Differential privacy and robust statistics in high dimensions

Xiyang Liu   XIYANGL@CS.WASHINGTON.EDU
Paul G. Allen School of Computer Science and Engineering, University of Washington

Weihao Kong   KWEIHAO@GMAIL.COM
Google Research

Sewoong Oh   SEWOONG@CS.WASHINGTON.EDU
Paul G. Allen School of Computer Science and Engineering, University of Washington

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We introduce a universal framework for characterizing the statistical efficiency of a statistical estimation problem with differential privacy guarantees. Our framework, which we call High-dimensional Propose-Test-Release (HPTR), builds upon three crucial components: the exponential mechanism from (McSherry and Talwar, 2007), robust statistics, and the Propose-Test-Release mechanism from (Dwork and Lei, 2009). Connecting all these together is the concept of resilience, which is central to robust statistical estimation. Resilience guides the design of the algorithm, the sensitivity analysis, and the success probability analysis of the test step in Propose-Test-Release. The key insight is that if we design an exponential mechanism that accesses the data only via one-dimensional and robust statistics, then the resulting local sensitivity can be dramatically reduced. Using resilience, we can provide tight local sensitivity bounds. These tight bounds readily translate into near-optimal utility guarantees in several cases. We give a general recipe for applying HPTR to a given instance of a statistical estimation problem and demonstrate it on canonical problems of mean estimation, linear regression, covariance estimation, and principal component analysis. We introduce a general utility analysis technique that proves that HPTR achieves near-optimal sample complexity under several scenarios studied in the literature.

Keywords: Differential privacy, robust statistics.

1. Introduction

Estimating a parameter of a distribution from i.i.d. samples is a canonical problem in statistics. For such problems, characterizing the computational and statistical cost of ensuring differential privacy (DP) has gained significant interest with the rise of DP as the de facto measure of privacy. This is spearheaded by exciting and foundational algorithmic advances, e.g., (Barber and Duchi, 2014; Karwa and Vadhan, 2017; Kamath et al., 2019, 2020; Cai et al., 2019). However, the computational efficiency of these algorithms often comes at the cost of requiring superfluous assumptions that are not necessary for statistical efficiency, such as known bounds on the parameters or knowledge of higher-order moments. Without such assumptions, the optimal sample complexity remains un-known even for canonical statistical estimation problems under differential privacy. Further, each algorithm needs to be customized to those assumptions or to the problem instances.

We take an alternative route of focusing only on the statistical cost of differential privacy without concerning computational efficiency. Our goal is to introduce a general unifying framework that (1) can be readily applied to each problem instance, (2) provides a tight characterization of the statistical complexity involved, and (3) requires minimal assumptions. Achieving this goal critically

relies on three key ingredients: the exponential mechanism introduced by McSherry and Talwar (2007), robust statistics, and the Propose-Test-Release mechanism introduced by Dwork and Lei (2009). We first explain these three components of our approach, and then demonstrate the utility of our proposed framework, called High-dimensional Propose-Test-Release (HPTR), in canonical problems of mean and covariance estimation, linear regression, and principal component analysis.

Exponential mechanism and sensitivity. Differential privacy (DP), introduced in the seminal paper of Dwork et al. (2006), is an agreed upon measure of privacy that provides plausible deniability to the individual entries. Given a dataset S of size n and its empirical distribution $\hat{p}_S = (1/n) \sum_{x_i \in S} \delta_{x_i}$ where δ_{x_i} is Dirac delta function at x_i, its *neighborhood* is defined as $\mathcal{N}_S = \{S' : |S'| = |S|, d_{\mathrm{TV}}(\hat{p}_S, \hat{p}_{S'}) \leq 1/n\}$, which is a set of datasets at Hamming distance[1] at most one from S, and $d_{\mathrm{TV}}(\cdot)$ is the total variation. Plausible deniability is achieved by introducing the right amount of randomness. A randomized estimator $\hat{\theta}(S)$ is said to be (ε, δ)-differentially private for some target $\varepsilon \geq 0$ and $\delta \in [0, 1]$ if $\mathbb{P}(\hat{\theta}(S) \in A) \leq e^{\varepsilon} \mathbb{P}(\hat{\theta}(S') \in A) + \delta$ for all neighboring datasets S, S' and all measurable subset $A \subseteq \mathbb{R}^p$.

The *sensitivity* plays a crucial role in designing DP estimators. Consider an example of mean estimation, where the error is measured in the Mahalanobis distance defined as $D_p(\hat{\mu}) = \|\Sigma_p^{-1/2}(\hat{\mu} - \mu_p)\|$, where μ_p and Σ_p are the mean and covariance of the sample-generating distribution p. This is a preferred error metric since it has unit variance in all directions and is invariant to a linear transformation of the samples. The first result in differentially private mean estimation under this Mahalanobis distance is provided in (Brown et al., 2021), which we discuss more in Section 1.1. A corresponding empirical loss is $D_{\hat{p}_S}(\hat{\mu}) = \|\Sigma_{\hat{p}_S}^{-1/2}(\hat{\mu} - \mu_{\hat{p}_S})\|$. The exponential mechanism from (McSherry and Talwar, 2007) produces an $(\varepsilon, 0)$-DP estimate $\hat{\mu}$ by sampling from a distribution that approximately and stochastically minimizes this empirical loss: $\hat{\mu} \sim (1/Z(S)) e^{-\frac{\varepsilon}{2\Delta} D_{\hat{p}_S}(\hat{\mu})}$, where $Z(S) = \int \exp\{-(\varepsilon/2\Delta) D_{\hat{p}_S}(\hat{\mu}) d\hat{\mu}$. The sensitivity of the score function $D_{\hat{p}_S}(\hat{\mu})$ is defined as $\Delta := \max_{\hat{\mu}, S, S' \in \mathcal{N}_S} |D_{\hat{p}_S}(\hat{\mu}) - D_{\hat{p}_{S'}}(\hat{\mu})|$, which is the influence of one data point on the loss. From this definition, the $(\varepsilon, 0)$-DP guarantee follows immediately (e.g., Lemma 3).

Using the exponential mechanism is crucial in HPTR for two reasons: adaptivity and flexibility. First, it naturally adapts to the geometry of the problem, which is encoded in the loss. For example, a more traditional Gaussian mechanism (Dwork and Roth, 2014) needs to estimate Σ_p privately in order to add a Gaussian noise tailored to that estimated Σ_p, which increases sample complexity significantly. On the other hand, the exponential mechanism seamlessly adapts to Σ_p without explicitly and privately estimating it. Further, the exponential mechanism allows us significant flexibility to design different loss functions, some of which can dramatically reduce the sensitivity. Discovering such a loss function is the main focus of this paper.

One major challenge is that the sensitivity is unbounded when the support of the distribution is unbounded. A common solution is to privately estimate a bounded domain that the samples lie in and use it to bound the sensitivity (e.g., (Karwa and Vadhan, 2017; Kamath et al., 2019)). We propose a fundamentally different approach using robust statistics.

Robust statistics and resilience. The concept of *resilience* defined in Steinhardt et al. (2018) (also known as stability in the literature, e.g., Diakonikolas and Kane (2019)) plays a critical role in robust statistics. For the mean, for example, a dataset S is said to be (α, ρ)-resilient for some $\alpha \in [0, 1]$ and $\rho > 0$ if for all $v \in \mathbb{R}^d$ with $\|v\| = 1$ and all subset $T \subseteq S$ of size at least

---

1. There are two notions of a neighborhood in DP, which are equally popular. We use the one based on exchanging an entry, but all the analyses can seamlessly be applied to the one that allows for insertion and deletion of an entry.

$|T| \geq \alpha n$, we have $\big| \langle v, \mu_{\hat{p}_T} \rangle - \langle v, \mu_{\hat{p}_S} \rangle \big| \leq \rho/\alpha$. A more precise statement is in Definition 8. This measures how resilient the empirical mean is to subsampling or deletion of a fraction of the samples. This resilience is a central concept in robust statistical estimation when a fraction of the dataset is arbitrarily corrupted by an adversary (Steinhardt et al., 2018; Zhu et al., 2019). We show and exploit the fact that resilience is fundamentally related to the sensitivity of robust statistics.

For each direction $v \in \mathbb{R}^d$ with $\|v\| = 1$, we construct a robust mean of a one-dimensional projected dataset, also known as trimmed mean, $S_v = \{\langle v, x_i \rangle \in \mathbb{R}\}_{x_i \in S}$, as follows. For some $\alpha \in [0, 1/2)$, remove αn data points corresponding to the largest entries in S_v and also remove the αn smallest entries. The mean of the remaining $(1 - 2\alpha)n$ points is the robust one-dimensional mean, which we denote by $\langle v, \mu_{\hat{p}_v}^{(robust)} \rangle \in \mathbb{R}$. From the resilience above, we know that the mean of the removed top part is upper bounded by $\langle v, \mu_{\hat{p}_S} \rangle + \rho/\alpha$. The mean of the removed bottom part is lower bounded by $\langle v, \mu_{\hat{p}_S} \rangle - \rho/\alpha$. Hence, the effective support of this robust one-dimensional mean estimator is upper and lower bounded by the same. This can be readily translated into a bound in sensitivity of the estimate, $\langle v, \mu_{\hat{p}_v}^{(robust)} \rangle$ (e.g., Lemma 16). A similar sensitivity bound holds for the robust one-dimensional variance estimator, $v^\top \Sigma_{\hat{p}_v}^{(robust)} v$, defined similarly. We propose an approach that critically relies on this observation that *one-dimensional robust statistics have low sensitivity on resilient datasets, i.e., datasets satisfying the resilience property with small ρ.*

This suggests that if we can design a score function that only depends on one-dimensional robust statistics of the data, it might inherit the low sensitivity of those robust statistics. To this end, we first transform the target error metric into an equivalent expression that only depends on one-dimensional (population) mean, $\langle v, \mu_p \rangle$, and variance, $v^\top \Sigma_p v$, i.e.,

$$\|\Sigma_p^{-1/2}(\hat{\mu} - \mu_p)\| = \max_{v \in \mathbb{R}^d, \|v\|=1} \frac{\langle v, \hat{\mu} \rangle - \langle v, \mu_p \rangle}{\sqrt{v^\top \Sigma_p v}} \, ,$$

which follows from Lemma 7. Next, we replace the population statistics with robust empirical ones to define a new empirical loss, $D_{\hat{p}_S}(\hat{\mu}) = \max_{v \in \mathbb{R}^d, \|v\|=1}(\langle v, \hat{\mu} \rangle - \langle v, \mu_{\hat{p}_v}^{(robust)} \rangle)/\sqrt{v^\top \Sigma_{\hat{p}_v}^{(robust)} v}$. Precise definitions of these robust statistics can be found in Eq. (4). For resilient datasets, such a score function has a dramatically smaller sensitivity compared to those that rely on high-dimensional robust statistics. For mean estimation under a sub-Gaussian distribution, the sensitivity of the proposed loss is $\tilde{O}(1/n)$, whereas a high-dimensional robust statistic has $\Omega(\sqrt{d}/n)$ sensitivity.

Such an improved sensitivity immediately leads to a better utility guarantee of the exponential mechanism. We explicitly prescribe such loss functions for the canonical problems of mean estimation, linear regression, covariance estimation, and principal component analysis. This leads to near-optimal utility in most cases and improves upon the state-of-the-art in others, as we demonstrate in Section 1.1. Further, this approach can potentially be more generally applied to a much broader class of problems. One remaining challenge is that the tight sensitivity bound we provide holds only for a resilient dataset. To reject bad datasets, we adopt the Propose-Test-Release (PTR) framework pioneered in the seminal work of Dwork and Lei (2009).

Propose-Test-Release and local sensitivity. The tight sensitivity bound we provide on the proposed exponential mechanism is *local* in the sense that it only holds for resilient datasets. However, differential privacy must be guaranteed for any input, whether it is resilient (with desired level of α and ρ) or not. We adopt Propose-Test-Release introduced in (Dwork and Lei, 2009) to handle such locality of sensitivity. In the first step, one proposes an upper bound on the sensitivity of the loss

$D_S(\hat{\theta})$, determined by the resilience of the dataset, which in turn is determined solely by the distribution family of interest and the target error rate. In the second step, one tests if the combination of the given dataset S, sensitivity bound Δ, and the exponential mechanism with loss $D_S(\hat{\theta})$ satisfy the DP conditions. A part of the privacy budget is used to test this in a differential private manner, such that the subsequent exponential mechanism can depend on the result of this test, i.e., we only proceed to the third step if S passes the test. Otherwise, the process stops and outputs a predefined symbol, $\perp$. In the third step, one releases the DP estimate via the exponential mechanism. This ensures DP for any input S. We are adopting the Propose-Test-Release mechanism pioneered in (Dwork and Lei, 2009), which we explain in detail in Section A. The resulting framework, which we call High-dimensional Propose-Test-Release (HPTR) is provided in Section 1.2.

Contributions. We introduce a novel (computationally inefficient) algorithm for differentially private statistical estimation, with the goal of characterizing the achievable sample complexity for various problems with minimal assumptions. The proposed framework, which we call High-dimensional Propose-Test-Release (HPTR), makes a fundamental connection between differential privacy and robust statistics, thus achieving a sample complexity that significantly improves upon other state-of-the-art approaches. HPTR is a generic framework that can be seamlessly applied to various statistical estimation problems, as we demonstrate for mean estimation, linear regression, covariance estimation, and principal component analysis. Further, our analysis technique, which requires minimal assumptions, also seamlessly generalizes to all problem instances of interest.

HPTR uses three crucial components: the exponential mechanism, robust statistics, and the Propose-Test-Release mechanism. Building upon the inherent adaptivity and flexibility of the exponential mechanism, we propose using a novel loss function (also called a score function in a typical design of exponential mechanisms) that accesses the data only via one-dimensional robust statistics. The use of 1-D robust statistics is critical, because it dramatically reduces the sensitivity. We prove this sensitivity bound using the fundamental concept of resilience, which is central in robust statistics. This novel robust exponential mechanism is incorporated within the PTR framework to ensure that differential privacy is guaranteed on all input datasets, including those that are not necessarily compliant with the statistical assumptions. One byproduct of using robust statistics is that robustness comes for free. HPTR is inherently robust to adversarial corruption of the data and achieves the optimal robust error rate under standard data corruption models.

We present informal versions of our main theoretical results in Section 1.1. We present HPTR algorithm in detail in Section 1.2. We provide a sketch of the proof and the main technical contributions in Section 1.3. Notations and background on DP are provided in Appendix A. Detailed explanations of the setting, main results, and the proofs for each instance of the problems are presented in Appendices B–E for mean estimation, linear regression, covariance estimation, and principal component analysis, respectively.

1.1. Main results and related work

For each canonical problem of interest in statistical estimation, HPTR can readily be applied to, in most cases, significantly improve upon known achievable sample complexity. Most of the lower bounds we reference are copied in Appendix H for completeness.

DP mean estimation. We apply our proposed HPTR framework to the standard DP mean estimation problem, where i.i.d. samples $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ are drawn from a distribution $P_{\mu,\Sigma}$ with an unknown mean μ (which corresponds to θ in the general notation) and an unknown covariance

$\Sigma \succ 0$, and we want to produce a DP estimate $\hat{\mu}$ of the mean. The resulting error is measured in Mahalanobis distance, $D_{P_{\mu,\Sigma}}(\hat{\mu}) = \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, which is scale-invariant and naturally captured the uncertainty in all directions. It has only recently been shown that optimal (inefficient) algorithm exists for private mean estimation with Mahalanobis distance by Brown et al. (2021), but the algorithm only applies to Gaussian data. This problem is especially challenging since we aim for a tight guarantee that adapts to the unknown Σ as measured in the Mahalanobis distance without sufficient samples to directly estimate Σ, as we explain below. Despite being a canonical problem in DP statistics, the optimal sample complexity is not known even for standard distributions: sub-Gaussian and heavy-tailed distributions. We characterize the optimal sample complexity of the two problems by providing the guarantee of HPTR and the matching sample complexity lower bounds. A precise definition of sub-Gaussian distributions is provided in Eq. (20).

Theorem 1 (DP sub-Gaussian mean estimation algorithm, Corollary B.2 informal) *Consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a sub-Gaussian distribution with mean μ and covariance Σ. There exists an (ε, δ)-differentially private algorithm $\hat{\mu}(S)$ that given $n = \tilde{O}_{\xi,\zeta}(d/\xi^2 + d/(\varepsilon\xi))$, achieves Mahalanobis error $\|\Sigma^{-1/2}(\hat{\mu}(S) - \mu)\| \leq \xi$ with probability $1 - \zeta$, where $\tilde{O}_{\xi,\zeta}$ hides the logarithmic dependency on ξ, ζ and we assume $\delta = e^{-O(d)}$.*

HPTR is the first algorithm for sub-Gaussian mean estimation with unknown covariance that matches the best known sample complexity lower bound of $n = \tilde{\Omega}(d/\xi^2 + d/(\xi\varepsilon))$ from (Karwa and Vadhan, 2017; Kamath et al., 2019) up to logarithmic factors in error ξ and failure probability ζ. Existing algorithms are suboptimal as they require either a larger sample size or strictly Gaussian assumptions. Advances in DP mean estimation started with computationally efficient approaches of (Karwa and Vadhan, 2017; Kamath et al., 2019; Barber and Duchi, 2014). We discuss the results as follows, and omit the polynomial factors in $\log(1/\delta)$. When the covariance Σ is known, Mahalanobis error ξ can be achieved with $n = \tilde{O}(d/\xi^2 + d/(\xi\varepsilon))$ samples. Under a relaxed assumption that $\mathbf{I}_{d \times d} \preceq \Sigma \preceq \kappa \mathbf{I}_{d \times d}$ with a known κ, $n = \tilde{O}(d/\xi^2 + d/(\xi\varepsilon) + d^{1.5}/\varepsilon)$ samples are required using a specific preconditioning approach tailored for the assumption and the knowledge of κ. For general unknown Σ, $O(d^2/\xi^2 + d^2/(\xi\varepsilon))$ samples are required using an explicit DP estimation of the covariance. Empirically, further gains can be achieved with CoinPress (Biswas et al., 2020).

Computationally inefficient approaches followed with a goal of identifying the fundamental optimal sample complexity with minimal assumptions (Bun et al., 2019; Aden-Ali et al., 2020). For the unknown covariance setting, the best known result under Mahalanobis error is achieved by Brown et al. (2021). This builds upon the differentially private Tukey median exponential mechanism introduced in (Liu et al., 2021) initially analyzed and proposed for identity covariance and a known bound on the mean. It is immediate that the Tukey depth used in the exponential mechanism is invariant under a linear transform, and hence small Tukey depth implies closeness in Mahalanobis distance. However, it is challenging to define an appropriate support of the exponential mechanism that adapts to the geometry of the covariance, which is critical in removing the unnecessary assumption on the knowledge of the bound on the mean. Brown et al. (2021) make significant advance by providing an adaptive support known as a safe set (which our algorithm builds upon as shown in Section 1.2) and providing the analysis techniques that shows that $n = \tilde{O}(d/\xi^2 + d/(\xi\varepsilon))$ is sufficient even when the covariance is unknown. However, Tukey depth heavily relies on the assumption that the distribution is strictly Gaussian. For sub-Gaussian distributions, Brown et al. (2021) propose a different approach achieving sample complexity of $n = \tilde{O}(d/\xi^2 + d/(\xi\varepsilon^2))$ samples with a sub-optimal $(1/\varepsilon^2)$ dependence.

Beyond the sub-Gaussian setting, it is natural to consider the DP mean estimation for distributions with heavier tails. We apply HPTR framework to the more general mean estimation problems for hypercontractive distributions. A distribution $P_{\mu,\Sigma}$ with mean μ and covariance Σ is (κ, k)-hypercontractive if for all $v \in \mathbb{R}^d$, $\mathbb{E}_{x \sim P_X}[|\langle v, (x - \mu)\rangle|^k] \leq \kappa^k (v^\top \Sigma v)^{k/2}$. The assumption of hypercontractivity is similar to the bounded k-th moment assumptions, except requiring an additional lower bound on the covariance. This additional assumption is necessary for our setting to make sure the Mahalanobis error guarantee is achievable. We state our main result for hypercontractive mean estimation as follows. For simplicity of the statement, we assume k, κ are constants.

Theorem 2 (DP hypercontractive mean estimation, Corollary B.3 informal) *Consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a (κ, k)-hypercontractive distribution with mean μ and covariance Σ. There exists an (ε, δ)-differentially private algorithm $\hat{\mu}(S)$ that given $n = \tilde{O}_d\left(\frac{d}{\xi^2} + \frac{d}{\varepsilon\xi^{1+1/(k-1)}}\right)$, achieves Mahalanobis error $\|\Sigma^{-1/2}(\hat{\mu}(S) - \mu)\| \leq \xi$ with probability at least 0.99, where $\tilde{O}_d$ hides a logarithmic factor on d, and we assumes $\delta = e^{-O(d)}$.*

We prove an $n = \Omega(d/\varepsilon\xi^{1+1/(k-1)})$ sample complexity lower bound for hypercontractive DP mean estimation in Proposition B.5 to show the optimality of our upper bound result. Notice that the first term $\tilde{O}_d(d/\xi^2)$ in the upper bound cannot be improved up to logarithmic factors even if we do not require privacy, thus HPTR is the first algorithm that achieves optimal sample complexity for hypercontractive mean estimation under Mahalanobis distance up to logarithmic factors in d. When the covariance is known, an existing DP mean estimator of Kamath et al. (2020) achieves a stronger $(\varepsilon, 0)$-DP with a similar sample size of $n = \tilde{O}(d/\xi^2 + d/(\varepsilon\xi^{1+1/(k-1)}))$, and no prior result is known for the unknown covariance case. When $k = 2$ and covaraince is known, Hopkins et al. (2021) achieve significantly improved guarantees with pure differential privacy using an innovative exponential mechanism with sum-of-squares proofs.

DP linear regression. We next apply HPTR framework to DP linear regression. Given i.i.d. samples $S = \{(x_i, y_i)\}_{i \in [n]}$ from a distribution $P_{\beta, \Sigma, \gamma^2}$ of a linear model: $y_i = x_i^\top \beta + \eta_i$, where the input $x_i \in \mathbb{R}^d$ has zero mean and covariance Σ and the noise $\eta_i \in \mathbb{R}$ has variance γ^2 satisfying $\mathbb{E}[x_i \eta_i] = 0$, the goal of DP linear regression is to output a DP estimate $\hat{\beta}$ of the unknown model parameter β, without knowledge about the covariance Σ. The resulting error is measured in $D_{P_{\beta,\Sigma,\gamma^2}}(\hat{\beta}) = (1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|$ which is equivalent to the standard *root excess risk* of the estimated predictor $\hat{\beta}$. Similar to Mahalanobis distance for mean estimation, this is challenging since we aim for a tight guarantee that adapts to the unknown Σ without having enough samples to directly estimate Σ.

Theorem 3 (DP sub-Gaussian linear regression, Corollary C.2 informal) *Consider a dataset $S = \{(x_i, y_i)\}_{i=1}^n$ generated from a linear model $y_i = x_i^\top \beta + \eta_i$ for some $\beta \in \mathbb{R}^d$, where $\{x_i\}_{i \in [n]}$ are i.i.d. sampled from zero-mean d-dimensional sub-Gaussian distribution with unknown covariance Σ, and $\{\eta_i\}_{i \in [n]}$ are i.i.d. sampled from zero mean one-dimensional sub-Gaussian with variance γ^2. We further assume the data x_i and the noise η_i are independent. There exists a (ε, δ)-differentially private algorithm $\hat{\beta}(S)$ that given $n = \tilde{O}_{\xi,\zeta}(d/\xi^2 + d/(\varepsilon\xi))$, achieves error $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta}(S) - \beta)\| \leq \xi$ with probability $1 - \zeta$, where $\tilde{O}_{\xi,\zeta}$ hides the logarithmic dependency on ξ, ζ and we assume $\delta = e^{-O(d)}$.*

HPTR significantly improves upon the best known algorithm for DP linear regression from (Cai et al., 2019), which requires Σ to be close to the identity matrix or, equivalently, requires that

we know Σ. Dwork and Lei (2009) propose to use PTR and B-robust regression algorithm from (Hampel et al., 1986) for differentially private linear regression under i.i.d. data assumptions (also exponential time), but only asymptotic consistency is proven as $n \to \infty$. Under an alternative setting where the data is deterministically given without any probabilistic assumptions, significant advances in DP linear regression has been made (Vu and Slavkovic, 2009; Kifer et al., 2012; Mir, 2013; Dimitrakakis et al., 2014; Bassily et al., 2014; Wang et al., 2015; Foulds et al., 2016; Minami et al., 2016; Wang, 2018; Sheffet, 2019). The state-of-the-art guarantee is achieved in (Wang, 2018; Sheffet, 2019) which under our setting translates into a sample complexity of $n = O(d^{1.5}/(\xi\varepsilon))$. The extra $d^{1/2}$ factor is due to the fact that no statistical assumption is made, and cannot be improved under the deterministic setting (not necessarily i.i.d.) that those algorithms are designed for. Similar to mean estimation, we also consider the DP linear regression for distributions with heavier tails, and apply HPTR framework to the linear regression problem under (k, κ)-hypercontractive distributions (see Definition 18). HPTR can handle both independent and dependent noise, and we state the independent noise case here the dependent noise case in Section C.3.3. For simplicity of the statement, we assume k, κ are constants.

Theorem 4 (DP hypercontractive linear regression for independent noise, Corollary C.3 informal) *Consider a dataset $S = \{(x_i, y_i)\}_{i=1}^n$ generated from a linear model $y_i = x_i^\top \beta + \eta_i$ for some $\beta \in \mathbb{R}^d$, where $\{x_i\}_{i \in [n]}$ are i.i.d. sampled from zero-mean d-dimensional (κ, k)-hypercontractive distribution with unknown covariance Σ and η_i are i.i.d. sampled from zero mean one-dimensional (κ, k)-hypercontractive distribution with variance γ^2. We further assume the data x_i and the noise $\{\eta_i\}_{i \in [n]}$ are independent. There exists an (ε, δ)-differentially private algorithm $\hat{\beta}(S)$ that given $n = \tilde{O}_d(d/\xi^2 + d/(\varepsilon\xi^{1+1/(k-1)}))$, achieves error $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta}(S) - \beta)\| \leq \xi$ with probability 0.99, where $\tilde{O}_d$ hides a logarithmic factor on d, and we assume $\delta = e^{-O(d)}$.*

To the best of our knowledge, HPTR is the first algorithm for linear regression that guarantees (ε, δ)-DP under hypercontractive distributions with independent noise. When applied to the setting where noise η_i is dependent on the input vector x_i, HPTR is the first algorithm for linear regression that guarantees (ε, δ)-DP. We refer the readers to Section C.3.3 for a more detailed description of our result.

DP covariance estimation. We present HPTR applied to covariance estimation from i.i.d. samples under a Gaussian distribution $\mathcal{N}(0, \Sigma)$. The reason for this choice is that the Mahalanobis error $\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F$ of the Kronecker product $x_i \otimes x_i$ is proportional to the natural error metric of total variation for Gaussian distributions. The strength of HPTR framework is that it can be seamlessly applied to general distributions, for example sub-Gaussian or heavytailed, but the resulting Mahalanobis error becomes harder to interpret as it involves fourth moment tensors.

Theorem 5 (DP Gaussian covariance estimation, Corollary D.3 informal) *Consider a dataset $S = \{x_i\}_{i=1}^n$ of n i.i.d. samples from $\mathcal{N}(0, \Sigma)$. There exists a (ε, δ)-differentially private estimator $\hat{\Sigma}$ that given $n = \tilde{O}_{\xi,\zeta}(d^2/\xi^2 + d^2/(\xi\varepsilon))$, achieves error $\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F \leq \xi$ with probability $1 - \zeta$, where $\tilde{O}_{\xi,\zeta}$ hides the logarithmic dependency on ξ, ζ and we assume $\delta = e^{-O(d)}$.*

This Mahalanobis distance guarantee (for the Kronecker product, $\{x_i \otimes x_i\}$, of the samples) implies that the estimated Gaussian distribution is close to the underlying one in total variation distance (see for example (Kamath et al., 2019, Lemma 2.9)): $d_{\text{TV}}(\mathcal{N}(0, \hat{\Sigma}), \mathcal{N}(0, \Sigma)) = O(\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F) = O(\xi)$. The sample complexity is near-optimal, matching a lower bound of $n =$

$\Omega(d^2/\xi^2 + \min\{d^2, \log(1/\delta)\}/(\varepsilon\xi))$ up to a logarithmic factor when $\delta = e^{-\Theta(d)}$. The first term follows from the classical estimation of the covariance without DP. The second term follows from extending the lower bound in (Kamath et al., 2019) constructed for pure differential privacy with $\delta = 0$ and matches the second term in our upper bound when $\delta = e^{-\Theta(d^2)}$. We note that a similar upper bound is achieved by the state-of-the-art (computationally inefficient) algorithm in (Aden-Ali et al., 2020), which improves over HPTR in the lower order terms not explicitly shown in this informal version of our theorem. Computationally efficient approaches of (Karwa and Vadhan, 2017; Kamath et al., 2019) require additional assumption that $\mathbf{I}_{d\times d} \preceq \Sigma \preceq \kappa\mathbf{I}_{d\times d}$ with a known κ. Kamath et al. (2021) introduced a novel preconditioning approach that is polynomial time and removes the upper and lower bounds on Σ completely, but requires sample complexity of $n = \tilde{O}(d^2/\xi^2 + d^2\mathrm{polylog}(1/\delta)/(\xi\varepsilon) + d^{5/2}\mathrm{polylog}(1/\delta)/\varepsilon)$. This gap was closed by Ashtiani and Liaw (2021) recently, who introduced a polynomial time algorithm achieving a similar optimal guarantee as HPTR.

DP principal component analysis. We next apply HPTR to the PCA problem.

Theorem 6 (DP sub-Gaussian principle component analysis, Corollary E.1) *Consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a zero-mean sub-Gaussian distribution with unknown covariance Σ. There exists an (ε, δ)-differentially private estimator $\hat{u}$ that given $n = \tilde{O}_{\xi,\zeta}(d/\xi^2 + d/(\varepsilon\xi))$, achieves error $1 - \hat{u}^\top\Sigma\hat{u}/\|\Sigma\| \leq \xi$ with probability $1 - \zeta$, where $\tilde{O}_{\xi,\zeta}$ hides the logarithmic dependency on ξ, ζ and we assume $\delta = e^{-O(d)}$.*

HPTR is the first estimator for sub-Gaussian distributions to nearly match the information-theoretic lower bound of $n = \Omega(d/\xi^2 + \min\{d, \log(1/\delta)\}/(\varepsilon\xi))$ as we showed in Proposition E.2. The first term $\Omega(d/\xi^2)$ is unavoidable even without DP (Proposition E.3). The second term in the lower bound follows from Proposition E.2, which matches the second term in the upper bound when $\delta = e^{-\Theta(d)}$. Existing DP PCA approaches from (Blum et al., 2005; Chaudhuri et al., 2013; Kapralov and Talwar, 2013; Dwork et al., 2014; Hardt and Roth, 2012, 2013) are designed for arbitrary samples not necessarily drawn i.i.d. and hence require a larger samples size of $n = \tilde{O}(d/\xi^2 + d^{1.5}/(\xi\varepsilon))$. This is also unavoidable for more general deterministic data, as it matches an information theoretic lower bound (Dwork et al., 2014) under weaker assumptions on the data than i.i.d. Gaussian (under the rescaling of each sample to have norm $O(\sqrt{d})$).

Theorem 7 (DP hypercontractive principle component analysis, Corollary E.5) *Consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a zero-mean (κ, k)-hypercontractive distribution with unknown covariance Σ. There exists an (ε, δ)-differentially private estimator $\hat{u}$ that given $n = \tilde{O}_{\xi,d}\left(\frac{d}{\xi^{(2k-2)/(k-2)}} + \frac{d}{\varepsilon\xi^{1+2/(k-2)}}\right)$, achieves error $1 - \frac{\hat{u}^\top\Sigma\hat{u}}{\|\Sigma\|} \leq \xi$ with probability 0.99, where $\tilde{O}_{\xi,d}$ hides the logarithmic dependency on ξ, d and we assume $\delta = e^{-O(d)}$.*

HPTR is the first estimator for hypercontractive distributions to guarantee differential privacy for PCA with sample complexity scaling as $O(d)$. As a complement of our algorithm, we proved a $n = \Omega(d/\xi^2 + \min\{d, \log(1/\delta)\}/(\xi^{1+2/(k-2)}\varepsilon))$ information-theoretic lower bound in Proposition E.6. The first term $\Omega(d/\xi^2)$ in the lower bound is needed even without DP, and there is a gap of factor $O(\xi^{-2/(k-2)})$ compared to the first term in our upper bound. The second term in the lower bound matches the last term in the upper bound when $\delta = e^{-\Theta(d)}$.

1.2. Algorithm

The proposed High-dimensional Propose-Test-Release (HPTR) is not computationally efficient, as the TEST step requires enumerating over a certain neighborhood of the input dataset and the RE-LEASE step requires enumerating over all directions in high dimension. The strengths of HPTR is that (i) the same framework can be seamlessly applied to many problems as we demonstrate in Sections B–E; (ii) a unifying recipe can be applied for all those instances to give tight utility guarantees as we explicitly prescribe in Section 1.2.1; and (iii) the algorithm is simple and the analysis is clear such that it is transparent how the distribution family of interest translates into the utility guarantee (via resilience). As a byproduct of the simplicity of the algorithm and clarity of the analysis, we achieve the state-of-the-art sample complexity for all those problem instances with minimal assumptions, oftentimes nearly matching the information theoretic lower bounds. As a byproduct of the use of robust statistics, we guarantee robustness against adversarial corruption for free (e.g., Theorems 10, 12, 14). We describe the framework for general statistical estimation problem where data is drawn i.i.d. from a distribution represented by two unknown parameters $\theta \in \mathbb{R}^p$ and ϕ and is coming from a known family of distributions. An example of a problem instance of this type would be mean estimation from heavy-tailed distributions that are (κ, k)-hypercontractive with unknown mean μ (which in the general notation is θ) and unknown covariance Σ (which corresponds to ϕ).

The main component is an exponential mechanism in RELEASE step below that uses a loss function $D_S(\hat{\theta})$ and a support $B_{\tau,S}$, defined as $B_{\tau,S} = \{\hat{\theta} \in \mathbb{R}^p : D_S(\hat{\theta}) \leq \tau\}$. Bounding the support of the exponential mechanism is important since the sensitivity also depends on $\hat{\theta}$ in many problems of interest. We build upon the algorithm and proof technique of (Brown et al., 2021) that first bounded the support of an exponential mechanism with the level set of the score function. We make several innovations in the analysis, as our setting is more challenging because we have local sensitivity (Definition 4) as opposed to a fixed sensitivity of one in the Tukey depth approach of (Brown et al., 2021). We discuss this in detail in the example of mean estimation in Section B.2.2. The specific choices of the threshold τ only depend on the tail of the distribution family of interest and not the parameters θ or ϕ or the data. In particular, we use the resilience property of the distribution family to prescribe the choice of τ for each problem instance that gives us the tight utility guarantees. As explained in Section 1, we use one-dimensional robust statistics to design the loss functions, which we elaborate for each problem instances in Sections B–E, where we also explain how to choose the sensitivity based on the resilience of the distribution family only.

After we PROPOSE the choice of the sensitivity Δ and threshold τ for the problem instance in hand, we TEST to make sure that the given dataset S is consistent with the assumptions made when selecting $D_S(\hat{\theta})$, Δ, and τ. This is done by testing the safety of the exponential mechanism, by privately checking the margin to safety, i.e., how many data points need to be changed from S for the exponential mechanism to violate differential privacy conditions. If the margin is large enough, HPTR proceeds to RELEASE. Otherwise, it halts and outputs $\perp$. A set of such unsafe datasets is defined in (Brown et al., 2021) as

$$\mathsf{UNSAFE}_{(\varepsilon,\delta,\tau)} = \Big\{ S' \subseteq \mathbb{R}^{d \times n} \,|\, \exists S'' \sim S' \text{ and } \exists E \subseteq \mathbb{R}^p \text{ such that}$$

$$\mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E) > e^\varepsilon \mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E) + \delta \text{ or } \mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E) > e^\varepsilon \mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E) + \delta \Big\},$$

$$\tag{1}$$

where $r_{(\varepsilon,\Delta,\tau,S)}$ denotes the pdf of the exponential mechanism in Eq. (2). Given a loss (or a distance) function, $D_S(\hat{\theta})$, which is a surrogate for the target measure of error, $D_\phi(\hat{\theta},\theta)$, High-dimensional Propose-Test-Release (HPTR) proceeds as follows:

1. **Propose**: Propose a target bound Δ on local sensitivity and a target threshold τ for safety.

2. **Test**:

   2.1. Compute the safety margin $m_\tau = \min_{S'} d_H(S,S')$ such that $S' \in \mathsf{UNSAFE}_{(\varepsilon/2,\delta/2,\tau)}$.

   2.2. If $\hat{m}_\tau = m_\tau + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon)\log(2/\delta)$, then output $\perp$, and otherwise continue.

3. **Release**: Output $\hat{\theta}$ sampled from a distribution with a pdf:

$$
r_{(\varepsilon,\Delta,\tau,S)}(\hat{\theta}) \;=\; \begin{cases} \frac{1}{Z}\exp\left\{-\frac{\varepsilon}{4\Delta}D_S(\hat{\theta})\right\} & \text{if } \hat{\theta} \in B_{\tau,S} \;, \\ 0 & \text{otherwise}\;, \end{cases} \tag{2}
$$

where $Z = \int_{B_{\tau,S}} \exp\{-(\varepsilon D_S(\hat{\theta}))/(4\Delta)\}d\hat{\theta}$.

