_2^2} \left[\left(\frac{48e^2\lambda_0}{d\varepsilon} - 1\right)^2 + \left(\frac{n_2}{d} - 1\right)\right]$.

Proof We start by noting that, by definition, it must be the case that $Q = \left(\mathbb{I}_{d \times d} \middle| 0_{d \times (n_2 - d)}\right)$. This implies that $QQ^\top = \mathbb{I}_{d \times d}$. Using this observation, we proceed to bound each of the three quantities of interest separately. Starting with the trace, we have:

$$
\begin{aligned}
\operatorname{tr}(A) &= \operatorname{tr}\left(Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) Q - \frac{\varepsilon}{4n_2} \mathbb{I}_{n_2 \times n_2}\right) \\
&= \operatorname{tr}\left(\left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) QQ^\top\right) - \frac{\varepsilon}{4n_2} \operatorname{tr}\left(\mathbb{I}_{n_2 \times n_2}\right) \\
&= \operatorname{tr}\left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) - \frac{\varepsilon}{4} \\
&\leqslant \frac{3}{2}\gamma - \frac{\varepsilon}{4},
\end{aligned}
$$

where, along the way, we appealed to the linearity and cyclic properties of the trace, as well as (15). The fact that the above bound is < 0 follows from our assumption about n_2 (recall that $\gamma := \frac{8e^2\lambda_0}{n_2}$).

Moving on to the spectral norm, its definition for symmetric matrices yields:

$$
\begin{aligned}
\|A\|_2 &= \sup_{v \in \mathbb{S}^{n_2 - 1}} \left| v^\top \left[Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) Q - \frac{\varepsilon}{4n_2} \mathbb{I}_{n_2 \times n_2}\right] v \right| \\
&= \sup_{v \in \mathbb{S}^{n_2 - 1}} \left| (Qv)^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) Qv - \frac{\varepsilon}{4n_2} \right| \\
&= \sup_{v \in \mathbb{S}^{n_2 - 1}} \left| v_{\leqslant d}^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) v_{\leqslant d} - \frac{\varepsilon}{4n_2} \right| \\
&= \max \left\{ \begin{array}{l} \sup\limits_{v \in \mathbb{S}^{n_2 - 1}} \left\{ v_{\leqslant d}^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) v_{\leqslant d} \right\} - \frac{\varepsilon}{4n_2}, \\ \frac{\varepsilon}{4n_2} - \inf\limits_{v \in \mathbb{S}^{n_2 - 1}} \left\{ v_{\leqslant d}^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) v_{\leqslant d} \right\} \end{array} \right\}. \qquad (17)
\end{aligned}
$$

We focus on the first term in the above $\max$. By assumption, we have $\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d} \succeq 0$. In this case, the worst-case upper bound is obtained when $\|v_{\leqslant d}\|_2 = 1$ and $v_{>d} = 0$. Consequently, in the following, we can treat v as a vector in $\mathbb{R}^d$ instead. We have:

$$\sup_{v\in\mathbb{S}^{d-1}}\left\{v^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)v\right\} - \frac{\varepsilon}{4n_2} = \sup_{v\in\mathbb{S}^{d-1}}\left\{v^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}v\right\} - \left(1 + \frac{\varepsilon}{4n_2}\right) \leqslant \frac{4}{3}\gamma - \frac{\varepsilon}{4n_2}, \tag{18}$$

where we relied on Lemma 39, properties of the PSD ordering (Fact 20) and our bounds on n_2 and ε to argue that:

$$\begin{aligned}(1-\gamma)\,\widehat{\Sigma} \preceq \widehat{\Sigma}' \preceq \frac{1}{1-\gamma}\widehat{\Sigma} &\iff (1-\gamma)\,\widehat{\Sigma}^{-1} \preceq \widehat{\Sigma}'^{-1} \preceq \frac{1}{1-\gamma}\widehat{\Sigma}^{-1} \\ &\iff (1-\gamma)\,\mathbb{I} \preceq \widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} \preceq \frac{1}{1-\gamma}\mathbb{I},\end{aligned} \tag{19}$$

and that $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$.

For the other term in (17), the assumption $\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} \succeq \mathbb{I}_{d\times d}$ implies that the worst-case upper bound is obtained when $v_{\leqslant d} = 0$ and $\|v_{>d}\|_2 = 1$. This yields the upper bound:

$$\frac{\varepsilon}{4n_2} - \inf_{v\in\mathbb{S}^{n_2-1}}\left\{v_{\leqslant d}^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)v_{\leqslant d}\right\} \leqslant \frac{\varepsilon}{4n_2}. \tag{20}$$

Comparing (18) with (20), we get $\frac{\varepsilon}{4n_2} \leqslant \frac{4}{3}\gamma - \frac{\varepsilon}{4n_2} \iff \varepsilon \leqslant \frac{64}{3}e^2\lambda_0$, which holds by our assumptions on the range of ε and λ_0. Going back to (17), we get:

$$\|A\|_2 \leqslant \frac{4}{3}\gamma - \frac{\varepsilon}{4n_2} = \frac{\varepsilon}{4n_2}\left(\frac{16}{3}\cdot\frac{n_2\gamma}{\varepsilon} - 1\right) = \frac{\varepsilon}{4n_2}\left(\frac{128e^2\lambda_0}{3} - 1\right).$$

Finally, we prove the upper bound on $\|A\|_F^2$. Let $\{\lambda_i\}_{i\in[d]}$ be the spectrum of $\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}$. The definition of the Frobenius norm and the cyclic property of the trace yield:

$$\begin{aligned}\|A\|_F^2 &= \operatorname{tr}\left(\left[Q^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)Q - \frac{\varepsilon}{4n_2}\mathbb{I}_{n_2\times n_2}\right]^2\right) \\ &= \operatorname{tr}\left(Q^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)QQ^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)Q\right) \\ &\quad - \frac{\varepsilon}{2n_2}\operatorname{tr}\left(Q^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)Q\right) + \frac{\varepsilon^2}{16n_2^2}\operatorname{tr}\left(\mathbb{I}_{n_2\times n_2}\right) \\ &\overset{(a)}{=} \operatorname{tr}\left(\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right)^2\right) - \frac{\varepsilon}{2n_2}\operatorname{tr}\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) + \frac{\varepsilon^2}{16n_2} \\ &= \sum_{i\in[d]}(\lambda_i - 1)^2 - \frac{\varepsilon}{2n_2}\sum_{i\in[d]}(\lambda_i - 1) + \frac{\varepsilon^2}{16n_2} \\ &= \sum_{i\in[d]}\left[(\lambda_i - 1)^2 - \frac{\varepsilon}{2n_2}(\lambda_i - 1) + \frac{\varepsilon^2}{16n_2^2}\right] - \frac{d\varepsilon^2}{16n_2^2} + \frac{\varepsilon^2}{16n_2} \\ &= \sum_{i\in[d]}\left[(\lambda_i - 1) - \frac{\varepsilon}{16n_2}\right]^2 + \frac{\varepsilon^2}{16n_2^2}(n_2 - d),\end{aligned} \tag{21}$$

where (a) used that $QQ^\top = \mathbb{I}_{d\times d}$ and the cyclic property of the trace.