It is straightforward to show that (ε,δ)-differential privacy is satisfied for all input S.

Theorem 8 *For any dataset $S \subset \mathcal{X}^n$, distance function $D_S : \mathbb{R}^p \to \mathbb{R}$ on that dataset, and parameters $\varepsilon, \delta, \Delta$ and τ, HPTR is (ε,δ)-differentially private.*

Proof The DP margin $\hat{m}_\tau$ is $(\varepsilon/2,0)$-differentially private, because the sensitivity of the margin is one, and we are adding a Laplace noise with parameter $2/\varepsilon$. The TEST step (together with the exponential mechanism) is $(0,\delta/2)$-differentially private since there is a probability $\delta/2$ event that a unsafe dataset S with a small margin m_τ is classified as a safe dataset and passes the test. On the complimentary event, namely, that the dataset that passed the TEST is indeed safe, the RELEASE step is $(\varepsilon/2,\delta/2)$-differentially private since we use $\mathsf{UNSAFE}_{(\varepsilon/2,\delta/2,\tau)}$ in the TEST step. ∎

1.2.1. UTILITY ANALYSIS OF HPTR FOR STATISTICAL ESTIMATION

We prescribe the following three-step recipe as a guideline for applying HPTR to each specific statistical estimation problem and obtaining a utility guarantee. Consider a problem of estimating an unknown θ from samples from a generative model $P_{\theta,\phi}$, where the error is measured in $D_\phi(\hat{\theta},\theta)$.

- Step 1: Design a surrogate $D_S(\hat{\theta})$ for the target error metric $D_\phi(\hat{\theta},\theta)$ using only one-dimensional robust statistics on S.

- Step 2: Assuming *resilience* of the dataset, propose an appropriate sensitivity bound Δ and threshold τ and analyze the utility of HPTR.

- Step 3: For each specific family of generative models $P_{\theta,\phi}$ with a known tail bound, characterize the resulting resilience and substitute it in the utility analysis from the previous step, which gives the final guarantee.

We demonstrate how to apply this recipe and carry out the utility analysis for mean estimation (Section B), linear regression (Section C), covariance estimation (Section D), and PCA (Section E). We explain and justify the use of one-dimensional robust statistics in Step 1 and the assumption on the resilience of the dataset in Step 2 in the next section using the mean estimation problem as a canonical example. It is critical to construct $D_S(\hat{\theta})$ using only one-dimensional and robust statistics; this ensures that $D_S(\hat{\theta})$ has a small sensitivity as demonstrated in Section B.1. We prove error bounds only assuming resilience of the dataset; this relies on a fundamental connection between sensitivity and resilience as explained in Section B.2.

1.3. Technical contributions and proof sketch

We use the canonical example of mean estimation to explain our proof sketch. For i.i.d. samples from a sub-Gaussian distribution $P_{\mu,\Sigma}$ with mean μ and covariance Σ, we show in Theorem 9 that HPTR achieves a near optimal sample complexity of $n = \tilde{O}(d/\alpha^2 + d/(\alpha\varepsilon))$ to get Mahalanobis error $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = \tilde{O}(\alpha)$ for some target accuracy $\alpha \in [0, 1]$.

Our proof strategy is to first show that the robust one-dimensional statistics have small sensitivity if the dataset is resilient. Consequently, the loss function $D_S(\hat{\mu})$ has a small *local* sensitivity, i.e. the sensitivity is small if restricted to $\hat{\mu}$ close to μ and if the dataset is resilient. To ensure DP, we run RELEASE only when those two locality conditions are satisfied; we first PROPOSE the sensitivity Δ and a threshold τ, and then we TEST that DP guarantees are met on the given dataset with those choices. We prove that resilient datasets pass this safety test with a high probability and achieve the desired accuracy. We give a sketch of the main steps below.

One-dimensional robust statistics have small sensitivity on resilient datasets. A set $S = \{x_i \in \mathbb{R}^d\}_{i \in [n]}$ of i.i.d. samples from a sub-Gaussian distribution has the following resilience property w.p. $1 - \zeta$ if $n = \tilde{\Omega}(d/\alpha^2)$, where $\tilde{\Omega}$ hides polylogarithmic factors α and the failure probability ζ:

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i - \mu \rangle \right| \le 2\sigma_v \sqrt{\log(1/\alpha)} , \text{ and } \left| \frac{1}{|T|} \sum_{x_i \in T} \left(\langle v, x_i - \mu \rangle^2 - \sigma_v^2 \right) \right| \le 2\sigma_v^2 \log(1/\alpha) ,$$

for any subset $T \subset S$ of size at least αn and for any unit norm $v \in \mathbb{R}^d$ where $\sigma_v^2 = v^\top \Sigma v$ (Lemma 17). This means that the α-tail of the distribution (when projected down to one dimension) cannot be too far from the true one in mean and variance. For mean estimation, we use the loss function of $D_S(\hat{\mu}) = \max_{v \in \mathbb{R}^d, \|v\|=1} \langle v, \hat{\mu} - \mu(\mathcal{M}_{v,\alpha}) \rangle / \sqrt{v^\top \Sigma(\mathcal{M}_{v,\alpha})v}$, where $\mu(T)$ and $\Sigma(T)$ are mean and covariance of a dataset T and $\mathcal{M}_{v,\alpha} \subset S$ is defined as follows. For each direction v, we partition S into three sets $\mathcal{T}_{v,\alpha}$, $\mathcal{M}_{v,\alpha}$, and $\mathcal{B}_{v,\alpha}$. $\mathcal{T}_{v,\alpha} \subset S$ is the subset of datapoints corresponding to the largest αn datapoints in $\{\langle v, x_i \rangle\}_{x_i \in S}$, the projected data points in the direction v. $\mathcal{B}_{v,\alpha} \subset S$ corresponds to the smallest αn values, and $\mathcal{M}_{v,\alpha} \subset S$ is the remaining $(1 - 2\alpha)n$ data points. We show that the robust projected mean, $\langle v, \mu(\mathcal{M}_{v,\alpha}) \rangle$ has sensitivity $O(\sigma_v \sqrt{\log(1/\alpha)}/n)$. Under the resilience above, the top α-tail, $\mathcal{T}_{v,\alpha}$, has the empirical mean that is no more than $O(\sigma_v \sqrt{\log(1/\alpha)})$ away from the true projected mean $\langle v, \mu \rangle$, and the same is true for $\mathcal{B}_{v,\alpha}$. It follows that there exists at least one data point in $\mathcal{T}_{v,\alpha}$ and one data point in $\mathcal{B}_{v,\alpha}$ that are no more than $O(\sigma_v \sqrt{\log(1/\alpha)})$ away from μ_v. The range of the middle subset $\mathcal{M}_{v,\alpha}$, therefore, is bounded by $O(\sigma_v \sqrt{\log(1/\alpha)})$, and the sensitivity of the robust mean $\langle v, \mu(\mathcal{M}_{v,\alpha}) \rangle$ is guaranteed to be $O(\sigma_v \sqrt{\log(1/\alpha)}/n)$. We can similarly show that $v^\top \Sigma(\mathcal{M}_{v,\alpha})v$ has sensitivity $O(\sigma_v^2 \log(1/\alpha)/n)$.

Under the above sensitivity bounds for the one dimensional statistics, it follows (for example, in Eq. (19)) that the sensitivity of the loss function $D_S(\hat{\mu})$ is bounded by $O(\sqrt{\log(1/\alpha)}/n)$ as

long as $D_S(\hat{\mu}) \leq \tau := C\alpha\sqrt{\log(1/\alpha)}$ and the dataset is resilient. It is worth noting here that since the sensitivity is only small when $D_S(\hat{\mu}) \leq \tau$, our exponential mechanism only samples from the set $B_{\tau,S}$, which contains only the hypotheses with small scores. We handle this locality with TEST step that checks that the DP conditions are satisfied for the given dataset and the choice of $\Delta := C'\sqrt{\log(1/\alpha)}/n$ and $\tau := C\alpha\sqrt{\log(1/\alpha)}$. It is critical for ensuring DP that these choices only depend on the resilience (which is the property of the distribution family of interest, which in this case is sub-Gaussian) and the target accuracy, and not on the dataset S.

Sample complexity analysis. Assuming the sensitivity is bounded by $\Delta = C'\sqrt{\log(1/\alpha)}/n$, as the safety test ensures, we analyze the utility of the exponential mechanism. For a target accuracy of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\alpha\sqrt{\log(1/\alpha)})$, we consider two sets, $B_{\text{out}} = \{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq c_0\alpha\sqrt{\log(1/\alpha)}\}$ and $B_{\text{in}} = \{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq c_1\alpha\sqrt{\log(1/\alpha)}\}$, for some $c_0 > c_1$. The exponential mechanism achieves accuracy $c_0\alpha\sqrt{\log(1/\alpha)}$ with probability $1 - \zeta$ if

$$\mathbb{P}(\hat{\mu} \notin B_{\text{out}}) \leq \frac{\mathbb{P}(\hat{\mu} \notin B_{\text{out}})}{\mathbb{P}(\hat{\mu} \in B_{\text{in}})} \lesssim \frac{\text{Vol}(B_{\tau,S})}{\text{Vol}(B_{\text{in}})} \frac{e^{-\frac{\varepsilon}{4\Delta}c_0\alpha\sqrt{\log(1/\alpha)}}}{e^{-\frac{\varepsilon}{4\Delta}c_1\alpha\sqrt{\log(1/\alpha)}}} \leq e^{O(d)}e^{-\frac{\varepsilon}{4\Delta}(c_0-c_1)\alpha\sqrt{\log(1/\alpha)}} \leq \zeta \,,$$

where the second inequality requires $D_S(\hat{\mu}) \simeq \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, which we show in Lemma 12. Since $\Delta = O(\sqrt{\log(1/\alpha)}/n)$, it is sufficient to have a large enough c_0 and $n = \tilde{O}((d+\log(1/\zeta))/(\alpha\varepsilon))$ with a large enough constant. Together with the sample size required to ensure resilience, this gives the desired sample complexity of $n = \tilde{O}(d/\alpha^2 + (d + \log(1/\zeta))/(\alpha\varepsilon))$ where $\tilde{O}$ hides polylogarithmic factors in $1/\alpha$ and $1/\delta$.

Safety test. We are left to show that for a resilient dataset, the failure probability of the safety test, $\mathbb{P}(m_\tau + \text{Lap}(2/\varepsilon) < (2/\varepsilon)\log(2/\delta))$, is less than ζ. This requires the safety margin to be large enough, i.e. $m_\tau \geq k^* = (2/\varepsilon)\log(4/(\delta\zeta))$. Recall that the safety margin is defined as the Hamming distance to the closest dataset to S where the $(\varepsilon/2, \delta/2)$-DP condition of the exponential mechanism is violated. We therefore need to show that the DP condition is satisfied for not only S but any dataset S' at Hamming distance at most k^* from S. We treat S' as a *corrupted* version of a resilient S by a fraction k^*/n-corruption. Since we are using robust statistics that are designed to be robust against data corruption, we can show that the corrupted resilient set still has a low sensitivity for $D_{S'}(\hat{\mu})$. Building upon the proof techniques introduced in (Brown et al., 2021) for a safety test for a Tukey median based exponential mechanism, we use the fact that S' is a corrupted version of a resilient dataset S to show that the safety test passes with high probability.

2. Conclusion

We provided a universal framework for characterizing the statistical efficiency of statistical estimation problems with differential privacy guarantees. Our framework, High-dimensional Propose-Test-Release (HPTR), is computationally inefficient and builds upon three key components: the exponential mechanism, robust statistics, and the Propose-Test-Release mechanism. Our key insight is that designing an exponential mechanism that accesses the data via only one-dimensional robust statistics can dramatically reduce the resulting local sensitivity. Using resilience, a central concept in robust statistics, we can provide tight local sensitivity bounds. These tight bounds readily translate into near-optimal utility guarantees in several statistical estimation problems of interest: mean estimation, linear regression, covariance estimation, and principal component analysis. Although our framework is written as a conceptual algorithm without a specific implementation, it is possible

to implement it with exponential computational complexity following the guidelines of Brown et al. (2021), where a similar exponential mechanism with PTR was proposed and an implementation was explicitly provided.

To protect against membership inference attacks, significant progress has been made in training DP models that are practical, e.g., Abadi et al. (2016); Yu et al. (2021); Anil et al. (2021). To protect against data poisoning attacks, a recent work utilizes robust statistics with great success, e.g., Hayase et al. (2021). In practice, however, we need to protect against both types of attacks to facilitate learning and analysis from shared data. Currently, there is an algorithmic deficiency in this space. Efficient algorithms achieving both DP and robustness against adversarial corruption are known only for mean estimation (Liu et al., 2021). We make a valuable contribution to the design of such algorithms for a broad class of problems, including covariance estimation, principal component analysis, and linear regression.

Further, these computationally efficient algorithms typically require more samples. For sub-Gaussian mean estimation with known covariance Σ, an efficient approach of Liu et al. (2021) requires $\tilde{O}(d/\alpha^2 + d^{3/2}/(\varepsilon\alpha))$ samples under α-corruption and (ε, δ)-DP to achieve an error of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = \tilde{O}(\alpha)$. HPTR requires only $O(d/\alpha^2 + d/(\varepsilon\alpha))$ samples. A significant open question is whether this $d^{1/2}$ gap is fundamental and cannot be improved.

Most of the lower bounds in this paper are only tight when $\delta = e^{-\Theta(d)}$. Proving tight lower bounds for approximate DP with $\delta > 0$ is a challenging task. Some recent techniques from Kamath et al. (2022) prove a tight lower bound for a wider regime of $\delta \leq \varepsilon$ for mean estimation, which could be applied to a broader class of the statistical estimation problems.

Acknowledgement

References

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318, 2016.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

Ishaq Aden-Ali, Hassan Ashtiani, and Gautam Kamath. On the sample complexity of privately learning unbounded high-dimensional gaussians. *arXiv preprint arXiv:2010.09929*, 2020.

Kareem Amin, Alex Kulesza, Andres Munoz, and Sergei Vassilvtiskii. Bounding user contributions: A bias-variance trade-off in differential privacy. In *International Conference on Machine Learning*, pages 263–271. PMLR, 2019.

Galen Andrew, Om Thakkar, Brendan McMahan, and Swaroop Ramaswamy. Differentially private learning with adaptive clipping. *Advances in Neural Information Processing Systems*, 34, 2021.

Rohan Anil, Badih Ghazi, Vineet Gupta, Ravi Kumar, and Pasin Manurangsi. Large-scale differentially private bert. *arXiv preprint arXiv:2108.01624*, 2021.

Hassan Ashtiani and Christopher Liaw. Private and polynomial time algorithms for learning gaussians and beyond. *arXiv preprint arXiv:2111.11320*, 2021.

Hilal Asi and John C Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. *Advances in neural information processing systems*, 33:14106–14117, 2020.

Marco Avella-Medina. The role of robust statistics in private data analysis. *CHANCE*, 33(4):37–42, 2020.

Marco Avella-Medina. Privacy-preserving parametric inference: a case for robust statistics. *Journal of the American Statistical Association*, 116(534):969–983, 2021.

Marco Avella-Medina and Victor-Emmanuel Brunel. Differentially private sub-gaussian location estimators. *arXiv preprint arXiv:1906.11923*, 2019.

Ainesh Bakshi and Adarsh Prasad. Robust linear regression: Optimal rates in polynomial time. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 102–115, 2021.

Rina Foygel Barber and John C Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*, 2014.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 464–473. IEEE, 2014.

Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. *arXiv preprint arXiv:2006.06618*, 2020.

Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005.

Gavin Brown, Marco Gaboardi, Adam Smith, Jonathan Ullman, and Lydia Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. *arXiv preprint arXiv:2106.13329*, 2021.

Victor-Emmanuel Brunel and Marco Avella-Medina. Propose, test, release: Differentially private estimation with high probability. *arXiv preprint arXiv:2002.08774*, 2020.

Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. *arXiv preprint arXiv:1906.02830*, 2019.

Mark Bun, Gautam Kamath, Thomas Steinke, and Steven Z Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems*, pages 156–167, 2019.

T Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *arXiv preprint arXiv:1902.04495*, 2019.

Kamalika Chaudhuri and Daniel Hsu. Convergence rates for differentially private statistical estimation. In *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, volume 2012, page 1327. NIH Public Access, 2012.

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

Mengjie Chen, Chao Gao, and Zhao Ren. Robust covariance and scatter matrix estimation under huber's contamination model. *The Annals of Statistics*, 46(5):1932–1960, 2018.

Yeshwanth Cherapanamjeri, Efe Aras, Nilesh Tripuraneni, Michael I Jordan, Nicolas Flammarion, and Peter L Bartlett. Optimal robust linear regression in nearly linear time. *arXiv preprint arXiv:2007.08137*, 2020.

Jules Depersin and Guillaume Lecué. Robust subgaussian estimation of a mean vector in nearly linear time. *arXiv preprint arXiv:1906.03058*, 2019.

Jules Depersin and Guillaume Lecué. On the robustness to adversarial corruption and to heavy-tailed data of the stahel-donoho median of means. *arXiv preprint arXiv:2101.09117*, 2021.

Ilias Diakonikolas and Daniel M Kane. Recent advances in algorithmic high-dimensional robust statistics. *arXiv preprint arXiv:1911.05911*, 2019.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Being Robust (in High Dimensions) Can Be Practical. *arXiv e-prints*, art. arXiv:1703.00893, March 2017.

Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.

Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2683–2702. SIAM, 2018.

Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019a.

Ilias Diakonikolas, Weihao Kong, and Alistair Stewart. Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2745–2754. SIAM, 2019b.

Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. Robust and private bayesian inference. In *International Conference on Algorithmic Learning Theory*, pages 291–305. Springer, 2014.

Wei Dong and Ke Yi. Universal private estimators. *arXiv preprint arXiv:2111.02598*, 2021.

Yihe Dong, Samuel Hopkins, and Jerry Li. Quantum entropy scoring for fast robust mean estimation and improved outlier detection. In *Advances in Neural Information Processing Systems*, pages 6067–6077, 2019.

David L Donoho. Breakdown properties of multivariate location estimators. Technical report, Technical report, Harvard University, Boston., 1982.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

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

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

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

James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. On the theory and practice of privacy-preserving bayesian data analysis. *arXiv preprint arXiv:1603.07294*, 2016.

Chao Gao. Robust regression via mutivariate regression depth. *Bernoulli*, 26(2):1139–1170, 2020.

Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons, 1986.

Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1255–1268, 2012.

Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 331–340, 2013.

Jonathan Hayase, Weihao Kong, Raghav Somani, and Sewoong Oh. Defense against backdoor attacks via robust covariance estimation. In *International Conference on Machine Learning*, pages 4129–4139. PMLR, 2021.

Sam Hopkins, Jerry Li, and Fred Zhang. Robust and heavy-tailed mean estimation made simple, via regret minimization. *Advances in Neural Information Processing Systems*, 33, 2020.

Samuel B Hopkins. Mean estimation with sub-gaussian rates in polynomial time. *Annals of Statistics*, 48(2):1193–1213, 2020.

Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. *arXiv preprint arXiv:2111.12981*, 2021.

Ziyue Huang, Yuting Liang, and Ke Yi. Instance-optimal mean estimation under differential privacy. *arXiv preprint arXiv:2106.00463*, 2021.

Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964.

Arun Jambulapati, Jerry Li, and Kevin Tian. Robust sub-gaussian principal component analysis and width-independent schatten packing. *Advances in Neural Information Processing Systems*, 33, 2020.

Arun Jambulapati, Jerry Li, Tselil Schramm, and Kevin Tian. Robust regression revisited: Acceleration and improved estimation rates. *arXiv preprint arXiv:2106.11938*, 2021.

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

Gautam Kamath, Vikrant Singhal, and Jonathan Ullman. Private mean estimation of heavy-tailed distributions. *arXiv preprint arXiv:2002.09464*, 2020.

Gautam Kamath, Argyris Mouzakis, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A private and computationally-efficient estimator for unbounded gaussians. *arXiv preprint arXiv:2111.04609*, 2021.

Gautam Kamath, Argyris Mouzakis, and Vikrant Singhal. New lower bounds for private estimation and a generalized fingerprinting lemma. *arXiv preprint arXiv:2205.08532*, 2022.

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

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *Conference on Learning Theory*, pages 25–1. JMLR Workshop and Conference Proceedings, 2012.

Adam Klivans, Pravesh K Kothari, and Raghu Meka. Efficient algorithms for outlier-robust regression. In *Conference On Learning Theory*, pages 1420–1430, 2018.

Weihao Kong, Raghav Somani, Sham Kakade, and Sewoong Oh. Robust meta-learning for mixed linear regression with small batches. *Advances in Neural Information Processing Systems*, 33, 2020.

Pravesh K Kothari, Pasin Manurangsi, and Ameya Velingker. Private robust estimation by stabilizing convex relaxations. *arXiv preprint arXiv:2112.03548*, 2021.

Kevin A Lai, Anup B Rao, and Santosh Vempala. Agnostic estimation of mean and covariance. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 665–674. IEEE, 2016.

Guillaume Lecué and Matthieu Lerasle. Robust machine learning by median-of-means: theory and practice. *The Annals of Statistics*, 48(2):906–931, 2020.

Jing Lei. Differentially private m-estimators. *Advances in Neural Information Processing Systems*, 24:361–369, 2011.

Jerry Li and Guanghao Ye. Robust gaussian covariance estimation in nearly-matrix multiplication time. *Advances in Neural Information Processing Systems*, 33, 2020.

Xiyang Liu, Weihao Kong, Sham Kakade, and Sewoong Oh. Robust and differentially private mean estimation. *arXiv preprint arXiv:2102.09159*, 2021.

H Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. Learning differentially private recurrent language models. *arXiv preprint arXiv:1710.06963*, 2017.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

Kentaro Minami, HItomi Arai, Issei Sato, and Hiroshi Nakagawa. Differential privacy without sensitivity. In *Advances in Neural Information Processing Systems*, pages 956–964, 2016.

Darakhshan J Mir. *Differential privacy: an exploration of the privacy-utility landscape*. Rutgers The State University of New Jersey-New Brunswick, 2013.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

Venkatadheeraj Pichapati, Ananda Theertha Suresh, Felix X Yu, Sashank J Reddi, and Sanjiv Kumar. Adaclip: Adaptive clipping for private sgd. *arXiv preprint arXiv:1908.07643*, 2019.

Adarsh Prasad, Arun Sai Suggala, Sivaraman Balakrishnan, and Pradeep Ravikumar. Robust estimation via robust gradient estimation. *arXiv preprint arXiv:1802.06485*, 2018.

Peter J Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, 8(37):283–297, 1985.

Or Sheffet. Old techniques in differentially private linear regression. In *Algorithmic Learning Theory*, pages 789–827. PMLR, 2019.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

Werner A Stahel. *Robuste schätzungen: infinitesimale optimalität und schätzungen von kovarianz-matrizen*. PhD thesis, ETH Zurich, 1981.

Jacob Steinhardt, Moses Charikar, and Gregory Valiant. Resilience: A criterion for learning in the presence of arbitrary outliers. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Eliad Tsfadia, Edith Cohen, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Friendlycore: Practical differentially private aggregation. *arXiv preprint arXiv:2110.10132*, 2021.

Jonathan Ullman and Adam Sealfon. Efficiently estimating erdos-renyi graphs with node differential privacy. *Advances in Neural Information Processing Systems*, 32, 2019.

Salil Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer, 2017.

Duy Vu and Aleksandra Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In *2009 IEEE International Conference on Data Mining Workshops*, pages 138–143. IEEE, 2009.

Yu-Xiang Wang. Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain. *arXiv preprint arXiv:1803.02596*, 2018.

Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *International Conference on Machine Learning*, pages 2493–2502. PMLR, 2015.

Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, et al. Differentially private fine-tuning of language models. *arXiv preprint arXiv:2110.06500*, 2021.

Banghua Zhu, Jiantao Jiao, and Jacob Steinhardt. Generalized resilience and robust statistics. *arXiv preprint arXiv:1909.08755*, 2019.

Appendix A. Notations, preliminary on differential privacy, and background on Propose-Test-Release

Notations. Let $[n] = \{1, 2, \ldots, n\}$. For $x \in \mathbb{R}^d$, we use $\|x\| = (\sum_{i \in [d]}(x_i)^2)^{1/2}$ to denote the Euclidean norm. For $X \in \mathbb{R}^{d_1 \times d_2}$, we use $\|X\| = \max_{\|v\|_2 = 1} \|Xv\|_2$ to denote the spectral norm. The $d \times d$ identity matrix is $\mathbf{I}_{d \times d}$. The Kronecker product is denoted by $x \otimes y$ for $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$, such that $(x \otimes y)_{(i-1)d+j} = x_i y_j$ for $i \in [d_1]$ and $j \in [d_2]$. Whenever it is clear from context, we use S to denote both a set of data points and also the set of indices of those data points. We use $S \sim S'$ to denote that two datasets S, S' of size n are neighbors, such that $d_{\mathrm{TV}}(\hat{p}_S, \hat{p}_{S'}) \leq 1/n$ where $d_{\mathrm{TV}}(\cdot)$ is the total variation and $\hat{p}_S$ is the empirical distribution of the data points in S. We use $\mu(S)$ and $\Sigma(S)$ to denote mean and covariance of the data points in a dataset S, respectively. We use μ_p and Σ_p to denote mean and covariance of the distribution p.

Differential privacy. We give the backgrounds on differential privacy and the Propose-Test-Release mechanism. We say two datasets S and S' of the same size are neighboring if the Hamming distance between them is at most one. There is another equally popular definition where injecting or deleting one data point to S is considered as a neighboring dataset. All our analysis generalizes to that definition also, but notations get slightly heavier.

Definition 1 ((Dwork et al., 2006)) *We say a randomized algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is (ε, δ)-differentially private if for all neighboring databases $S \sim S' \in \mathcal{X}^n$, and all $Y \subseteq \mathcal{Y}$, we have $\mathbb{P}(M(S) \in Y) \leq e^{\varepsilon}\mathbb{P}(M(S') \in Y) + \delta$.*

HPTR relies on the exponential mechanism for its adaptivity and flexibility.

Definition 2 (Exponential mechanism (McSherry and Talwar, 2007)) *The exponential mechanism $M_{\exp} : \mathcal{X}^n \to \Theta$ takes database $S \in \mathcal{X}^n$, candidate space Θ, score function $D_S(\hat{\theta})$ and sensitivity Δ as input, and select output with probability proportional to $\exp\{-\varepsilon D_S(\hat{\theta})/2\Delta\}$.*

The exponential mechanism is $(\varepsilon, 0)$-DP if the sensitivity of $D_S(\hat{\theta})$ is bounded by Δ.

Lemma 3 ((McSherry and Talwar, 2007)) *If $\max_{\hat{\theta} \in \Theta} \max_{S \sim S'} |D_S(\hat{\theta}) - D_{S'}(\hat{\theta})| \leq \Delta$, then the exponential mechanism is $(\varepsilon, 0)$-DP.*

Starting from the seminal paper (Dwork and Lei, 2009), there are increasing efforts to apply differential privacy to statistical problems, where the dataset consists of i.i.d. samples from a distribution. There are two main challenges. First, the support is typically not bounded, and hence, the sensitivity is unbounded. Dwork and Lei (2009) proposed to resolve this by using robust statistics, such as median, to estimate the mean. The second challenge is that while median is quite insensitive on i.i.d. data, this low sensitivity is only local and holds only for i.i.d. data from a certain class of distributions. This led to the original definition of local sensitivity in the following.

Definition 4 (Local Sensitivity) *We define local sensitivity of dataset $S \in \mathcal{X}^n$ and function $f : \mathcal{X}^n \to \mathbb{R}$ as $\Delta_f(S) := \max_{S' \sim S} |f(S) - f(S')|$.*

Dwork and Lei (2009) introduced the Propose-Test-Release mechanism to resolve both issues. First, a certain robust statistic $f(S)$, such as median, mode, Inter-Quantile Range (IQR), or B-robust

regression model (Hampel et al., 1986) is chosen as a query. It can be used to approximate a target statistic of interest, such as mean, range, or linear regression model, or the robust statistic itself could be the target. Then, the PTR mechanism proceeds in three steps. In the propose step, a local sensitivity Δ is proposed such that $\Delta_f(S) \leq \Delta$ for all S that belongs to a certain family. In the test step, a safety margin m, which is how many data points have to be changed to violate the local sensitivity, is computed and a private version of the safety margin, $\hat{m}$, is compared with a threshold. If the safety margin is large enough, then the algorithm outputs $f(S)$ via a Laplace mechanism with parameter $2\Delta/\varepsilon$. Otherwise, the algorithm halts and outputs $\perp$.

Definition 5 (Propose-Test-Release (PTR) (Dwork and Lei, 2009; Vadhan, 2017)) *For a query function* $f : \mathcal{X}^n \to \mathbb{R}$*, the* PTR *mechanism* $M_{\mathrm{PTR}} : \mathcal{X}^n \to \mathbb{R}$ *proceeds as follows:*

1. **Propose***: Propose a target bound* $\Delta \geq 0$ *on local sensitivity.*

2. **Test***:*

    2.1. *Compute* $m = \min_{S'} d_H(S, S')$ *such that local sensitivity of* S' *satisfies* $\Delta_f(S') \geq \Delta$.
    2.2. *If* $\hat{m} = m + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon) \log(1/\delta)$ *then output* $\perp$*, and otherwise continue.*

3. **Release***: Output* $f(S) + \mathrm{Lap}(2\Delta/\varepsilon)$.

It immediately follows that PTR is (ε, δ)-differentially private for any input dataset.

Lemma 6 ((Dwork and Lei, 2009; Vadhan, 2017)) M_{PTR} *is* (ε, δ)-*DP*

Given a robust statistic of interest, the art is in identifying the family of datasets with small local sensitivity and showing that the sensitivity is small enough to provide good utility. For example, for privately releasing the mode, for the family of distributions whose occurrences of the mode is at least $(4/\varepsilon) \log(1/\delta)$ larger than the occurrences of the second most frequent value, the local sensitivity is zero and PTR outputs the true mode with probability at least $1 - \delta$ (Vadhan, 2017). Such a specialized PTR mechanism for zero local sensitivity is also called the stability based method.

In general, a naive method of computing m in the TEST step requires enumerating over all possible databases $S \in \mathcal{X}^n$. For typical one-dimensional data/statistics, for example median estimation, this step can be computed efficiently. This led to a fruitful line of research in DP statistics on one-dimensional data. (Dwork and Lei, 2009; Brunel and Avella-Medina, 2020) propose PTR mechanisms for the range and the median of of a 1-D smooth distribution and (Avella-Medina and Brunel, 2019; Avella-Medina, 2020; Brunel and Avella-Medina, 2020) propose PTR mechanisms that can estimating median and mean of a 1-D sub-Gaussian distribution. The stability-based method introduced in (Vadhan, 2017) can be used to release private histograms, among other things, which can be subsequently used as a black box to solve several important problems including range estimation of a 1-D sub-Gaussian distribution (Karwa and Vadhan, 2017; Kamath et al., 2019; Liu et al., 2021) or a 1-D heavy-tailed distribution (Kamath et al., 2020; Liu et al., 2021), and general counting queries. PTR and stability-based mechanisms are powerful tools when estimating robust statistics of a distribution from i.i.d. samples.

Even if computational complexity is not concerned, however, directly applying PTR to high dimensional distributions can increase the statistical cost significantly, which has limited the application of PTR. One exception is the recent work of (Brown et al., 2021). For the mean estimation

problem with Mahalanobis error metric of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, the private Tukey median mechanism introduced in (Liu et al., 2021) is studied. One major limitation of the utility analysis is that private Tukey median requires the support to be bounded. In (Liu et al., 2021), this is circumvented by assuming the covariance Σ is known, in which case one can find a support with, for example, the private histogram of (Vadhan, 2017). Instead, Brown et al. (2021) proposed using private Tukey median inside the PTR mechanism and designed an advanced safety test for high-dimensional problems. This naturally bounds the support that adapts to the geometry of the problem without explicitly and privately estimating Σ. One notable byproduct of this approach is that the resulting exponential mechanism is no longer pure DP, but rather (ε, δ)-DP. This is because the resulting exponential mechanism has a support that depends on the dataset S, and hence two exponential mechanisms on two neighboring datasets have different supports. The limitations of the private Tukey median are that (i) it requires symmetric distributions, like Gaussian distributions, and do not generalize to even sub-Gaussian distributions, and (ii) it only works for mean estimation. To handle the first limitation, Brown et al. (2021) propose another PTR mechanism using Gaussian noise, which works for more general sub-Gaussian distributions but achieves sub-optimal sample complexity.

HPTR builds upon this advanced PTR with the high-dimensional safety test from (Brown et al., 2021). However, there are major challenges in applying this safety test to HPTR, which we overcome with the resilience property of the dataset and the robustness of the loss function. For private Tukey median, the sensitivity is always one for any $\hat{\mu}$ and any S, and the only purpose of the safety test is to ensure that the support is not too different between two neighboring datasets. For HPTR, the sensitivity is local in two ways: it requires S to be resilient and the estimate $\hat{\mu}$ to be sufficiently close to μ. To ensure a large enough margin when running the safety test, HPTR requires this local sensitivity to hold not just for the given S but for all S' within some Hamming distance from S. We use the fact that this larger neighborhood is included in an even larger set of databases that are adversarial corruption of the α-fraction of the original resilient dataset S with a certain choice of α. The robustness of our loss function implies that the bounded sensitivity is preserved under such corruption of a resilient dataset. This is critical in proving that a resilient dataset passes the safety test with high probability.

We take a first-principles approach to design a universal framework for DP statistical estimation that blends exponential mechanism, robust statistics, and PTR. The exponential mechanism in HPTR adapts to the geometry of the problem without explicitly estimating any other parameters and also gives us the flexibility to apply to a wide range of problems. The choice of the loss functions that only depend on one-dimensional statistics is critical in achieving the low sensitivity, which directly translates into near optimal utility guarantees for several canonical problems. Ensuring differential privacy is achieved by building upon the advanced PTR framework of (Brown et al., 2021), with a few critical differences. Notably, the safety analysis uses the resilience of robust statistics in a fundamental way.

On the other hand, there is a different way of handling local sensitivity, which is known as smooth sensitivity. Introduced in (Nissim et al., 2007), smooth sensitivity is a smoothed version of local sensitivity on the neighborhood of the dataset, defined as

$$\Delta_f^{\text{smooth}}(S) \;=\; \max_{S' \in \mathcal{X}^n} \{\Delta_f(S') e^{-\varepsilon d_{\text{H}}(S,S')}\}$$

Note that, in general, computing smooth sensitivity is also computationally inefficient with exceptions of (Avella-Medina, 2021; Ullman and Sealfon, 2019). Using smooth sensitivity, (Lei, 2011; Smith, 2011; Chaudhuri and Hsu, 2012; Avella-Medina, 2021) leverage robust M-estimators for

differentially private estimation and inference. The intuition is based on the fact that the influence function of the M-estimators can be used to bound the smooth sensitivity. The applications include: linear regression, location estimation, generalized linear models, private testing. However, these approaches require restrictive assumptions on the dataset that need to be checked (for example via PTR) and fine-grained analyses on the statistical complexity is challenging; there is no sample complexity analysis comparable to ours. One exception is (Bun and Steinke, 2019), which proposes a smooth sensitivity based approach and gives an upper bound on the sub-Gaussian mean estimation error for a finite n, but only for one-dimensional data.

Starting with (Nissim et al., 2007; Dwork and Lei, 2009), robust statistics has been used previously in private estimation, but mostly in one-dimensional problems. (McMahan et al., 2017; Pichapati et al., 2019; Amin et al., 2019; Andrew et al., 2021) propose practical differentially private estimators for one-dimensional median, mean, and interquartile range estimations, but without the analysis on the utility. Some recent works by (Asi and Duchi, 2020; Huang et al., 2021; Dong and Yi, 2021) further studied these problems with more fine-grained analysis that provides instance optimality. (Avella-Medina and Brunel, 2019; Avella-Medina, 2020; Brunel and Avella-Medina, 2020) developed private algorithms for estimating the mean and median of unbounded 1-dimensional sub-Gaussian and heavy tail distributions. In a different direction, Tsfadia et al. (2021) privately finds subset of data points that includes all points, except possibly few outliers, and is certified to have diameter same as the effective diameter of the dataset. This can be used as a pre-processing to solve, for example, mean estimation and clustering robustly and privately.

More recently, (Hopkins et al., 2021; Kothari et al., 2021) investigate the use of convex relaxations for robust estimation to design polynomial time algorithms with privacy guarantees. In a breakthrough result at the intersection of robust estimation and private estimation, Hopkins et al. (2021) introduces an exponential mechanism based on sum-of-squares proofs that can be implemented in polynomial time. For mean estimation with covariance bounded distributions, this achieves a significantly improved result with optimal sample complexity, optimal robustness, pure DP, and a sub-Gaussian error rate. Kothari et al. (2021) introduces convex relaxations for robust estimation to satisfy strong worst-case stability and prove guarantees on estimating higher-order moments beyond covariance matrices.

Appendix B. Mean estimation

In a standard mean estimation, we are given i.i.d. samples $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ drawn from a distribution $P_{\mu,\Sigma}$ with an unknown mean μ (which corresponds to θ in the general notation) and an unknown covariance $\Sigma \succ 0$ (which corresponds to ϕ in the general notation), and we want to produce a DP estimate $\hat{\mu}$ of the mean. The resulting error is best measured in Mahalanobis distance, $D_\Sigma(\hat{\mu}, \mu) = \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, because this is a scale-invariant distance; every direction has unit variance after whitening by Σ.