Our goal now is to derive a worst-case upper bound on (21). By assumption, we have that $\lambda_i \geqslant 1, \forall i \in [d]$. Due to this, Lemmas 39 and 25 imply:

$$\begin{aligned}\left\|\widehat{\Sigma}^{-\frac{1}{2}}\widehat{\Sigma}'\widehat{\Sigma}^{-\frac{1}{2}} - \mathbb{I}\right\|_{\mathrm{tr}} \leqslant \gamma\left(1+2\gamma\right) &\implies \left\|\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right\|_{\mathrm{tr}} \leqslant \gamma\left(1+2\gamma\right) \\ &\iff \sum_{i\in[d]}\left(\lambda_i - 1\right) \leqslant \gamma\left(1+2\gamma\right) \leqslant \frac{3}{2}\gamma, \end{aligned} \tag{22}$$

where the last inequality used the assumption that n_2 is large enough to ensure $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$.

Combining (21) and (22), along with the assumption $\lambda_i \geqslant 1, \forall i \in [d]$, we get a constrained convex maximization problem that is solvable exactly via the KKT conditions. The upper bound we obtain that way is:

$$\begin{aligned}\|A\|_F^2 \leqslant d\left(\frac{3\gamma}{2d} - \frac{\varepsilon}{4n_2}\right)^2 + \frac{\varepsilon^2}{16n_2^2}\left(n_2 - d\right) &= d\left(\frac{3}{2d}\cdot\frac{8e^2\lambda_0}{n_2} - \frac{\varepsilon}{4n_2}\right)^2 + \frac{\varepsilon^2}{16n_2^2}\left(n_2 - d\right) \\ &= \frac{d}{n_2^2}\left(\frac{12e^2\lambda_0}{d} - \frac{\varepsilon}{4}\right)^2 + \frac{\varepsilon^2}{16n_2^2}\left(n_2 - d\right) \\ &= \frac{d\varepsilon^2}{16n_2^2}\left[\left(\frac{48e^2\lambda_0}{d\varepsilon} - 1\right)^2 + \left(\frac{n_2}{d} - 1\right)\right],\end{aligned}$$

completing the calculation. ■

As a consequence of the previous lemma, we obtain the following corollary:

Corollary 53 *In the setting of Lemmas 48 and 52, we have* $\min\left\{\frac{-\mathrm{tr}(A)}{\|A\|_2}, \frac{\mathrm{tr}(A)^2}{\|A\|_F^2}\right\} \geqslant \frac{3C_2}{256e^2}\log\left(\frac{1}{\delta}\right)$.

Proof We will show separately for each term that it is lower-bounded by the target bound. For the first term, Lemma 52, as well as our bound on n_2 and the definition of γ, imply:

$$\begin{aligned}\frac{-\mathrm{tr}\left(A\right)}{\|A\|_2} \geqslant \frac{\frac{\varepsilon}{4} - \frac{3}{2}\gamma}{\frac{\varepsilon}{4n_2}\left(\frac{128e^2\lambda_0}{3}\right) - 1} = \frac{n_2 - \frac{6n_2\gamma}{\varepsilon}}{\frac{128e^2\lambda_0}{3} - 1} = \frac{n_2 - \frac{48e^2\lambda_0}{\varepsilon}}{\frac{128e^2\lambda_0}{3} - 1} &\geqslant \frac{C_2\frac{\lambda_0\log\left(\frac{1}{\delta}\right)}{\varepsilon} - \frac{48e^2\lambda_0}{\varepsilon}}{\frac{128e^2\lambda_0}{3}} \\ &\overset{(a)}{\geqslant} \frac{\frac{C_2}{2}\cdot\frac{\lambda_0\log\left(\frac{1}{\delta}\right)}{\varepsilon}}{\frac{128e^2\lambda_0}{3}} \\ &= \frac{3C_2}{256e^2}\cdot\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon} \\ &\overset{(b)}{\geqslant} \frac{3C_2}{256e^2}\log\left(\frac{1}{\delta}\right),\end{aligned} \tag{23}$$

where (a) relies on the assumption that C_2 is appropriately large, and (b) uses that $\varepsilon \leqslant 1$.

For the second term, working similarly to before yields:

$$\begin{aligned}\frac{\operatorname{tr}(A)^2}{\|A\|_F^2} \geqslant \frac{\left(\frac{\varepsilon}{4}-\frac{3}{2}\gamma\right)^2}{\frac{d\varepsilon^2}{16n_2^2}\left[\left(\frac{48e^2\lambda_0}{d\varepsilon}-1\right)^2+\left(\frac{n_2}{d}-1\right)\right]} &= \frac{\left(n_2-\frac{6n_2\gamma}{\varepsilon}\right)^2}{d\left[\left(\frac{48e^2\lambda_0}{d\varepsilon}-1\right)^2+\left(\frac{n_2}{d}-1\right)\right]} \\ &= \frac{\left(n_2-\frac{48e^2\lambda_0}{\varepsilon}\right)^2}{d\left[\left(\frac{48e^2\lambda_0}{d\varepsilon}-1\right)^2+\left(\frac{n_2}{d}-1\right)\right]}. \end{aligned} \tag{24}$$

We bound the numerator and the denominator of (24) separately. For the numerator, by our bound on n_2, we get that:

$$n_2 - \frac{48e^2\lambda_0}{\varepsilon} \geqslant \frac{1}{2}n_2. \tag{25}$$

For the denominator, again by our bound on n_2, we have:

$$\left(\frac{48e^2\lambda_0}{d\varepsilon}-1\right)^2+\left(\frac{n_2}{d}-1\right) \leqslant \left(\frac{48e^2}{C_2}\cdot\frac{n_2}{d\log\left(\frac{1}{\delta}\right)}\right)^2+\frac{n_2}{d}. \tag{26}$$