This problem is especially challenging since we aim for a tight guarantee that adapts to the unknown Σ as measured in the Mahalanobis distance without enough samples to directly estimate Σ (see Section 1.1 for a survey). Despite being a canonical problem in DP statistics, the optimal sample complexity is not known even for standard sub-Gaussian and heavy-tailed distributions. We characterize the optimal sample complexity by showing that HPTR matches the known lower bounds in Section B.3. This follows directly from the general three-step strategy outlined in Section 1.2.1.

B.1. Step 1: Designing the surrogate $D_S(\hat{\mu})$ for the Mahalanobis distance

We want to privately release $\hat{\mu}$ with small Mahalanobis distance $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$. In the exponential mechanism in RELEASE step, we propose using the surrogate distance,

$$D_S(\hat{\mu}) \quad = \quad \max_{v:\|v\|\leq 1} \frac{\langle v, \hat{\mu}\rangle - \mu_v(\mathcal{M}_{v,\alpha})}{\sigma_v(\mathcal{M}_{v,\alpha})} \; , \tag{3}$$

where the robust one-dimensional mean $\mu_v(\mathcal{M}_{v,\alpha})$ and variance $\sigma_v^2(\mathcal{M}_{v,\alpha})$ are defined as follows. We partition $S = \{x_i\}_{i=1}^n$ into three sets, $\mathcal{B}_{v,\alpha}$, $\mathcal{M}_{v,\alpha}$, and $\mathcal{T}_{v,\alpha}$, by considering a set of projected data points $S_v = \{\langle v, x_i\rangle\}_{x_i \in S}$ and letting $\mathcal{B}_{v,\alpha}$ be the data points corresponding to the subset of bottom $(2/5.5)\alpha n$ data points with the smallest values in S_v, $\mathcal{T}_{v,\alpha}$ be the subset of the top $(2/5.5)\alpha n$ data points with the largest values, and $\mathcal{M}_{v,\alpha}$ be the subset of remaining $(1-(4/5.5)\alpha)n$ data points. For a fixed direction v, define

$$\mu_v(\mathcal{M}_{v,\alpha}) \quad = \quad \frac{1}{|\mathcal{M}_{v,\alpha}|} \sum_{x_i \in \mathcal{M}_{v,\alpha}} \langle v, x_i\rangle \; , \text{ and } \; \sigma_v^2(\mathcal{M}_{v,\alpha}) = \frac{1}{|\mathcal{M}_{v,\alpha}|} \sum_{x_i \in \mathcal{M}_{v,\alpha}} (\langle v, x_i\rangle - \mu_v(\mathcal{M}_{v,\alpha}))^2 \; , \tag{4}$$

which are robust estimates of the population projected mean $\mu_v = \langle v, \mu\rangle$ and the population projected variance $\sigma_v^2 = v^\top \Sigma v$.

General guiding principles for designing $D_S(\hat{\mu})$. We propose the following three design principles that apply more generally to all problem instances of interest. The first guideline is that it should recover the target error metric $D_\Sigma(\hat{\mu}, \mu) = \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$ when we substitute the population statistics, e.g., μ_v and σ_v for mean estimation, for their robust counterparts: $\mu_v(\mathcal{M}_{v,\alpha})$ and $\sigma_v(\mathcal{M}_{v,\alpha})$. This ensures that minimizing $D_S(\hat{\mu})$ is approximately equivalent to minimizing the target metric $D_\Sigma(\hat{\mu}, \mu) = \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$ (Lemma 12). For mean estimation, this equivalence is shown in the following lemma.

Lemma 7 *For any $\mu \in \mathbb{R}^d$ and $0 \prec \Sigma \in \mathbb{R}^{d\times d}$, let $\mu_v = \langle v, \mu\rangle$ and $\sigma_v^2 = v^\top \Sigma v$. Then, we have*

$$\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \quad = \quad \max_{v:\|v\|\leq 1} \frac{\langle v, \hat{\mu}\rangle - \mu_v}{\sigma_v} \; .$$

Proof Let $\hat{\mu} - \mu = \sum_{\ell=1}^d a_\ell u_\ell$ with $a_\ell = \langle u_\ell, \hat{\mu} - \mu\rangle$, $\|a\| = \|\hat{\mu} - \mu\|$ and u_ℓ's are the singular vectors of Σ. Similarly, let $v = \sum_{\ell=1}^d b_\ell u_\ell$ with $\|b\| = 1$. Then, we have

$$\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|^2 = \sum(a_\ell^2/\sigma_\ell) \text{ and } \frac{\langle v, (\hat{\mu} - \mu)\rangle}{\sigma_v} = \frac{\langle a, b\rangle}{\sqrt{\sum b_\ell^2 \sigma_\ell}} \; .$$

From Cauchy-Schwarz, we have $\langle a, b\rangle^2 \leq (\sum b_\ell^2 \sigma_\ell)(\sum a_\ell^2 \sigma_\ell^{-1})$, which proves that

$$\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \geq \max_{v:\|v\|=1} (1/\sigma_v)\langle v, (\hat{\mu} - \mu)\rangle \; .$$

To show equality, we find v that makes Cauchy-Schwarz inequality tight. Let $v = \sum_{\ell=1}^d b_\ell u_\ell$ with a choice of $b_\ell = (1/Z)a_\ell \sigma_\ell^{-1}$ and $Z = \sqrt{\sum_\ell a_\ell^2 \sigma_\ell^{-2}}$. This implies $\|b\| = 1$ and

$$\langle a, b\rangle \quad = \quad \frac{1}{Z} \sum_{\ell=1}^d (1/\sigma_\ell)a_\ell^2, \text{ and } \sqrt{\sum b_\ell^2 \sigma_\ell} = \frac{1}{Z}\sqrt{\sum_{\ell=1}^d (1/\sigma_{u_\ell})a_\ell^2} \; ,$$

which implies that there exists a v such that $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = (1/\sigma_v)\langle v, \hat{\mu} - \mu\rangle$ and $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq \max_{v:\|v\|=1}(1/\sigma_v)\langle v, \hat{\mu} - \mu\rangle$. ∎

The second guideline is that $D_S(\hat{\mu})$ should depend only on the *one-dimensional* statistics of the data. This is critical since the sensitivity of high-dimensional statistics increases with the ambient dimension d. For example, consider using the robust mean estimate $\hat{\mu}_{\text{robust}}(S) \in \mathbb{R}^d$ from (Dong et al., 2019) and using the Euclidean distance $D_S(\hat{\mu}) = \|\hat{\mu} - \hat{\mu}_{\text{robust}}(S)\|$ in the exponential mechanism, where we are assuming $\Sigma = \mathbf{I}$ for simplicity. It can be shown that, even for Gaussian distributions, this requires $n = \widetilde{\Omega}(d^{3/2}/(\varepsilon\alpha) + d/\alpha^2)$ samples to achieve an accuracy of $\|\hat{\mu} - \mu\| = \widetilde{O}(\alpha)$. This is significantly sub-optimal compared to what HPTR achieves in Corollary B.2, which leverages the fact that sensitivity of one-dimensional statistics is dimension-independent.

The last guideline is to use robust statistics. Robust statistics have small sensitivity on *resilient* datasets, which is critical in achieving the near-optimal guarantees. We elaborate on it in Section B.2.2.

B.2. Step 2: Utility analysis under resilience

For utility, we prefer smaller Δ and τ to ensure that the exponential mechanism samples $\hat{\mu}$ closer to the minimum of $D_S(\hat{\mu}) \approx \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$. However, aggressive choices can violate the DP condition and hence fail the safety test. Near-optimal utility can be achieved by selecting Δ and τ based on the *resilience* of the dataset, defined as follows.

Definition 8 (Resilience for mean estimation (Steinhardt et al., 2018; Zhu et al., 2019)) *For some $\alpha \in (0,1)$, $\rho_1 \in \mathbb{R}_+$, and $\rho_2 \in \mathbb{R}_+$, we say a set of n data points S_{good} is (α, ρ_1, ρ_2)-resilient with respect to (μ, Σ) if for any $T \subset S_{\text{good}}$ of size $|T| \geq (1-\alpha)n$, the following holds for all $v \in \mathbb{R}^d$ with $\|v\| = 1$:*

$$\left|\frac{1}{|T|}\sum_{x_i \in T}\langle v, x_i\rangle - \mu_v\right| \leq \rho_1\,\sigma_v \,, \text{ and} \tag{5}$$

$$\left|\frac{1}{|T|}\sum_{x_i \in T}\left(\langle v, x_i\rangle - \mu_v\right)^2 - \sigma_v^2\right| \leq \rho_2\,\sigma_v^2 \,, \tag{6}$$

where $\mu_v = \langle v, \mu\rangle$ and $\sigma_v^2 = v^\top\Sigma v$.

Originally, resilience is introduced in the context of robust statistics. Resilience measures how sensitive the sample statistics are to removing an α-fraction of the data points. A dataset from a distribution with a lighter tail has smaller resilience (ρ_1, ρ_2). For example, sub-Gaussian distributions have $\rho_1 = O(\alpha\sqrt{\log(1/\alpha)})$ and $\rho_2 = O(\alpha\log(1/\alpha))$ (Lemma 17), which are smaller than the resilience of heavy-tailed distributions with bounded k-th moment, i.e., $\rho_1 = O(\alpha^{1-1/k})$ and $\rho_2 = O(\alpha^{1-2/k})$ (Lemma 19). Resilience plays a crucial role in robust statistics, where the resilience of a dataset determines the minimax sample complexity of estimating population statistics from adversarially corrupted samples (Steinhardt et al., 2018; Zhu et al., 2019).

In the context of differential privacy, our design of HPTR is guided by our analysis showing that the sensitivity of one-dimensional robust statistics is fundamentally governed by resilience. Leveraging this three-way connection between the use of robust statistics in the algorithm, the resilience of the data, and the sensitivity of the distance $D_S(\hat{\mu})$ is crucial in achieving near-optimal utility.

Concretely, we consider α as a free parameter that we can choose depending on the target accuracy. For example, let $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = 32\rho_1$ be our target accuracy. Note that we did not optimize the constants in our analysis, and they can be further tightened. In the case of sub-Gaussian distributions, we have $\rho_1 = C'\alpha\sqrt{\log(1/\alpha)}$ w.h.p. when the sample size is large enough. This determines the value of α that achieves the target accuracy and also the choice of Δ and τ, as follows.

The robust statistics of a resilient dataset (i.e., one with small resilience) cannot change too much when a small fraction of the dataset is changed. This is made precise in Lemma 16, which shows, for example, that the robust mean $\mu_v(\mathcal{M}_{v,\alpha})$ can change only by $O(\rho_1/(\alpha n))$ when one data point is arbitrarily changed. This implies the sensitivity of $D_S(\hat{\mu})$ is also small: $\Delta = O(\rho_1/(\alpha n))$. Choosing $\tau = 42\rho_1$ to be larger by a constant factor from the target accuracy, we show that a sample size of $n = O(d/(\varepsilon\alpha))$ is sufficient to achieve the desired utility.

Theorem 9 (Utility guarantee for mean estimation) *There exist positive constants c and C such that for any (α, ρ_1, ρ_2)-resilient set S with respect to some $(\mu \in \mathbb{R}^d, \Sigma \succ 0)$ satisfying $\alpha \in (0, c)$, $\rho_1 < c$, $\rho_2 < c$, and $\rho_1^2 \leq c\alpha$, HPTR with the choices of the distance function in Eq. (3), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq 32\rho_1$ with probability $1 - \zeta$, if*

$$n \geq C\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} \; .$$

This theorem shows how a resilient dataset (which is a deterministic condition) implies small error for HPTR. We make formal connections to standard assumptions on the sample generating distributions and their respective resiliences in Section B.3, where we also discuss the optimality of this utility guarantee. For example, sub-Gaussian distributions have $\rho_1 = O(\alpha\sqrt{\log(1/\alpha)})$ when $n \geq C'd/(\alpha\log(1/\alpha))^2$ for any α smaller than a universal constant. This implies that HPTR achieves a target accuracy of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq \tilde{\alpha}$ with sample size $\tilde{O}(\frac{d}{\tilde{\alpha}^2} + \frac{d}{\tilde{\alpha}\varepsilon})$, where $\tilde{O}$ hides logarithmic factors in $1/\alpha$, δ, and ζ. We explain the intuition behind our analysis and provide a complete proof in Sections B.2.2–B.2.6. One by-product of using robust statistics is that we get robustness for free, as we next show.

B.2.1. ROBUSTNESS OF HPTR

One by-product of using robust statistics is that HPTR is also robust to adversarial corruption. We therefore provide a more general guarantee that simultaneously achieves DP and robustness. Suppose we are given a dataset S that is a corrupted version of a resilient dataset S_{good}.

Assumption 1 (α_{corrupt}-corruption) *Given a set $S_{\text{good}} = \{\tilde{x}_i \in \mathbb{R}^d\}_{i=1}^n$ of n data points, an adversary inspects all data points, selects $\alpha_{\text{corrupt}}n$ of the data points, and replaces them with arbitrary dataset S_{bad} of size $\alpha_{\text{corrupt}}n$. The resulting corrupted dataset is called $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$.*

This adaptive adversary is strong since the corruption can adapt to the entire dataset (for example, it covers the Huber contamination model (Huber, 1964) and the non-adaptive adversarial model (Lecué and Lerasle, 2020)). This threat model is now standard in robust statistics literature (Steinhardt et al., 2018). If the original S_{good} is resilient, we show that the same guarantee as Theorem 9 holds under corruption up to an α_{corrupt} fraction of S_{good} for sufficiently small $\alpha_{\text{corrupt}} \leq (1/5.5)\alpha$.

The factor $1/5.5$ is due to the fact that the algorithm treats some of the good data points as outliers (which is at most $4\alpha_{\text{corrupt}}$ due to the top and bottom tails cut in the definition of $\mathcal{M}_{v,(2/5.5)\alpha}$), and we need to handle neighboring datasets up to $(0.5/5.5)\alpha n$ Hamming distance. Hence, we need to ensure resilience for α that is at least 5.5 times larger than the corruption α_{corrupt}.

Definition 9 (Corrupt good set) *We say a dataset S is $(\alpha_{\text{corrupt}}, \alpha, \rho_1, \rho_2)$-corrupt good with respect to (μ, Σ) if it is an α_{corrupt}-corruption of an (α, ρ_1, ρ_2)-resilient dataset S_{good}.*

We get the following theorem showing that HPTR can tolerate up to $(1/5.5)\alpha$ fraction of the data being arbitrarily corrupted.

Theorem 10 (Robustness) *There exist positive constants c and C such that for any $((2/11)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S with respect to $(\mu \in \mathbb{R}^d, \Sigma \succ 0)$ satisfying $\alpha < c$, $\rho_1 < c$, $\rho_2 < c$, and $\rho_1^2 \leq c\alpha$, HPTR with the distance function in Eq. (3), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq 32\rho_1$ with probability $1 - \zeta$, if*

$$n \geq C\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} \ .$$

In Sections B.2.2–B.2.6, we prove this more general result. When there is no adversarial corruption, Theorem 9 immediately follows as a special case by selecting α as a free parameter depending on the target accuracy. The constants in all the theorems can be improved if we track them more carefully, and we did not attempt to optimize them in this paper.

B.2.2. PROOF STRATEGY FOR THEOREM 10

We show in Section B.2.5 that the robust one-dimensional statistics, $\mu_v(\mathcal{M}_{v,\alpha})$ and $\sigma_v^2(\mathcal{M}_{v,\alpha})$, have small sensitivity if the dataset is resilient. Consequently, $D_S(\hat{\mu})$ has a small *local* sensitivity, i.e., the sensitivity is small if restricted to $\hat{\mu}$ close to μ and if the dataset is resilient. To ensure DP, we run RELEASE only when those two locality conditions are satisfied; we first PROPOSE the sensitivity Δ and a threshold τ, and then we TEST that DP guarantees are met on the given dataset with those choices. Resilient datasets (i) pass this safety test with a high probability and (ii) achieve the desired accuracy, both of which rely on our general analysis of HPTR with a general distance function (Theorem 15). We give sketches of the main steps below.

One-dimensional robust statistics have small sensitivity on resilient datasets. Consider the robust projected mean $\mu_v(\mathcal{M}_{v,\alpha})$ for some small enough $\alpha > 0$. If S is (α, ρ_1, ρ_2)-resilient, then the following technical lemma shows that the top and bottom $(2/5.5)\alpha$-tails cannot deviate too much from the mean.

Lemma 10 (Lemma 10 from (Steinhardt et al., 2018)) *For a (α, ρ_1, ρ_2)-resilient dataset S with respect to (μ, Σ) and any $0 \leq \tilde{\alpha} \leq \alpha$, the following holds for any subset $T \subset S$ of size at least $\tilde{\alpha}n$ and for any unit norm $v \in \mathbb{R}^d$:*

$$\left|\frac{1}{|T|}\sum_{x_i \in T}\langle v, x_i - \mu\rangle\right| \ \leq \ \frac{2 - \tilde{\alpha}}{\tilde{\alpha}}\rho_1\sigma_v \ , \text{ and} \tag{7}$$

$$\left|\frac{1}{|T|}\sum_{x_i \in T}\left(\langle v, x_i - \mu\rangle^2 - \sigma_v^2\right)\right| \ \leq \ \frac{2 - \tilde{\alpha}}{\tilde{\alpha}}\rho_2\sigma_v^2 \ . \tag{8}$$

Under the definitions in Eq. (3), the top $(2/5.5)\alpha$-tail denoted by $\mathcal{T}_{v,\alpha}$ and bottom $(2/5.5)\alpha$-tail denoted by $\mathcal{B}_{v,\alpha}$ have the empirical means that are no more than $O(\sigma_v\rho_1/\alpha)$ away from the true projected mean μ_v, respectively. It follows that there exists at least one data point in $\mathcal{T}_{v,\alpha}$ and one data point in $\mathcal{B}_{v,\alpha}$ that are no more than $O(\sigma_v\rho_1/\alpha)$ away from μ_v. This implies that the range of the middle subset $\mathcal{M}_{v,\alpha}$ is provably bounded by $O(\sigma_v\rho_1/\alpha)$, and the sensitivity of the robust mean $\mu_v(\mathcal{M}_{v,\alpha})$ is guaranteed to be $O(\sigma_v\rho_1/(\alpha n))$. We can similarly show that $\sigma_v^2(\mathcal{M}_{v,\alpha})$ has sensitivity $O(\sigma_v^2\rho_1^2/(\alpha^2 n))$, as shown in Eq. (18). Note that these sensitivity bounds are *local* in the sense that they require the data to be (α, ρ_1, ρ_2)-resilient.

Small local sensitivity of $D_S(\hat{\mu})$. Under the above sensitivity bounds for $\mu_v(\mathcal{M}_{v,\alpha})$ and $\sigma_v^2(\mathcal{M}_{v,\alpha})$, it follows after some calculations as shown in Eq. (19) that the sensitivity for a resilient dataset S is bounded by

$$| D_S(\hat{\mu}) - D_{S'}(\hat{\mu}) | \ \leq \ C' \frac{\rho_1}{\alpha n}\Big(1 + \frac{\rho_1\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|}{\alpha}\Big) , \tag{9}$$

for some constant C' and all neighboring datasets S', assuming ρ_2 is sufficiently small. Note that this sensitivity bound is *local* for two reasons; for this sensitivity to be small (i.e. $O(\rho_1/(\alpha n))$), we require S to be resilient and $\hat{\mu}$ to be close to μ. Thus, the meaning of *local* here is two fold, while traditionally local sensitivity in the privacy literature only concerns the sensitivity of a particular dataset S. We handle these two localities with the T‍EST step, among other things, checks that the DP conditions are satisfied for the given dataset and the choice of Δ and τ, which bounds the support of the exponential mechanism to be within $B_{\tau,S} = \{\hat{\mu} : D_S(\hat{\mu}) \leq \tau\}$ with a choice of $\tau = O(\rho_1)$. Consequently, we require $\rho_1^2/\alpha \ll 1$ for the second term in Eq. (9) to be dominated by the first. Fortunately, this is indeed true for all scenarios of interests to us. For sub-Gaussian distributions, $\rho_1^2 = \alpha^2 \log(1/\alpha) \ll \alpha$. For k-th moment bounded distributions with $k > 3$, $\rho_1^2 = \alpha^{2-2/k} \ll \alpha$. For covariance bounded distributions, we do not hope to get a Mahalanobis distance guarantee. Instead, we aim for a Euclidean distance guarantee whose sensitivity does not depend on $\hat{\mu}$, and we do not require $\rho_1^2/\alpha \ll 1$ (Section B.3.3).

Sample complexity analysis. Assuming the sensitivity of $D_S(\hat{\mu})$ is bounded by $\Delta = O(\rho_1/(\alpha n))$, which we ensure with the safety test, we analyze the utility of the exponential mechanism. For a target accuracy of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\rho_1)$, we consider two sets $B_{\mathrm{out}} = \{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq c_0\rho_1\}$ and $B_{\mathrm{in}} = \{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq c_1\rho_1\}$ for some $c_0 > c_1$. The exponential mechanism achieves accuracy $c_0\rho_1$ with probability $1 - \zeta$ if

$$\mathbb{P}(\hat{\mu} \notin B_{\mathrm{out}}) \ \leq \ \frac{\mathbb{P}(\hat{\mu} \notin B_{\mathrm{out}})}{\mathbb{P}(\hat{\mu} \in B_{\mathrm{in}})} \ \lesssim \ \frac{\mathrm{Vol}(B_{\tau,S})}{\mathrm{Vol}(B_{\mathrm{in}})} \frac{e^{-\frac{\varepsilon}{4\Delta}c_0\rho_1}}{e^{-\frac{\varepsilon}{4\Delta}c_1\rho_1}} \leq e^{O(d)}e^{-\frac{\varepsilon}{4\Delta}(c_0-c_1)\rho_1} \leq \zeta ,$$

where the second inequality requires $D_S(\hat{\mu}) \simeq \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, which we show in Lemma 12. Since the volume ratio is $\mathrm{Vol}(B_{\tau,S})/\mathrm{Vol}(B_{\mathrm{out}}) = e^{O(d)}$, $\tau = O(\rho_1)$, and $\Delta = O(\rho_1/(\alpha n))$, it is sufficient to have a large enough c_0 and $n = O((d + \log(1/\zeta))/(\alpha\varepsilon))$ with a large enough constant.

Safety test. We are left to show that for a resilient dataset, the failure probability of the safety test, $\mathbb{P}(m_\tau + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon)\log(2/\delta))$, is less than ζ. This requires the safety margin to be large enough, i.e., $m_\tau \geq k^* = (2/\varepsilon)\log(4/(\delta\zeta))$. Recall that the safety margin is defined as the Hamming distance to the closest dataset to S where the $(\varepsilon/2, \delta/2)$-DP condition of the exponential mechanism is violated. We therefore need to show that the DP condition is satisfied not only for S but for any dataset S' at Hamming distance at most k^* from S.

Consider two exponential mechanisms, $r_{(\varepsilon,\Delta,\tau,S')}$ and $r_{(\varepsilon,\Delta,\tau,S'')}$, on neighboring datasets S' and S''. Since $B_{\tau,S'} \neq B_{\tau,S''}$, we separately analyze the intersection $B_{\tau,S'} \cap B_{\tau,S''}$ and the differences $B_{\tau,S'} \setminus B_{\tau,S''}$ and $B_{\tau,S''} \setminus B_{\tau,S'}$. In the intersection, we show that the two probability distributions are within a multiplicative factor $e^{\varepsilon/2}$ of each other:

$$\mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\mu} \in A) \ \leq \ e^{\varepsilon/2}\mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\mu} \in A)$$

for all $A \subseteq B_{\tau,S'} \cap B_{\tau,S''}$, S' within Hamming distance k^* from a resilient dataset S, and $S'' \sim S'$. The main challenge is that S' is no longer a resilient dataset but a k^*-neighbor of a resilient dataset. Since such S' is $(k^*/n, \alpha, \rho_1, \rho_2)$-corrupt good (Definition 9), we show that corrupt good sets also inherit the bounded local sensitivity of a resilient dataset seamlessly, as shown in Lemma 16.

In the set difference, we show that the total probability mass, $\mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S)}}(\hat{\mu} \in B_{\tau,S} \setminus B_{\tau,S'})$ and $\mathbb{P}_{r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\mu} \in B_{\tau,S'} \setminus B_{\tau,S})$, are bounded by δ, respectively, as long as the overlap of the two supports are large enough. This requires $\tau \gg \Delta k^*$, as we show in Appendix F.1, which is satisfied for $n \geq (\log(1/(\delta\zeta))/(\alpha\varepsilon))$.

Outline. The analyses for the accuracy and the safety test build upon a universal analysis of HPTR in Theorem 15, which holds more generally for any distance function $D_\phi(\hat{\theta})$ in the estimation problems of interest. For mean estimation, we show in Sections B.2.3-B.2.5 that the sufficient conditions of Theorem 15 are met for the choices of constants and parameters: $\rho = \rho_1$, $c_0 = 31.8$, $c_1 = 10.2$, $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$, $\tau = 42\rho_1$, and $\Delta = 110\rho_1/(\alpha n)$. We can set c_2 to be a large constant and will only change the constant factor in the sample complexity, which we do not track. A proof of Theorem 10 is provided in Section B.2.6, from which Theorem 9 follows immediately. All the lemmas assume $((1/5.5)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S, $\alpha \leq 0.015$, $\rho_1 \leq 0.013$, and $\rho_2 \leq 0.0005$. We omit this assumption in stating the lemmas for brevity.

B.2.3. RESILIENCE IMPLIES ROBUSTNESS

For the assumption (d) in Theorem 15, we show that $D_S(\hat{\mu})$ is a good approximation of the true distance $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$ in Lemma 12. We first show that the one-dimensional mean and the variance of the filtered out $\mathcal{M}_{v,\alpha}$ are robust.

Lemma 11 *For any unit norm $v \in \mathbb{R}^d$, $|\langle v, \mu - \mu(\mathcal{M}_{v,\alpha})\rangle| \leq 6\rho_1 \sigma_v$ and $0.9\sigma_v \leq \sigma_v(\mathcal{M}_{v,\alpha}) \leq 1.1\sigma_v$.*

Proof For the mean bound,

$$|\langle v, \mu - \mu(\mathcal{M}_{v,\alpha})\rangle|$$
$$\leq \frac{|\mathcal{M}_{v,\alpha} \cap S_{\text{bad}}|}{|\mathcal{M}_{v,\alpha}|}|\langle v, \mu(S_{\text{bad}} \cap \mathcal{M}_{v,\alpha}) - \mu\rangle| + \frac{|\mathcal{M}_{v,\alpha} \cap S_{\text{good}}|}{|\mathcal{M}_{v,\alpha}|}|\langle v, \mu(S_{\text{good}} \cap \mathcal{M}_{v,\alpha}) - \mu\rangle|$$
$$\leq \frac{(1/5.5)\alpha}{1 - (4/5.5)\alpha}\frac{2\rho_1\sigma_v}{(1/5.5)\alpha} + \frac{1 - (1/5.5)\alpha}{1 - (4/5.5)\alpha}\rho_1\sigma_v$$
$$\leq (2\rho_1 + \rho_1)\sigma_v/(1 - (4/5.5)\alpha) . \tag{10}$$

The second inequality follows from the following. First, $|\langle v, \mu(S_{\text{good}} \cap \mathcal{M}_{v,\alpha}) - \mu\rangle| \leq \sigma_v\rho_1$ by the definition of resilience and that fact that $|S_{\text{good}} \cap \mathcal{M}_{v,\alpha}| \geq (1 - (5/5.5)\alpha)n$. Next, since $|\langle v, \mu(S_{\text{bad}} \cap \mathcal{M}_{v,\alpha}) - \mu\rangle|$ is less than $|\langle v, \mu(S_{\text{good}} \cap \mathcal{T}_{v,\alpha}) - \mu\rangle|$ or $|\langle v, \mu(S_{\text{good}} \cap \mathcal{B}_{v,\alpha}) - \mu\rangle|$, both

of which are at most $2\rho_1\sigma_v/(1/5.5)\alpha$, from applying Lemma 10 with a set size at least $(1/5.5)\alpha n$, we have

$$|\langle v, \mu(S_{\text{bad}} \cap \mathcal{M}_{v,\alpha}) - \mu\rangle| \leq \frac{2}{(1/5.5)\alpha}\rho_1\sigma_v .$$

The mean bound follows from (10) and $\alpha \leq 0.1$. For the variance upper bound,

$$\sigma_v(\mathcal{M}_{v,\alpha})^2 = \frac{1}{(1-(4/5.5)\alpha)n}\sum_{x_i\in\mathcal{M}_{v,\alpha}}\langle v, x_i - \mu(\mathcal{M}_{v,\alpha})\rangle^2 \leq \frac{1}{(1-(4/5.5)\alpha)n}\sum_{x_i\in\mathcal{M}_{v,\alpha}}\langle v, x_i - \mu\rangle^2 ,$$

where the first inequality follows from the fact that subtracting the empirical mean $\mu(\mathcal{M}_{v,\alpha})$ minimizes the second moment. We can decompose the empirical deviation and show an upper bound first:

$$\frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}}(\langle v, x_i - \mu\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n}$$

$$= \frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}\cap S_{\text{bad}}}(\langle v, x_i - \mu\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n} + \frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}\cap S_{\text{good}}}(\langle v, x_i - \mu\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n}$$

$$\leq \frac{(1/5.5)\alpha(2\rho_2/(1/5.5)\alpha)\sigma_v^2 + (1-(4/5.5)\alpha)\rho_2\sigma_v^2}{1-(4/5.5)\alpha} \leq 6\rho_2\sigma_v^2 , \tag{11}$$

where in the second inequality we used resilience on $\mathcal{M}_{v,\alpha} \cap S_{\text{good}}$ of size at least $1 - (5/5.5)\alpha$. For $x_i \in S_{\text{bad}} \cap \mathcal{M}_{v,\alpha}$, we use the fact that

$$\left|\langle v, x_i - \mu\rangle^2 - \sigma_v^2\right| \leq \max\left\{\frac{\sum_{j\in S_{\text{good}}\cap\mathcal{T}_{v,\alpha}}(\langle v, x_j - \mu\rangle^2 - \sigma_v^2)}{|S_{\text{good}} \cap \mathcal{T}_{v,\alpha}|}, \frac{\sum_{j\in S_{\text{good}}\cap\mathcal{B}_{v,\alpha}}(\langle v, x_j - \mu\rangle^2 - \sigma_v^2)}{|S_{\text{good}} \cap \mathcal{B}_{v,\alpha}|}\right\}$$

$$\leq \frac{2\rho_2\sigma_v^2}{(1/5.5)\alpha} ,$$

where we used Eq. (8) in Lemma 10 for sets with size at least $(1/5.5)\alpha n$. For the variance deviation lower bound,

$$\frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}}(\langle v, x_i - \mu(\mathcal{M}_{v,\alpha})\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n} = \frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}}\left(\langle v, x_i - \mu\rangle^2 - \sigma_v^2 - \langle v, \mu - \mu(\mathcal{M}_{v,\alpha})\rangle^2\right)}{(1-(4/5.5)\alpha)n}$$

$$\geq \frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}\cap S_{\text{bad}}}(\langle v, x_i - \mu\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n} + \frac{\sum_{x_i\in\mathcal{M}_{v,\alpha}\cap S_{\text{good}}}(\langle v, x_i - \mu\rangle^2 - \sigma_v^2)}{(1-(4/5.5)\alpha)n} - 36\rho_1^2\sigma_v^2 ,$$

$$\geq -\frac{2\rho_2\sigma_v^2}{1-(4/5.5)\alpha} - \frac{1-(4/5.5)\alpha}{1-(4/5.5)\alpha}\rho_2\sigma_v^2 - 36\rho_1^2\sigma_v^2 \geq -(3.2\rho_2 + 36\rho_1^2)\sigma_v^2 , \tag{12}$$

where we used $\alpha \leq 0.1$, the first term only uses the fact that $|S_{\text{bad}}| \leq (1/5.5)\alpha n$, the second term uses resilience, and the last term uses the mean bound we proved earlier. In (11) and (12), assuming $\rho_1 \leq 0.04$, and $\rho_2 \leq 0.035$, we have $\sqrt{1 + 6\rho_2} \leq 1.1$ and $\sqrt{1 - 3.2\rho_2 - 36\rho_1^2} \geq 0.9$. ∎

We show that resilience implies our estimate of the distance is robust.

Lemma 12 *If $\hat{\mu} \in B_{\tau,S}$ and $\tau = 42\rho_1$, then $\left|\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| - D_S(\hat{\mu})\right| \leq 6\rho_1 + 0.1\tau \leq 10.2\rho_1$.*

Proof From Lemma 11, we know that for all $\hat{\mu} \in B_{t,S}$,

$$D_S(\hat{\mu}) = \max_{\|v\|=1} \frac{\langle v, \hat{\mu} - \mu(\mathcal{M}_{v,\alpha})\rangle}{\sigma_v(\mathcal{M}_{v,\alpha})} \geq \max_{\|v\|=1} \frac{\langle v, \hat{\mu} - \mu\rangle - 6\rho_1\sigma_v}{1.1\sigma_v} . \tag{13}$$

and

$$D_S(\hat{\mu}) = \max_{\|v\|=1} \frac{\langle v, \hat{\mu} - \mu(\mathcal{M}_{v,\alpha})\rangle}{\sigma_v(\mathcal{M}_{v,\alpha})} \leq \max_{\|v\|=1} \frac{\langle v, \hat{\mu} - \mu\rangle + 6\rho_1\sigma_v}{0.9\sigma_v} . \tag{14}$$

Applying Lemma 7, we get $0.9D_S(\hat{\mu}) - 6\rho_1 \leq \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq 1.1D_S(\hat{\mu}) + 6\rho_1$. Since $D_S(\hat{\mu}) \leq \tau$, we get the desired bound.

∎

B.2.4. BOUNDED VOLUME

We show that the assumption (a) in Theorem 15 is satisfied for robust estimate $D_S(\hat{\mu})$.

Lemma 13 *For $\rho = \rho_1$, $c_1 = 10.2$, $\tau = 42\rho_1$, $\Delta = 110\rho_1/(\alpha n)$, and $c_2 \geq \log(67/12) + \log((c_0 + 2c_1)/c_1)$, we have $(7/8)\tau - (k^* + 1)\Delta > 0$,*

$$\frac{\text{Vol}(B_{\tau+(k^*+1)\Delta+c_1\rho,S})}{\text{Vol}(B_{(7/8)\tau-(k^*+1)\Delta-c_1\rho,S})} \leq e^{c_2 d} , \text{ and}$$

$$\frac{\text{Vol}(\{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq (c_0 + 2c_1)\rho\})}{\text{Vol}(\{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq c_1\rho\})} \leq e^{c_2 d} .$$

Proof The second part of assumption (a) follows from the fact that

$$\text{Vol}(\{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq r\}) = c_d|\Sigma|r^d ,$$

where $|\Sigma| = \prod_{j=1}^d \sigma_j(\Sigma)$ is the determinant of Σ and $\sigma_j(\Sigma)$ is the j-th singular value for some constant c_d that depends only on the dimension and selecting $c_2 \geq \log((c_0 + 2c_1)/c_1)$.

The first part is tricky since we do not yet have a handle on the set $B_{t,S}$ for $t > \tau$. In particular, we do not know how $D_S(\hat{\mu})$ relates to $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$ for such a $\hat{\mu}$ outside of $B_{\tau,S}$. To this end, we use the following corollary.

Corollary B.1 (Corollary of Lemma 12) *If $\hat{\mu} \in B_{2\tau,S}$ and $\tau = 42\rho_1$, then $\left|\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| - D_S(\hat{\mu})\right| \leq 14.2\rho_1$.*

We will show that $(7/8)\tau - (k^* + 1)\Delta > 0$. Since this implies that $\tau + (k^* + 1)\Delta \leq 2\tau$, we can use the above corollary to show that

$$\frac{\text{Vol}(B_{\tau+(k^*+1)\Delta+c_1\rho,S})}{\text{Vol}(B_{(7/8)\tau-(k^*+1)\Delta-c_1\rho,S})} \leq \frac{\text{Vol}\big(\{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq \tau + (k^* + 1)\Delta + c_1\rho + 14.2\rho_1\}\big)}{\text{Vol}\big(\{\hat{\mu} : \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq (7/8)\tau - (k^* + 1)\Delta - c_1\rho - 14.2\rho_1\}\big)}$$

$$= \left(\frac{\tau + (k^* + 1)\Delta + c_1\rho + 14.2\rho_1}{(7/8)\tau - (k^* + 1)\Delta - c_1\rho - 14.2\rho_1}\right)^d$$

$$\leq (67/12)^d \leq e^{c_2 d} ,$$

for the choices of $\rho = \rho_1$, $c_1 = 10.2$, $\tau = 42\rho_1$, $\Delta = 110\rho_1/(\alpha n)$, and $c_2 \geq \log(67/12)$, where we used the fact that for $n \geq C\log(1/(\delta\zeta))/(\alpha\varepsilon)$ with a large enough constant C, we have $(k^* + 1)\Delta \leq 0.3\rho_1$. It follows that the condition $(7/8)\tau - (k^* + 1)\Delta > 0$ is also satisfied. ∎

B.2.5. RESILIENCE IMPLIES BOUNDED LOCAL SENSITIVITY

We show that resilience implies the assumption (b) in Theorem 15 (Lemma 16). However, since local sensitivity needs to be established first for not just the given set S but also Hamming distance $k^* + 1$ neighborhood of S, we need robustness results for this broader regime. Assuming $(k^* + 1)/n \leq \alpha/11$, we can extend robustness results analogously, as follows. We consider a set S' with k data points arbitrarily changed from S. This implies that S' is a $((1/5.5)\alpha + (k/n), \alpha, \rho_1, \rho_2)$-corrupt good set with respect to (μ, Σ). We first prove the analogous bounds to Lemma 11 for this S'.