Bounding (24) using (25) and (26), we get:

$$\frac{\operatorname{tr}(A)^2}{\|A\|_F^2} \geqslant \frac{\left(\frac{1}{2}n_2\right)^2}{d\left[\left(\frac{48e^2}{C_2}\cdot\frac{n_2}{d\log\left(\frac{1}{\delta}\right)}\right)^2+\frac{n_2}{d}\right]} \geqslant \frac{\frac{1}{4}}{\left(\frac{48e^2}{C_2}\right)^2\cdot\frac{1}{d\log^2\left(\frac{1}{\delta}\right)}+\frac{1}{n_2 d}}. \tag{27}$$

We have again for the denominator:

$$\begin{aligned} & n_2 \geqslant C_2\frac{\lambda_0\log\left(\frac{1}{\delta}\right)}{\varepsilon} \\ \iff & \frac{1}{n_2 d} \leqslant \frac{\varepsilon}{C_2\lambda_0 d\log\left(\frac{1}{\delta}\right)} \leqslant \frac{1}{C_2 d^2\log\left(\frac{1}{\delta}\right)} \\ \implies & \left(\frac{48e^2}{C_2}\right)^2\cdot\frac{1}{d\log^2\left(\frac{1}{\delta}\right)}+\frac{1}{n_2 d} \leqslant \left[\left(\frac{48e^2}{C_2}\right)^2+\frac{1}{C_2}\right]\frac{1}{d\log\left(\frac{1}{\delta}\right)} \leqslant \frac{2}{C_2}\cdot\frac{1}{d\log\left(\frac{1}{\delta}\right)}, \end{aligned} \tag{28}$$

where the last inequality again relied on C_2 being appropriately large.

Thus, lower-bounding (27) using (28) yields:

$$\frac{\operatorname{tr}(A)^2}{\|A\|_F^2} \geqslant \frac{C_2}{8}d\log\left(\frac{1}{\delta}\right). \tag{29}$$

Taking into account (23) and (29) yields the desired result. ■

We now have everything we need to establish the main result of this section.

Proof [Proof of Proposition 16]. The proof will follow a structure similar to that of Proposition 15, i.e., establishing the desired upper bound amounts to upper-bounding the probability of the conditions identified in Lemma 51. Indeed, by Lemma 48, (8), and Lemma 51, we get that:

$$\begin{aligned}
&D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{\left(1-\frac{1}{n_1}\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1-\frac{1}{n_1}\right) n_2} W' z\right) \\
&\leqslant D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d} \| T\right) \\
&\leqslant \mathop{\mathbb{P}}_{t \sim z_{\leqslant d}}\left[f_{z_{\leqslant d}}(t) > e^{\frac{\varepsilon}{2}} f_T(t)\right] \\
&\leqslant \mathop{\mathbb{P}}_{s \sim z_{\leqslant d}}\left[\left\{s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s > \frac{1}{2}\right\} \cup \left\{s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) s > \frac{\varepsilon}{4 n_2}\right\}\right] \\
&\leqslant \mathop{\mathbb{P}}_{s \sim z_{\leqslant d}}\left[s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s > \frac{1}{2}\right] + \mathop{\mathbb{P}}_{s \sim z_{\leqslant d}}\left[s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d \times d}\right) s > \frac{\varepsilon}{4 n_2}\right],
\end{aligned} \tag{30}$$

where the fact that $s \sim z_{\leqslant d}$ follows from the definition of s (see Lemma 49) and rotational invariance for the uniform distribution over the unit sphere (Fact 33).

We will analyze each of the terms of (30) separately. For the first term, observe that, by properties of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30), $\|s\|_2^2 \sim \mathrm{Beta}\left(\frac{d}{2}, \frac{n_2-d}{2}\right)$. Thus, by the previous and Beta concentration (Fact 31), we get:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{s \sim z_{\leqslant d}}\left[s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s > \frac{1}{2}\right] \\
&\leqslant \mathop{\mathbb{P}}_{s \sim z_{\leqslant d}}\left[\|s\|_2^2 > \frac{3}{8}\right] \\
&\leqslant 2 \exp\left(-C_4 \min\left\{\frac{(n_2-d)^2}{2d}\left(\frac{3}{8} - \frac{d}{n_2}\right)^2, \frac{n_2-d}{2}\left(\frac{3}{8} - \frac{d}{n_2}\right)\right\}\right) \\
&\overset{(a)}{=} 2 \exp\left(-C_4 \frac{n_2-d}{2}\left(\frac{3}{8} - \frac{d}{n_2}\right)\right) \\
&\overset{(b)}{\leqslant} \frac{\delta}{2},
\end{aligned} \tag{31}$$

where (a) and (b) are implied by our bound on n_2 (assuming C_2 to be large enough).

We now turn our attention to the second term of (30). To analyze this term, we will make two assumptions, without loss of generality. The first assumption is that z is generated according to the method of Fact 34 (i.e., by normalizing a standard Gaussian random vector). We note that this is a reasonable assumption to make. Indeed, it is always possible to replace z in $D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}' + \sqrt{\left(1-\frac{1}{n_1}\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1-\right.}\right.$ with a z' that has been generated according to that method, and the value of the divergence term would still be the same, since z and z' have the same density function. Moreover, it is always within our power as algorithm designers to assume that, when z is drawn in Line 8, it is generated by a subroutine that implements the method (especially given the fact that the only resources we are looking to minimize are sample and time complexity, but not randomness complexity).

The second assumption we will make is that $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} \succeq \mathbb{I}_{d \times d}$. Again, this is reasonable because, if $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}$ had an eigenvalue that is < 1, this would have a negative contribution to the

LHS of $s^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) s > \frac{\varepsilon}{4n_2}$, thus making it harder to satisfy the inequality (i.e., the conditions imposed on s would have to be more restrictive). Thus, our assumption can only lead to an increase in the probability of the condition being satisfied, which justifies our choice.