Lemma 14 *For an $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2)$-corrupt good set S' with respect to (μ, Σ), $\tilde{\alpha} \leq (1/11)\alpha$, and any unit norm $v \in \mathbb{R}^d$, $|\langle v, \mu - \mu(\mathcal{M}_{v,\alpha})\rangle| \leq 14\rho_1\,\sigma_v$ and $0.9\sigma_v \leq \sigma_v(\mathcal{M}_{v,\alpha}) \leq 1.1\sigma_v$.*

Proof Analogous to (10), we have

$$
\begin{aligned}
|\langle v, \mu - \mu(\mathcal{M}_{v,\alpha})\rangle| &\leq \frac{(1/5.5)\alpha + \tilde{\alpha}}{1 - (4/5.5)\alpha}\frac{2\rho_1\sigma_v}{(1/5.5)\alpha - \tilde{\alpha}} + \frac{1 - (1/5.5)\alpha - \tilde{\alpha}}{1 - (4/5.5)\alpha}\rho_1\sigma_v \\
&\leq 14\rho_1\sigma_v ,
\end{aligned}
$$

where we used the fact that $(5/5.5)\alpha + \tilde{\alpha} \leq \alpha$. Analogous to (11), we have

$$
\begin{aligned}
\frac{\sum_{x_i \in \mathcal{M}_{v,\alpha}}(\langle v, x_i - \mu(\mathcal{M}_{v,\alpha})\rangle^2 - \sigma_v^2)}{(1 - (4/5.5)\alpha)n} &\leq \frac{((1/5.5)\alpha + \tilde{\alpha})(\frac{2\rho_2}{(1/5.5)\alpha - \tilde{\alpha}})\sigma_v^2 + (1 - (1/5.5)\alpha - \tilde{\alpha})\rho_2\sigma_v^2}{1 - (4/5.5)\alpha} \\
&\leq 14\rho_2\sigma_v^2 .
\end{aligned}
$$

Analogous to (12), we have

$$
\begin{aligned}
\frac{\sum_{x_i \in \mathcal{M}_{v,\alpha}}(\langle v, x_i - \mu(\mathcal{M}_{v,\alpha})\rangle^2 - \sigma_v^2)}{(1 - (4/5.5)\alpha)n} &\geq -\frac{((1/5.5)\alpha + \tilde{\alpha})2\rho_2\sigma_v^2}{(1 - (4/5.5)\alpha)((1/5.5)\alpha - \tilde{\alpha})} - \rho_2\sigma_v^2 - 14^2\rho_1^2\sigma_v^2 \\
&\geq -(7.3\rho_2 + 196\rho_1^2)\sigma_v^2 .
\end{aligned}
$$

For $\alpha \leq 0.045$, $\rho_1 \leq 0.013$, and $\rho_2 \leq 0.0005$, we have the desired bounds. ∎

Lemma 15 *For an $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2)$-corrupt good set S' with respect to (μ, Σ) and $\tilde{\alpha} \leq (1/11)\alpha$, if $\hat{\mu} \in B_{t,S'}$ for some $t > 0$ then we have $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| \leq 14\rho_1 + 1.1t$ and $|D(\hat{\mu}, S') - \|\Sigma^{-1/2}(\hat{\mu} - \mu)\|| \leq 14\rho_1 + 0.1t$.*

Proof Analogously to the proof of Lemma 12, we have

$$
\begin{aligned}
1.1D(\hat{\mu}, S') &\geq -14\rho_1 + \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| , \text{ and} \\
0.9D(\hat{\mu}, S') &\leq 14\rho_1 + \|\Sigma^{-1/2}(\hat{\mu} - \mu)\| .
\end{aligned}
$$

This gives the desired bound. ∎

The sensitivity of $D_S(\hat{\mu})$ is *local* in two ways. First, we get the desired sensitivity bound for a dataset S that behaves nicely, which is captured by the notion of a $((1/5.5)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S. Second, the sensitivity bound requires the estimate parameter $\hat{\mu}$ to be close to μ in $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$. Both *locality in dataset* and *locality in estimate* are ensured by the safety test (Test step in HPTR). To show that corrupt good datasets pass the safety test, the following lemma establishes that those datasets have small local sensitivity.

Lemma 16 *For $\Delta = 110\rho_1/(\alpha n)$, $\tau = 42\rho_1$, and an $((1/5.5\alpha), \alpha, \rho_1, \rho_2)$-corrupt good S, if*

$$n = \Omega\Big(\frac{\log(1/(\delta\zeta))}{\alpha\varepsilon}\Big), \tag{15}$$

then the local sensitivity in assumption (b) is satisfied.

Remark. Note that to keep $\Delta = O(\rho_1/(\alpha n))$ that we want (and is critical in getting the final utility guarantee), we need the extra corruption to be $k^*/n = O(\alpha)$. This implies $n = \Omega(k^*/\alpha) = \Omega(\log(1/(\delta\zeta))/(\varepsilon\alpha))$. Further, $k^* = \Omega(\log(1/(\delta\zeta))/\varepsilon)$ cannot be improved since it is critical in achieving a small failure probability in the Testing step. Hence, the sample complexity of $\Omega(\log(1/(\delta\zeta))/(\varepsilon\alpha))$ cannot be improved under the current proof strategy.

Proof Since S is $((1/5.5)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good and $d_H(S, S') \leq k^*$, it follows that S' is $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2)$-corrupt good with $\tilde{\alpha} = (k^*/n)$. We further assume that $\tilde{\alpha} \leq (1/11)\alpha$, which follows from $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$ and $n = \Omega(\log(1/\delta\zeta)/(\varepsilon\alpha))$ with a large enough constant. We show that this resilience implies that S' is dense around the boundary of $\mathcal{M}_{v,\alpha}$, which in turn implies low sensitivity.

Recall that $\mathcal{T}_{v,\alpha} \subset S$ is the set of data points corresponding to the largest $(2/5.5)\alpha n$ data points in the projected set $S'_{(v)} = \{\langle v, x_i\rangle\}_{x_i \in S'}$, and $\mathcal{B}_{v,\alpha} \subset S$ is the bottom set. Let S_{good} denote the original uncorrupted resilient dataset. Applying Lemma 10 to $S_{\text{good}} \cap \mathcal{T}_{v,\alpha}$ (and $S_{\text{good}} \cap \mathcal{B}_{v,\alpha}$) of size at least $(1/11)\alpha$ (since the corruption fraction is at most $(1/5.5)\alpha + \tilde{\alpha} \leq (1.5/5.5)\alpha$),

$$\big|\langle v, \mu(S_{\text{good}} \cap \mathcal{T}_{v,\alpha}) - \mu\rangle\big| \leq \frac{2\rho_1\sigma_v}{(1/11)\alpha}, \text{ and } \big|\langle v, \mu(S_{\text{good}} \cap \mathcal{B}_{v,\alpha}) - \mu\rangle\big| \leq \frac{2\rho_1\sigma_v}{(1/11)\alpha}.$$

This implies that there is at least one good data point that is closer to the center than the means of the upper tail and the bottom tail:

$$\min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{v,\alpha}} \big|\langle v, x_i - \mu\rangle\big| \leq \frac{2\rho_1\sigma_v}{(1/11)\alpha}, \text{ and } \min_{x_i \in S_{\text{good}} \cap \mathcal{B}_{v,\alpha}} \big|\langle v, x_i - \mu\rangle\big| \leq \frac{2\rho_1\sigma_v}{(1/11)\alpha}.$$

It follows that the distance between two closest points in $\mathcal{T}_{v,\alpha}$ and $\mathcal{B}_{v,\alpha}$ is bounded by

$$\min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{v,\alpha}} \langle v, x_i\rangle - \max_{x_i \in S_{\text{good}} \cap \mathcal{B}_{v,\alpha}} \langle v, x_i\rangle \leq (44/\alpha)\rho_1\sigma_v, \tag{16}$$

when $\mu \in \mathcal{M}_{v,\alpha}$. When $\mu \in \mathcal{T}_{v,\alpha}$ or $\mu \in \mathcal{B}_{v,\alpha}$, it is straightforward that the above inequality holds. This implies low sensitivity as follows.

Recall that $\mathcal{M}_{v,\alpha}(S')$ denotes the middle part after filtering out the top and bottom $(2/5.5)\alpha$ quantiles from $\{\langle v, x_i\rangle\}_{x_i \in S'}$. For a neighboring dataset S'' and the corresponding $S''_{(v)}$, consider a scenario where one point x_i in $\mathcal{M}_{v,\alpha}(S')$ is replaced by another point $\tilde{x}_i$. If $\langle v, \tilde{x}_i\rangle \in [\max_{x_i \in S_{\text{good}} \cap \mathcal{B}_{v,\alpha}} \langle v, x_i\rangle, \min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{v,\alpha}} \langle v, x_i\rangle]$, then Eq. (16) implies that $|\langle v, x_i - \tilde{x}_i\rangle| \leq (44/\alpha)\rho_1\sigma_v$. Otherwise, $\mathcal{M}_{v,\alpha}(S'')$ will have x_i replaced by either $\arg\min_{j \in S_{\text{good}} \cap \mathcal{T}_{v,\alpha}} \langle v, x_j\rangle$ or $\arg\max_{j \in S_{\text{good}} \cap \mathcal{B}_{v,\alpha}} \langle v, x_j\rangle$. In either case, Eq. (16) implies that $|\langle v, x_i - \tilde{x}_i\rangle| \leq (44/\alpha)\rho_1\sigma_v$. The other case of when the replaced sample $x_i \in S$ is not in $\mathcal{M}_{v,\alpha}$ follows similarly.

From this, we get the following bounds on the sensitivity of the robust mean and robust variance. Note that using robust statistics is critical in getting such small sensitivity bounds. Let $\mu' = \mu(\mathcal{M}_{v,\alpha}(S'))$ and $\mu'' = \mu(\mathcal{M}_{v,\alpha}(S''))$, where we write the dataset S' in $\mathcal{M}_{v,\alpha}(S')$ explicitly,

$$\big|\langle v, \mu' - \mu''\rangle\big| \leq \frac{44\rho_1\sigma_v}{\alpha(1 - (4/5.5)\alpha)n}. \tag{17}$$

For the variance bound, let $\sigma_v'^2 = \sigma_v^2(\mathcal{M}_{v,\alpha}(S')) = (1/|\mathcal{M}_{v,\alpha}(S')|)\sum_{x_i' \in \mathcal{M}_{v,\alpha}(S')}\langle v, x_i' - \mu'\rangle^2$ and $\sigma_v''^2 = \sigma_v^2(\mathcal{M}_{v,\alpha}(S''))$. Since $(1 - (4/5.5)\alpha)n\sigma_v'^2 = \sum_{x_i' \in \mathcal{M}_{v,\alpha}(S')}\langle v, x_i' - \mu'\rangle^2 = \sum_{x_i' \in \mathcal{M}_{v,\alpha}(S')}(\langle v, x_i' - \mu''\rangle^2 - \langle v, \mu'' - \mu'\rangle^2)$, we have $(1-(4/5.5)\alpha)n(\sigma_v'^2 - \sigma_v''^2) = \sum_{x_i' \in \mathcal{M}_{v,\alpha}(S')}\langle v, x_i' - \mu''\rangle^2 - \sum_{x_i'' \in \mathcal{M}_{v,\alpha}(S'')}\langle v, x_i'' - \mu''\rangle^2 - (1-(4/5.5)\alpha)n\langle v, \mu'' - \mu'\rangle^2$. We bound each term separately. Note that $\mathcal{M}_{v,\alpha}(S')$ and $\mathcal{M}_{v,\alpha}(S'')$ only differ in at most one data point. We denote those by x' and x'', respectively. Then,

$$
\begin{aligned}
\Big|\sum_{x_i' \in \mathcal{M}_{v,\alpha}(S')}\langle v, x_i' - \mu''\rangle^2 - \sum_{x_i'' \in \mathcal{M}_{v,\alpha}(S'')}\langle v, x_i'' - \mu''\rangle^2\Big| &= \big|\langle v, x' - \mu''\rangle^2 - \langle v, x'' - \mu''\rangle^2\big| \\
&= \big|\langle v, x' + x'' - 2\mu''\rangle\langle v, x' - x''\rangle\big| \\
&= \big|\langle v, x' - \mu'\rangle + \langle v, \mu' - \mu''\rangle + \langle v, x'' - \mu''\rangle\big|\,\big|\langle v, x' - x''\rangle\big| \\
&\leq 3\Big(\frac{44\rho_1\sigma_v}{\alpha}\Big)^2,
\end{aligned}
$$

and

$$
(1 - (4/5.5)\alpha)n\langle v, \mu' - \mu''\rangle^2 \leq (1 - (4/5.5)\alpha)n\frac{(44\rho_1\sigma_v)^2}{(\alpha(1 - (4/5.5)\alpha)n)^2}.
$$

This implies that

$$
|\sigma_v'^2 - \sigma_v''^2| \leq \frac{(44\rho_1(\alpha/2)\sigma_v)^2}{(1 - (4/5.5)\alpha)n\alpha^2}\Big(3 + \frac{1}{(1 - (4/5.5)\alpha)n}\Big) \leq \frac{4(44\rho_1\sigma_v)^2}{(1 - (4/5.5)\alpha)n\alpha^2}. \tag{18}
$$

Together, we get the following bound on the sensitivity of $D(\hat\mu, S')$. Since $\max_v a_v - \max_v b_v \leq \max_v |a_v - b_v|$, we have

$$
\begin{aligned}
\big|D_{S'}(\hat\mu) - D_{S''}(\hat\mu)\big| &\leq \max_{v:\|v\|=1}\Big|\frac{\langle v, \hat\mu - \mu'\rangle}{\sigma_v'} - \frac{\langle v, \hat\mu - \mu''\rangle}{\sigma_v''}\Big| \\
&\leq \max_{v:\|v\|=1}\frac{|\langle v, \mu' - \mu''\rangle|}{\sigma_v'} + \frac{|\langle v, \hat\mu - \mu''\rangle|}{\sigma_v}\Big|\frac{\sigma_v}{\sigma_v'} - \frac{\sigma_v}{\sigma_v''}\Big| \\
&\leq \frac{44\rho_1}{0.9\alpha(1 - (4/5.5)\alpha)n} + \|\Sigma^{-1/2}(\hat\mu - \mu'')\|\max_v\frac{\sigma_v}{\sigma_v'\sigma_v''(\sigma_v' + \sigma_v'')}|\sigma_v'^2 - \sigma_v''^2| \\
&\leq \frac{44\rho_1}{0.9\alpha(1 - (4/5.5)\alpha)n} + \frac{5312\rho_1^2}{\alpha^2(1 - (4/5.5)\alpha)n}\|\Sigma^{-1/2}(\hat\mu - \mu'')\|,
\end{aligned}
$$

where we used triangle inequality in the second inequality and the third inequality follows from $\sigma_v' \geq 0.9\sigma_v$ (Lemma 14), Eqs. (17), and Lemma 7, and the last inequality follows from $\sigma_v'' \geq 0.9\sigma_v$ and (18).

From Lemma 15, $\hat\mu \in B_{\tau+(k^*+3)\Delta,S}$ implies $\|\Sigma^{-1/2}(\hat\mu - \mu)\| \leq 14\rho_1 + 1.1(\tau + (k^* + 3)\Delta)$. From Lemma 14, $\|\Sigma^{-1/2}(\mu-\mu'')\| \leq 14\rho_1$. We apply triangle inequality and show that $\|\Sigma^{-1/2}(\hat\mu - \mu'')\| \leq c\alpha/\rho_1$ for the choices of Δ, k^*, τ and n, with an arbitrarily small constant c:

$$
\begin{aligned}
\|\Sigma^{-1/2}(\hat\mu - \mu'')\| &\leq 28\rho_1 + 1.1(\tau + (k^* + 3)\Delta) \\
&\leq C\rho_1 + C\frac{\rho_1\log(1/(\delta\zeta))}{\varepsilon\alpha n} \\
&\leq 2C\rho_1,
\end{aligned}
$$

for some constant $C > 0$, where $\Delta = 110\rho_1/(\alpha n)$, $\tau = 42\rho_1$, $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$, and $n \geq C'\log(1/(\delta\zeta))/(\varepsilon\,\alpha)$. Under the assumption that $\rho_1^2 \leq c\alpha$ and $\alpha \leq c$ for some small enough c, this implies

$$
\begin{aligned}
|D_{S'}(\hat{\mu}) - D_{S''}(\hat{\mu})| \quad &\leq \quad \frac{44\rho_1}{0.9(1 - (4/5.5)\alpha)\alpha n}\Big(1 + \frac{121\rho_1}{\alpha}2C\rho_1\Big) \\
&\leq \quad \frac{(44/0.9)\rho_1}{\alpha n}\frac{1 + 44c}{1 - (4/5.5)c} \leq \Delta = \frac{110\rho_1}{\alpha n} \; .
\end{aligned}
\tag{19}
$$

∎

B.2.6. PROOF OF THEOREM 10

We show that the sufficient conditions of Theorem 15 are met for the following choices of constants and parameters: $p = d$, $\rho = \rho_1$, $c_0 = 31.8$, $c_1 = 10.2$, $\tau = 42\rho_1$, and $\Delta = 110\rho_1/(\alpha n)$. We can set c_2 to be a large constant and will only change the constant factor in the sample complexity.

The assumptions (a), (b), and (d) follow from Lemmas 13, 16, and 12, respectively. Assumption (c) follows from

$$
\Delta = \frac{110\rho_1}{\alpha n} \leq \frac{1.2\rho_1\varepsilon}{32(c_2 d + (\varepsilon/2) + \log(16/(\delta\zeta)))} = \frac{(c_0 - 3c_1)\rho\varepsilon}{32(c_2 d + (\varepsilon/2) + \log(16/(\delta\zeta)))} \; ,
$$

for a large enough $n \geq C'(d + \log(1/(\delta\zeta)))/(\alpha\varepsilon)$. This finishes the proof of Theorem 10, from which Theorem 9 immediately follows.

B.3. Step 3: Near-optimal guarantees

We provide utility guarantees for popular families of distributions in private or robust mean estimation literature: sub-Gaussian (Barber and Duchi, 2014; Lai et al., 2016; Steinhardt et al., 2018; Zhu et al., 2019; Karwa and Vadhan, 2017; Kamath et al., 2019; Cai et al., 2019; Bun et al., 2019; Biswas et al., 2020; Aden-Ali et al., 2020; Brown et al., 2021; Diakonikolas et al., 2019a; Diakonikolas et al., 2017; Dong et al., 2019; Hopkins, 2020; Diakonikolas et al., 2018; Huang et al., 2021), k-th moment bounded (Barber and Duchi, 2014; Lai et al., 2016; Steinhardt et al., 2018; Zhu et al., 2019; Kamath et al., 2020), and covariance bounded (Barber and Duchi, 2014; Lai et al., 2016; Steinhardt et al., 2018; Zhu et al., 2019; Kamath et al., 2020; Dong et al., 2019; Hopkins et al., 2020; Depersin and Lecué, 2019, 2021). We apply known resilience bounds for each family of distributions and substitute them in Theorems 9 and 10. In all cases, the resulting sample complexity is near-optimal, which follows from matching information-theoretic lower bounds.

Since we aim for Mahalanobis distance error bounds, the corresponding mean resilience we need in Definition 8 scales linearly in the projected standard deviation. For sub-Gaussian distributions, this requires the projected variance $v^\top \Sigma v$ to be lower bounded by how fast the tail is decreasing, as captured by the sub-Gaussian proxy $\Omega(v^\top \Gamma v)$ in Eq. (20) (Section B.3.1). For k-th moment bounded distributions with $k > 3$, this requires the projected variance to be lower bounded by $\Omega(\mathbb{E}[|\langle v, x - \mu\rangle|^k]^{2/k})$, a condition known as hypercontractivity (Section B.3.2). When we do not have such lower bounds on the covariance, HPTR can only hope to achieve Euclidean distance error bounds. Under our design principle, this translates into the choice of $D_S(\hat{\mu}) = \max_{\|v\| \leq 1}\langle v, \hat{\mu}\rangle - \mu_v(\mathcal{M}_{v,\alpha})$. We give an example of this scenario with covariance bounded distributions (Section B.3.3).

B.3.1. SUB-GAUSSIAN DISTRIBUTIONS

We say a distribution P is sub-Gaussian with proxy Γ if for all $\|v\| = 1$ and $t \in \mathbb{R}$,

$$\mathbb{E}_{x \sim P}\big[\exp(t\langle v, x\rangle)\big] \leq \exp\Big(\frac{t^2\, v^\top \Gamma v}{2}\Big). \tag{20}$$

Under this standard sub-Gaussianity, we are only guaranteed mean resilience of Eq. (5), for example, with R.H.S scaling as $\rho_1\sqrt{v^\top \Gamma v}$ instead of $\rho_1\sqrt{v^\top \Sigma v}$. This implies that the Mahalanobis distance of any robust estimate can be made arbitrarily large by shrinking the covariance in one direction such that $v^\top \Sigma v \ll v^\top \Gamma v$. To avoid such degeneracy, we add an additional assumption that $\Sigma \succeq c\Gamma$, which is also common in robust statistics literature, e.g., (Jambulapati et al., 2020). With this definition, it is known that sub-Gaussian samples are $(\alpha, O(\alpha\sqrt{\log(1/\alpha)}), O(\alpha\log(1/\alpha)))$-resilient.

Lemma 17 (Resilience of sub-Gaussian samples (Zhu et al., 2019) and (Jambulapati et al., 2020, Corollary 4)) *For any fixed $\alpha \in (0, 1/2)$, consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a sub-Gaussian distribution with mean μ, covariance Σ, and a sub-Gaussian proxy $0 \prec \Gamma \preceq c_1\Sigma$ for a constant c_1. There exist constants c_2 and $c_3 > 0$ such that if $n \geq c_2((d + \log(1/\zeta))/(\alpha\log(1/\alpha))^2)$, then S is $(\alpha, c_3\alpha\sqrt{\log(1/\alpha)}, c_3\alpha\log(1/\alpha))$-resilient with respect to (μ, Σ) with probability $1 - \zeta$.*

This lemma and Theorem 9 imply the following utility guarantee. Further, from Theorem 10, the guarantee also holds under α-corruption of the i.i.d. samples from a sub-Gaussian distribution.

Corollary B.2 *Under the hypothesis of Lemma 17, there exists a constant $c > 0$ such that for any $\alpha \in (0, c)$, a dataset of size*

$$n = O\Big(\frac{d + \log(1/\zeta)}{(\alpha\log(1/\alpha))^2} + \frac{d + \log(1/(\delta\zeta))}{\alpha\varepsilon}\Big),$$

sensitivity of $\Delta = O((1/n)\sqrt{\log(1/\alpha)})$, and threshold of $\tau = O(\alpha\sqrt{\log(1/\alpha)})$, with large enough constants are sufficient for $\mathrm{HPTR}(S)$ with the distance function in Eq. (3) to achieve

$$\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\alpha\sqrt{\log(1/\alpha)}) \tag{21}$$

with probability $1 - \zeta$. Further, the same guarantee holds even if α-fraction of the samples is arbitrarily corrupted, as shown in Assumption 1.

This sample complexity is near-optimal up to logarithmic factors in $1/\alpha$ and $1/\zeta$ for $\delta = e^{-O(d)}$. Even for DP mean estimation without corrupted samples, HPTR is the first algorithm for sub-Gaussian distributions with unknown covariance that nearly matches the lower bound of $n = \widetilde{\Omega}(d/\alpha^2 + d/(\alpha\varepsilon) + \log(1/\delta)/\varepsilon)$ from (Karwa and Vadhan, 2017; Kamath et al., 2019), where $\widetilde{\Omega}$ hides polylogarithmic terms in $1/\zeta, 1/\alpha, d, 1/\varepsilon$ and $\log(1/\delta)$. The third term has a gap of $1/\alpha$ factor to our upper bound, but this term is dominated by other terms under the assumption that $\delta = e^{-O(d)}$. For completeness, we state the lower bound in Appendix H. Existing algorithms are suboptimal since they require either $n = \widetilde{O}((d/\alpha^2) + (d(\log(1/\delta)^3)/(\alpha\varepsilon^2)))$ samples with $(1/\varepsilon^2)$ dependence to achieve the error rate of Eq. (21) (Brown et al., 2021) or extra conditions, such as

strictly Gaussian distributions (Brown et al., 2021; Bun et al., 2019) or known covariance matrices (Kamath et al., 2019; Aden-Ali et al., 2020; Barber and Duchi, 2014).

The error bound is near-optimal in its dependence in α under α-corruption. HPTR is the first estimator that is both (ε, δ)-DP and also achieves the robust error rate of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\alpha\sqrt{\log(1/\alpha)})$, nearly matching the known information-theoretic lower bound of $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = \Omega(\alpha)$ (Chen et al., 2018). This lower bound holds for any estimator that is not necessarily private and regardless of how many samples are available. In comparison, the existing robust and DP estimator from (Liu et al., 2021), which runs in polynomial time, requires the knowledge of the covariance matrix Σ and a larger sample complexity of $n = \widetilde{\Omega}((d/\alpha^2) + (d^{3/2}\log(1/\delta))/(\alpha\varepsilon))$. If privacy is not required (i.e., $\varepsilon = \infty$), a robust mean estimator from (Zhu et al., 2019) achieves the same error bound and sample complexity as ours.

B.3.2. HYPERCONTRACTIVE DISTRIBUTIONS

For an integer $k \geq 3$, a distribution $P_{\mu,\Sigma}$ is k-th moment bounded with a mean μ and covariance Σ if for all $\|v\| = 1$, we have $\mathbb{E}_{x\sim P_X}[|\langle v, (x - \mu)\rangle|^k] \leq \kappa^k$ for some $\kappa > 0$. However, similar to the sub-Gaussian case, Mahalanobis distance guarantees require an additional lower bound on the covariance. To this end, we assume hypercontractivity, which is common in robust statistics literature, e.g., (Klivans et al., 2018).

Definition 18 *A distribution $P_{\mu,\Sigma}$ is (κ, k)-hypercontractive if for all $v \in \mathbb{R}^d$, $\mathbb{E}_{x\sim P_X}[|\langle v, (x - \mu)\rangle|^k] \leq \kappa^k(v^\top\Sigma v)^{k/2}$.*

Although samples from such heavy-tailed distributions are known to be not resilient, it is known that they are $O(\alpha)$-close in total variation distance to an $(\alpha, O(\alpha^{1-1/k}), O(\alpha^{1-2/k}))$-resilient dataset. This means that the resulting dataset is $((1/11)\alpha, \alpha, O(\alpha^{1-1/k}), O(\alpha^{1-2/k}))$-corrupt good, for example. Note that hypercontractivity is invariant under affine transformations, and κ does not depend on the condition number of the covariance.

Lemma 19 (Resilience of k-th moment bounded samples (Zhu et al., 2019, Lemma G.10)) *For any fixed $\alpha \in (0, 1/2)$, consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a (κ, k)-hypercontractive distribution with mean μ and covariance $\Sigma \succ 0$ for some $k \geq 3$. For any $c_3 > 0$, there exist constants c_1 and $c_2 > 0$ that depend only on c_3 such that if*

$$n \geq c_1\left(\frac{d}{\zeta^{2(1-1/k)}\alpha^{2(1-1/k)}} + \frac{k^2\alpha^{2-2/k}d\log d}{\zeta^{2-4/k}\kappa^2} + \frac{\kappa^2 d\log d}{\alpha^{2/k}}\right),$$

then S is $(c_3\alpha, \alpha, c_2 k\kappa\alpha^{1-1/k}\zeta^{-1/k}, c_2 k^2\kappa^2\alpha^{1-2/k}\zeta^{-2/k})$-corrupt good with respect to (μ, Σ) with probability $1 - \zeta$.

This lemma and Theorem 9 imply the following utility guarantee. Further, from Theorem 10, the guarantee also holds under $(1/5.5 - c_3)\alpha$-corruption of the i.i.d. samples from a (κ, k)-hypercontractive distribution. Choosing appropriate constants, we get the following result.

Corollary B.3 *Under the hypothesis of Lemma 19, there exists a constant $c_{\kappa,k,\zeta}$ that depends only on k, κ, and ζ such that for any $\alpha \in (0, c_{\kappa,k,\zeta})$, a dataset of size*

$$n = O\left(\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} + \frac{d}{\zeta^{2(1-1/k)}\alpha^{2(1-1/k)}} + \frac{k^2\alpha^{2-2/k}d\log d}{\zeta^{2-4/k}\kappa^2} + \frac{\kappa^2 d\log d}{\alpha^{2/k}}\right),$$

sensitivity of $\Delta = O(1/(n\alpha^{1/k}))$, and threshold of $\tau = O(\alpha^{1-1/k})$ with large enough constants are sufficient for HPTR(S) *with the distance function in Eq.* (3) *to achieve* $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(k\kappa\zeta^{-1/k}\alpha^{1-1/k})$ *with probability* $1 - \zeta$*. Further, the same guarantee holds even if* α*-fraction of the samples is arbitrarily corrupted, as shown in Assumption* 1.

This sample complexity is near-optimal in its dependence in d, $1/\varepsilon$, and $1/\alpha$ when $\delta = e^{-\Theta(d)}$. Suppose ζ, k, and κ are $\Theta(1)$. Even for DP mean estimation without robustness, HPTR is the first algorithm that achieves $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\alpha^{1-1/k})$ with $n = \widetilde{O}(\frac{d}{\alpha^{2(1-1/k)}} + \frac{d+\log(1/\delta)}{\varepsilon\alpha})$ samples, which nearly matches the known lower bounds. The first term $O(d/\alpha^{2(1-1/k)})$ cannot be improved even if we do not require privacy. The second term $O((d + \log(1/\delta))/\varepsilon\alpha)$ nearly matches the lower bound of $n = \Omega(\min\{d, \log((1 - e^{-\varepsilon})/\delta)\}/(\varepsilon\alpha))$ for DP mean estimation that we show in Proposition B.5. If $0 < \varepsilon \leq 1$ and $\delta = e^{-\Theta(d)}$ (as assumed in (Barber and Duchi, 2014)), the upper and lower bounds match. An existing DP mean estimator (without robustness) of (Kamath et al., 2020) achieves a stronger $(\varepsilon, 0)$-DP and similar accuracy but in Euclidean distance with a similar sample size of $n = \widetilde{O}(\frac{d}{\alpha^{2(1-1/k)}} + \frac{d}{\varepsilon\alpha})$. However, it requires a known or identity covariance matrix and a known bound on the unknown mean of the form $\mu \in [-R, R]^d$. Such a bounded search space is critical in achieving a stronger *pure* privacy guarantee with $\delta = 0$.

The error bound is optimal in its dependence in α under α-corruption. The error bound $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\| = O(\alpha^{1-1/k})$ matches the following information-theoretic lower bound in Proposition B.4; no algorithm can distinguish two distributions whose means are at least $O(\alpha^{1-1/k})$ apart from α-fraction of samples corrupted, even with infinite samples. HPTR is the first algorithm that guarantees both differential privacy and robustness (i.e., the error depends only on α and not in d) for k-th moment bounded distributions. If privacy is not required (i.e., $\varepsilon = \infty$), a robust mean estimator from (Zhu et al., 2019) achieves a similar error bound and sample complexity as ours.

Proposition B.4 (Lower bound for robust mean estimation)
For any $\alpha \in (0, 1/2)$*, there exist two distributions* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *satisfying the hypotheses of Lemma* 19 *such that* $d_{\text{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \alpha$*, and*

$$\|\Sigma^{-1/2}(\mu_1 - \mu_2)\| = \Omega(\alpha^{1-1/k}).$$

Proof We construct two scalar distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ with $d_{\text{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \alpha$ as follows:

$$\mathcal{D}_1(x) = \begin{cases} (1-\alpha)/2, & \text{if } x \in \{-1, 1\} \\ \alpha & \text{if } x = -\alpha^{1/k} \end{cases}, \quad \text{and} \quad \mathcal{D}_2(x) = \begin{cases} (1-\alpha)/2, & \text{if } x \in \{-1, 1\} \\ \alpha & \text{if } x = \alpha^{1/k} \end{cases}$$

The variance is $\Omega(1)$ for both distributions, and $|\mathbb{E}_{x\sim\mathcal{D}_1}[x] - \mathbb{E}_{x\sim\mathcal{D}_2}[x]| = 2\alpha^{1-1/k}$. Then, it suffices to show that $\mathcal{D}_1$ and $\mathcal{D}_2$ are both $(O(1), k)$-hypercontractive. In fact, we know $\mathbb{E}_{x\sim\mathcal{D}_1}[x] = -\alpha^{1-1/k}$, $\mathbb{E}_{x\sim\mathcal{D}_1}[x^2] = \mathbb{E}_{x\sim\mathcal{D}_2}[x^2] = 1 - \alpha + \alpha^{1-2/k}$, and $\mathbb{E}_{\mathcal{D}_1}[|x|^k] = 2 - \alpha$. Since $\alpha \in (0, 1/2)$, there exists a constant c such that $\mathbb{E}_{x\sim\mathcal{D}_1}[|x - \mu_1|^k] \leq c$, which concludes the proof. $\blacksquare$

Proposition B.5 (Lower bound for DP mean estimation) *Let* $\mathcal{P}_{\mu,\Sigma,k}$ *be the set of* $(1, k)$*-hypercontractive distributions with mean* $\mu \in \mathbb{R}^d$ *and covariance* $\Sigma \in \mathbb{R}^{d\times d}$*. Let* $\mathcal{M}_{\varepsilon,\delta}$ *be a class of* (ε, δ)*-DP estimators using* n *i.i.d. samples from* $P \in \mathcal{P}_{\mu,\Sigma,k}$*. Then, for* $\varepsilon \in (0, 10)$*, there exists a constant* c *such*

that

$$\inf_{\hat{\mu}\in\mathcal{M}_{\varepsilon,\delta}} \sup_{\mu\in\mathbb{R}^d,\Sigma\succ0,P\in\mathcal{P}_{\mu,\Sigma,k}} \mathbb{E}_{S\sim P^n}[\|\Sigma^{-1/2}(\hat{\mu}(S)-\mu)\|^2] \geq c\min\left\{\left(\frac{d\wedge\log((1-e^{-\varepsilon})/\delta)}{n\varepsilon}\right)^{2-2/k},1\right\}.$$

Proof We extend the proof of (Barber and Duchi, 2014, Proposition 4) to hypercontractive distributions. Before we prove the lower bound, we first establish the private version of the standard statistical estimation problem. Specifically, let $\mathcal{P}$ denote a family of distributions of interest and $\theta : \mathcal{P} \to \Theta$ denote the population parameter. The goal is to estimate θ from i.i.d. samples $x_1, x_2, \ldots, x_n \sim \mathcal{P}$. Let $\hat{\theta}$ be an (ε, δ)-differentially private estimator. Furthermore, let $\rho : \Theta \times \Theta \to \mathbb{R}^+$ be a (semi)metric on parameter space Θ and $\ell : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing loss function with $\ell(0) = 0$.

To measure the performance of our (ε, δ)-DP estimator $\hat{\theta}$, we define the *minimax risk* as follows:

$$\inf_{\hat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{x_1,x_2,\ldots,x_n\sim P}\left[\ell\left(\rho\left(\hat{\theta}\left(x_1,\ldots,x_n\right),\theta(P)\right)\right)\right]. \tag{22}$$

To prove the lower bound of the minimax risk, we construct a well-separated family of distributions and convert the estimation problem into a testing problem. Specifically, let $\mathcal{V}$ be an index set of finite cardinality. Define $\mathcal{P}_{\mathcal{V}} = \{P_v, v \in \mathcal{V}\} \subset \mathcal{P}$ to be an indexed family of distributions. If for all $v \neq v' \in \mathcal{V}$ we have $\rho(P_v, P_{v'}) \geq 2t$, we say $\mathcal{P}_{\mathcal{V}}$ is *2t-packing* of Θ.

The proof of (Barber and Duchi, 2014, Proposition 4) is based on following lemma.

Lemma 20 ((Barber and Duchi, 2014, Theorem 3)) *Fix $p \in [0, 1]$ and let $\mathcal{P}_{\mathcal{V}}$ be a 2t-packing of Θ such that $P_v = (1 - p)Q_0 + pQ_v$. Let $\hat{\theta}$ be a (ε, δ) differentially private estimator. Then,*

$$\frac{1}{|\mathcal{V}|}\sum_{\nu\in\mathcal{V}} P_v\left(\rho\left(\hat{\theta},\theta(P_v)\right)\geq t\right) \geq \frac{(|\mathcal{V}|-1)\cdot\left(\frac{1}{2}e^{-\varepsilon\lceil np\rceil}-\delta\frac{1-e^{-\varepsilon\lceil np\rceil}}{1-e^{-\varepsilon}}\right)}{1+(|\mathcal{V}|-1)\cdot e^{-\varepsilon\lceil np\rceil}}. \tag{23}$$

In our problem, we set $\mathcal{P}$ to be $\mathcal{P} = \mathcal{P}_{\mu,\Sigma,k}$. It suffices to construct such an index set $\mathcal{V}$ and indexed family of distributions $\mathcal{P}_{\mathcal{V}}$. We construct a packing set similar to that defined in the proof of (Barber and Duchi, 2014, Proposition 4). By (Acharya et al., 2021, Lemma 6), there exists a finite set $\mathcal{V} \subset \mathbb{R}^d$ with cardinality $|\mathcal{V}| = 2^{\Omega(d)}$, $\|v\| = 1$ for all $v \in \mathcal{V}$, and $\|v - v'\| \geq 1/2$ for all $v \neq v' \in \mathcal{V}$. Define Q_0 as $Q_0 = \mathcal{N}(0, \mathbf{I}_{d\times d})$ and Q_v as a point mass on $x = \alpha^{-1/k}cv$, where $v \in \mathcal{V}$. We construct P_v as $P_v = \alpha Q_v + (1 - \alpha)Q_0$.

We first verify that $\mathcal{P}_{\mathcal{V}} \subset \mathcal{P}$. It is easy to see $\mu(P_v) = \mathbb{E}_{x\sim P_v}[x] = \alpha^{1-1/k}v$ and $\Sigma(P_v) = \mathbb{E}_{x\sim P_v}[(x - \mu(P_v))(x - \mu(P_v))^\top] = (1 - \alpha)\mathbf{I}_{d\times d} + \alpha(1 - \alpha)\alpha^{-2/k}vv^\top$. This implies $\frac{1}{2}\mathbf{I}_{d\times d} \preceq \Sigma(P_v) \preceq \mathbf{I}_{d\times d}$. Since $\mathbb{E}\left[(X - \mathbb{E}[X])^k\right] \leq \mathbb{E}\left[X^k\right]$ for any $X \geq 0$, it suffices to show that $\mathbb{E}_{x\sim P_v}[|\langle u, x\rangle|^k] \leq C^k$ for some constant $C > 0$ and any $\|u\| = 1$. In fact, letting c_k denote the k-th moment of standard Gaussian, we have

$$\mathbb{E}_{x\sim P_v}[|\langle u, x\rangle|^k] = (1 - \alpha)c_k + \alpha\left|\left\langle u, \alpha^{-1/k}v\right\rangle\right|^k = O(1).$$

It is also easy to see that $d_{\mathrm{TV}}(P_v, P_{v'}) = \alpha$. Let $\rho(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|$. We also have

$$t = \min_{v\neq v'\in\mathcal{V}} \alpha^{1-1/k}\|v - v'\| \geq \frac{1}{2}\alpha^{1-1/k}.$$

Next, we apply the reduction of estimation to testing with this packing $\mathcal{V}$. For (ε, δ)-DP estimator $\hat{\mu}$, using Lemma 20, we have

$$
\begin{aligned}
\sup_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n}[\|\Sigma(P)^{-1/2}(\hat{\mu}(S) - \mu(P))\|^2] &\geq \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \mathbb{E}_{S \sim P_v^n}[\|\Sigma(P_v)^{-1/2}(\hat{\mu}(S) - \mu(P_v))\|^2] \\
&= t^2 \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} P_v \left(\|\Sigma(P_v)^{-1/2}(\hat{\mu}(S) - \theta(P_v))\| \geq t \right) \\
&\asymp t^2 \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} P_v \left(\|\hat{\mu}(S) - \theta(P_v)\| \geq t \right) \\
&\gtrsim t^2 \frac{e^{d/2} \cdot \left(\frac{1}{2} e^{-\varepsilon \lceil n\alpha \rceil} - \frac{\delta}{1 - e^{-\varepsilon}} \right)}{1 + e^{d/2} e^{-\varepsilon \lceil n\alpha \rceil}} ,
\end{aligned}
$$

where the last inequality follows from the fact that $d \geq 2$.

The rest of the proof follows from (Barber and Duchi, 2014, Proposition 4). We choose

$$
\alpha = \frac{1}{n\varepsilon} \min \left\{ \frac{d}{2} - \varepsilon, \log \left(\frac{1 - e^{-\varepsilon}}{4\delta e^{\varepsilon}} \right) \right\}
$$

so that

$$
\sup_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n}[\|\Sigma(P)^{-1/2}(\hat{\mu}(S) - \mu(P))\|^2] \gtrsim \alpha^{2 - 2/k} .
$$

This means, for $\varepsilon \in (0, 1)$,

$$
\inf_{\hat{\mu} \in \mathcal{M}_{\varepsilon, \delta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n}[\|\Sigma(P)^{-1/2}(\hat{\mu}(S) - \mu(P))\|^2] \gtrsim \min \left\{ \left(\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon} \right)^{2 - 2/k}, 1 \right\} ,
$$

which completes the proof. ∎

B.3.3. COVARIANCE BOUNDED DISTRIBUTIONS

A distribution $P_{\mu, \Sigma}$ is covariance bounded with mean μ and covariance Σ if $\|\Sigma\| \leq 1$. Contrary to the previous cases, the sample variance is not resilient since $\{\langle v, x_i - \mu \rangle^2\}$ do not concentrate. To get around this issue, we use the Euclidean distance: $D_\phi(\hat{\mu}, \mu) = \|\hat{\mu} - \mu\|$. This leads to the surrogate Euclidean distance of

$$
D_S(\hat{\mu}) = \max_{\|v\| \leq 1} \langle v, \hat{\mu} \rangle - \mu_v(\mathcal{M}_{v, \alpha}) . \tag{24}
$$

Since this does not depend on the robust variance $\sigma_v^2(\mathcal{M}_{v, \alpha})$, we only require the following first order resilience.

Lemma 21 (Resilience of covariance bounded samples (Zhu et al., 2019, Lemma G.3)) *For any fixed $\alpha \in (0, 1/2)$, consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from a covariance bounded distribution with mean μ and covariance $\Sigma \succ 0$. If $n = \Omega(d \log(d/\zeta)/(\alpha))$, then with*

probability $1 - 3\zeta$, for any subset $T \subset S$ of size $|T| \geq (1 - \alpha)n$, there exists a constant $C > 0$ such that the following holds for all $\alpha \in (0, 1/2)$ and for all $v \in \mathbb{R}^d$ with $\|v\| = 1$:

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i \rangle - \mu_v \right| \leq C\alpha^{1/2} \, ,$$

where $\mu_v = \langle v, \mu \rangle$.

This lemma and Theorem 10, adapted for the new $D_S(\hat{\mu}) = \max_{\|v\| \leq 1} \langle v, \hat{\mu} \rangle - \mu_v(\mathcal{M}_{v,\alpha})$, imply the following utility guarantee.

Corollary B.6 *Under the hypothesis of Lemma 21, there exists a constant c_ζ that only depends on ζ such that for $\alpha \in (0, c_\zeta)$, a dataset of size*

$$n = O\left(\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} + \frac{d \log(d/\zeta)}{\alpha} \right) ,$$

sensitivity of $\Delta = O(1/(n\sqrt{\alpha}))$, and threshold of $\tau = O(\sqrt{\alpha})$ with large enough constants are sufficient for $\mathrm{HPTR}(S)$ with the distance function in Eq. (24) to achieve $\|\hat{\mu} - \mu\| = O(\alpha^{1/2})$ with probability $1 - 3\zeta$. Further, the same guarantee holds even if a α-fraction of the samples is arbitrarily corrupted, as shown in Assumption 1.

This sample complexity is near-optimal in its dependence on d, $1/\varepsilon$, and $1/\alpha$ for $\delta = e^{-O(d)}$. It matches the information-theoretic lower bound of $n = \Omega(d/\varepsilon\alpha)$ from (Kamath et al., 2020). For completeness, we write the lower bound in Appendix H. This problem is easier than the sub-Gaussian or k-th moment bounded settings, since the error is measured in Euclidean distance and hence one does not need to adapt to the unknown covariance. Therefore, there exist other algorithms achieving near-optimality and even runs in polynomial time (Kamath et al., 2020).

The error rate is near-optimal under α-corruption, matching the information-theoretic lower bound of $\|\hat{\mu} - \mu\| = \Omega(\alpha^{1/2})$ (Dong et al., 2019). Note that there exists an DP and robust algorithm from (Liu et al., 2021) that achieves near-optimality in both error rate and sample complexity but requires an additional assumption that the spectral norm of the covariance is known and the unknown mean is in a bounded set, $[-R, R]^d$, with a known R.

Remark. Corollary B.6 is suboptimal since (1) the error metric is Euclidean $\|\hat{\mu} - \mu\|$ instead of Mahalanobis $\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|$, and (2) sample complexity scales as $1/\zeta$ instead of $\log(1/\zeta)$. It remains an open problem if these gaps can be closed. For the former, one could use the Stahel-Donoho outlyingness (Stahel, 1981; Donoho, 1982),

$$D_S(\hat{\mu}) = \sup_{v \in \mathbb{R}^d, \|v\|=1} \frac{|\langle v, \hat{\mu} \rangle - \mathrm{Med}(\langle v, S \rangle)|}{\mathrm{Med}(|\langle v, S \rangle - \mathrm{Med}(\langle v, S \rangle)|)} \, ,$$

in the exponential mechanism, which replaces second moment based normalization by a first moment based one that is resilient. Here, $\mathrm{Med}(\langle v, S \rangle)$ is the median of $\{\langle v, x_i \rangle\}_{x_i \in S}$. Further, replacing the median by the median of means can improve the dependence on ζ. Such directions have been fruitful for robust but non-private mean estimation (Depersin and Lecué, 2021).

Appendix C. Linear regression

In a standard linear regression, we have i.i.d. samples $S = \{(x_i \in \mathbb{R}^d, y_i \in \mathbb{R})\}_{i=1}^n$ from a distribution $P_{\beta,\Sigma,\gamma^2}$ of a linear model:

$$y_i = x_i^\top \beta + \eta_i,$$

where the input $x_i \in \mathbb{R}^d$ has zero mean and covariance Σ and the noise $\eta_i \in \mathbb{R}$ has variance γ^2. We further assume $\mathbb{E}[x_i \eta_i] = 0$, which is equivalent to assuming that the true parameter $\beta = \Sigma^{-1} \mathbb{E}[y_i x_i]$. In DP linear regression, we want to output a DP estimate $\hat{\beta}$ of the unknown model parameter β (which corresponds to $\theta = \mu$ in the general notation), assuming that both covariance $\Sigma \succ 0$ and noise variance γ^2 (corresponding to $\phi = (\Sigma, \gamma)$ in the general notation) are unknown. The resulting error is measured in $D_{\Sigma,\gamma}(\hat{\beta}, \beta) = (1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|$, which is equivalent to the (re-scaled) root excess prediction risk of the estimated predictor $\hat{\beta}$. Similar to Mahalanobis distance for mean estimation, this is challenging since we aim for a tight guarantee that adapts to the unknown Σ without having enough samples to directly estimate Σ. We follow the three-step strategy of Section 1.2.1 and provide utility guarantees.

C.1. Step 1: Designing the surrogate $D_S(\hat{\beta})$ for the error metric $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|$

In the RELEASE step of HPTR, we propose the following surrogate error metric for the exponential mechanism:

$$D_S(\hat{\beta}) = \max_{v:\|v\|\leq 1} \frac{\frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{x_i \in \mathcal{N}_{v,\hat{\beta},\alpha}} \langle v, x_i(y_i - x_i^\top \hat{\beta}) \rangle}{\sigma_v(\mathcal{M}_{v,\alpha})\hat{\gamma}}, \tag{25}$$

where $\hat{\gamma}^2$ is defined as

$$\hat{\gamma}^2 = \min_{\bar{\beta}} \frac{1}{|\mathcal{B}_{\bar{\beta},\alpha}|} \sum_{i \in \mathcal{B}_{\bar{\beta},\alpha}} (y_i - x_i^\top \bar{\beta})^2. \tag{26}$$

We define $\mathcal{N}_{v,\hat{\beta},\alpha}$, $\mathcal{M}_{v,\alpha}$ and $\mathcal{B}_{\bar{\beta},\alpha}$ as follows. For a fixed v, $\mathcal{M}_{v,\alpha}$ is defined in Section B.1 as a subset of S with size $(1 - (4/5.5)\alpha)n$ that remains after removing $(4/5.5)\alpha n$ data points corresponding to the top $(2/5.5)\alpha n$ and the bottom $(2/5.5)\alpha n$ of samples when projected down to $S_v = \{\langle v, x_i \rangle\}_{i \in [n]}$. We denote a robust estimate of the variance in direction v as $\sigma_v(\mathcal{M}_{v,\alpha})^2 = (1/|\mathcal{M}_{v,\alpha}|) \sum_{x_i \in \mathcal{M}_{v,\alpha}} \langle v, x_i \rangle^2$ since x_i's are zero mean. Similarly, for a fixed $\hat{\beta}$ and v, we consider a set of projected data points $S_{v,\hat{\beta}} = \{\langle v, x_i(y_i - x_i^\top \hat{\beta}) \rangle\}_{i \in [n]}$ and partition S into three disjoint sets, $\mathcal{B}_{v,\hat{\beta},\alpha}$, $\mathcal{N}_{v,\hat{\beta},\alpha}$, and $\mathcal{T}_{v,\hat{\beta},\alpha}$, where $\mathcal{B}_{v,\hat{\beta},\alpha}$ is the subset of S corresponding to the bottom $(2/5.5)\alpha n$ data points with the smallest values in $S_{v,\hat{\beta}}$, $\mathcal{T}_{v,\hat{\beta},\alpha}$ corresponds to the top $(2/5.5)\alpha n$ data points, and $\mathcal{N}_{v,\hat{\beta},\alpha}$ corresponds to the remaining $(1-(4/5.5)\alpha)n$ middle data points. We use $\mathcal{T}_{v,\hat{\beta},\alpha}$, $\mathcal{N}_{v,\hat{\beta},\alpha}$, and $\mathcal{B}_{v,\hat{\beta},\alpha}$ to denote both the set of paired examples $\{(x_i, y_i)\}$ and the set of indices of those examples, and it should be clear from the context which one we mean.

For a fixed $\bar{\beta}$, $\mathcal{B}_{\bar{\beta},\alpha}$ is defined as a subset of S with size $(1 - (3.5/5.5)\alpha)n$ that remains after removing the largest $(2/5.5)\alpha n$ data points in set $S_{\bar{\beta}} = \{(y_i - x_i^\top \bar{\beta})^2\}_{i \in [n]}$.

This choice is justified by Lemma 22, which shows that if we replace the robust one-dimensional statistics by the true ones, we recover the target error metric. Hence, the exponential mechanism

with distance $D_S(\hat{\beta})$ is approximately and stochastically minimizing $\|\Sigma^{1/2}(\hat{\beta} - \beta)\|$. For a more elaborate justification of using $D_S(\hat{\beta})$, we refer to a similar choice for mean estimation in Section B.1.

Lemma 22 *For any $\beta \in \mathbb{R}^d$, $0 \prec \Sigma \in \mathbb{R}^{d \times d}$, and $\gamma > 0$, let $\sigma_v^2 = v^\top \Sigma v$. If $\mathbb{E}[\eta_i x_i] = 0$, $y_i = x_i^\top \beta + \eta_i$ and $(x_i, y_i) \sim P_{\beta, \Sigma, \gamma^2}$, then we have*

$$\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = \max_{v: \|v\| \le 1} \frac{\mathbb{E}_{P_{\beta, \Sigma, \gamma^2}}[\langle v, x_i(y_i - x_i^\top \hat{\beta})\rangle]}{\sigma_v} , \text{ and}$$

$$\gamma^2 = \min_{\bar{\beta} \in \mathbb{R}^d} \mathbb{E}[(y_i - x_i^\top \bar{\beta})^2] .$$

Proof We have

$$\max_{v: \|v\| \le 1} \frac{\mathbb{E}_{P_{\beta, \Sigma, \gamma^2}}[\langle v, x_i(y_i - x_i^\top \hat{\beta})\rangle]}{\sigma_v} = \max_{v: \|v\| \le 1} \frac{\mathbb{E}_{P_{\beta, \Sigma, \gamma^2}}[\langle v, x_i(x_i^\top(\beta - \hat{\beta}) + \eta_i)\rangle]}{\sigma_v}$$

$$= \max_{v: \|v\| \le 1} \frac{\langle v, \Sigma(\beta - \hat{\beta})\rangle}{\sigma_v} = \|\Sigma^{1/2}(\beta - \hat{\beta})\| ,$$

where the second equality uses the fact that η_i has zero mean and x_i has covariance Σ. The last equality follows from Lemma 52. For the noise, we have $\mathbb{E}[(y_i - x_i^\top \bar{\beta})^2] = \mathbb{E}[(x_i^\top \beta + \eta_i - x_i^\top \bar{\beta})^2] = \mathbb{E}[\eta_i^2] + \mathbb{E}[(\beta - \bar{\beta})x_i x_i^\top(\beta - \bar{\beta})]$, which follows from $\mathbb{E}[\eta_i x_i] = 0$. This is minimized when $\bar{\beta} = \beta$, and the minimum is γ^2. ∎

C.2. Step 2: Utility analysis under resilience

The following resilience is a fundamental property of the dataset that determines the sensitivity of $D_S(\hat{\beta})$. We refer to Section B.2 for a detailed explanation of how resilience relates to sensitivity.

Definition 23 (Resilience for linear regression) *For some $\alpha \in (0, 1)$, $\rho_1 \in \mathbb{R}_+$, $\rho_2 \in \mathbb{R}_+$, and $\rho_3 \in \mathbb{R}_+$, we say a set of n labelled data points $S_{\text{good}} = \{(x_i \in \mathbb{R}^d, y_i \in \mathbb{R})\}_{i=1}^n$ is $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient with respect to (β, Σ, γ) for some $\beta \in \mathbb{R}^d$, positive definite $\Sigma \in \mathbb{R}^{d \times d}$, and $\gamma > 0$ if for any $T \subset S_{\text{good}}$ of size $|T| \ge (1 - \alpha)n$, the following holds for all $v \in \mathbb{R}^d$ with $\|v\| = 1$:*

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} \langle v, x_i\rangle(y_i - x_i^\top \beta) \right| \le \rho_1 \sigma_v \gamma , \tag{27}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i\rangle^2 - \sigma_v^2 \right| \le \rho_2 \sigma_v^2 , \tag{28}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i\rangle \right| \le \rho_3 \sigma_v , \text{ and} \tag{29}$$

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} (y_i - x_i^\top \beta)^2 - \gamma^2 \right| \le \rho_4 \gamma^2 , \tag{30}$$

where $\sigma_v^2 = v^\top \Sigma v$.

For example, n i.i.d. samples from sub-Gaussian x_i's and sub-Gaussian η_i's (independent of x_i's) is $\left(\alpha, O(\alpha \log(1/\alpha)), O(\alpha \log(1/\alpha)), O(\alpha\sqrt{\log(1/\alpha)}), O(\alpha \log(1/\alpha))\right)$-resilient. A resilient dataset implies a sensitivity of $\Delta = O(\rho_1/(\alpha n)) = O(\log(1/\alpha)/n)$, where α is a free parameter determined by the target accuracy $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\alpha \log(1/\alpha))$. We show that a sample size of $O((d + \log(1/\delta))/(\varepsilon\alpha))$ is sufficient to achieve the target accuracy for any resilient dataset. In Section C.3, we apply this theorem to resilient datasets from several sampling distributions of interest and characterize the trade-offs.

Theorem 11 (Utility guarantee for linear regression) *There exist positive constants c and C such that for any $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient set S with respect to $(\beta, \Sigma \succ 0, \gamma > 0)$ satisfying $\alpha \in (0, c), \rho_1 < c$, $\rho_2 < c$, $\rho_3^2 \leq c\alpha$ and $\rho_4 < c$, HPTR with the distance function in Eq. (25), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| \leq 32\rho_1$ with probability $1 - \zeta$ if*

$$n \geq C\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} . \tag{31}$$

C.2.1. ROBUSTNESS OF HPTR

One by-product of using robust statistics in $D_S(\hat{\beta})$ is that robustness for HPTR comes for free under a standard data corruption model.

Assumption 2 (α_{corrupt}-corruption) *Given a set $S_{\text{good}} = \{(\tilde{x}_i \in \mathbb{R}^d, \tilde{y}_i \in \mathbb{R})\}_{i=1}^n$ of n data points, an adversary inspects all data points, selects $\alpha_{\text{corrupt}}n$ of the data points, and replaces them with arbitrary dataset S_{bad} of size $\alpha_{\text{corrupt}}n$. The resulting corrupted dataset is called $S = \{(x_i \in \mathbb{R}^d, y_i \in \mathbb{R})\}_{i=1}^n$.*

The same guarantee as Theorem 11 holds under corruption up to a corruption of $\alpha_{\text{corrupt}} < (1/5.5)\alpha$ fraction of a $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient dataset S_{good}. The factor $(1/5.5)$ is due to the fact that the algorithm can remove $(4/5.5)\alpha$ fraction of the good points and a slack of $(0.5/5.5)\alpha$ fraction is needed to resilience of neighboring datasets.

Definition 24 (Corrupt good set) *We say a dataset S is $(\alpha_{\text{corrupt}}, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good with respect to (β, Σ, γ) if it is an α_{corrupt}-corruption of an $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient dataset S_{good}.*

Theorem 12 (Robustness) *There exist positive constants c and C such that for any $((2/11)\alpha, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S with respect to $(\beta, \Sigma \succ 0, \gamma > 0)$ satisfying $\alpha < c$, $\rho_1 < c$, $\rho_2 < c$, $\rho_3^2 \leq c\alpha$ and $\rho_4 < c$, HPTR with the distance function in Eq. (25), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| \leq 32\rho_1$ with probability $1 - \zeta$, if*

$$n \geq C\frac{d + \log(1/(\delta\zeta))}{\varepsilon\alpha} . \tag{32}$$

We provide a proof in Sections C.2.2-C.2.6. When there is no adversarial corruption, Theorem 11 immediately follows by selecting α as a free parameter.

C.2.2. PROOF STRATEGY FOR THEOREM 12

The overall proof strategy follows that of Section B.2.2 for mean estimation. We highlight the differences here.

Lemma 25 (Lemma 10 from (Steinhardt et al., 2018)) *For a $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient set S with respect to (β, Σ, γ) and any $0 \leq \tilde{\alpha} \leq \alpha$, the following holds for any subset $T \subset S$ of size at least $\tilde{\alpha}n$ and for any unit vector $v \in \mathbb{R}^d$:*

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} \langle v, x_i \rangle (y_i - x_i^\top \beta) \right| \leq \frac{2 - \tilde{\alpha}}{\tilde{\alpha}} \rho_1 \sigma_v \gamma \ , \tag{33}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i \rangle^2 - \sigma_v^2 \right| \leq \frac{2 - \tilde{\alpha}}{\tilde{\alpha}} \rho_2 \sigma_v^2 \ , \tag{34}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i \rangle \right| \leq \frac{2 - \tilde{\alpha}}{\tilde{\alpha}} \rho_3 \sigma_v \ , \ and \tag{35}$$

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} (y_i - x_i^\top \beta)^2 - \gamma^2 \right| \leq \frac{2 - \tilde{\alpha}}{\tilde{\alpha}} \rho_4 \gamma^2 \ . \tag{36}$$

This technical lemma is critical in showing that the sensitivity of one-dimensional statistics is bounded by the resilience of the dataset, such that the sensitivity of $D_S(\hat{\beta})$ for a resilient S is bounded by

$$|D_S(\hat{\beta}) - D_{S'}(\hat{\beta})| \leq C'\left(1 + \frac{\rho_3^2}{\alpha}\right) \frac{\rho_1 + (1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|}{\alpha n}$$

for some constant C' and for any neighboring dataset S'. The desired sensitivity bound is local in two ways: it requires S to be resilient and $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\rho_1)$. Under the assumption that $\rho_3^2/\alpha = O(1)$ with a small enough constant, this achieves the desired bound $\Delta = O(\rho_1/(\alpha n))$ with $\hat{\beta} \in B_{\tau,S}$ and $\tau = O(\rho_1)$. The standard utility analysis of exponential mechanisms shows that the error of $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\rho_1)$ can be achieved when $e^{O(d) - c\frac{\varepsilon}{\Delta}\rho_1} \leq \zeta$, which happens if $n = \Omega((d + \log(1/\zeta))/(\varepsilon\alpha))$ with a large enough constant. The TEST step checks the two localities by ensuring that DP conditions are met for the given dataset.

Outline. Analogous to the mean estimation proof, the analyses of utility and the safety test build upon the universal analysis of HPTR in Theorem 15. For linear regression, we show in Sections C.2.3-C.2.5 that the assumptions of Theorem 15 are met for a resilient dataset and the choices of constants and parameters: $\rho = \rho_1$, $c_0 = 31.8$, $c_1 = 10.2$, $\tau = 42\rho_1$, $\Delta = 110\rho_1/(\alpha n)$, $\tau = 42\rho_1$, $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$, and a large enough constant c_2. We assume that $\alpha < c$ and $\rho_1 < c$ for a small enough constant c. A proof of Theorem 12 is shown in Section C.2.6, and Theorem 11 immediately follows by selecting α as a free parameter.

The above resilience properties also imply the following useful resilience on the $S_{\bar{\beta}} = \{(y_i - \bar{\beta}^\top x_i)^2\}_{i=[n]}$ for any vector $\bar{\beta}$.

Lemma 26 (Resilience of residual square) *Let $S_{good} = \{(x_i \in \mathbb{R}^d, y_i \in \mathbb{R})\}_{i=[n]}$ be $(\alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-resilient with respect to (β, Σ, γ). Let $\rho^* = \max\{\rho_1, \rho_2, \rho_4\}$. Then, we have*

1. *for any $T \in S_{\text{good}}$ of size $|T| \geq (1 - \alpha)n$ and any vector $\bar{\beta} \in \mathbb{R}^d$,*

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} (y_i - \bar{\beta}^\top x_i)^2 - (\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 \right| \leq \rho^*(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 , \quad (37)$$

2. *and for any $0 \leq \tilde{\alpha} \leq \alpha$ and $T \in S_{\text{good}}$ of size $|T| \geq \tilde{\alpha} n$, we have*

$$\left| \frac{1}{|T|} \sum_{(x_i, y_i) \in T} (y_i - \bar{\beta}^\top x_i)^2 - (\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 \right| \leq \frac{2 - \tilde{\alpha}}{\tilde{\alpha}} \rho^*(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 . (38)$$

Proof The proof follows directly from resilience properties of Eq. (27), (28) and (30). ■

C.2.3. RESILIENCE IMPLIES ROBUSTNESS

To show that the assumption (d) in Theorem 15 is satisfied, we use the robustness of one-dimensional variance $\sigma_v(\mathcal{M}_{v,\alpha})$ (Lemma 27) and show that $D_S(\hat{\beta})$ is a good approximation of $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|$ (Lemma 29).

Lemma 27 *For an $((2/11)\alpha, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S with respect to (β, Σ, γ) and any unit norm vector $v \in \mathbb{R}^d$, we have $0.9\sigma_v \leq \sigma_v(\mathcal{M}_{v,\alpha}) \leq 1.1\sigma_v$.*

Proof This follows from Lemma 11. ■

Lemma 28 *For a $((2/11)\alpha, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S with respect to (β, Σ, γ) and any unit norm vector $v \in \mathbb{R}^d$, we have $0.99\gamma \leq \hat{\gamma} \leq 1.01\gamma$.*

Proof Analogous to the proof of Lemma 25, for any fixed $\bar{\beta}$, we have

$$\left| \frac{1}{|\mathcal{B}_{\bar{\beta},\alpha}|} \sum_{i \in \mathcal{B}_{\bar{\beta},\alpha}} (y_i - x_i^\top \bar{\beta})^2 - (\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 \right|$$

$$\leq \frac{|\sum_{\mathcal{B}_{\bar{\beta},\alpha} \cap S_{\text{good}}} (y_i - x_i^\top \bar{\beta})^2 - (\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2|}{(1 - (2/5.5)\alpha)n}$$

$$+ \frac{|\sum_{\mathcal{B}_{\bar{\beta},\alpha} \cap S_{\text{bad}}} (y_i - x_i^\top \bar{\beta})^2 - (\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2|}{(1 - (2/5.5)\alpha)n}$$

$$\overset{(a)}{\leq} \frac{(1 - (2/5.5)\alpha)n\rho^*(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2}{(1 - (2/5.5)\alpha)n} + \frac{(2/11)\alpha n \cdot 2\rho^*(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2/((2/11)\alpha)}{(1 - (2/5.5)\alpha)n}$$

$$\overset{(b)}{\leq} 4\rho^*(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 , \tag{39}$$

where (a) follows from Lemma 26, and (b) follows from our assumption that $\alpha \leq c$ for some small enough constant c.

Let $F(\bar{\beta}) = \frac{1}{|\mathcal{B}_{\bar{\beta},\alpha}|} \sum_{i \in \mathcal{B}_{\bar{\beta},\alpha}} (y_i - x_i^\top \bar{\beta})^2$. We know that $\hat{\gamma}^2 = \min_{\bar{\beta}} F(\bar{\beta}) \leq F(\beta)$, which, together with Eq. (39), implies

$$\hat{\gamma}^2 \leq (1 + 4\rho^*)\gamma^2 \leq 1.0201\gamma^2 \ ,$$

when $\rho^* \leq c$ for some c small enough.

Also, we have

$$\hat{\gamma}^2 \geq (1 - 4\rho^*)(\gamma + \|\Sigma^{1/2}(\beta - \bar{\beta})\|)^2 \geq (1 - 4\rho^*)\gamma^2 \geq 0.9801\gamma^2.$$

when $\rho^* \leq c$ for some c small enough. ∎

Lemma 29 *For a $((2/11)\alpha, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S with respect to (β, Σ, γ), if $\hat{\beta} \in B_{\tau,S}$ and $\tau = 42\rho_1$, then $\big| \|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma - D_S(\hat{\beta}) \big| \leq 0.15\tau + 1.1\rho_1 \leq 10.2\rho_1$.*

Proof

By Lemma 22, Lemma 53 and resilience Eq. (27) and Eq. (28), we have

$$
\left| \max_{v:\|v\| \leq 1} \frac{\frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} \langle v, x_i(y_i - x_i^\top \hat{\beta}) \rangle}{\sigma_v} - \left\| \Sigma^{1/2}(\beta - \hat{\beta}) \right\| \right|
$$

$$
= \left| \max_{v:\|v\| \leq 1} \frac{\frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} \left(v^\top x_i x_i^\top (\beta - \hat{\beta}) + v^\top x_i \eta_i \right)}{\sigma_v} - \max_{v:\|v\| \leq 1} \frac{v^\top \Sigma(\beta - \hat{\beta})}{\sigma_v} \right|
$$

$$
\leq \max_{v:\|v\| \leq 1} \left| \frac{v^\top \left(\frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} x_i x_i^\top - \Sigma \right) (\beta - \hat{\beta})}{\sigma_v} + \frac{v^\top \frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} x_i \eta_i}{\sigma_v} \right|
$$

$$
\leq \left\| \Sigma^{-1/2} \left(\frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} x_i x_i^\top - \Sigma \right) (\beta - \hat{\beta}) \right\| + \left\| \Sigma^{-1/2} \frac{1}{|\mathcal{N}_{v,\hat{\beta},\alpha}|} \sum_{i \in \mathcal{N}_{v,\hat{\beta},\alpha}} x_i \eta_i \right\|
$$

$$
\leq \rho_2 \|\Sigma^{1/2}(\beta - \hat{\beta})\| + \rho_1 \gamma \ .
$$

Together with Lemma 27, this implies

$$
\frac{0.9 D_S(\hat{\beta})\hat{\gamma} - \rho_1 \gamma}{1 + \rho_2} \ \leq \ \left\| \Sigma^{1/2}(\beta - \hat{\beta}) \right\| \ \leq \ \frac{1.1 D_S(\hat{\beta})\hat{\gamma} + \rho_1 \gamma}{1 - \rho_2} \ .
$$

Assuming $\rho_2 \leq 0.013$, we have $0.86 D_S(\hat{\beta}) - 1.1\rho_1 \leq \left\| \Sigma^{1/2}(\beta - \hat{\beta}) \right\|/\gamma \leq 1.15 D_S(\hat{\beta}) + 1.1\rho_1$. Since $D_S(\hat{\beta}) \leq \tau$, we get the desired bound. ∎

47

C.2.4. BOUNDED VOLUME

We show that the assumption (a) in Theorem 15 is satisfied for robust estimate $D_S(\hat{\beta})$.

Lemma 30 *For $\rho = \rho_1$, $c_0 = 31.8$, $c_1 = 10.2$, $\tau = 42\rho_1$, $\Delta = 110\rho_1/(\alpha n)$, and $c_2 \geq \log(67/12) + \log((c_0 + 2c_1)/c_1)$, we have $(7/8)\tau - (k^* + 1)\Delta > 0$,*

$$\frac{\text{Vol}(B_{\tau + (k^*+1)\Delta + c_1\rho, S})}{\text{Vol}(B_{(7/8)\tau - (k^*+1)\Delta - c_1\rho, S})} \leq e^{c_2 d} \text{ , and}$$

$$\frac{\text{Vol}(\{\hat{\beta} : \|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma \leq (c_0 + 2c_1)\rho\})}{\text{Vol}(\{\hat{\beta} : \|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma \leq c_1\rho\})} \leq e^{c_2 d} \text{ .}$$

Proof
    The proof is similar to the proof of Lemma 13. The second part of assumption (a) follows from the fact that

$$\text{Vol}(\{\hat{\beta} : \|\Sigma^{1/2}(\hat{\beta} - \beta)\| \leq r\}) = c_d |\Sigma| r^d \text{ ,}$$

for some constant c_d that depends only on the dimension and selecting $c_2 \geq \log((c_0 + 2c_1)/c_1)$. The first part follows from our choices of c_0, c_1, τ, Δ and the following corollary.

Corollary C.1 (Corollary of Lemma 29) *If $\hat{\beta} \in B_{2\tau, S}$ and $\tau = 42\rho_1$, then $\big| \|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma - D_S(\hat{\beta}) \big| \leq 14.2\rho_1$.*

∎

C.2.5. RESILIENCE IMPLIES BOUNDED LOCAL SENSITIVITY

We show that resilience implies the assumption (b) in Theorem 15 (Lemma 34). Assuming $(k^* + 1)/n \leq \alpha/2$, we show a set S' with at most k^* data points arbitrarily changed from S has bounded local sensitivity. This implies that S' is a $((1/5.5)\alpha + (k^*/n), \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set with respect to (β, Σ, γ).

Lemma 31 *For an $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S' with respect to (β, Σ, γ), $\tilde{\alpha} \leq (1/11)\alpha$, and any unit norm $v \in \mathbb{R}^d$, we have $0.9\sigma_v \leq \sigma_v(\mathcal{M}_{v,\alpha}) \leq 1.1\sigma_v$.*

Proof This follows from Lemma 14. ∎

Lemma 32 *For a $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S' with respect to (β, Σ, γ), and any unit norm vector $v \in \mathbb{R}^d$, we have $0.99\gamma \leq \hat{\gamma} \leq 1.01\gamma$.*

Proof This proof follows from the proof of Lemma 28. ∎

Lemma 33 *For a $((1/5.5)\alpha + \tilde{\alpha}, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good set S' with respect to (β, Σ, γ) and $\tilde{\alpha} \leq (1/11)\alpha$, if $\hat{\beta} \in B_{t, S'}$ then we have $\|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma \leq 1.1\rho_1 + 1.15t$ and $\big|D_{S'}(\hat{\beta}) - \|\Sigma^{1/2}(\hat{\beta} - \beta)\|/\gamma\big| \leq 1.1\rho_1 + 0.15t$.*

Proof This proof follows from the proof of Lemma 29. ∎

Lemma 34 *For $\Delta = 110\rho_1/(\alpha n)$, $\tau = 42\rho_1$, and an $((1/5.5)\alpha, \alpha, \rho_1, \rho_2, \rho_3, \rho_4)$-corrupt good S, if*

$$n \; = \; \Omega\Big(\frac{\log(1/(\delta\zeta))}{\alpha\varepsilon}\Big)$$

with a large enough constant, then the local sensitivity in assumption (b) is satisfied.