Under the above two assumptions, we note that there will exist a random vector $G \sim \mathcal{N}(0,1)^{\otimes n_2}$ such that $s_i = \frac{G_i}{\|G\|_2}, \forall i \in [d]$.[9] This follows from the fact that $s \sim z_{\leqslant d}$. Furthermore, for $Q := \left(\mathbb{I}_{d\times d}\middle|0_{d\times(n_2-d)}\right)$ (as in Lemma 52), we get $s = Q\frac{G}{\|G\|_2}$. We substitute based on this in the second term of (30), and write $A := Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) Q - \frac{\varepsilon}{4n_2}\mathbb{I}_{n_2\times n_2}$ (again as in Lemma 52 and Corollary 53). Then, the Hanson-Wright inequality (Fact 28) and Corollary 53 yield:

$$
\begin{aligned}
&\underset{s\sim z_{\leqslant d}}{\mathbb{P}}\left[s^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) s > \frac{\varepsilon}{4n_2}\right] \\
&= \underset{G\sim\mathcal{N}(0,1)^{\otimes n_2}}{\mathbb{P}}\left[\frac{G^\top}{\|G\|_2} Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) Q\frac{G}{\|G\|_2} > \frac{\varepsilon}{4n_2}\right] \\
&= \underset{G\sim\mathcal{N}(0,1)^{\otimes n_2}}{\mathbb{P}}\left[G^\top Q^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}_{d\times d}\right) QG > \frac{\varepsilon}{4n_2}\|G\|_2^2\right] \\
&= \underset{G\sim\mathcal{N}(0,1)^{\otimes n_2}}{\mathbb{P}}\left[G^\top A G > 0\right] \\
&= \underset{G\sim\mathcal{N}(0,1)^{\otimes n_2}}{\mathbb{P}}\left[G^\top A G - \operatorname{tr}(A) > -\operatorname{tr}(A)\right] \\
&\leqslant \underset{G\sim\mathcal{N}(0,1)^{\otimes n_2}}{\mathbb{P}}\left[\left|G^\top A G - \operatorname{tr}(A)\right| \geqslant -\operatorname{tr}(A)\right] \\
&\leqslant 2\exp\left(-C_3 \min\left\{\frac{\operatorname{tr}(A)^2}{\|A\|_F^2}, \frac{-\operatorname{tr}(A)}{\|A\|_2}\right\}\right) \\
&\leqslant 2\exp\left(-\frac{3C_2C_3}{256e^2}\log\left(\frac{1}{\delta}\right)\right) \\
&\leqslant \frac{\delta}{2},
\end{aligned} \tag{32}
$$

where, in the last inequality, we assume that C_2 is appropriately large.

Upper-bounding (30) using (31) and (32) completes the proof. ■

Appendix G. Alternative Version of Proposition 16

In this appendix, we present our original version and proof of Proposition 16. We start by sketching a naïve approach, and explaining why it fails, thus motivating the proof that we will present here.

The steps up to Lemma 51 are the same as in Section 3.2.2. Our goal from this point on is to identify a sufficient condition that implies:

$$
s^\top \left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right) s \leqslant \frac{\varepsilon}{4n_2}. \tag{33}
$$

9. This gives some additional insight as to why a random linear combination with uniformly distributed vector on the unit sphere yields the desired privacy guarantee. Considering a random linear combination can be interpreted as adding normalized Gaussian noise which has been rescaled using a subset of the input datapoints.

Observe that, by (19), we have:

$$(1-\gamma)\,\mathbb{I} \preceq \widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} \preceq \frac{1}{1-\gamma}\mathbb{I} \iff -\gamma\mathbb{I} \preceq \widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I} \preceq \frac{\gamma}{1-\gamma}\mathbb{I}$$
$$\implies \left\|\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right\|_2 \leqslant \frac{\gamma}{1-\gamma}. \tag{34}$$

Based on the above, a naïve approach involves observing that, by the definition of the spectral norm for symmetric matrices and the fact that $\gamma := \frac{8e^2\lambda_0}{n_2} \leqslant \frac{1}{4}$, we have:

$$s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right)s \leqslant \left\|\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right\|_2 \|s\|_2^2 \leqslant \frac{\gamma}{1-\gamma}\|s\|_2^2 \leqslant \frac{4}{3}\gamma\|s\|_2^2 = \frac{32e^2\lambda_0}{3n_2}\|s\|_2^2. \tag{35}$$

Thus, a sufficient condition that would lead to (33) being satisfied is $\|s\|_2^2 \leqslant \frac{3\varepsilon}{128e^2\lambda_0}$. However, issues arise when upper-bounding the probability of this condition failing. Specifically, properties of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) yield $\|s\|_2^2 \sim \text{Beta}\left(\frac{d}{2}, \frac{n_2-d}{2}\right)$, so the natural next step here is to apply Beta concentration (Fact 31). However, that requires $\frac{3\varepsilon}{128e^2\lambda_0} > \frac{d}{n_2} \iff n_2 \geqslant \frac{128e^2\lambda_0 d}{3\varepsilon}$. Since $\lambda_0 \geqslant d$, this leads to a sample complexity bound that is sub-optimal in the dimension dependence by a quadratic factor, which is clearly prohibitive.

The failure of the naïve approach described above prompts us to re-examine (35). Observe that the way we leveraged the definition of the spectral norm in the upper bound is not tight in general. Indeed, for the inequality to be tight, it must be the case that s is parallel to the eigenvector of $\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}$ that corresponds to the largest in absolute value eigenvalue. This significantly restricts the realizations of s, which is a factor that was not taken into account in our analysis.

At this point, we recall the second consequence of Lemma 39. We denote the spectrum of $\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}$ by $\{\lambda_i\}_{i\in[d]}$. Without loss of generality, we can assume that $\lambda_i \geqslant 1, \forall i \in [d]$. Indeed, having an eigenvalue that is < 1 would make it easier to satisfy (33), since the corresponding term would contribute negatively to the LHS. Then, (22) yields:

$$\sum_{i\in[d]}(\lambda_i - 1) \leqslant \gamma\,(1+2\gamma)\,. \tag{36}$$

By our bound on n_2, we know that $1+2\gamma = \Theta(1)$, yielding $\sum\limits_{i\in[d]}(\lambda_i - 1) \leqslant \Theta(\gamma)$. Thus, the above implies that not all terms $\lambda_i - 1$ can be $\approx \gamma$. Two extreme cases would involve having either one large term $\lambda_i - 1 = \Theta(\gamma)$ and all the other terms being negligible, or having most (if not all) terms being $\Theta\left(\frac{\gamma}{d}\right)$. We stress that using the definition of the spectral norm as we did in (35) is tight only in the latter case, the former requiring us to reason about the angle between s and the eigenvector that corresponds to the eigenvalue for which $\lambda_i - 1 = \Theta(\gamma)$.[10]

Based on the above intuition, our proof will try to interpolate between the two previous regimes. We ignore the smallest eigenvalues (these would contribute little to $s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right)s$ anyway), and partition the rest into sets that consist of eigenvalues whose magnitude is within a constant

10. We note that this observation is standard, and can be interpreted as a strong form of the pigeonhole principle. That is, consider an example where we have numbers $x_1, \ldots, x_n > 0$ such that $\sum\limits_{i\in[n]} x_i = m$. Then, the two extreme cases are that there must either be multiple numbers which are $\Omega\left(\frac{m}{n}\right)$ or one number which is $\Omega(m)$. Recently, this trick has appeared in the context of the privacy literature in the proof of Lemma 4.9 in Agarwal et al. (2025).

factor of each other. Then, we get different conditions on different components of s, depending what the magnitude of the corresponding eigenvalue is.[11] This leads to a sample complexity of $\widetilde{\mathcal{O}}\left(\lambda_0\right)$ and no explicit dependence on d, thus resolving the previous issue.

We break the proof into two statements. The first statement can be considered to be analogous to Lemma 51. Indeed, we showed in Lemma 51 that:

$$s^\top \widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} s \leqslant \frac{1}{2} \text{ and } s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right) s \leqslant \frac{\varepsilon}{4n_2} \implies \ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_T(t)}\right) \leqslant \frac{\varepsilon}{2}.$$

Now, we will identify a set of sufficient conditions on s that imply $s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right) s \leqslant \frac{\varepsilon}{4n_2}$. In accordance with the sketch given above, we set $\ell_{\max} := \left\lceil \log\left(\frac{8n_2\gamma(1+2\gamma)}{\varepsilon}\right)\right\rceil = \left\lceil \log\left(\frac{64e^2\lambda_0(1+2\gamma)}{\varepsilon}\right)\right\rceil$ (recall that $\gamma := \frac{8e^2\lambda_0}{n_2} \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$), and define the following partition of the indices $i \in [d]$:

$$\mathcal{B}_\ell := \left\{ i \in [d] : \frac{\gamma(1+2\gamma)}{2^\ell} < \lambda_i - 1 \leqslant \frac{\gamma(1+2\gamma)}{2^{\ell-1}} \right\}, \forall \ell \in [\ell_{\max}],$$

$$\mathcal{B}_0 := [d] \setminus \bigcup_{\ell \in [\ell_{\max}]} \mathcal{B}_\ell.$$

The key property of this partition is that:

$$|\mathcal{B}_\ell| < 2^\ell, \forall \ell \in [\ell_{\max}], \tag{37}$$

since, otherwise, (36) would be violated.

We now give the formal statement and proof of our lemma.

Lemma 54 *Let $\{v_i\}_{i\in[d]}$ be an orthonormal set of eigenvectors that correspond to the spectrum $\{\lambda_i\}_{i\in[d]}$ of $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}$. Then, for vectors $s \in \mathbb{R}^d$ with $\|s\|_2 \leqslant 1$, we have:*

$$\sum_{i \in \mathcal{B}_\ell} \langle s, v_i \rangle^2 \leqslant \frac{2^{\ell-1}\varepsilon}{96e^2\lambda_0\ell_{\max}}, \forall \ell \in [\ell_{\max}] \implies s^\top \left(\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}} - \mathbb{I}\right) s \leqslant \frac{\varepsilon}{4n_2}.$$

Proof Using the spectral decomposition (Fact 22), the target inequality can be written as:

$$\sum_{i\in[d]} (\lambda_i - 1) \langle v_i, s \rangle^2 \leqslant \frac{\varepsilon}{4n_2}. \tag{38}$$

We will leverage the assumption to upper-bound the LHS, and show that the resulting bound does not exceed the RHS. To obtain the upper bound, we note that, for every $i \in \mathcal{B}_0$, we must have $\lambda_i - 1 \leqslant \frac{\varepsilon}{8n_2}$. Using this observation and the definition of the sets $\mathcal{B}_\ell, \forall \ell \in [\ell_{\max}]$, we get:

$$\begin{aligned}
\sum_{i\in[d]} (\lambda_i - 1) \langle v_i, s \rangle^2 &= \sum_{\ell \in [\ell_{\max}]} \sum_{i \in \mathcal{B}_\ell} (\lambda_i - 1) \langle s, v_i \rangle^2 + \sum_{i \in \mathcal{B}_0} (\lambda_i - 1) \langle s, v_i \rangle^2 \\
&\leqslant \sum_{\ell \in [\ell_{\max}]} \sum_{i \in \mathcal{B}_\ell} (\lambda_i - 1) \langle s, v_i \rangle^2 + \frac{\varepsilon}{8n_2} \sum_{i \in \mathcal{B}_0} \langle s, v_i \rangle^2 \\
&\leqslant \gamma(1+2\gamma) \sum_{\ell \in [\ell_{\max}]} \sum_{i \in \mathcal{B}_\ell} \frac{\langle s, v_i \rangle^2}{2^{\ell-1}} + \frac{\varepsilon}{8n_2},
\end{aligned} \tag{39}$$

11. By components of s here we refer to its projections along the eigenbasis of $\widehat{\Sigma}^{\frac{1}{2}} \widehat{\Sigma}'^{-1} \widehat{\Sigma}^{\frac{1}{2}}$. Implicit here is a change of basis argument.

where we used the fact that $\{v_i\}_{i\in[d]}$ is an orthonormal basis of $\mathbb{R}^d$ to argue that:

$$\sum_{i\in\mathcal{B}_0}\langle s,v_i\rangle^2 \leqslant \sum_{i\in[d]}\langle s,v_i\rangle^2 = \|s\|_2^2 \leqslant 1.$$

Thus, based on (39), to satisfy (38), it suffices to have:

$$\sum_{\ell\in[\ell_{\max}]}\sum_{i\in\mathcal{B}_\ell}\frac{\langle s,v_i\rangle^2}{2^{\ell-1}} \leqslant \frac{\varepsilon}{8\gamma\left(1+2\gamma\right)n_2} = \frac{\varepsilon}{64e^2\lambda_0\left(1+2\gamma\right)},$$

where the equality used the fact that $\gamma := \frac{8e^2\lambda_0}{n_2}$.

Our assumption suffices for this to hold. Indeed, we have:

$$\sum_{\ell\in[\ell_{\max}]}\sum_{i\in\mathcal{B}_\ell}\langle s,v_i\rangle^2 \leqslant \frac{2^{\ell-1}\varepsilon}{96e^2\lambda_0} \leqslant \frac{\varepsilon}{64e^2\lambda_0\left(1+2\gamma\right)},$$

where the last inequality follows from the assumption that $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$. ■

We are now ready to prove our original version of Proposition 16. The proof has virtually the same structure as for that proposition. However, instead of upper-bounding the probability of $s^\top\left(\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}-\mathbb{I}\right)s > \frac{\varepsilon}{4n_2}$ occurring (which would require the argument of Section 3.2.2), we will instead use Beta concentration (Fact 31) to upper-bound the probability of the event that corresponds to the assumption in Lemma 54 failing.