Proof We follow the proof strategy of Lemma 16 in Section B.2.5. Consider a dataset S' that is at Hamming distance at most $(1/11)\alpha n$ from S and corresponding partition $(\mathcal{T}'_{v,\hat{\beta},\alpha}, \mathcal{N}'_{v,\hat{\beta},\alpha}, \mathcal{B}'_{v,\hat{\beta},\alpha})$ of S' for a specific direction v. By the resilience property of the tails in Eq. (33) and Eq. (34), Lemma 52, and Lemma 53, we have for any $v \in \mathbb{R}^d$ with unit norm $\|v\| = 1$ and any $\hat{\beta} \in \mathbb{R}^d$,

$$\frac{v^\top \frac{1}{|\mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}|} \sum_{i \in \mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} \left(\left(x_i x_i^\top - \Sigma\right)(\beta - \hat{\beta}) + x_i \eta_i \right)}{\sigma_v}$$

$$\leq \; \left\| \Sigma^{-1/2} \left(\frac{1}{|\mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}|} \sum_{i \in \mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} \left(x_i x_i^\top - \Sigma\right)(\beta - \hat{\beta}) \right) \right\| + \tag{40}$$

$$\left\| \Sigma^{-1/2} \left(\frac{1}{|\mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}|} \sum_{i \in \mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} x_i \eta_i \right) \right\|$$

$$\leq \; \frac{2\rho_2}{(1/11)\alpha} \|\Sigma^{1/2}(\beta - \hat{\beta})\| + \frac{2\rho_1}{(1/11)\alpha} \gamma \, , \tag{41}$$

where S_{good} is the original uncorrupted resilient dataset. Similarly, we have

$$\frac{v^\top \frac{1}{|\mathcal{B}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}|} \sum_{i \in \mathcal{B}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} \left(\left(x_i x_i^\top - \Sigma\right)(\beta - \hat{\beta}) + x_i \eta_i \right)}{\sigma_v} \; \leq \; \frac{2\rho_2}{(1/11)\alpha} \|\Sigma^{1/2}(\beta - \hat{\beta})\| + \frac{2\rho_1}{(1/11)\alpha} \gamma \, .$$

This implies

$$\min_{i \in \mathcal{T}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} \frac{v^\top \left(x_i x_i^\top (\beta - \hat{\beta}) + x_i \eta_i \right)}{\sigma_v} - \max_{i \in \mathcal{B}'_{v,\hat{\beta},\alpha} \cap S_{\text{good}}} \frac{\tilde{v}^\top \left(x_i x_i^\top (\beta - \hat{\beta}) + x_i \eta_i \right)}{\sigma_v}$$

$$\leq \; \frac{44\rho_1}{\alpha} \gamma + \frac{44\rho_2}{\alpha} \|\Sigma^{1/2}(\beta - \hat{\beta})\| \, . \tag{42}$$

Analogous to Lemma 16, for a neighboring databases S' and S'', the corresponding middle sets $\mathcal{N}'_{v,\hat{\beta},\alpha}$ and $\mathcal{N}''_{v,\hat{\beta},\alpha}$ differ by at most one entry. Denote those entries by x'_i and $\eta'_i = y'_i - \langle \beta, x'_i \rangle$ in $\mathcal{N}'_{v,\hat{\beta},\alpha}$ and x''_j and η''_j in $\mathcal{N}''_{v,\hat{\beta},\alpha}$. Then, from Eq. (42), we have

$$\left| v^\top \left(\left(x'_i x_i'^\top - x''_j x_j''^\top\right)(\beta - \hat{\beta}) + x'_i \eta'_i - x''_j \eta''_j \right) \right| \leq \left(\frac{44\rho_1}{\alpha} \gamma + \frac{44\rho_2}{\alpha} \|\Sigma^{1/2}(\beta - \hat{\beta})\| \right) \sigma_v \, ,$$

which implies that

$$\left| v^\top \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}'_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right) - v^\top \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right) \right|$$

$$\leq \quad \frac{\sigma_v}{(1-(4/5.5)\alpha)n}\left(\frac{44\rho_1}{\alpha}\gamma + \frac{44\rho_2}{\alpha}\|\Sigma^{1/2}(\beta-\hat{\beta})\| \right) . \tag{43}$$

By the resilience properties in Eq. (27) and Eq. (28), and Lemma 53, Lemma 22, and the fact that $\mathcal{N}''_{v,\hat{\beta},\alpha} \cap S_{\text{good}}$ is at least of size $(1-\alpha)n$, we have for the data points in $\mathcal{N}''_{v,\hat{\beta},\alpha} \cap S_{\text{good}}$,

$$\frac{v^\top \frac{1}{|\mathcal{N}''_{v,\hat{\beta},\alpha}\cap S_{\text{good}}|} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}\cap S_{\text{good}}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v} \leq (1+\rho_2)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| + \rho_1\gamma .$$

By Eq. (41), for any $x_i'' \in \mathcal{N}''_{v,\hat{\beta},\alpha} \cap S_{\text{bad}}$ (where $S_{\text{bad}} = S'' \setminus S_{\text{good}}$), we have

$$\frac{v^\top \left(x_i'' x_i''^\top(\beta-\hat{\beta}) + x_i''\eta_i'' \right)}{\sigma_v} \quad \leq \quad \frac{v^\top \frac{1}{|\mathcal{T}''_{v,\hat{\beta},\alpha}\cap S_{\text{good}}|} \sum_{i\in\mathcal{T}''_{v,\hat{\beta},\alpha}\cap S_{\text{good}}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v}$$

$$\leq \quad \left(\frac{22\rho_2}{\alpha} + 1 \right)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| + \frac{22\rho_1}{\alpha}\gamma .$$

Since $|S_{\text{bad}}| \leq (1.5/5.5)\alpha n$ and $\alpha < c$ for some small enough constant c, we have

$$\frac{v^\top \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v}$$

$$= \quad \frac{v^\top \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}\cap S_{\text{bad}}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v} +$$

$$\frac{v^\top \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}\cap S_{\text{good}}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v}$$

$$\leq \quad \frac{(6\rho_2 + (1.5/5.5)\alpha)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| + 6\rho_1\gamma}{1-(4/5.5)\alpha} + \left((1+\rho_2)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| + \rho_1\gamma \right)$$

$$\leq \quad 7\rho_1\gamma + (1 + \alpha + 7\rho_2)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| . \tag{44}$$

Analogous to Eq. (18), by using the resilience properties in Eqs. (28) and (29), we have

$$|\sigma_v'^2 - \sigma_v''^2| \quad = \quad \frac{1}{(1-(4/5.5)\alpha)n}\left| \sum_{x_i\in\mathcal{N}'_{v,\hat{\beta},\alpha}} \langle v,x_i\rangle^2 - \sum_{x_i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \langle v,x_i\rangle^2 \right|$$

$$\leq \quad \frac{64\cdot 11^2\cdot\rho_3^2\sigma_v^2}{\alpha^2(1-(4/5.5)\alpha)n} . \tag{45}$$

By Eqs. (44), (43), and (45), we have

$$
\left| D_{S'}(\hat{\beta}) - D_{S''}(\hat{\beta}) \right|
$$

$$
\leq \max_{v:\|v\|=1} \left| \frac{v^\top \frac{1}{|\mathcal{N}'_{v,\hat{\beta},\alpha}|} \sum_{i\in\mathcal{N}'_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma'_v \hat{\gamma}'} - \frac{v^\top \frac{1}{|\mathcal{N}''_{v,\hat{\beta},\alpha}|} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma''_v \hat{\gamma}''} \right|
$$

$$
\leq \max_{v:\|v\|=1} \left| \frac{v^\top \left(\frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}'_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right) - \frac{1}{(1-(4/5.5)\alpha)n} \sum_{i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right) \right)}{\sigma'_v \hat{\gamma}'} \right|
$$

$$
+ \max_{v:\|v\|=1} \left| \frac{v^\top \frac{1}{|\mathcal{N}''_{v,\hat{\beta},\alpha}|} \sum_{x_i\in\mathcal{N}''_{v,\hat{\beta},\alpha}} \left(x_i x_i^\top(\beta-\hat{\beta}) + x_i\eta_i \right)}{\sigma_v} \right| \left| \frac{\sigma_v}{\sigma'_v \hat{\gamma}'} - \frac{\sigma_v}{\sigma''_v \hat{\gamma}''} \right|
$$

$$
\leq \frac{44\rho_1}{0.9\cdot 0.99(1-(4/5.5)\alpha)n\alpha} + \frac{44\rho_2}{0.9\cdot 0.99(1-(4/5.5)\alpha)n\alpha} \frac{\|\Sigma^{1/2}(\beta-\hat{\beta})\|}{\gamma}
$$

$$
+ \frac{64\cdot 11^2\cdot \rho_3^2\cdot 0.02\gamma}{0.9^3\alpha^2(1-(4/5.5)\alpha)n\cdot 0.99^2\gamma^2} \left(7\rho_1\gamma + (1+\alpha+7\rho_2)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| \right)
$$

$$
\leq \left(\frac{0.12}{\alpha n} + \frac{0.016}{\alpha n} \right) \frac{\|\Sigma^{1/2}(\hat{\beta}-\beta)\|}{\gamma} + \left(\frac{9\rho_1}{\alpha n} + \frac{0.07\rho_1}{\alpha n} \right)
$$

$$
\leq \frac{0.2}{\alpha n} \frac{\|\Sigma^{1/2}(\hat{\beta}-\beta)\|}{\gamma} + \frac{50\rho_1}{\alpha n}
$$

where the last three inequalities follow from our assumptions that $\alpha \leq c$ and $\rho_2 \leq c$, $\rho_3^2 \leq c\alpha$, $\rho_4 \leq c$ with a small enough constant c and Lemma 32. From Lemma 33, we know that if $\hat{\beta} \in B_{\tau+(k^*+3)\Delta,S}$, we have $\|\Sigma^{1/2}(\hat{\beta}-\beta)\|/\gamma \leq 1.1\rho_1 + 1.15(\tau+(k^*+3)\Delta)$. We show that $\|\Sigma^{1/2}(\hat{\beta}-\beta)\| \leq 50\rho_1\gamma$ for the choices of Δ, k^*, τ and n:

$$
\begin{aligned}
1.1\rho_1 + 1.15(\tau+(k^*+3)\Delta) &\leq 49\rho_1 + \frac{50\rho_1 \log(1/(\delta\zeta))}{\varepsilon\alpha n} \\
&\leq 50\rho_1 ,
\end{aligned}
$$

where $\Delta = 110\rho_1/(\alpha n)$, $\tau = 42\rho_1$, $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$, $\varepsilon \leq \log(4/\delta\zeta)$ and $n \geq C'\log(1/(\delta\zeta))/(\varepsilon\,\alpha)$ for some large enough universal constant $C' > 0$. This implies that

$$
\left| D_{S'}(\hat{\beta}) - D_{S''}(\hat{\beta}) \right| \leq \frac{110\rho_1}{\alpha n} = \Delta .
$$

∎

C.2.6. PROOF OF THEOREM 12

We show that the sufficient conditions of Theorem 15 are met for the following choices of constants and parameters: $p = d$, $\rho = \rho_1$, $c_0 = 31.8$, $c_1 = 10.2$, $\tau = 42\rho_1$, and $\Delta = 110\rho_1/(\alpha n)$. We set c_2 to be a large constant and change only the constant factor in the sample complexity. The

assumptions (a), (b), and (d) follow from Lemmas 30, 34, and 29, respectively. The assumption (c) follows from

$$\Delta = \frac{110\rho_1}{\alpha n} \le \frac{1.2\rho_1\varepsilon}{32(c_2d + (\varepsilon/2) + \log(16/(\delta\zeta)))} = \frac{(c_0 - 3c_1)\rho\varepsilon}{32(c_2p + (\varepsilon/2) + \log(16/(\delta\zeta)))}$$

for a large enough $n \ge C'(d + \log(1/(\delta\zeta)))/(\alpha\varepsilon)$. This finishes the proof of Theorem 12, from which Theorem 11 follows immediately.

C.3. Step 3: Achievability guarantees

We provide utility guarantees for popular families of distributions studied in the private or robust linear regression literature: sub-Gaussian (Diakonikolas et al., 2019b; Gao, 2020; Zhu et al., 2019; Cai et al., 2019; Wang, 2018) and hypercontractive (Zhu et al., 2019; Klivans et al., 2018; Cherapanamjeri et al., 2020; Jambulapati et al., 2021; Bakshi and Prasad, 2021; Prasad et al., 2018). Similar to mean estimation, the resilience we need scales with the variance. For sub-Gaussian distributions, this requires a lower bound on the variance of the form $\sigma \preceq c\Gamma$ for the sub-Gaussian proxy Γ. For the k-th moment bounded distributions, we require hypercontractivity.

C.3.1. SUB-GAUSSIAN DISTRIBUTIONS

The most common scenario in linear regression is when both input x_i and noise η_i are sub-Gaussian (as defined in Eq. (20)) and independent of each other. The next lemma shows that the resulting dataset is $(O(\alpha\log(1/\alpha)), O(\alpha\log(1/\alpha)), O(\alpha\sqrt{\log(1/\alpha)}), O(\alpha\log(1/\alpha)))$-resilient, which follows from the covariance resilience of sub-Gaussian distributions.

Lemma 35 (Resilience for sub-Gaussian samples) *Let $\mathcal{D}_1$ be a distribution of $x_i \in \mathbb{R}^d$, which is zero mean sub-Gaussian with covariance Σ and sub-Gaussian proxy $0 \prec \Gamma \preceq c\Sigma$ for some constant c. Let $\mathcal{D}_2$ be a distribution of $\eta_i \in \mathbb{R}$, which is a zero mean one-dimensional sub-Gaussian with variance γ^2 and sub-Gaussian proxy $\gamma_0^2 \le c\gamma^2$ for some constant c. A multiset of i.i.d. labeled samples $S = \{(x_i, y_i)\}_{i=1}^n$ is generated from a linear model with noise η_i independent of x_i: $y_i = x_i^\top\beta + \eta_i$, where the input x_i and the independent noise η_i are i.i.d. samples from $\mathcal{D}_1$ and $\mathcal{D}_2$. There exist constants c_1 and $c_2 > 0$ such that, for any $\alpha \in (0, 1/2)$, if $n \ge c_1((d + \log(1/\zeta))/(\alpha\log(1/\alpha))^2)$, then, with probability $1 - \zeta$, S is $(\alpha, c_2\alpha\log(1/\alpha), c_2\alpha\log(1/\alpha), c_2\alpha\sqrt{\log(1/\alpha)}, c_2\alpha\log(1/\alpha))$-resilient with respect to (β, Σ, γ).*

Proof This follows from (Jambulapati et al., 2020, Corollary 4). Let $\tilde{x}_i := \begin{bmatrix} \Sigma^{-1/2}x_i \\ \eta_i/\gamma \end{bmatrix} \in \mathbb{R}^{d+1}$.

By definition, we know that $\tilde{x}_i$ can be seen as samples from a zero mean sub-Gaussian distribution with covariance $\mathbf{I}_{(d+1)\times(d+1)}$. By (Jambulapati et al., 2020, Corollary 4) and a union bound, we know that if $n = \Omega(d + \log(1/\zeta))/(\alpha\log(1/\alpha))^2$, then there exists a constant C_1 such that with probability $1 - \zeta$, for any $T \subset S$ and $|T| \ge (1 - \alpha)n$ and any unit vector $u \in \mathbb{R}^{d+1}$, $v \in \mathbb{R}^d$, we

have

$$\left| u^\top \left(\frac{1}{|T|} \sum_{x_i \in T} \tilde{x}_i \tilde{x}_i^\top - \mathbf{I}_{(d+1) \times (d+1)} \right) u \right| \leq C_1 \alpha \log(1/\alpha) , \tag{46}$$

$$\left| v^\top \left(\frac{1}{|T|} \sum_{x_i \in T} \Sigma^{-1/2} x_i x_i^\top \Sigma^{-1/2} - \mathbf{I}_{d \times d} \right) v \right| \leq C_1 \alpha \log(1/\alpha) , \text{ and} \tag{47}$$

$$\left| \frac{1}{|T|} \sum_{\eta_i \in T} \frac{\eta_i^2}{\gamma^2} - 1 \right| \leq C_1 \alpha \log(1/\alpha) . \tag{48}$$

Let $u := \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, where $u_1 \in \mathbb{R}^d$ and $u_2 \in \mathbb{R}$ and $\|u_1\|^2 + u_2^2 = 1$. Then, Eq. (46) is equivalent to

$$\left| u_1^\top \left(\frac{1}{|T|} \sum_{i \in T} \Sigma^{-1/2} x_i x_i^\top \Sigma^{-1/2} - \mathbf{I}_{d \times d} \right) u_1 + \frac{2u_2}{\gamma} u_1^\top \frac{1}{|T|} \sum_{i \in T} \Sigma^{-1/2} x_i \eta_i + \frac{u_2^2}{\gamma^2} \frac{1}{|T|} \sum_{i \in T} (\eta_i^2 - \gamma^2) \right|$$

$$\leq \quad C_1 \alpha \log(1/\alpha) . \tag{49}$$

By Eq. (47) and (48), we know

$$\left| u_1^\top (\frac{1}{|T|} \sum_{i \in T} \Sigma^{-1/2} x_i x_i^\top \Sigma^{-1/2} - \mathbf{I}_{d \times d}) u_1 \right| \leq C_1 \alpha \log(1/\alpha) \|u_1\|^2$$

$$\left| \frac{u_2^2}{\gamma^2} \frac{1}{|T|} \sum_{i \in T} (\eta_i^2 - \gamma^2) \right| \leq C_1 \alpha \log(1/\alpha) u_2^2 .$$

This means that

$$-C_1 \alpha \log(1/\alpha)(1 + \|u_1\|^2 + u_2^2) \leq \frac{2u_2}{\gamma} u_1^\top \frac{1}{|T|} \sum_{i \in T} \Sigma^{-1/2} x_i \eta_i \leq C_1 \alpha \log(1/\alpha)(1 + \|u_1\|^2 + u_2^2) . \tag{50}$$

For any unit vector $w \in \mathbb{R}^d$, let $u_1 = 0.5w$. Thus, we have $u_2^2 = 0.75$. Eq. (50) implies

$$\left| \frac{1}{\gamma} w^\top \frac{1}{|T|} \sum_{i \in T} \Sigma^{-1/2} x_i \eta_i \right| \leq C_2 \alpha \log(1/\alpha) \tag{51}$$

for some constant C_2. This proves the first resilience property in Eq. (27). The second, third and fourth resilience properties in Eqs. (28), (29) and (30) follow from (Dong et al., 2019, Lemma 4.1), (Jambulapati et al., 2020, Corollary 4) and a union bound. ∎

The preceding resilience lemma and Theorem 12 imply the following optimal utility guarantee.

Corollary C.2 *Under the hypothesis of Lemma 35, there exists a constant $c > 0$ such that for any $\alpha \in (0, c)$, a sample size of*

$$n = O\Big(\frac{d + \log(1/\zeta)}{(\alpha \log(1/\alpha))^2} + \frac{d + \log(1/(\delta\zeta))}{\alpha\varepsilon} \Big) ,$$

a sensitivity of $\Delta = O(\log(1/\alpha)/n)$, and a threshold of $\tau = O(\alpha \log(1/\alpha))$ with large enough constants are sufficient for $\mathrm{HPTR}(S)$ *with the distance function in Eq. (25) to achieve*

$$\frac{1}{\gamma}\|\Sigma^{1/2}(\hat{\beta} - \beta)\| \;=\; O(\alpha \log(1/\alpha)) \tag{52}$$

with probability $1 - \zeta$. Further, the same guarantee holds even if α-fraction of the samples is arbitrarily corrupted, as in Assumption 2.

The best known algorithm for DP linear regression from (Cai et al., 2019) requires Σ to be close to the identity matrix, which is equivalent to assuming that we know Σ. The error bound is nearly optimal under α-corruption, i.e., HPTR is the first robust estimator that is both differentially private and also achieves the near-optimal error rate of $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\alpha \log(1/\alpha))$, matching the known information-theoretic lower bound of $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = \Omega(\alpha)$ (Gao, 2020) up to a log factor. This lower bound holds for any robust estimator that is not necessarily private and regardless of how many samples are available. If privacy is not required (i.e., $\varepsilon = \infty$), a similar guarantee can be achieved by, for example, (Diakonikolas et al., 2019b).

C.3.2. Hypercontractive distributions with independent noise

We assume that x_i and η_i are independent and (κ, k)-hypercontractive and $(\tilde{\kappa}, k)$-hypercontractive, respectively, as in Definition 18. For the necessity of hypercontractive conditions for robust linear regression, we refer to (Zhu et al., 2019, Section F.5). The next lemma shows that the resulting dataset has a subset of size at least $(1-\alpha)n$ that is $(O(\alpha), O(\alpha^{1-1/k}), O(\alpha^{1-2/k}), O(\alpha^{1-1/k}), O(\alpha^{1-2/k}))$-resilient.

Lemma 36 (Resilience for hypercontractive samples) *For some integer $k \geq 4$ and positive scalar parameters κ and $\tilde{\kappa}$, let $\mathcal{D}_1$ be a (κ, k)-hypercontractive distribution on $x_i \in \mathbb{R}^d$ with zero mean and covariance $\Sigma \succ 0$. Let $\mathcal{D}_2$ be a $(\tilde{\kappa}, k)$-hypercontractive distribution on $\eta_i \in \mathbb{R}$ with zero mean and variance γ^2. A multiset of labeled samples $S = \{(x_i, y_i)\}_{i=1}^{n}$ is generated from the linear model $y_i = x_i^\top \beta + \eta_i$, where the input x_i and the independent noise η_i are i.i.d. samples from $\mathcal{D}_1$ and $\mathcal{D}_2$. For any $\alpha \in (0, 1/2)$ and any constant $c_3 > 0$, there exist constants c_1 and $c_2 > 0$ that depend only on c_3 such that if*

$$n \;\geq\; c_1 \Big(\frac{d}{\zeta^{2(1-1/k)}\alpha^{2(1-1/k)}} + \frac{k^2\alpha^{2-2/k}(1+1/\tilde{\kappa}^2)d\log d}{\zeta^{2-4/k}\kappa^2} + \frac{\kappa^2(1+\tilde{\kappa}^2)d\log d}{\alpha^{2/k}} \Big) , \tag{53}$$

then S is $(c_3\alpha, \alpha, c_2 k\kappa\tilde{\kappa}\alpha^{1-1/k}\zeta^{-1/k}, c_2 k^2\kappa^2\alpha^{1-2/k}\zeta^{-2/k}, c_2 k\kappa\alpha^{1-1/k}\zeta^{-1/k}, c_2 k^2\tilde{\kappa}^2\alpha^{1-2/k}\zeta^{-2/k})$-corrupt good with respect to (β, Σ, γ) with probability $1 - \zeta$.

Proof Since x_i and η_i are independent, we know

$$\mathbb{E}\left[\left| \left\langle v, \gamma^{-1}\Sigma^{-1/2}x\eta \right\rangle \right|^k \right] = \mathbb{E}\left[\left| \left\langle v, \Sigma^{-1/2}x \right\rangle \right|^k \right] \mathbb{E}\left[|\gamma^{-1}\eta|^k \right] \leq \kappa^k \tilde{\kappa}^k .$$

This implies that $\gamma^{-1}\Sigma^{-1/2}x\eta$ is a k-th moment bounded distribution with covariance $\mathbf{I}_{d\times d}$. By Lemma 19, under the sample complexity of (53), with probability $1-8\zeta$, there exists a subset $S_{\text{good}} \subset S$ such that $|S_{\text{good}}| \geq (1-\alpha)n$, and there exists a constant C such that for any subset $T \subset S_{\text{good}}$ and $|T| \geq (1-10\alpha)|S_{\text{good}}|$, we have

$$\left\| \frac{1}{|T|} \sum_{i\in T} \frac{1}{\gamma} \Sigma^{-1/2} x_i \eta_i \right\| \leq Ck\kappa\tilde{\kappa}\gamma\alpha^{1-1/k}\zeta^{-1/k} . \tag{54}$$

This proves the first resilience property in Eq. (27). The second resilience property in Eq. (28), the third in Eq. (29) and the fourth in Eq. (30) follow directly from Lemma 19. ∎

The preceding resilience lemma and Theorem 12 imply the following utility guarantee. HPTR is naturally robust against $(1/5.5 - c_3)\alpha$-corruption of the data. Choosing appropriate constants, we get the following result.

Corollary C.3 *Under the hypothesis of Lemma 36, there exists a constant $c > 0$ such that for any $\alpha \leq c$ and $k^2\kappa^2\alpha^{1-2/k} \leq c$, it is sufficient to have a dataset of size*

$$n = O\Big(\frac{d}{\zeta^{2(1-1/k)}\alpha^{2(1-1/k)}} + \frac{k^2\alpha^{2-2/k}(1+1/\tilde{\kappa}^2)d\log d}{\zeta^{2-4/k}\kappa^2} + \frac{\kappa^2(1+\tilde{\kappa}^2)d\log d}{\alpha^{2/k}} + \frac{d+\log(1/\delta)}{\alpha\varepsilon} \Big) , \tag{55}$$

a sensitivity of $\Delta = O(1/(n\alpha^{1/k}))$, and a threshold of $\tau = O(\alpha^{1-1/k})$ with large enough constants for $\text{HPTR}(S)$ with the distance function in Eq. (25) to achieve $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(k\kappa\tilde{\kappa}\alpha^{1-1/k}\zeta^{-1/k})$ with probability $1-\zeta$. Further, the same guarantee holds even if α-fraction of the samples is arbitrarily corrupted, as in Assumption 2.

The error bound is optimal under α-corruption; namely, the error bound $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta}-\beta)\| = O(\alpha^{1-1/k})$ matches the lower bound $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = \Omega(\alpha^{1-1/k})$ by Bakshi and Prasad (2021), where the noise η_i is $(1,k)$-hypercontractive and independent of x_i, which is also $(1,k)$-hypercontractive. For completeness, we provide the lower bound in Appendix H. HPTR is the first algorithm that guarantees both differential privacy and an optimal robust error bound of $O(\alpha^{1-1/k})$ for hypercontractive distributions. If only robust error bound under α-corruption is at issue, (Zhu et al., 2019) also achieves the same optimal error bound but does not provide differential privacy. Further, in this robust but not private case with $\varepsilon = \infty$, our sample complexity improves by a factor of $\alpha^{2/k}$ upon the state-of-the-art sample complexity of (Zhu et al., 2019, Theorem 3.3), which shows that $n = O(d/\alpha^2)$ is sufficient to achieve $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\alpha^{1-1/k})$.

Suppose $k, \kappa, \tilde{\kappa}$, and ζ are $\Theta(1)$. HPTR achieves $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\alpha^{1-1/k})$ with $n = \tilde{O}(d/(\alpha^{2-2/k}) + (d + \log(1/\delta))/(\alpha\varepsilon))$ samples, where $\tilde{O}$ hides logarithmic factors in d. To the best of our knowledge, HPTR is the first algorithm for linear regression that guarantees (ε, δ)-DP under hypercontractive distributions with independent noise.

C.3.3. HYPERCONTRACTIVE DISTRIBUTIONS WITH DEPENDENT NOISE

We assume x_i and η_i may be dependent and marginally (κ, k)-hypercontractive and $(\tilde{\kappa}, k)$-hypercontractive, respectively, as defined in Definition 18. In this case, the first resilience ρ_1 that determines the error rate increases from $O(\alpha^{1-1/k})$ to $O(\alpha^{1-2/k})$ as a result of the potential correlation between input and noise. The next lemma shows that the the resulting dataset has a subset of size at least $(1-\alpha)n$ that is $(O(\alpha), O(\alpha^{1-2/k}), O(\alpha^{1-2/k}), O(\alpha^{1-1/k}), O(\alpha^{1-2/k}))$-resilient.

Lemma 37 (Resilience for hypercontractive samples with dependent noise) *For some integer $k \geq 4$ and positive scalar parameters κ and $\tilde{\kappa}$, let $\mathcal{D}_1$ be a (κ, k)-hypercontractive distribution on $x_i \in \mathbb{R}^d$ with zero mean and covariance $\Sigma \succ 0$. Let $\mathcal{D}_2$ be a $(\tilde{\kappa}, k)$-hypercontractive distribution on $\eta_i \in \mathbb{R}$ with variance γ^2. A multiset of labeled samples $S = \{(x_i, y_i)\}_{i=1}^n$ is generated from a linear model as follows: $y_i = x_i^\top \beta + \eta_i$, where $\{(x_i, \eta_i)\}_{i \in [n]}$ are i.i.d. samples from some distribution $\mathcal{D}$ whose marginal distribution for x_i is $\mathcal{D}_1$, the marginal distribution for η_i is $\mathcal{D}_2$, and $\mathbb{E}[x_i \eta_i] = 0$. For any $\alpha \in (0, 1/2)$ and $c_3 > 0$, there exist constants c_1 and $c_2 > 0$ that depend only on c_3 such that if*

$$n \geq c_1 \left(\frac{d}{\zeta^{2(1-1/k)} \alpha^{2(1-1/k)}} + \frac{k^2 \alpha^{2-4/k}(1 + 1/\tilde{\kappa}^2) d \log d}{\zeta^{2-4/k} \kappa^2 \tilde{\kappa}^2} + \frac{\kappa^2(\tilde{\kappa}^2 + 1) d \log d}{\alpha^{4/k}} \right), \quad (56)$$

then S is $(c_3\alpha, \alpha, c_2 k \kappa \tilde{\kappa} \alpha^{1-2/k} \zeta^{-2/k}, c_2 k^2 \kappa^2 \alpha^{1-2/k} \zeta^{-2/k}, c_2 k \kappa \alpha^{1-1/k} \zeta^{-1/k}, c_2 k^2 \tilde{\kappa}^2 \alpha^{1-2/k} \zeta^{-2/k})$-corrupt good with respect to (β, Σ, γ) with probability $1 - \zeta$.

Proof Since η_i and x_i are dependent, we can bound only the $k/2$-th moment of $\gamma^{-1} \Sigma^{-1/2} x \eta$. By the Holder inequality, we have

$$\mathbb{E}\left[\left| \left\langle v, \Sigma^{-1/2} \gamma^{-1} x \eta \right\rangle \right|^{k/2} \right] \leq \sqrt{\mathbb{E}\left[\left| \left\langle v, \Sigma^{-1/2} x \right\rangle \right|^k \right] \mathbb{E}\left[|\gamma^{-1} \eta|^k \right]} \leq \kappa^{k/2} \tilde{\kappa}^{k/2} .$$

The rest of the proof follows similarly to the proof of Lemma 36. ∎

The preceding resilience lemma and Theorem 12 imply the following optimal utility guarantee, which achieves an error rate of $O(\alpha^{1-2/k})$.

Corollary C.4 *Under the hypothesis of Lemma 37, there exists a constant $c > 0$ such that for any $\alpha \leq c$ and $k^2 \kappa^2 \alpha^{1-2/k} \leq c$, it is sufficient to have a dataset of size*

$$n = O\left(\frac{d + \log(1/\delta)}{\alpha \varepsilon} + \frac{d}{\zeta^{2(1-1/k)} \alpha^{2(1-1/k)}} + \frac{k^2 \alpha^{2-4/k}(1 + 1/\tilde{\kappa}^2) d \log d}{\zeta^{2-4/k} \kappa^2 \tilde{\kappa}^2} + \frac{\kappa^2(\tilde{\kappa}^2 + 1) d \log d}{\alpha^{4/k}} \right),$$

a sensitivity $\Delta = O(1/(n\alpha^{2/k}))$, and a threshold $\tau = O(\alpha^{1-2/k})$, with large enough constants for $\mathrm{HPTR}(S)$ with the distance function in Eq. (25) to achieve $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(k \kappa \tilde{\kappa} \alpha^{1-2/k} \zeta^{-2/k})$ with probability $1 - \zeta$. Further, the same guarantee holds even if an α-fraction of the samples is arbitrarily corrupted, as in Assumption 2.

This error rate is optimal in its dependence on α under α-corruption. When η_i and x_i are dependent, (Bakshi and Prasad, 2021) gives a lower bound of error rate $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = \Omega(\tilde{\kappa} \alpha^{1-2/k})$ that holds regardless of how many samples we have and without the privacy constraints. For completeness, we provide the lower bound in Appendix H. If only a robust error bound under α-corruption is at issue, (Zhu et al., 2019) also achieves the same optimal error bound but does not provide differential privacy. Further, in this robust but not private case with $\varepsilon = \infty$, our sample complexity improves by a factor of $\alpha^{2/k}$ upon the state-of-the-art sample complexity of (Zhu et al., 2019, Theorem 3.3), which shows that $n = O(d/\alpha^2)$ is sufficient to achieve $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\| = O(\alpha^{1-2/k})$.

Remark. Suppose $\zeta, \kappa, \tilde{\kappa}$, and k are $\Theta(1)$. The sample complexity of HPTR is $n = \tilde{O}((d + \log(1/\delta))/\alpha^{2(1-1/k)} + d/(\alpha\varepsilon))$. The first term in the upper bound might be loose, since we ensure stronger resilience than we need. From Theorem 11, we know that we require $\rho_1 \leq c$ and $\rho_3^2 \leq c\alpha$, and from the optimal error rate, we want $\rho_1 \leq c\alpha^{1-2/k}$. The resilience we ensure in Lemma 37 is $(\alpha, \rho_1 = \alpha^{1-2/k}, \rho_2 = \alpha^{1-2/k}, \rho_3 = \alpha^{1-1/k})$, which guarantees an unnecessarily small ρ_2 and ρ_3. A similar slack was also in the mean estimation, which did not affect the final sample complexity. In this case, i.e., with linear regression and hypercontractive distributions, it enlarges sample complexity. Tighter analysis of the resilience, which guarantees a larger ρ_2 and ρ_3, can improve the first term in the sample complexity in its dependence on α.

For the second term, we provide a nearly matching lower bound of $n = \Omega(\min\{d, \log(1/\delta)\}/\alpha\varepsilon)$ to achieve $(1/\gamma)\|\Sigma^{1/2}(\hat{\beta} - \beta)\|^2 \leq O(\alpha^{2-4/k})$ in Proposition C.5, proving that it is tight when $\delta = \exp(-\Theta(d))$. To the best of our knowledge, HPTR is the first algorithm for linear regression that guarantees (ε, δ)-DP under hypercontractive distributions with dependent noise.

Proposition C.5 (Lower bound of hypercontractive linear regression with dependent noise)
For any $k \geq 4$, let $\mathcal{P}_{\kappa,k,\Sigma,\gamma^2}$ be a distribution over $(x_i, \eta_i) \in \mathbb{R}^d \times \mathbb{R}$, where x_i is (κ, k)-hypercontractive with zero mean and covariance Σ, and η_i is (κ, k)-hypercontractive with zero mean and variance γ^2. We observe labelled examples a linear model $y_i = x_i^\top \beta + \eta_i$ with $\mathbb{E}[x_i\eta_i] = 0$ such that $\beta = \Sigma^{-1}\mathbb{E}[y_i x_i]$. Let $\mathcal{M}_{\varepsilon,\delta}$ denote a class of (ε, δ)-DP estimators that are measurable functions over n i.i.d. samples $S = \{(x_i, y_i)\}_{i=1}^n$ from a distribution. There exist positive constants $c, \gamma, \kappa = O(1)$ such that, for $\varepsilon \in (0, 10)$,

$$\inf_{\hat{\beta} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{\Sigma \succ 0, P \in \mathcal{P}_{\kappa,k,\Sigma,\gamma^2}} \frac{1}{\gamma}\mathbb{E}_{P^n}[\|\Sigma^{1/2}(\hat{\beta}(S) - \beta)\|^2] \geq c \min\left\{\left(\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon}\right)^{2-4/k}, 1\right\}.$$

Proof We adopt the same framework as used in the proof of Proposition B.5. We choose $\mathcal{P}$ to be $\mathcal{P} = \mathcal{P}_{\Sigma,k}$. It suffices to construct index set $\mathcal{V}$ and indexed family of distributions $\mathcal{P}_\mathcal{V}$ such that $d_{\mathrm{TV}}(P_v, P_{v'}) = \alpha$ and $\rho(\beta_v, \beta_{v'}) \geq t$, where β_v is the least square solution of P_v. By (Acharya et al., 2021, Lemma 6), there exists a finite set $\mathcal{V} \subset \mathbb{R}^d$ with cardinality $|\mathcal{V}| = 2^{\Omega(d)}$, $\|v\| = 1$ for all $v \in \mathcal{V}$, and $\|v - v'\| \geq 1/2$ for all $v \neq v' \in \mathcal{V}$. Let $f_{\mu,\Sigma}(x)$ be a density function of $\mathcal{N}(\mu, \Sigma)$. We construct a marginal distribution over $\mathbb{R}^d$ as follows

$$D_x^v(x) = \begin{cases} \alpha/2, & \text{if } x = -\alpha^{-1/k}v, \\ \alpha/2, & \text{if } x = \alpha^{-1/k}v, \\ (1 - \alpha)f_{0,\mathbf{I}_{d \times d}}(x) & \text{otherwise}, \end{cases} \quad . \tag{57}$$

It is straightforward to verify that $\mathbb{E}_{P_x^v}[x] = 0$, $\mathbb{E}_{P_x^v}[xx^\top] = (1 - \alpha)\mathbf{I}_{d \times d} + \alpha^{1-2/k}vv^\top$ and thus $\frac{1}{2}\mathbf{I}_{d \times d} \preceq \mathbb{E}_{P_x^v}[xx^\top] \preceq 2\mathbf{I}_{d \times d}$ for $\alpha \leq 1/2$. Furthermore, we have

$$\mathbb{E}_{x \sim P_x^v}[|\langle u, x\rangle|^k] \leq \langle u, v\rangle^k + (1 - \alpha)c_k^k = O(1),$$

where we use the fact that there exists a constant $c_k > 0$ such that the k-th moment of Gaussian distribution is bounded by c_k^k. Since $\frac{1}{2}\mathbf{I}_{d \times d} \preceq \mathbb{E}_{P_x^v}[xx^\top] \preceq 2\mathbf{I}_{d \times d}$, we know that x is $(O(1), k)$-hypercontractive. We construct conditional distribution $D^v(y|x)$ as follows

$$y|x = \begin{cases} -\alpha^{-1/k} & \text{if } x = -\alpha^{-1/k}v \\ \alpha^{-1/k} & \text{if } x = \alpha^{-1/k}v \\ \mathcal{N}(0, 1) & \text{otherwise} \end{cases} \quad .$$

Then, we have

$$
\begin{aligned}
\beta_v &= \mathbb{E}_{x\sim P_x^v}[xx^\top]^{-1}\mathbb{E}_{x,y\sim P_{x,y}^v}[xy]\\
&= \mathbb{E}_{x\sim P_x^v}[xx^\top]^{-1}\alpha^{1-2/k}v\ .
\end{aligned}
$$

This implies that $t = \min_{v\neq v'\in\mathcal{V}}\|\beta_v - \beta_{v'}\| \geq 1/2\alpha^{1-2/k}\min_{v\neq v'\in\mathcal{V}}\|v - v'\| = \Omega(\alpha^{1-2/k})$. We are left to verify that $\eta = y - \langle\beta_v, x\rangle$ is also hypercontractive:

$$
\mathbb{E}[|\eta|^k] = \alpha\left|\alpha^{-1/k} - v^\top\mathbb{E}_{x\sim P_x^v}[xx^\top]^{-1}v\alpha^{1-3/k}\right|^k + (1-\alpha)\mathbb{E}_{x\sim\mathcal{N}(0,2\mathbf{I}_{d\times d})}[|x|^k] = O(1)\ ,
$$

where we use the fact that the k-th moment of standard Gaussian is bounded by some constants $C_k > 0$ and $k = O(1)$. It is straightforward to see that the total variation distance $d_{\mathrm{TV}}(P_{x,y}^v, P_{x,y}^{v'}) = \alpha$.