Proposition 55 *Let $\varepsilon\in[0,1]$ and $\delta\in\left[0,\frac{\varepsilon}{10}\right]$, $\lambda_0\geqslant d$. Let us assume that:*

$$n_2 \geqslant C_2\frac{\lambda_0\log\left(\lambda_0\right)\left(\log\left(\log\left(\lambda_0\right)\right)+\log\left(\frac{1}{\delta}\right)\right)}{\varepsilon},$$

for some appropriately large absolute constant $C_2\geqslant 1$. Finally, let $X,X'\in\mathbb{R}^{n\times d}$ be adjacent datasets such that $\Psi(X)=\Psi(X')=\texttt{Pass}$. Then, for $z\sim\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:

$$D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\middle\|\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}W'z\right)\leqslant\delta.$$

Proof As in the proof of Proposition 16, we have:

$$\begin{aligned}
&D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\middle\|\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}W'z\right)\\
&\leqslant D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d}\|T\right)\\
&\leqslant \mathop{\mathbb{P}}_{t\sim z_{\leqslant d}}\left[f_{z_{\leqslant d}}(t) > e^{\frac{\varepsilon}{2}}f_T(t)\right]\\
&\leqslant \mathop{\mathbb{P}}_{s\sim z_{\leqslant d}}\left[\left\{s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s > \frac{1}{2}\right\}\cup\left\{\exists\ell\in[\ell_{\max}] : \sum_{i\in\mathcal{B}_\ell}\langle s,v_i\rangle^2 > \frac{2^{\ell-1}\varepsilon}{96e^2\lambda_0\ell_{\max}}\right\}\right]\\
&\leqslant \mathop{\mathbb{P}}_{s\sim z_{\leqslant d}}\left[s^\top\widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}'^{-1}\widehat{\Sigma}^{\frac{1}{2}}s > \frac{1}{2}\right] + \sum_{\ell\in[\ell_{\max}]}\mathop{\mathbb{P}}_{s\sim z_{\leqslant d}}\left[\sum_{i\in\mathcal{B}_\ell}\langle s,v_i\rangle^2 > \frac{2^{\ell-1}\varepsilon}{96e^2\lambda_0\ell_{\max}}\right].
\end{aligned} \tag{40}$$

The process to upper-bound the first term is exactly the same as the one we followed in the proof of Proposition 16, so we point readers to (31), and do not repeat it here. It remains to show that the sum over ℓ in (40) is upper-bounded by $\frac{\delta}{2}$. To establish that, we will show that individual terms of the sum are upper-bounded by $\frac{\delta}{2\ell_{\max}}$.

We note that, for any $\ell \in [\ell_{\max}]$, the sum $\sum_{i \in \mathcal{B}_\ell} \langle s, v_i \rangle^2$ is distributed as $\text{Beta}\left(\frac{|\mathcal{B}_\ell|}{2}, \frac{n_2 - |\mathcal{B}_\ell|}{2}\right)$. This follows from the density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ and rotational invariance (Facts 30 and 33).. Observe that the mean of the distribution $\text{Beta}\left(\frac{|\mathcal{B}_\ell|}{2}, \frac{n_2 - |\mathcal{B}_\ell|}{2}\right)$ is $\frac{|\mathcal{B}_\ell|}{n_2}$. Because of (37), this is upper-bounded by $\frac{2^\ell}{n_2}$ for all $\ell \in [\ell_{\max}]$. By our sample complexity bound, this is at most $\frac{2^{\ell-1}\varepsilon}{192 e^2 \lambda_0 \ell_{\max}}$, $\forall \ell \in [\ell_{\max}]$. Thus, we can appeal to Beta concentration (Fact 31), and obtain the bound:

$$
\begin{aligned}
&\underset{s \sim z_{\leqslant d}}{\mathbb{P}}\left[\sum_{i \in \mathcal{B}_\ell} \langle s, v_i \rangle^2 > \frac{2^{\ell-1}\varepsilon}{96 e^2 \lambda_0 \ell_{\max}}\right] \\
&\leqslant 2\exp\left(-C_4 \min\left\{\frac{\left(n_2 - 2^\ell\right)^2}{2 \cdot 2^\ell}\left(\frac{2^{\ell-1}\varepsilon}{96 e^2 \lambda_0 \ell_{\max}} - \frac{2^\ell}{n_2}\right)^2, \frac{n_2 - 2^\ell}{2}\left(\frac{2^{\ell-1}\varepsilon}{96 e^2 \lambda_0 \ell_{\max}} - \frac{2^\ell}{n_2}\right)\right\}\right) \\
&\leqslant 2\exp\left(-C_4 \frac{n_2 - 2^\ell}{2} \cdot \frac{2^{\ell-1}\varepsilon}{192 e^2 \lambda_0 \ell_{\max}}\right) \\
&\leqslant \frac{\delta}{2\ell_{\max}},
\end{aligned}
$$

where the last inequality follows from our bound on n_2. ■

Appendix H. Omitted Proofs from Section 3.2.2

We start with the lemma that implements the first step sketched in Section 3.2.3.

Lemma 56 *Let $\varepsilon \in [0,1]$ and $\delta \in \left[0, \frac{\varepsilon}{10}\right]$, $\lambda_0 \geqslant d$. Also, assume that $n_1 \geqslant C_1 \frac{\sqrt{\lambda_0} \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ and $n_2 \geqslant C_2 \frac{\lambda_0 \log\left(\frac{1}{\delta}\right)}{\varepsilon}$ for appropriately large absolute constants $C_1, C_2 \geqslant 1$. Finally, for $n := n_1 + 2n_2$, let $X, X' \in \mathbb{R}^{n \times d}$ be adjacent datasets and $R \subseteq [n_1]$ be a representation set such that $\Psi(X, R) = \Psi(X', R) = 1$. Then, for $z \sim \mathcal{U}\left(\mathbb{S}^{n_2-1}\right)$, we have:*

$$
D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu} + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W z \middle\| \widehat{\mu}' + \sqrt{\left(1 - \frac{1}{n_1}\right) n_2} W z\right) = D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d} \| z_{\leqslant d} + \ell\right),
$$

for $\ell := \sqrt{\frac{n_1}{n_1 - 1} \cdot \frac{1}{n_2}} \Lambda^{-\frac{1}{2}} U^\top \left(\widehat{\mu}' - \widehat{\mu}\right)$ where $\Lambda \succ 0$ is a diagonal matrix and U is a rotation matrix such that $\widehat{\Sigma} = U \Lambda U^\top$. Furthermore, we have $\|\ell\|_2 \leqslant \frac{\sqrt{114 \lambda_0 e}}{n_1 \sqrt{n_2}}$.

Proof We start by repeating a number of arguments from the beginning of the proof of Lemma 48, namely the DPI-based argument of (7), and the SVD-based argument of (8). This yields:

$$\begin{aligned} & D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\middle\|\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\right) \\ &= D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d}\middle\|\sqrt{\frac{n_1}{n_1-1}\cdot\frac{1}{n_2}}\Lambda^{-\frac{1}{2}}U^\top\left(\widehat{\mu}'-\widehat{\mu}\right)+z_{\leqslant d}\right), \end{aligned} \tag{41}$$

where $U \in \mathbb{R}^{d\times d}$ is a rotation matrix and $\Lambda \in \mathbb{R}^{d\times d}$ is a diagonal positive-definite matrix such that $U\Lambda U^\top = WW^\top = \widehat{\Sigma}$. We set $\ell := \sqrt{\frac{n_1}{n_1-1}\cdot\frac{1}{n_2}}\Lambda^{-\frac{1}{2}}U^\top\left(\widehat{\mu}'-\widehat{\mu}\right)$. By our bounds on n_1 and n_2, and the assumption that $\Psi\left(X,R\right) = \Psi\left(X',R\right) = 1$, the conditions of Lemma 41 are satisfied, so the lemma yields:

$$\begin{aligned} \|\ell\|_2 = \sqrt{\frac{n_1}{n_1-1}\cdot\frac{1}{n_2}}\left\|\Lambda^{-\frac{1}{2}}U^\top\left(\widehat{\mu}'-\widehat{\mu}\right)\right\|_2 &\overset{(a)}{=} \sqrt{\frac{n_1}{n_1-1}\cdot\frac{1}{n_2}}\left\|\widehat{\mu}'-\widehat{\mu}\right\|_\Sigma \\ &\leqslant \sqrt{\frac{n_1}{n_1-1}}\cdot\frac{\sqrt{\left(1+2\gamma\right)38\lambda_0 e}}{n_1\sqrt{n_2}} \\ &\overset{(b)}{\leqslant} \frac{\sqrt{114\lambda_0 e}}{n_1\sqrt{n_2}}, \end{aligned}$$

where (a) used the observation that:

$$\left\|\Lambda^{-\frac{1}{2}}U^\top\left(\widehat{\mu}'-\widehat{\mu}\right)\right\|_2 = \sqrt{\left(\widehat{\mu}'-\widehat{\mu}\right)^\top U\Lambda^{-1}U^\top\left(\widehat{\mu}'-\widehat{\mu}\right)} = \sqrt{\left(\widehat{\mu}'-\widehat{\mu}\right)^\top\widehat{\Sigma}^{-1}\left(\widehat{\mu}'-\widehat{\mu}\right)} = \left\|\widehat{\mu}'-\widehat{\mu}\right\|_{\widehat{\Sigma}},$$

and (b) used our bounds on n_1, n_2, and ε to argue that $n_1 \geqslant 2 \iff \frac{n_1}{n_1-1} \leqslant 2$ and $\gamma \leqslant \frac{\varepsilon}{4} \leqslant \frac{1}{4}$. ∎

As with the other term of (1), we must now reason about the quantity:

$$D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d}\|z_{\leqslant d}+\ell\right) \leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[f_{z_{\leqslant d}}\left(t\right) > e^{\frac{\varepsilon}{2}}f_{z_{\leqslant d}+\ell}\left(t\right)\right]. \tag{42}$$

Thus, we need to identify the density of $z_{\leqslant d}+\ell$, and identify sufficient conditions for the log-density ratio to satisfy $\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_{z_{\leqslant d}+\ell}(t)}\right) \leqslant \frac{\varepsilon}{2}$. We do so in the following lemma:

Lemma 57 *In the setting of Lemma 56, we have that:*

$$\|t\|_2 \leqslant 0.9 \text{ and } \left|\langle t,\ell\rangle\right| \leqslant \frac{1}{50}\cdot\frac{\varepsilon}{n_2} \leqslant 0.01 \implies \ln\left(\frac{f_{z_{\leqslant d}}\left(t\right)}{f_{z_{\leqslant d}+\ell}\left(t\right)}\right) \leqslant \frac{\varepsilon}{2}.$$

Proof We start by identifying the density of $z_{\leqslant d}+\ell$. Standard theory about transformations of random vectors (Fact 32) yields that $f_{z_{\leqslant d}+\ell}\left(t\right) = f_{z_{\leqslant d}}\left(t-\ell\right)$. Additionally, by the conclusion of Lemma 56 and our bounds on n_1, n_2, and δ, we get:

$$\|\ell\|_2 \leqslant \frac{1}{5}\cdot\frac{\varepsilon}{\sqrt{n_2}} \leqslant 0.1. \tag{43}$$

We now work towards upper-bounding the log-density ratio $\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_{z_{\leqslant d}+\ell}(t)}\right)$ using our assumptions. We highlight that $z_{\leqslant d}$ is supported on the unit ball in d dimensions, whereas $z_{\leqslant d}+\ell$ is supported on the ball of radius 1 centered at the point ℓ (which (43) implies has small distance from the origin). Observe that, by our assumptions and (43), we have:

$$\|t-\ell\|_2^2 = \|t\|_2^2 - 2\langle t,\ell\rangle + \|\ell\|_2^2 \leqslant \|t\|_2^2 + 2\left|\langle t,\ell\rangle\right| + \|\ell\|_2^2 \leqslant 0.81 + 2\cdot 0.01 + 0.01 < 0.9. \quad (44)$$