Next, we apply a reduction of estimation to testing with this packing $\mathcal{V}$ similar to that we used in the proof of Proposition B.5. For (ε, δ)-DP estimator $\hat{\beta}$, using Theorem 20, we have

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}}\mathbb{E}_{P^n}[\|\Sigma(P)^{1/2}(\hat{\beta}(S) - \beta(P))\|^2]\\
&\geq \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}\mathbb{E}_{P_v^n}[\|\Sigma(P_v)^{1/2}(\hat{\beta}(S) - \beta(P_v))\|^2]\\
&= t^2\frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}P_v\left(\|\Sigma(P_v)^{1/2}(\hat{\beta}(S) - \beta(P_v))\| \geq t\right)\\
&\asymp t^2\frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}P_v\left(\|\hat{\beta}(S) - \beta(P_v)\| \geq t\right)\\
&\gtrsim t^2\frac{e^{d/2}\cdot\left(\frac{1}{2}e^{-\varepsilon\lceil n\alpha\rceil} - \frac{\delta}{1-e^{-\varepsilon}}\right)}{1 + e^{d/2}e^{-\varepsilon\lceil n\alpha\rceil}}\ ,
\end{aligned}
$$

where $\beta(P)$ is the least squares solution of the distribution P, $\Sigma(P)$ is the covariance of x from P, and the last inequality follows from the fact that $d \geq 2$. The rest of the proof follows from (Barber and Duchi, 2014, Proposition 4). We choose

$$
\alpha = \frac{1}{n\varepsilon}\min\left\{\frac{d}{2} - \varepsilon, \log\left(\frac{1 - e^{-\varepsilon}}{4\delta e^\varepsilon}\right)\right\}
$$

and $t = \Omega(\alpha^{1-2/k})$ for $\varepsilon \in (0, 10)$ so that

$$
\sup_{P\in\mathcal{P}}\mathbb{E}_{P^n}[\|\Sigma(P)(\hat{\beta}(S) - \beta(P))\|^2] \gtrsim \alpha^{2-4/k}\ .
$$

This means that for all $k \geq 4$, there exist some $\kappa, \gamma = O(1)$ such that

$$
\inf_{\hat{\beta}\in\mathcal{M}_{\varepsilon,\delta}}\sup_{\Sigma\succ 0, P\in\mathcal{P}_{\kappa,k,\Sigma,\gamma^2}}\mathbb{E}_{P^n}[\|\Sigma^{1/2}(\hat{\beta}(S) - \beta(P))\|^2] \gtrsim \min\left\{\left(\frac{d\wedge\log(1 - e^{-\varepsilon}/\delta)}{n\varepsilon}\right)^{2-4/k}, 1\right\}\ ,
$$

which completes the proof by noting that $\gamma = \Theta(1)$. $\blacksquare$

Appendix D. Covariance estimation

In a standard covariance estimation, we are given i.i.d. samples $S = \{x_i \in \mathbb{R}^d\}_{i \in [n]}$ drawn from a distribution $P_{\Sigma, \Psi}$ with zero mean, an unknown covariance matrix $0 \prec \Sigma \in \mathbb{R}^{d \times d}$, and an unknown positive semidefinite matrix $\Psi := \mathbb{E}[(x_i \otimes x_i - \Sigma^\flat)(x_i \otimes x_i - \Sigma^\flat)^\top] \in \mathbb{R}^{d^2 \times d^2}$, where $\otimes$ denotes the Kronecker product. We treat the fourth moment matrix Ψ as a linear operator on a subspace $\mathcal{S}_{\text{sym}} \subset \mathbb{R}^{d^2}$, defined as $\mathcal{S}_{\text{sym}} := \{M^\flat \in \mathbb{R}^{d^2} : M \text{ is symmetric}\}$ following the definitions and notations from (Diakonikolas et al., 2018).

Definition 38 *For any matrix $M \in \mathbb{R}^{d \times d}$, let $M^\flat \in \mathbb{R}^{d^2}$ denote its canonical flattening into a vector in $\mathbb{R}^{d^2}$, and for any vector $v \in \mathbb{R}^{d^2}$, let $v^\sharp$ denote the unique matrix $M \in \mathbb{R}^{d \times d}$ such that $M^\flat = v$.*

This definition of Ψ as an operator on $\mathcal{S}_{\text{sym}}$ is without loss of generality since here we apply Ψ only to flattened symmetric matrices, which significantly lightens the notations, for example, for Gaussian distributions. We consider all $d^2 \times d^2$ matrices in this section to be linear operators on $\mathcal{S}_{\text{sym}}$, and we restrict our support of the exponential mechanism in RELEASE to be the set of positive definite matrices $\{\hat{\Sigma} \in \mathbb{R}^{d \times d} : \hat{\Sigma} \succ 0\}$.

Lemma 39 *((Diakonikolas et al., 2018, Theorem 4.12)) If $P_{\Sigma, \Psi} = \mathcal{N}(0, \Sigma)$, then $\mathbb{E}[x_i \otimes x_i] = \Sigma^\flat$, and, as a matrix in $\mathbb{R}^{d^2 \times d^2}$, we have $\Psi_{n(i-1)+j, n(k-1)+\ell} = \Sigma_{i,k}\Sigma_{j,\ell} + \Sigma_{i,\ell}\Sigma_{j,k}$ for all $(i, j, k, \ell) \in [d]^4$; as an operator on $\mathcal{S}_{\text{sym}}$, we can equivalently write it as $\Psi = 2(\Sigma \otimes \Sigma)$.*

Further, we can assume an invertible operator Ψ and define the Mahalanobis distance for $x_i \otimes x_i$, which is $D_\Psi(\hat{\Sigma}, \Sigma) = \|\Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat)\|$. For Gaussian distributions, for example, we have $D_\Psi(\hat{\Sigma}, \Sigma) = (1/\sqrt{2})\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. This is a natural choice of a distance because the total variation distance between two Gaussian distributions is $d_{\text{TV}}(\mathcal{N}(0, \Sigma), \mathcal{N}(0, \Sigma')) = O(\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F)$ (see, for example, (Kamath et al., 2019, Lemma 2.9)). We want a DP estimate of the covariance Σ with a small Mahalanobis distance $D_\Psi(\hat{\Sigma}, \Sigma)$. If the sample-generating distribution is not zero-mean, we can either apply a robust mean estimation with a subset of samples to estimate the mean or estimate the covariance using zero mean samples of the form $\{x_i - x_{i+\lceil n/2 \rceil}\}_{i \in [n/2]}$.

D.1. Step 1: Designing the surrogate $D_S(\hat{\Sigma})$ for the Mahalanobis distance

To sample only positive definite matrices, we restrict the domain of our score function to be $D_\Sigma : \{\hat{\Sigma} \in \mathbb{R}^{d \times d} : \hat{\Sigma} \succ 0\} \to \mathbb{R}_+$ and assume $D_\Sigma(\hat{\Sigma}) = \infty$ for non positive definite $\hat{\Sigma}$:

$$D_S(\hat{\Sigma}) = \max_{V \in \mathbb{R}^{d \times d} : V^\top = V, \|V\|_F = 1} \frac{\langle V, \hat{\Sigma} \rangle - \Sigma_V(\mathcal{M}_{V,\alpha})}{\psi_V(\mathcal{M}_{V,\alpha})} , \tag{58}$$

where we define the set $\mathcal{M}_{V,\alpha}$ similarly to the definition in Section B.1. We consider a projected dataset $\{\langle V, x_i x_i^\top \rangle\}_{i \in S}$ and partition S into three sets, $\mathcal{B}_{V,\alpha}$, $\mathcal{M}_{V,\alpha}$ and $\mathcal{T}_{V,\alpha}$, where $\mathcal{B}_{V,\alpha}$ corresponds to the subset of $(2/5.5)\alpha n$ data points with smallest values in $\{\langle V, x_i x_i^\top \rangle\}_{i \in S}$, $\mathcal{T}_{V,\alpha}$ is the subset of top $(2/5.5)\alpha n$ data points with the largest values, and $\mathcal{M}_{V,\alpha}$ is the subset of remaining $1 - (4/5.5)\alpha n$ data points. For a fixed symmetric matrix $V \in \mathbb{R}^{d \times d}$ with $\|V\|_F = 1$, we define $\Sigma_V(\mathcal{M}_{V,\alpha}) = \frac{1}{|\mathcal{M}_{V,\alpha}|} \sum_{x_i \in \mathcal{M}_{V,\alpha}} \langle V, x_i x_i^\top \rangle$ and $\psi_V(\mathcal{M}_{V,\alpha})^2 = \frac{1}{|\mathcal{M}_{V,\alpha}|} \sum_{x_i \in \mathcal{M}_{V,\alpha}} (\langle V, x_i x_i^\top \rangle - \Sigma_V(\mathcal{M}_{V,\alpha}))^2$,

which are robust estimates of the population projected covariance $\Sigma_V = \langle V, \Sigma \rangle$ and projected fourth moment $\psi_V^2 = (V^\flat)^\top \Psi V^\flat$. Next, we show that this score function $D_S(\hat{\Sigma})$ recovers our target error metric $D_\Psi(\hat{\Sigma}, \Sigma) = \|\Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat)\|$ when we substitute $\Sigma_V(\mathcal{M}_{V,\alpha})$ and $\psi_V(\mathcal{M}_{V,\alpha})$ with population statistics Σ_V and ψ_V, respectively. This justifies the choice of $D_S(\hat{\Sigma})$, as discussed in Section B.1.

Lemma 40 *For any $0 \prec \Sigma \in \mathbb{R}^{d \times d}$, $0 \prec \hat{\Sigma}$ and any invertible linear operator $\Psi \in \mathbb{R}^{d^2 \times d^2}$ on $\mathcal{S}_{\mathrm{sym}}$, we have*

$$\max_{V \in \mathbb{R}^{d \times d}: V^\top = V, \|V\|_F = 1} \frac{\langle V, \hat{\Sigma} \rangle - \Sigma_V}{\psi_V} = \left\| \Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat) \right\| , \tag{59}$$

where $\Sigma_V = \langle V, \Sigma \rangle$ and $\psi_V^2 = (V^\flat)^\top \Psi V^\flat$.

This follows immediately from Lemma 7.

D.2. Step 2: Utility analysis under resilience

The following resilience property of the dataset is critical in selecting Δ and τ and analyzing utility.

Definition 41 (Resilience) *For some $\alpha \in (0, 1)$, $\rho_1 \in \mathbb{R}_+$, and $\rho_2 \in \mathbb{R}_+$, we say a set of n data points S_{good} is (α, ρ_1, ρ_2)-resilient with respect to (Σ, Ψ) if for any $T \subset S_{\mathrm{good}}$ of size $|T| \geq (1 - \alpha)n$, the following holds for all symmetric matrices $V \in \mathbb{R}^{d \times d}$ with $\|V\|_F = 1$:*

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \left\langle V, x_i x_i^\top \right\rangle - \langle V, \Sigma \rangle \right| \leq \rho_1 \psi_V , \text{ and} \tag{60}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \left(\left\langle V, x_i x_i^\top \right\rangle - \langle V, \Sigma \rangle \right)^2 - \psi_V^2 \right| \leq \rho_2 \psi_V . \tag{61}$$

Note that covariance estimation for $\{x_i\}$ is equivalent to mean estimation for $\{x_i \otimes x_i\}$. We can immediately apply the mean estimation utility guarantee in Theorem 9 to show that $\|\Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat)\| = O(\rho_1)$ can be achieved with $n = O(d^2/\varepsilon\alpha)$ samples.

Corollary D.1 (Corollary of Theorem 9) *There exist positive constants c and $C > 0$ such that for any (α, ρ_1, ρ_2)-resilient dataset S with respect to (Σ, Ψ) satisfying $\alpha < c$, $\rho_1 < c$ and $\rho_2 < c$, and $\rho_1^2 \leq c\alpha$, HPTR with the distance function in Eq. (58), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $\|\Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat)\| \leq 32\rho_1$ with probability $1 - \zeta$ if*

$$n \geq C \frac{d^2 + \log(1/(\delta\zeta))}{\varepsilon\alpha} . \tag{62}$$

Under Assumption 1 on $\alpha_{\mathrm{corrupt}}$-corruption and Definition 9 on corrupt good sets extended to $\{x_i \otimes x_i\}_{i=1}^n$, it follows from Theorem 10 that the same guarantee holds under an adversarial corruption.

Corollary D.2 (Corollary of Theorem 10) *There exist positive constants c and $C > 0$ such that for any $((1/11)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S with respect to (Σ, Ψ) satisfying $\alpha < c$, $\rho_1 < c$ and $\rho_2 < c$, and $\rho_1^2 \leq c\alpha$, HPTR with the distance function in Eq. (58), $\Delta = 110\rho_1/(\alpha n)$, and $\tau = 42\rho_1$ achieves $\|\Psi^{-1/2}(\hat{\Sigma}^\flat - \Sigma^\flat)\| \leq 32\rho_1$ with probability $1 - \zeta$ if*

$$n \geq C \frac{d^2 + \log(1/(\delta\zeta))}{\varepsilon\alpha} . \tag{63}$$

D.3. Step 3: Near-optimal guarantees

Covariance estimation has been studied for Gaussian distributions under differential privacy (Karwa and Vadhan, 2017; Kamath et al., 2019; Aden-Ali et al., 2020) and robust estimation under α-corruption (Li and Ye, 2020; Diakonikolas et al., 2019a; Chen et al., 2018; Rousseeuw, 1985; Zhu et al., 2019). Note that from Lemma 39, we know that $\Psi = 2(\Sigma \otimes \Sigma)$ and the Mahalanobis distance simplifies to $D_\Psi(\hat{\Sigma}, \Sigma) = \|\Sigma^{1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F$ for Gaussian distributions.

D.3.1. GAUSSIAN DISTRIBUTIONS

For Gaussian distributions, the second moment resilience in Eq. (60) is satisfied with $\rho_1 = O(\alpha \log(1/\alpha))$, and the 4th moment resilience in Eq. (61) is satisfied with $\rho_2 = O(\alpha \log^2(1/\alpha))$.

Lemma 42 (Resilience for Gaussian) *Consider a dataset $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ of n i.i.d. samples from $\mathcal{N}(0, \Sigma)$. If $n = \Omega\left((d^2 + \log(1/\zeta))/(\alpha^2 \log(1/\alpha))\right)$ with a large enough constant, then there exists a constant $C > 0$ such that S is $(\alpha, C\alpha \log(1/\alpha), C\alpha \log^2(1/\alpha))$-corrupt good with respect to $(\Sigma, \Psi = 2\Sigma \otimes \Sigma)$ with probability $1 - \zeta$.*

Proof Since x is Gaussian, by Lemma 39, we have $\Psi = \mathbb{E}[(x \otimes x - \Sigma^\flat)(x \otimes x - \Sigma^\flat)^\top] = 2\Sigma \otimes \Sigma$. We can write $\psi_V^2 = 2\operatorname{Tr}(V^\top \Sigma V \Sigma) = 2\langle V, \Sigma V \Sigma \rangle$.

Lemma 43 ((Li and Ye, 2020, Lemma B.1) and (Dong et al., 2019, Fact 4.2)) *Let $\delta > 0$ and $\alpha \in (0, 0.5)$. A dataset $S = \{x_1, x_2, \cdots, x_n\}$ consists of n i.i.d. samples from $\mathcal{N}(0, \mathbf{I}_{d \times d})$. If $n = \Omega\left((d^2 + \log(1/\zeta))/(\alpha^2 \log(1/\alpha))\right)$ with a large enough constant, then there exists a universal constant $C_1 > 0$ and $C_2 > 0$ such that with probability $1 - \zeta$, for any subset $T \subset S$ and $|T| \geq (1 - \alpha)n$, we have*

$$\left\| \frac{1}{|T|} \sum_{x_i \in T} x_i \otimes x_i - \mathbf{I}_{d \times d}^\flat \right\| \leq C_1 \alpha \log(1/\alpha) \text{ , and}$$

$$\left\| \frac{1}{|T|} \sum_{x_i \in T} \left(x_i \otimes x_i - \mathbf{I}_{d \times d}^\flat\right)\left(x_i \otimes x_i - \mathbf{I}_{d \times d}^\flat\right)^\top - 2\mathbf{I}_{d \times d} \otimes \mathbf{I}_{d \times d} \right\| \leq C_2 \alpha \log(1/\alpha)^2 \text{ .}$$

By Lemma 43, we know with probability $1 - \zeta$ that for any subset $T \subset S$ and $|T| \geq (1 - \alpha)n$, we have

$$\left\| \frac{1}{|T|} \sum_{x_i \in T} (\Sigma^{-1/2}x_i) \otimes (\Sigma^{-1/2}x_i) - \mathbf{I}_{d \times d}^\flat \right\| \leq C_1 \alpha \log(1/\alpha) \text{ .}$$

This is equivalent to

$$\left| (V^\flat)^\top \frac{1}{|T|} \sum_{x_i \in T} (\Sigma^{-1/2} \otimes \Sigma^{-1/2})(x_i \otimes x_i) - (V^\flat)^\top \mathbf{I}_{d \times d}^\flat \right| \leq C_1 \alpha \log(1/\alpha) \text{ ,}$$

for any $\|V\|_F = 1$. This implies that

$$\left| (V^\flat)^\top \frac{1}{|T|} \sum_{x_i \in T} (x_i \otimes x_i) - (V^\flat)^\top (\Sigma \otimes \Sigma)^{1/2} \mathbf{I}_{d \times d}^\flat \right| \leq C_1 \alpha \log(1/\alpha) \sqrt{(V^\flat)^\top (\Sigma \otimes \Sigma) V^\flat} \text{ ,}$$

which is also equivalent to, for some constant C,

$$\left| \left\langle V, \frac{1}{|T|} \sum_{x_i \in T} x_i x_i^\top \right\rangle - \langle V, \Sigma \rangle \right| \leq C\alpha \log(1/\alpha) \sqrt{2 \langle V, \Sigma V \Sigma \rangle} \,,$$

which proves the first resilience Eq. (60) in Definition 41.

Similarly, by Lemma 43, we have

$$\left\| \frac{1}{|T|} \sum_{x_i \in T} \left(\Sigma^{-1/2} x_i \otimes \Sigma^{-1/2} x_i - \mathbf{I}_{d\times d}^\flat \right) \left(\Sigma^{-1/2} x_i \otimes \Sigma^{-1/2} x_i - \mathbf{I}_{d\times d}^\flat \right)^\top - 2\mathbf{I}_{d\times d} \otimes \mathbf{I}_{d\times d} \right\| \leq C_2 \alpha \log(1/\alpha)^2 \,.$$

This is equivalent, for any $\|V\|_F = 1$, to

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \left\langle V^\flat, \Sigma^{-1/2} x_i \otimes \Sigma^{-1/2} x_i - \mathbf{I}_{d\times d}^\flat \right\rangle^2 - 2 \right| \leq C_2 \alpha \log(1/\alpha)^2 \,.$$

This implies

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \left\langle V^\flat, x_i \otimes x_i - \Sigma^\flat \right\rangle^2 - 2(V^\flat)^\top (\Sigma \otimes \Sigma) V^\flat \right| \leq C_2 \alpha \log(1/\alpha)^2 \langle V, \Sigma V \Sigma \rangle \,,$$

which is also equivalent, for some constant C, to

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \left(\left\langle V, x_i x_i^\top \right\rangle - \langle V, \Sigma \rangle \right)^2 - 2 \operatorname{Tr}(V^\top \Sigma V \Sigma) \right| \leq 2C\alpha \log(1/\alpha)^2 \langle V, \Sigma V \Sigma \rangle \,,$$

which proves the second resilience Eq. (61) in Definition 41.

∎

The second and fourth moment resilience properties of Gaussian distributions in Lemma 42, together with the utility analysis of HPTR in Corollary. D.2, imply the following utility guarantee.

Corollary D.3 *Under the hypotheses of Lemma 42, there exists a constant $c > 0$ such that for any $\alpha \in (0, c)$, a dataset of size*

$$n = O\left(\frac{d^2 + \log(1/\zeta)}{\alpha^2 \log(1/\alpha)} + \frac{d^2 + \log(1/(\delta\zeta))}{\alpha\varepsilon} \right) \,,$$

a sensitivity of $\Delta = O(\log(1/\alpha)/n)$, and a threshold $\tau = O(\alpha \log(1/\alpha))$ with large enough constants are sufficient for $\mathrm{HPTR}(S)$ with a choice of distance function in Eq. (58) to achieve

$$\|\Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} - \mathbf{I}_{d\times d}\|_F = O(\alpha \log(1/\alpha)) \,, \tag{64}$$

with probability $1 - \zeta$. Further, the same guarantee holds even if an α-fraction of the samples is arbitrarily corrupted, as in Assumption 1.

This Mahalanobis distance guarantee (for the Kronecker product, $\{x_i \otimes x_i\}$, of the samples) implies that the predicted Gaussian distribution is close to the sample generating one in total variation distance (see, for example, (Kamath et al., 2019, Lemma 2.9)) $d_{\text{TV}}(\mathcal{N}(0, \hat{\Sigma}), \mathcal{N}(0, \Sigma)) = O(\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\|_F) = O(\alpha \log(1/\alpha))$. This relation also implies that the error bound is near-optimal under α-corruption, matching a lower bound up to a factor of $O(\log(1/\alpha))$. Even if DP is not required and we are given infinite samples, an adversary can move an α fraction of the probability mass to switch a Gaussian distribution into another one at Mahalanobis distance $\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbf{I}_{d \times d}\|_F = \Omega(\alpha)$. Hence, we cannot tell which of the two distributions the (potentially infinite) samples came from.

The sample complexity is near-optimal, matching a lower bound up to a factor of $O(\log(1/\alpha))$ when $\delta = e^{-\Theta(d^2)}$. For a constant ζ, HPTR requires $n = O(d^2/(\alpha^2 \log(1/\alpha)) + d^2/(\alpha\varepsilon) + \log(1/\delta)/(\alpha\varepsilon))$. This nearly matches a lower bound (that holds even if there is no corruption) on n to achieve the guarantee of Eq. (64) of $n = \Omega(d^2/(\alpha \log(1/\alpha))^2 + \min\{d^2, \log(1/\delta)\}/(\varepsilon\alpha \log(1/\alpha)) + \log(1/\delta)/\varepsilon)$. The first term follows from the classical estimation of the covariance without DP and matches the first term in our upper bound up to a $O(\log(1/\alpha))$ factor. The second term follows from extending the lower bound in (Kamath et al., 2019), constructed for pure differential privacy with $\delta = 0$, and matches the second term in our upper bound up to a $O(\log(1/\alpha))$ factor when $\delta = e^{-\Theta(d^2)}$. The last term, from (Karwa and Vadhan, 2017), has a gap of $O(1/\alpha)$ factor compared to the third term in our upper bound, but this term is typically not dominating when δ is large enough, i.e., $\delta = e^{-O(d^2)}$. We note that a slightly tighter upper bound is achieved by the state-of-the-art algorithm in (Aden-Ali et al., 2020) that requires only $O(d^2/(\alpha \log(1/\alpha))^2 + d^2/(\varepsilon\alpha \log(1/\alpha)) + \log(1/\delta)/\varepsilon)$. The state-of-the-art polynomial time algorithm in Ashtiani and Liaw (2021) requires no assumptions on Σ and achieves a similar optimal guarantee as HPTR. This improves upon a previous polynomial time algorithm of (Kamath et al., 2021) that also requires no assumptions on Σ but has a larger sample complexity: $n = \tilde{O}(d^2/(\alpha \log(1/\alpha))^2 + d^2 \text{polylog}(1/\delta)/(\varepsilon\alpha \log(1/\alpha)) + d^{5/2} \text{polylog}(1/\delta)/\varepsilon)$.

If privacy is not an issue (i.e., $\varepsilon = \infty$), HPTR achieves the error in Eq. (64) with $n = O(d^2/\alpha^2 \log(1/\alpha))$ samples. There are polynomial time estimators that achieve the same guarantee (Li and Ye, 2020; Diakonikolas et al., 2019a). The gap of $\log(1/\alpha)$ to the lower bound in the error can be tightened using algorithms that are not computationally efficient, as shown in (Chen et al., 2018; Rousseeuw, 1985).

Remark. When we have a sample size of only $n = O(d/\alpha^2)$, our analysis provides no guarantees. However, for robust covariance estimation under α-corruption, one can still guarantee a bound on a weaker error metric in the spectral norm: $\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - \mathbf{I}_{d \times d}\| = O(\alpha \log(1/\alpha))$ (Zhu et al., 2019, Theorem 3.4). There is no corresponding DP covariance estimator in that small sample regime. A promising direction is to apply the HPTR framework, but it remains challenging to design a score function for this spectral norm distance that depends only on one-dimensional robust statistics.

Appendix E. Principal component analysis

In Principal Component Analysis (PCA), we are given i.i.d. samples $S = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ drawn from a zero mean distribution P_Σ with an unknown covariance matrix Σ. We want to find a top eigenvector of Σ, $u \in \arg\max_{\|v\|=1} v^\top \Sigma v$, privately. The performance of our estimate $\hat{u}$ is measured by how much of the covariance is captured in the direction $\hat{u}$ relative to that of u,

$D_\Sigma(\hat{u}) = 1 - (\hat{u}^\top \Sigma \hat{u} / u^\top \Sigma u)$, where u is one of the top eigenvectors of Σ. When the mean is not zero, this can be handled similarly to covariance estimation in Section D.

E.1. Step 1: Designing the surrogate score function $D_S(\hat{u})$

It is straightforward to design a score function of $D_S : \mathbb{S}^{(d-1)} \to \mathbb{R}_+$, where $\mathbb{S}^{(d-1)}$ is the unit sphere in $\mathbb{R}^d$

$$D_S(\hat{u}) \;=\; 1 - \frac{\hat{u}^\top \Sigma(\mathcal{M}_{\hat{u},\alpha}) \hat{u}}{\max_{v \in \mathbb{R}^d : \|v\|=1} v^\top \Sigma(\mathcal{M}_{v,\alpha}) v} \; , \tag{65}$$

where $\mathcal{M}_{\hat{u},\alpha} \subset S$ is the subset of data points corresponding to the smallest $(1 - (2/3.5)\alpha)n$ values in the projected set $S_{\hat{u}} = \{\langle \hat{u}, x_i \rangle^2\}_{x_i \in S}$ and $\Sigma(\mathcal{M}_{\hat{u},\alpha}) = (1/|\mathcal{M}_{\hat{u},\alpha}|) \sum_{x_i \in \mathcal{M}_{\hat{u},\alpha}} x_i x_i^\top$. Note that when we replace $\Sigma(\mathcal{M}_{\hat{u},\alpha})$ with the population covariance matrix Σ, we recover the target error metric of $D_\Sigma(\hat{u}) = 1 - (\hat{u}^\top \Sigma \hat{u} / \max_{\|v\|=1} v^\top \Sigma v)$. For this choice of $D_S(\hat{u})$, the support of the exponential mechanism is already compact, and we do not restrict it any further, say, to be in $B_{\tau,S}$. This simplifies the HPTR algorithm and also the analysis, as follows. We define

$$\mathsf{UNSAFE}_\varepsilon \;=\; \Big\{ S' \subset \mathbb{R}^{d \times n} \,|\, \exists S'' \sim S' \text{ and } \exists E \text{ such that } \mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S'')}}(\hat{u} \in E) > e^\varepsilon \mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S')}}(\hat{u} \in E)$$
$$\text{or } \mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S')}}(\hat{u} \in E) > e^\varepsilon \mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S'')}}(\hat{u} \in E) \Big\} \, .$$

Note that since the support is the same for all S, we can achieve a stronger pure DP with $\delta = 0$ in the exponential mechanism. However, we still need $\delta > 0$ in the TEST step. HPTR for PCA proceeds as follows.

1. PROPOSE: Propose a target sensitivity bound $\Delta = 80\rho_2/(\alpha n)$.

2. TEST:

   2.1. Compute the safety margin $m = \min_{S'} d_H(S, S')$ such that $S' \in \mathsf{UNSAFE}_{\varepsilon/2}$.

   2.2. If $\hat{m} = m + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon) \log(2/\delta)$, then output $\perp$; otherwise, continue.

3. RELEASE: Output $\hat{u}$ sampled from a distribution with a pdf:

$$r_{(\varepsilon,\Delta,S)}(\hat{u}) \;=\; \frac{1}{Z} \exp\left(-\frac{\varepsilon}{4\Delta} D_S(\hat{u})\right) \, ,$$

from $\mathbb{S}^{(d-1)} = \{\hat{u} \in \mathbb{R}^d : \|\hat{u}\| = 1\}$ where $Z = \int_{\mathbb{S}^{(d-1)}} \exp\{-(\varepsilon D_S(\hat{u}))/(4\Delta)\} \, d\hat{u}$.

The choice of ρ_2 depends on the hypothesis on the tail of the sample-generating distribution, and α depends on the target accuracy as guided by Theorem 13 (or the fraction of adversarial corruption in the case of the outlier robust PCA setting in Theorem 14). The target privacy guarantee determines (ε, δ).

E.2. Step 2: Utility analysis under resilience

The following resilience properties are critical in selecting the sensitivity Δ and in analyzing the utility.

Definition 44 (Resilience for PCA) *For some $\rho_1 \in \mathbb{R}_+, \rho_2 \in \mathbb{R}_+$, we say a set of n data points $S_{\text{good}} = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ is (α, ρ_1, ρ_2)-resilient with respect to Σ for some positive semidefinite $\Sigma \in \mathbb{R}^{d \times d}$ if for any $T \subset S_{\text{good}}$ of size $|T| \geq (1 - \alpha)n$, the following holds for all $v \in \mathbb{R}^d$ with $\|v\| = 1$:*

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i \rangle \right| \quad \leq \quad \rho_1 \, \sigma_v \text{ and} \tag{66}$$

$$\left| \frac{1}{|T|} \sum_{x_i \in T} \langle v, x_i \rangle^2 - \sigma_v^2 \right| \quad \leq \quad \rho_2 \, \sigma_v^2 \, . \tag{67}$$

where $\sigma_v^2 = v^\top \Sigma v$.

We refer to Section B.2 for the explanation of how resilience is fundamentally connected to sensitivity. For an example of a Gaussian distribution, the samples are $(\alpha, O(\alpha \sqrt{\log(1/\alpha)}), O(\alpha \log(1/\alpha)))$-resilient (with a large enough n). We show next how resilience implies an error bound for HPTR, which is $O(\alpha \log(1/\alpha))$ for Gaussian distributions.

Theorem 13 *There exist positive constants c and C such that for any (α, ρ_1, ρ_2)-resilient set S with respect to some Σ and satisfying $\alpha < \rho_2 < c$, HPTR Section E.1 for PCA with the choices of the distance function in Eq. (65) and $\Delta = 80\rho_2/(\alpha n)$ achieves $1 - (\hat{u}^\top \Sigma \hat{u} / \|\Sigma\|) \leq 20\rho_2$ with probability $1 - \zeta$ if*

$$n \; \geq \; C \left(\frac{\log(1/(\delta\zeta)) + d \log(1/\rho_2)}{\varepsilon\alpha} \right) \, . \tag{68}$$

We discuss the implications of this result in Section E.3 for specific instances of the problem. Under Assumption 1 on α_{corrupt}-corruption of the data and Definition 9 on the corrupt good sets, we show that HPTR is also robust against corruption.

Theorem 14 *There exist positive constants c and C such that for any $((2/7)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S with respect to some Σ satisfying $\alpha < rho_2 < c$, HPTR in Section E.1 for PCA with the choices of the distance function in Eq. (65) and $\Delta = 80\rho_2/(\alpha n)$ achieves $1 - (\hat{u}^\top \Sigma \hat{u} / \|\Sigma\|) \leq 20\rho_2$ with probability $1 - \zeta$ if*

$$n \; \geq \; C \left(\frac{\log(1/(\delta\zeta)) + d \log(1/\rho_2)}{\varepsilon\alpha} \right) \, . \tag{69}$$

We provide a proof of the robust and DP PCA in Section E.2.2, where Theorem 13 follows immediately by selecting α as a free parameter. As the HPTR Section E.1 for PCA is significantly simpler, we do not apply the general analysis in Theorem 15; instead, we prove the preceding theorem directly. To this end, we first show a bound on sensitivity and next show that the safety test succeeds with high probability in Section E.2.1.

E.2.1. RESILIENCE IMPLIES BOUNDED LOCAL SENSITIVITY

Given the resilience properties of a corrupt good set S, we show that the sensitivity of $D_S(\hat{u})$ is bounded by Δ.

Lemma 45 *Suppose $\alpha \leq c$ for some small enough constant c. For $\Delta = 80\rho_2/(\alpha n)$ and a $((2/7)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good S, if*

$$n = \Omega\Big(\frac{\log(1/(\delta\zeta))}{\alpha\varepsilon}\Big) ,$$

with a large enough constant, then for all S' within a Hamming distance $k^* = (2/\varepsilon)\log(4/(\zeta\delta))$ from S, we have

$$\max_{S'' \sim S'} |D_{S''}(\hat{u}) - D_{S'}(\hat{u})| \leq \Delta , \tag{70}$$

for all unit vectors $\hat{u}$ and all neighboring datasets S''.

Proof The proof is similar to the proof of Lemma 16. We first assume $(k^* + 1)/n \leq \alpha/7$, which requires $n = \Omega(\log(1/\delta\zeta))/(\alpha\varepsilon)$ with a large enough constant. This implies that S' is a $((3/7)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set. The rest of this proof uses this assumption. Let $\mathcal{T}_{\hat{u},\alpha}(S') \subset S$ be the subset of data points corresponding to the largest $(2/3.5)\alpha n$ values in the projected set $S'_{\hat{u}} = \{\langle \hat{u}, x_i \rangle^2\}_{x_i \in S'}$. Recall that S_{good} is the original resilient dataset before corruption by an adversary. From Lemma 10 and the fact that $|S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}(S')| \geq (1/7)\alpha n$, it follows that $(1/|S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}(S')|) \sum_{x_i \in S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}} \langle \hat{u}, x_i \rangle^2 \leq (1 + (2\rho_2)/((1/7)\alpha))\sigma_{\hat{u}}^2$, where $\sigma_{\hat{u}} = \sqrt{\hat{u}^\top \Sigma \hat{u}}$. This implies that

$$\min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}} \langle \hat{u}, x_i \rangle^2 \leq \Big(1 + \frac{2\rho_2}{(1/7)\alpha}\Big)\sigma_{\hat{u}}^2 . \tag{71}$$

Let $\mathcal{M}_{\hat{u},\alpha}(S')$ be the remaining subset of S', with $(1-(2/3.5)\alpha)n$ smallest values in $\{(\langle \hat{u}, x_i \rangle)^2\}_{i \in [n]}$. $\mathcal{M}_{\hat{u},\alpha}(S')$ and $\mathcal{M}_{\hat{u},\alpha}(S'')$ can differ by at most one data point. Let x' and x'' be the unique pair of data points that are in $\mathcal{M}_{\hat{u},\alpha}(S')$ and $\mathcal{M}_{\hat{u},\alpha}(S'')$, respectively. If there is no such pair, then the two filtered subsets are the same, and the following claims are trivially true.

If $\langle \hat{u}, x'' \rangle^2 \leq \max_{x_i \in \mathcal{M}_{\hat{u},\alpha}(S')} \langle \hat{u}, x_i \rangle^2 \leq \min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}(S')} \langle \hat{u}, x_i \rangle^2$, we have $|\langle \hat{u}, x' \rangle^2 - \langle \hat{u}, x'' \rangle^2| \leq (1 + 14\rho_2/\alpha)\sigma_{\hat{u}}^2$, where $\sigma_{\hat{u}}^2 = \hat{u}^\top \Sigma \hat{u}$. If $\langle \hat{u}, x'' \rangle^2 > \max_{x_i \in \mathcal{M}_{\hat{u},\alpha}(S')} \langle \hat{u}, x_i \rangle^2$, then x'' is at most $\langle \hat{u}, x'' \rangle^2 \leq \min_{x_i \in S_{\text{good}} \cap \mathcal{T}_{\hat{u},\alpha}(S')} \langle \hat{u}, x_i \rangle^2$, where equality holds if the smallest point in the top subset enters $\mathcal{M}_{\hat{u},\alpha}(S'')$. This also implies $|\langle \hat{u}, x' \rangle^2 - \langle \hat{u}, x'' \rangle^2| \leq (1 + 14\rho_2/\alpha)\sigma_{\hat{u}}^2$. Let $\sigma_v'^2 = v^\top \Sigma(\mathcal{M}_{v,\alpha}(S'))v$ and $\sigma_v''^2 = v^\top \Sigma(\mathcal{M}_{v,\alpha}(S''))v$. Then, for any $\|v\| = 1$,

$$
\begin{aligned}
|\sigma_v'^2 - \sigma_v''^2| &= \left| v^\top \Big(\frac{1}{(1 - (2/3.5)\alpha)n} \sum_{x_i \in \mathcal{M}_{v,2\alpha}(S')} x_i x_i^\top - \frac{1}{(1 - (2/3.5)\alpha)n} \sum_{x_i \in \mathcal{M}_{v,2\alpha}(S'')} x_i x_i^\top \Big) v \right| \\
&\leq \frac{2}{n}|\langle v, x' \rangle^2 - \langle v, x'' \rangle^2| \leq \frac{2}{n}\Big(1 + \frac{14\rho_2}{\alpha}\Big)v^\top \Sigma v ,
\end{aligned}
$$

for $\alpha \leq c$ small enough. Then, for the local sensitivity, we have

$$
\begin{aligned}
|D_{S'}(\hat{u}) - D_{S''}(\hat{u})| &\leq \left| \frac{\sigma_{\hat{u}}'^2 - \sigma_{\hat{u}}''^2}{\max_{\|v\|=1} \sigma_v'^2} \right| + \left| \frac{\sigma_{\hat{u}}''^2}{\max_{\|v\|=1} \sigma_v'^2} - \frac{\sigma_{\hat{u}}''^2}{\max_{\|v\|=1} \sigma_v''^2} \right| \\
&\leq \frac{2}{n}\Big(1 + \frac{14\rho_2}{\alpha}\Big)\frac{\hat{u}^\top \Sigma \hat{u}}{0.9\|\Sigma\|} + \frac{1.1\hat{u}^\top \Sigma \hat{u}}{0.9^2\|\Sigma\|^2}\frac{2}{n}\Big(1 + \frac{14\rho_2}{\alpha}\Big)\|\Sigma\| ,
\end{aligned}
$$

where we used the resilience in Eq. (67) with a small enough $\rho_2 \leq c$ such that $0.9 v^\top \Sigma v \leq \sigma_v'^2 \leq 1.1 v^\top \Sigma v$ and $0.9 v^\top \Sigma v \leq \sigma_v''^2 \leq 1.1 v^\top \Sigma v$ (which follow from Lemma 47). When $\rho_2 \leq \alpha$, this is bounded by $|D_{S'}(\hat{u}) - D_{S''}(\hat{u})| \leq 80 \rho_2 / (\alpha n) = \Delta$.