The above implies that our conditions ensure that t will be within the support of $z_{\leqslant d}+\ell$, so $\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_{z_{\leqslant d}+\ell}(t)}\right)$ will be finite. Building on this observation, the density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) and the standard inequality $\ln(x) \leqslant x-1, \forall x>0$ yield for the log-density ratio:

$$\begin{aligned}
\ln\left(\frac{f_{z_{\leqslant d}}(t)}{f_{z_{\leqslant d}+\ell}(t)}\right) = \frac{n_2-d-2}{2}\ln\left(\frac{1-\|t\|_2^2}{1-\|t-\ell\|_2^2}\right) &< \frac{n_2}{2}\left(\frac{1-\|t\|_2^2}{1-\|t-\ell\|_2^2}-1\right)\\
&= \frac{n_2}{2}\cdot\frac{\|\ell\|_2^2 - 2\langle t,\ell\rangle}{1-\|t-\ell\|_2^2}\\
&\leqslant \frac{n_2}{2}\cdot\frac{\|\ell\|_2^2 + 2\left|\langle t,\ell\rangle\right|}{1-\|t-\ell\|_2^2}\\
&\overset{(a)}{\leqslant} \frac{n_2}{2}\cdot 10\left(\frac{1}{25}\cdot\frac{\varepsilon^2}{n_2}+\frac{1}{25}\cdot\frac{\varepsilon}{n_2}\right)\\
&\overset{(b)}{\leqslant} \frac{n_2}{2}\cdot 10\cdot\frac{2}{25}\cdot\frac{\varepsilon}{n_2}\\
&= \frac{10}{25}\varepsilon\\
&\leqslant \frac{\varepsilon}{2},
\end{aligned}$$

where (a) used (43) and (44), as well as our assumption about $\left|\langle t,\ell\rangle\right|$, and (b) used that $\varepsilon \leqslant 1$. ■

We conclude by proving the main result of the section.

Proof [Proof of Proposition 17]. We work as in the proofs of Propositions 15 and 16. Lemmas 56 and 57, as well as (42), yield:

$$\begin{aligned}
& D_{e^{\frac{\varepsilon}{2}}}\left(\widehat{\mu}+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\middle\|\widehat{\mu}'+\sqrt{\left(1-\frac{1}{n_1}\right)n_2}Wz\right) \\
&= D_{e^{\frac{\varepsilon}{2}}}\left(z_{\leqslant d}\|z_{\leqslant d}+\ell\right) \\
&\leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[f_{z_{\leqslant d}}(t) > e^{\frac{\varepsilon}{2}} f_{z_{\leqslant d}+\ell}(t)\right] \\
&\leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\{\|t\|_2 > 0.9\} \cup \left\{|\langle \ell, t\rangle| > \frac{1}{50}\cdot\frac{\varepsilon}{n_2}\right\}\right] \\
&\leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2 > 0.9\right] + \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[|\langle \ell, t\rangle| > \frac{1}{50}\cdot\frac{\varepsilon}{n_2}\right] \\
&= \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2 > 0.9\right] + \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\left|\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle\right| > \frac{1}{50}\cdot\frac{\varepsilon}{n_2}\cdot\frac{1}{\|\ell\|_2}\right] \\
&\leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2 > 0.9\right] + \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\left|\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle\right| > \frac{1}{50}\cdot\frac{\varepsilon}{n_2}\cdot\frac{n_1\sqrt{n_2}}{\sqrt{114\lambda_0 e}}\right] \\
&\leqslant \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2 > 0.9\right] + \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\left|\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle\right| > \frac{C_1}{50\sqrt{114}}\cdot\frac{\log\left(\frac{1}{\delta}\right)}{\sqrt{n_2}}\right], \qquad (45)
\end{aligned}$$

where the last inequality used our bound on n_1.

We handle each term of (45) as we did in the conclusion of Proposition 15. For the first term, properties of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_{\leqslant d}$ (Fact 30) yield that $\|t\|_2^2 \sim \mathrm{Beta}\left(\frac{d}{2}, \frac{n_2-d}{2}\right)$. Additionally, our bound on n_2 implies that $n_2 > 4d$. Thus, we get from Beta concentration (Fact 31):

$$\begin{aligned}
\underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2 > 0.9\right] &= \underset{t\sim z_{\leqslant d}}{\mathbb{P}}\left[\|t\|_2^2 > 0.81\right] \\
&\leqslant 2\exp\left(-C_4 \min\left\{\frac{(n_2-d)^2}{2d}\left(0.81-\frac{d}{n_2}\right)^2, \frac{n_2-d}{2}\left(0.81-\frac{d}{n_2}\right)\right\}\right) \\
&= 2\exp\left(-C_4\frac{n_2-d}{2}\left(0.81-\frac{d}{n_2}\right)\right) \\
&\leqslant 2\exp\left(-\frac{C_4 n_2}{10}\right) \\
&\leqslant \frac{\delta}{2}. \qquad (46)
\end{aligned}$$

For the second term of (45), rotational invariance for uniform distributions over the unit sphere (Fact 33) and the density of $\mathcal{U}\left(\mathbb{S}^{n_2-1}\right)_1$ (Fact 30) imply that $\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle \sim z_1$, yielding that $\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle^2 \sim$

$\mathrm{Beta}\left(\frac{1}{2}, \frac{n_2-1}{2}\right)$. Thus, by our bounds on n_2 and δ, Beta concentration (Fact 31) yields:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{t\sim z_{\leqslant d}}\left[\left|\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle\right| > \frac{C_1}{50\sqrt{114}} \cdot \frac{\log\left(\frac{1}{\delta}\right)}{\sqrt{n_2}}\right] \\
&= \mathop{\mathbb{P}}_{t\sim z_{\leqslant d}}\left[\left\langle \frac{\ell}{\|\ell\|_2}, t\right\rangle^2 > \frac{C_1^2}{285000} \cdot \frac{\log^2\left(\frac{1}{\delta}\right)}{n_2}\right] \\
&\leqslant 2\exp\left(-C_4 \min\left\{\frac{(n_2-1)^2}{2}\left(\frac{C_1^2}{285000}\cdot\frac{\log^2\left(\frac{1}{\delta}\right)}{n_2} - \frac{1}{n_2}\right)^2, \frac{n_2-1}{2}\left(\frac{C_1^2}{285000}\cdot\frac{\log^2\left(\frac{1}{\delta}\right)}{n_2} - \frac{1}{n_2}\right)\right\}\right) \\
&= 2\exp\left(-C_4\frac{n_2-1}{2}\left(\frac{C_1^2}{285000}\cdot\frac{\log^2\left(\frac{1}{\delta}\right)}{n_2} - \frac{1}{n_2}\right)\right) \\
&\leqslant 2\exp\left(-C_4\frac{n_2-1}{2}\cdot\frac{C_1^2}{570000}\cdot\frac{\log^2\left(\frac{1}{\delta}\right)}{n_2}\right) \\
&= 2\exp\left(-C_1^2 C_4\frac{n_2-1}{n_2}\cdot\frac{\log^2\left(\frac{1}{\delta}\right)}{1140000}\right) \\
&\leqslant 2\exp\left(-\frac{C_1^2 C_4}{2280000}\log^2\left(\frac{1}{\delta}\right)\right) \\
&\leqslant \frac{\delta}{2}.
\end{aligned} \tag{47}$$

Upper-bounding (45) using (46) and (47) yields the desired result. ■