$\blacksquare$

Since the support is the same for all exponential mechanisms regardless of the dataset, the sensitivity bound immediately implies safety. The following lemma shows that we have a sufficient safety margin to succeed with probability of at least $1 - \zeta$ since $k^* = (2/\varepsilon) \log(4/(\delta\zeta))$ and the threshold is $(2/\varepsilon) \log(2/\delta)$.

Lemma 46 *Under the hypothesis of Lemma 45, for any S' at Hamming distance at most k^* from S, we have $S' \in \mathsf{SAFE}_{\varepsilon/2}$.*

E.2.2. PROOF OF THEOREM 14

This proof is similar to the proof of a universal utility analysis in Theorem 15. First, we show that we pass the safety test with high probability. By Lemma 46, we know $m > k^* = 2/\varepsilon \log(4/(\zeta\delta))$. Then, we have

$$\mathbb{P}\left(\text{output } \bot\right) = \mathbb{P}\left(m + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon) \log(2/\delta)\right) \ \leq \ \frac{\zeta}{2} \ .$$

Next, we assume the dataset passed the safety test and show that $\mathbb{P}_{\hat{u} \sim r_{(\varepsilon, \Delta, S)}}(\hat{u}^\top \Sigma \hat{u} \geq (1 - 4\rho_2)\|\Sigma\|) \geq 1 - \zeta/2$.

Lemma 47 *For an $((2/7)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S with respect to Σ, then $|\hat{u}^\top \Sigma \hat{u} - \hat{u}^\top \Sigma(\mathcal{M}_{\hat{u},\alpha})\hat{u}| \leq 4\rho_2 \hat{u}^\top \Sigma \hat{u}$.*

Proof

We have

$$
\begin{aligned}
|\hat{u}^\top \Sigma \hat{u} - \hat{u}^\top \Sigma(\mathcal{M}_{\hat{u},\alpha})\hat{u}| &= \frac{|\sum_{i \in \mathcal{M}_{\hat{u},\alpha}}(\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2)|}{(1 - (2/3.5)\alpha)n} \\
&\leq \frac{|\sum_{i \in \mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{good}}}(\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2)|}{(1 - (2/3.5)\alpha)n} + \frac{|\sum_{i \in \mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{good}}}(\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2)|}{(1 - (2/3.5)\alpha)n}
\end{aligned} \quad (72)
$$

For $i \in \mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{bad}}$, by Lemma 10, we have

$$
\begin{aligned}
|\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2| &\leq \max\left\{ \frac{\sum_{i \in \mathcal{T}_{\hat{u},\alpha} \cap S_{\mathrm{good}}}(\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2)}{|\mathcal{T}_{\hat{u},\alpha} \cap S_{\mathrm{good}}|}, \sigma_{\hat{u}}^2 \right\} \\
&\leq \frac{2\rho_2 \sigma_{\hat{u}}^2}{(1/3.5)\alpha} ,
\end{aligned} \quad (73)
$$

where in the last inequality, we applied our assumption that $\rho_2 \geq \alpha$.

By the resilience property in Eq. (67) on $\mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{good}}$, we also have

$$\frac{|\sum_{i \in \mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{good}}}(\langle \hat{u}, x_i \rangle^2 - \sigma_{\hat{u}}^2)|}{|\mathcal{M}_{\hat{u},\alpha} \cap S_{\mathrm{good}}|} \leq \rho_2 \sigma_{\hat{u}}^2 \ . \quad (74)$$

Plugging Eq. (73) and (74) into (72), we have

$$|\hat{u}^\top \Sigma \hat{u} - \hat{u}^\top \Sigma(\mathcal{M}_{\hat{u},\alpha})\hat{u}| \leq \frac{2\rho_2 \sigma_{\hat{u}}^2 + (1 - (2/3.5)\alpha)\rho_2 \sigma_{\hat{u}}^2}{1 - (2/3.5)\alpha} \leq 4\rho_2 \sigma_{\hat{u}}^2 \ ,$$

for $\alpha \leq c$ small enough. ∎

This implies that $|D_\Sigma(\hat{u}) - D_S(\hat{u})| \leq 4\rho_2$ for an $((2/7)\alpha, \alpha, \rho_1, \rho_2)$-corrupt good set S.

Let $\mu(\cdot)$ denote the uniform measure on the unit sphere. By the fact that for any $0 < r < 2$, a cap of radius r on the $(d-1)$-dimensional unit sphere $\mathbb{S}^{(d-1)}$ has a measure of at least $(1/2)(r/2)^{d-1}$ from, for example, (Kapralov and Talwar, 2013, Fact 3.1), we have for some constant $c_2 > 0$ and $\rho_2 \leq 1/8$,

$$\mu(\{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1 - 4\rho_2)\|\Sigma\|, \|v\| = 1\}) \geq \left(\cos^{-1}(1 - 4\rho_2)/2\right)^{d-1} \geq e^{-c_2 d \log(1/\rho_2)} \ . \quad (75)$$

By Lemma 47, the choice of $\Delta = 80\rho_2/(\alpha n)$, we have

$$\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\|\Sigma\| - \hat{u}^\top \Sigma \hat{u} \leq 4\rho_2\|\Sigma\|\right)$$

$$= \int_{\{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1-4\rho_2)\|\Sigma\|, \|v\|=1\}} r_{(\varepsilon,\Delta,S)}(\hat{u}) \, d\hat{\mu}$$

$$\geq \text{Vol}(\{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1 - 4\rho_2)\|\Sigma\|, \|v\| = 1\}) \min_{\hat{\mu} \in \{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1-4\rho_2)\|\Sigma\|, \|v\|=1\}} r_{(\varepsilon,\Delta,S)}(\hat{u})$$

$$\geq \text{Vol}(\mathbb{S}^{(d-1)}) \, \mu(\{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1 - 4\rho_2)\|\Sigma\|, \|v\| = 1\}) \min_{\hat{u} \in \{v \in \mathbb{R}^d : v^\top \Sigma v \geq (1-4\rho_2)\|\Sigma\|, \|v\|=1\}} r_{(\varepsilon,\Delta,S)}(\hat{u})$$

$$\geq \text{Vol}(\mathbb{S}^{(d-1)}) \, e^{-c_2 d \log(1/\rho_2)} \frac{1}{Z} \exp\left\{ - \frac{\varepsilon}{4\Delta} \max_{\|\hat{u}\|=1, 4\rho_2 \geq 1 - \frac{\hat{u}^\top \Sigma \hat{u}}{\|\Sigma\|}} 1 - \frac{\hat{u}^\top \Sigma(\mathcal{M}_{\hat{u},\alpha})\hat{u}}{\|\Sigma\|} \right\}$$

$$\geq \text{Vol}(\mathbb{S}^{(d-1)}) \, e^{-c_2 d \log(1/\rho_2)} \frac{1}{Z} \exp\left\{ - \frac{\alpha \varepsilon n}{40} \right\} \ ,$$

and similarly,

$$\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\|\Sigma\| - \hat{u}^\top \Sigma \hat{u} \geq 20\rho_2\|\Sigma\|\right) \leq \text{Vol}(\mathbb{S}^{(d-1)}) \max_{\hat{u} \in \{v \in \mathbb{R}^d : v^\top \Sigma v \leq (1-20\rho_2)\|\Sigma\|, \|v\|=1\}} r_{(\varepsilon,\Delta,S)}(\hat{u})$$

$$\leq \text{Vol}(\mathbb{S}^{(d-1)}) \frac{1}{Z} e^{-\varepsilon \alpha n(20\rho_2 - 4\rho_2)\|\Sigma\|/(320\rho_2\|\Sigma\|)}$$

$$\leq \text{Vol}(\mathbb{S}^{(d-1)}) \frac{1}{Z} \exp\left\{ - \frac{\alpha \varepsilon n}{20} \right\}$$

This implies that

$$\log\left(\frac{\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\lambda_1 - \hat{u}^\top \Sigma \hat{u} \leq 4\rho_2\|\Sigma\|\right)}{\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\lambda_1 - \hat{u}^\top \Sigma \hat{u} \geq 20\rho_2\|\Sigma\|\right)} \right) \geq \frac{\varepsilon \alpha n}{40} - c_2 d \log(1/\rho_2) \ .$$

If we set $n = \Omega\left(\frac{\log(1/\zeta) + d \log(1/\rho_2)}{\varepsilon \alpha} \right)$, we get

$$\frac{\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\lambda_1 - \hat{u}^\top \Sigma \hat{u} \leq 4\rho_2\lambda_1\right)}{\mathbb{P}_{\hat{u} \sim r_{(\varepsilon,\Delta,S)}} \left(\lambda_1 - \hat{u}^\top \Sigma \hat{u} \geq 20\rho_2\lambda_1\right)} \geq \frac{2}{\zeta} \ ,$$

which completes the proof.

E.3. Step 3: Achievability guarantees

We provide utility guarantees for a private PCA for sub-Gaussian and hypercontractive distributions.

E.3.1. SUB-GAUSSIAN DISTRIBUTIONS

Using the resilience of sub-Gaussian distributions with respect to $(\mu = 0, \Sigma)$ in Lemma 17, which is the same as the resilience properties we need for the PCA in Definition 44, Theorem 14 implies the following corollary.

Corollary E.1 *Under the hypothesis of Lemma 17 with $\mu = 0$ and any PSD matrix $\Sigma \in \mathbb{R}^{d \times d}$, there exist universal constants c and $C > 0$ such that for any $\alpha \in (0, c)$, a dataset of size*

$$n = O\left(\frac{d + \log(1/\zeta)}{(\alpha \log(1/\alpha))^2} + \frac{\log(1/(\delta\zeta)) + d\log(1/(\alpha \log(1/\alpha)))}{\varepsilon\alpha}\right),$$

and sensitivity of $\Delta = O(\log(1/\alpha)/n)$ with large enough constants are sufficient for $\mathrm{HPTR}(S)$ in Section E.1 for a PCA with the choices of the distance function in Eq. (65) to achieve

$$1 - \frac{\hat{u}^\top \Sigma \hat{u}}{\|\Sigma\|} \leq C\alpha \log(1/\alpha), \tag{76}$$

with probability $1 - \zeta$. Further, the same guarantee holds even if an α-fraction of the samples is arbitrarily corrupted, as in Assumption 1.

The error bound is near-optimal under α-corruption, matching a lower bound up to a factor of $O(\log(1/\alpha))$. HPTR is the first estimator that guarantees (ε, δ)-DP and also achieves the robust error rate of $1 - \hat{u}^\top \Sigma \hat{u}/\|\Sigma\| = O(\alpha \log(1/\alpha))$, nearly matching the information-theoretic lower bound of $1 - \hat{u}^\top \Sigma \hat{u}/\|\Sigma\| = \Omega(\alpha)$. This lower bound, which can be easily constructed using $\mathcal{N}(0, \mathbf{I} + \alpha e_1 e_1^\top)$ and $\mathcal{N}(0, \mathbf{I} + \alpha e_2 e_2^\top)$, holds for any estimator that is not necessarily private and regardless of how many samples are available. If privacy is not required, a near-optimal robust error rate can be achieved by outlier-robust PCA approaches in (Kong et al., 2020; Jambulapati et al., 2020).

The sample complexity is near-optimal, matching a lower bound up to a factor of $O(\log(1/\alpha))$ when $\delta = e^{-\Theta(d)}$. Even for a DP PCA without corrupted samples, HPTR is the first estimator for sub-Gaussian distributions to nearly match the information-theoretic lower bound of $n = \Omega(d/(\alpha \log(1/\alpha))^2 + \min\{d, \log((1 - e^{-\varepsilon})/\delta)\}/(\varepsilon\alpha \log(1/\alpha)))$ to achieve the error in Eq. (76). The first term is unavoidable since even without DP and robustness, when the data comes from a Gaussian distribution, estimating the principal component up to error $\alpha \log(1/\alpha)$ requires $\Omega(d/(\alpha \log(1/\alpha))^2)$ samples (Proposition E.3). The second term in the lower bound follows from Proposition E.2, which matches the second term in the upper bound up to a factor of $O(\log(1/\alpha))$ when $\delta = e^{-\Theta(d)}$ and $\varepsilon > 0$. Existing DP PCA approaches from (Chaudhuri et al., 2013; Kapralov and Talwar, 2013; Dwork et al., 2014) are designed for arbitrary samples not necessarily drawn i.i.d., and hence they require a larger sample size of $n = \tilde{O}(d/\alpha^2 + d^{1.5}\sqrt{\log(1/\delta)}/(\alpha\varepsilon))$ i.i.d. samples from a Gaussian distribution to achieve the guarantee in Eq. (76), where $\tilde{O}$ hides polylogarithmic terms in $1/\alpha$ and $1/\zeta$.

Remark. Rank-k PCA under α-corruption from a Gaussian dataset is of great practical interest. An outlier-robust PCA algorithm in (Kong et al., 2020, Appendix D) outputs an orthonormal matrix $\hat{U} \in \mathbb{R}^{d \times k}$ achieving

$$\text{Tr}(U_k^\top \Sigma U_k) - \text{Tr}(\hat{U}^\top \Sigma \hat{U}) \;=\; O\big(\alpha \text{Tr}(U_k^\top \Sigma U_k) + \nu k^{1/2} \alpha \log(1/\alpha)\big)\,,$$

where $U_k \in \arg\max_{U^\top U = \mathbf{I}_{k \times k}} U^\top \Sigma U$ and $\nu^2 = \max_{V \in \mathbb{R}^{d \times d}, \|V\|_F = 1, V = V^\top, \text{rank}(V) \le k} \langle V, \Sigma V \Sigma \rangle$. It is a promising direction to design a DP rank-k PCA algorithm by applying the HPTR framework that can achieve a similar error rate. It is not immediately clear how to design an appropriate score function for general rank k, and a simple technique of peeling off rank-one components one-by-one (using the rank-one PCA with HPTR) will not achieve the target error bound.

Proposition E.2 (Lower bound for private sub-Gaussian PCA) *Let $\mathcal{P}_\Sigma$ be the set of zero-mean sub-Gaussian distributions with covariance $\Sigma \in \mathbb{R}^{d \times d}$. Let $\mathcal{M}_{\varepsilon,\delta}$ be a class of (ε, δ)-DP, d-dimensional estimators of the top principal component of Σ using n i.i.d. samples from $P \in \mathcal{P}_\Sigma$. Then, for $\varepsilon \in (0, 10)$, there exists a universal constant $c > 0$ such that*

$$\inf_{\hat{u} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} \left[1 - \frac{\hat{u}(S)^\top \Sigma \hat{u}(S)}{\|\Sigma\|}\right] \;\ge\; c \cdot \min\left\{\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon}, 1\right\}\,.$$

Proof We adopt the same proof strategy as the proof of Proposition B.5 for mean estimation. By (Acharya et al., 2021, Lemma 6), there exists a finite index set $\mathcal{V} \subset \mathbb{R}^d$ with cardinality $|\mathcal{V}| = 2^{\Omega(d)}$, $\|v\| = 1$ for all $v \in \mathcal{V}$ and $\|v - v'\| \ge 1/2$ for all $v \ne v' \in \mathcal{V}$. Let $Q_0 = \mathcal{N}(0, \mathbf{I}_d)$. For each $v \in \mathcal{V}$, let Q_v be a uniform distribution on $\{\pm v\}$. We define $P_v := (1 - \alpha)Q_0 + \alpha Q_v$. Then P_v is sub-Gaussian with covariance $\Sigma_v := \mathbf{I}_{d \times d} + \alpha vv^\top$. It is straightforward to see that $0.5\mathbf{I}_{d \times d} \preceq \Sigma_v \preceq \mathbf{I}_{d \times d}$ and the top eigenvector of Σ_v is v.

Since $\|v - v'\| \ge 1/2$, we have

$$D_{\Sigma_{v'}}(v) = 1 - \frac{v^\top \Sigma_v' v}{\|\Sigma_{v'}\|} = 1 - (1 - \alpha + \alpha \langle v, v' \rangle^2) = \alpha(1 - \langle v, v' \rangle^2) \ge \frac{\alpha}{8}\,.$$

The principal component estimation problem can be reduced to a testing problem with this packing $\mathcal{V}$. For the (ε, δ)-DP estimator $\hat{u}$, using Lemma 20, let $t = \frac{\alpha}{8}$, we have

$$\begin{aligned}
\sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n}[D_\Sigma(\hat{u})] &\ge \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \mathbb{E}_{S \sim P_v^n}[D_{\Sigma_v}(\hat{u})] \\
&= \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} P_v\left(D_{\Sigma_v}(\hat{u}) \ge t\right) \\
&\gtrsim t \frac{e^{d/2} \cdot \left(\frac{1}{2} e^{-\varepsilon \lceil n\alpha \rceil} - \frac{\delta}{1 - e^{-\varepsilon}}\right)}{1 + e^{d/2} e^{-\varepsilon \lceil n\alpha \rceil}}\,,
\end{aligned}$$

where the last inequality follows from the fact that $d \ge 2$. The rest of the proof follows from (Barber and Duchi, 2014, Proposition 4). We choose

$$\alpha = \frac{1}{n\varepsilon} \min\left\{\frac{d}{2} - \varepsilon, \log\left(\frac{1 - e^{-\varepsilon}}{4\delta e^\varepsilon}\right)\right\}$$

so that

$$\sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n}[D_{\Sigma_v}(\hat{u})] \gtrsim \alpha .$$

This implies, for $t = \alpha/12$ and $\varepsilon \in (0, 10)$, that

$$\inf_{\hat{u} \in \mathcal{M}_{\varepsilon, \delta}} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n}[D_\Sigma(\hat{u})] \gtrsim \min \left\{ \frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon}, 1 \right\} ,$$

which completes the proof. ∎

It is well known that even for Gaussian distributions, learning the principal component up to error α requires $\Omega(d/\alpha^2)$. We provide a lower bound proof here for completeness.

Proposition E.3 (Sample Complexity Lower bound for PCA) *Let $\mathcal{P}_\Sigma$ be the set of zero-mean Gaussian distributions with covariance $\Sigma \in \mathbb{R}^{d \times d}$. Let $\mathcal{M}_d$ be the class of estimators of the d-dimensional top principal component of Σ using n i.i.d. samples from $P \in \mathcal{P}_\Sigma$. There exists a universal constant $c > 0$ such that*

$$\inf_{\hat{u} \in \mathcal{M}_d} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} \left[1 - \frac{\hat{u}(S)^\top \Sigma \hat{u}(S)}{\|\Sigma\|} \right] \geq c \cdot \min \left\{ \sqrt{\frac{d}{n}}, 1 \right\} .$$

Proof

The following proposition helps to prove a minimax lower bound on estimating $\|\Sigma\|$. We first define some notations.

Definition 48 (Definition 3.1 in (Diakonikolas et al., 2017)) *For a distribution A on the real line with probability density function $A(x)$ and a unit vector $v \in \mathbb{R}^d$, consider the distribution over $\mathbb{R}^n$ with probability density function $P_v(x) = A(v^\top x) \exp(-\|x - (v^\top x)v\|_2^2/2) \cdot (2\pi)^{-(d-1)/2}$.*

Proposition E.4 (Proposition 7.1 in (Diakonikolas et al., 2017)) *Let A be a distribution on $\mathbb{R}$ such that A has a mean 0 and $\chi^2(A, N(0, 1))$ is finite. Then, there is no algorithm for any d, given $n < d/(8\chi^2(A, N(0, 1)))$ samples from a distribution D over $\mathbb{R}^d$ which is either $N(0, I)$ or P_v for some unit vector $v \in \mathbb{R}^d$, that correctly distinguishes between the two cases with probability at least $2/3$.*

To apply Proposition E.4, let A be Gaussian distribution $\mathcal{N}(0, 1 + \alpha)$. Through simple calculation, it can be shown that $\chi^2(\mathcal{N}(0, 1), \mathcal{N}(0, 1 + \alpha)) = \frac{1}{\sqrt{1 - \alpha^2}} - 1 \leq \alpha^2$ whenever $\alpha^2 \leq 1/2$. Then, for the first case in Proposition E.4, $\|\Sigma\| = \|I\| = 1$, the second case has $\|\Sigma\| = 1 + \alpha$, and Proposition E.4 implies that there exists absolute constant c such that

$$\inf_{\hat{\lambda}} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} \left[1 - \frac{\hat{\lambda}(S)}{\|\Sigma\|} \right] \geq c \cdot \min \left\{ \sqrt{\frac{d}{n}}, 1 \right\} .$$

Since we can turn a principal component estimator $u(S)$ into an estimator of $\|\Sigma\|$ through n additional fresh samples to estimate $u(S)^\top \Sigma u(S)$ up to a minor multiplicative error $O(1/\sqrt{n})$. This implies there exists a universal constant $c > 0$ such that

$$\inf_{\hat{u} \in \mathcal{M}_d} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} \left[1 - \frac{\hat{u}(S)^\top \Sigma \hat{u}(S)}{\|\Sigma\|} \right] \geq c \cdot \min \left\{ \sqrt{\frac{d}{n}}, 1 \right\} .$$

∎

E.3.2. HYPERCONTRACTIVE DISTRIBUTIONS

In this section, we apply our results on hypercontractive distributions in Definition 18. Using the resilience of hypercontractive distributions with respect to $(\mu = 0, \Sigma)$ in Lemma 19, which is the same as the resilience properties we need for PCA in Definition 44, Theorem 14 implies the following corollary.

Corollary E.5 *Under the hypothesis of Lemma 19 with $k \geq 3$, $\mu = 0$ and any PSD matrix $\Sigma \in \mathbb{R}^{d \times d}$, there exist universal constants c and $C > 0$ such that for any $\alpha \in (0, c)$, a dataset of size*

$$n = O\left(\frac{d}{\zeta^{2(1-1/k)}\alpha^{2(1-1/k)}} + \frac{k^2\alpha^{2-2/k}d\log d}{\zeta^{2-4/k}\kappa^2} + \frac{\kappa^2 d\log d}{\alpha^{2/k}} + \frac{\log(1/(\delta\zeta)) + d\log(1/\alpha^{1-2/k})}{\varepsilon\alpha} \right) ,$$

and sensitivity of $\Delta = O(\alpha^{1-2/k}/n)$ with large enough constants are sufficient for $\mathrm{HPTR}(S)$ in Section E.1 for PCA with the choices of the distance function in Eq. (65) to achieve

$$1 - \frac{\hat{u}^\top \Sigma \hat{u}}{\|\Sigma\|} \leq C\alpha^{1-2/k} , \tag{77}$$

with probability $1 - \zeta$. Further, the same guarantee holds even if an α-fraction of the samples is arbitrarily corrupted, as in Assumption 1.

The error bound is optimal under α-corruption up to a constant factor. HPTR is the first estimator that guarantees (ε, δ)-DP and also achieves the robust error rate of $1 - \hat{u}^\top \Sigma \hat{u}/\|\Sigma\| = O(\alpha^{1-2/k})$, matching the information-theoretic lower bound of $1 - \hat{u}^\top \Sigma \hat{u}/\|\Sigma\| = \Omega(\alpha^{1-2/k})$. This lower bound can be easily constructed using Eq. (57), where two hypercontractive distributions are at total variation distance $O(\alpha)$ and the top principal component of one distribution achieves an error lower bounded by $1 - \hat{u}^\top \Sigma \hat{u}/\|\Sigma\| = \Omega(\alpha^{1-2/k})$. Even if privacy is not required, there is no outlier-robust PCA estimator matching this optimal error rate for a general k.

The sample complexity is $n = \tilde{O}(d/\alpha^{2(1-1/k)} + (d + \log(1/\delta))/(\varepsilon\alpha))$ for a constant ζ, k, and κ, where $\tilde{O}$ hides logarithmic factors in $1/\alpha$ and d. Even for DP PCA without corrupted samples, HPTR is the first estimator for hypercontractive distributions to guarantee differential privacy. The information-theoretic lower bound is $n = \Omega(d/\alpha^{2(1-2/k)} + \min\{d, \log((1 - e^{-\varepsilon})/\delta)\}/(\alpha\varepsilon))$ to achieve the error in Eq. (77). The first term is unavoidable, even without DP and robustness, when the data comes from a Gaussian distribution because estimating the principal component up to error $\alpha^{1-2/k}$ requires $\Omega(d/\alpha^{2(1-2/k)})$ samples (Proposition E.3). There is a gap of factor $O(\alpha^{-2/k})$ compared to the first term in our upper bound. Since the sample complexity lower bound in Proposition E.3 is constructed using Gaussian distributions, it might be possible to tighten it further using hypercontractive distributions. The second term in the lower bound follows from Proposition E.6, which matches the last term in the upper bound up to a factor of $O(\log(1/\alpha))$ when $\delta = e^{-\Theta(d)}$ and $\varepsilon > 0$. To the best of our knowledge, HPTR is the first algorithm for PCA that guarantees (ε, δ)-DP under hypercontractive distributions.

Proposition E.6 (Lower bound for hypercontractive private PCA) *Let $\mathcal{P}_\Sigma$ be the set of zero-mean hypercontractive distributions with covariance $\Sigma \in \mathbb{R}^{d \times d}$. Let $\mathcal{M}_{\varepsilon, \delta}$ be a class of (ε, δ)-DP estimators using n i.i.d. samples from $P \in \mathcal{P}_\Sigma$. Then, for $\varepsilon \in (0, 10)$, there exists a constant c such that*

$$\inf_{\hat{u} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{\Sigma \succ 0, P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} \left[1 - \frac{\hat{u}^\top \Sigma \hat{u}}{\|\Sigma\|} \right] \geq c \min \left\{ \left(\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon} \right)^{1-2/k}, 1 \right\} . \tag{78}$$

Proof We use the same construction as used in the distribution of x in the proof of Proposition C.5. By (Acharya et al., 2021, Lemma 6), there exists a finite index set $\mathcal{V} \subset \mathbb{R}^d$ with cardinality $|\mathcal{V}| = 2^{\Omega(d)}$, $\|v\| = 1$ for all $v \in \mathcal{V}$ and $\|v - v'\| \geq 1/2$ for all $v \neq v' \in \mathcal{V}$. For each $v \in \mathcal{V}$ and $\alpha \in (0, 1/2)$, we construct the density function of distribution P_v as defined in Eq. (57). Let Σ_v denote the covariance matrix of P_v. The proof of Proposition C.5 shows that $\Sigma_v = (1 - \alpha)\mathbf{I}_{d \times d} + \alpha^{1-2/k}vv^\top$, $d_{\mathrm{TV}}(P_v, P_v') = \alpha$ and that P_v is $(O(1), k)$-hypercontractive.

Since $\|v - v'\| \geq 1/2$, we know that $\langle v, v'\rangle \leq 7/8$, and we have

$$D_{\Sigma_{v'}}(v) = 1 - \frac{v^\top \Sigma'_v v}{\|\Sigma_{v'}\|} = 1 - \frac{1 - \alpha + \alpha^{1-2/k}\langle v, v'\rangle^2}{1 - \alpha + \alpha^{1-2/k}} \geq \frac{\alpha^{1-2/k}}{8(1 - \alpha + \alpha^{1-2/k})} > \frac{\alpha^{1-2/k}}{12} ,$$

for $\alpha < c$ small enough.

Next, we apply the reduction of estimation to testing with this packing $\mathcal{V}$. For a (ε, δ)-DP estimator $\hat{u}$, using Lemma 20, let $t = \frac{\alpha^{1-2/k}}{12}$. Then, we have

$$
\begin{aligned}
\sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n}[D_\Sigma(\hat{u})] &\geq \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \mathbb{E}_{S \sim P_v^n}[D_{\Sigma_v}(\hat{u})] \\
&= \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} P_v\left(D_{\Sigma_v}(\hat{u}) \geq t\right) \\
&\gtrsim t \frac{e^{d/2} \cdot \left(\frac{1}{2}e^{-\varepsilon\lceil n\alpha\rceil} - \frac{\delta}{1-e^{-\varepsilon}}\right)}{1 + e^{d/2}e^{-\varepsilon\lceil n\alpha\rceil}} ,
\end{aligned}
$$

where the last inequality follows from the fact that $d \geq 2$.

The rest of the proof follows from (Barber and Duchi, 2014, Proposition 4). We choose

$$\alpha = \frac{1}{n\varepsilon} \min\left\{\frac{d}{2} - \varepsilon, \log\left(\frac{1 - e^{-\varepsilon}}{4\delta e^\varepsilon}\right)\right\}$$

so that

$$\sup_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n}[D_{\Sigma_v}(\hat{u})] \gtrsim \alpha^{1-2/k} .$$

This means, for $t = (1/12)\alpha^{1-2/k}$ and $\varepsilon \in (0, 10)$, that

$$\inf_{\hat{u} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n}[D_\Sigma(\hat{u})] \gtrsim \min\left\{\left(\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{n\varepsilon}\right)^{1-2/k}, 1\right\} ,$$

which completes the proof. ∎

Appendix F. General case: utility analysis of HPTR

We prove the following theorem that provides a utility guarantee for HPTR output $\hat{\theta}$ measured in $D_\phi(\hat{\theta}, \theta)$.

Theorem 15 *For a given dataset S, a target error function $D_\phi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_+$, probability $\zeta \in (0,1)$, and privacy (ε, δ), HPTR achieves $D_\phi(\hat{\theta}, \theta) = c_0 \rho$ for some $\rho > 0$ and any constant $c_0 > 3c_1$ with probability $1 - \zeta$ if there exist constants $c_1, c_2 > 0$ and $(\Delta \in \mathbb{R}^+, \rho \in \mathbb{R}^+)$ such that with the choice of $k^* = (2/\varepsilon) \log(4/(\delta\zeta))$, $\tau = (c_0 + c_1)\rho$, the following assumptions are satisfied:*

(a) *(Bounded volume)* $(7/8)\tau - (k^* + 1)\Delta > 0$,

$$\frac{\mathrm{Vol}(B_{\tau + (k^*+1)\Delta + c_1\rho, S})}{\mathrm{Vol}(B_{(7/8)\tau - (k^*+1)\Delta - c_1\rho, S})} \leq e^{c_2 p} \text{, and}$$

$$\frac{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq (c_0 + 2c_1)\rho\})}{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_1\rho\})} \leq e^{c_2 p},$$

(b) *(Local sensitivity) For all S' within Hamming distance k^* from S, $\max_{S'' \sim S'} \|D_{S''}(\hat{\mu}) - D_{S'}(\hat{\mu})\| \leq \Delta$ for all $\hat{\mu} \in B_{\tau + (k^*+3)\Delta, S}$,*

(c) *(Bounded sensitivity)* $\Delta \leq \frac{(c_0 - 3c_1)\rho\varepsilon}{32(c_2 p + (\varepsilon/2) + \log(16/\delta\zeta))}$, *and*

(d) *(Robustness)* $|D_\phi(\hat{\theta}, \theta) - D_S(\hat{\theta})| \leq c_1 \rho$ *for all $\hat{\theta} \in B_{\tau, S}$.*

The parameter $\rho \in \mathbb{R}_+$ represents the target error up to a constant factor and depends on the resilience of the underlying distribution $P_{\theta, \phi}$ that the samples are drawn from. We explicitly prescribe how to choose the parameter ρ for each problem instance in Sections B, C, D, and E. Following the standard analysis techniques for exponential mechanisms, we show that the output concentrates around an inner set $\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_0\rho\}$, by comparing its probability mass with an outer set $\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \geq c_1\rho\}$. This uses the ratio of the volumes in the assumption (a) and the closeness of the error metric and $D(\hat{\theta})$ in the assumption (d). When there is a strict gap between the two, which happens if $\varepsilon\rho/\Delta \gg p + \log(1/\zeta)$ as in the assumption (c), this implies $D_\phi(\hat{\theta}, \theta) \leq c_0\rho$ with probability $1 - \zeta$. We provide a proof in Section F.2.

A major challenge in analyzing HPTR is in showing that the safety test threshold $k^* = (2/\varepsilon) \log(4/(\delta\zeta))$ is not only large enough to ensure that datasets with safety violation is screened with probability $1 - \delta/2$ but also small enough such that good datasets satisfying the assumptions (a), (b), and (c) pass the test with probability $1 - \zeta/2$. We establish this first in Section F.1.

F.1. Large safety margin

In this section, we show in Lemma 51 that under the assumptions of Theorem 15, we get a large enough margin for safety such that we pass the safety test with high probability. We follow the proof strategy introduced in (Brown et al., 2021) adapted to our more general framework. A major challenge is the lack of a uniform bound on the sensitivity, which the analysis of (Brown et al., 2021) relies on. We generalize the analysis by showing that while the data does not satisfy uniform sensitivity bound, we can still exploit its *local* sensitivity bound in the assumption (b).

The following main technical lemma is a counter part of (Brown et al., 2021, Lemma 3.7), where we have an extra challenge that the sensitivity bound is only local; there exists $\hat{\theta}$ far from θ where the sensitivity bound fails. We rely on the assumption (b) to resolve it. Let $w_S(B) \triangleq \int_B \exp\{-(\varepsilon/4\Delta)D_S(\hat{\mu})\}d\hat{\mu}$ be the weight of a subset $B \subset \mathbb{R}^p$. The following lemma will be used

to show that the denominator of the exponential distribution in RELEASE step does not change too fast between two neighboring datasets.

Lemma 49 *Under the assumption [(b)] and $\delta \in (0, 1/2)$, for a dataset S' at Hamming distance at most k^* from S, if $w_{S'}(B_{\tau-\Delta,S'}) \geq (1-\delta)w_{S'}(B_{\tau+\Delta,S'})$ then $S' \in \mathsf{SAFE}_{\varepsilon,4e^{2\varepsilon}\delta,\tau}$.*

Proof We follow the proof strategy of ([Brown et al., 2021], Lemma 3.7) but there are key differences due to the fact that we do not have a universal sensitivity bound, but only local bound. In particular, we first establish that under the local sensitivity assumption, $B_{\tau,S''} \subseteq B_{\tau+\Delta,S'}$ for all $S'' \sim S'$, which will be used heavily throughout the proof. Since $D_{S''}(\hat{\theta}) \leq D_{S'}(\hat{\theta}) + \Delta$ for all $\hat{\theta} \in B_{\tau+(k^*+3)\Delta,S}$, we have $B_{\tau,S''} \cap B_{\tau+(k^*+3)\Delta,S} \subseteq B_{\tau+\Delta,S'}$. We are left to show that $B_{\tau,S''} \setminus B_{\tau+(k^*+3)\Delta,S} = \emptyset$, which follows from the fact that $(B_{\tau,S''} \setminus B_{\tau+(k^*+1.5)\Delta,S}) \cap B_{\tau+(k^*+3)\Delta,S} = \emptyset$ and $D_{S''}(\hat{\theta})$ is a Lipschitz continuous function. Similarly, it follows that $B_{\tau-\Delta,S'} \subseteq B_{\tau,S''}$. In particular, this implies that $B_{\tau,S'} \subseteq B_{\tau+(k^*+3)\Delta,S}$ for any S' with $d_H(S', S) \leq k^*$.

We first show that for any $E \subset B_{\tau,S'}$ one side of the $(\varepsilon/2, 4e^{\varepsilon/2}\delta)$-DP condition is met: $\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E) \leq e^{\varepsilon/2}\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E) + 4e^{\varepsilon/2}\delta$ for all $S'' \sim S'$ where $r_{(\varepsilon,\Delta,\tau,S')}$ and $r_{(\varepsilon,\Delta,\tau,S'')}$ are the distributions used in the exponential mechanism as defined in (2) respectively. For $B = B_{\tau,S'} \cap B_{\tau,S''}$, we have

$$
\begin{aligned}
\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E) &= \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \cap B) + \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \setminus B) \\
&= \frac{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \cap B)}{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E \cap B)} \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E \cap B) + \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \setminus B) \\
&\leq \frac{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \cap B)}{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E \cap B)} \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E) + \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \notin B_{\tau,S''}) \, .
\end{aligned}
$$

The ratio is bounded due to the local sensitivity bound at S' as

$$
\begin{aligned}
\frac{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \in E \cap B)}{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S'')}}(\hat{\theta} \in E \cap B)} &\leq e^{\varepsilon/4}\frac{w_{S''}(B_{\tau,S''})}{w_{S'}(B_{\tau,S'})} \\
&\leq e^{\varepsilon/2}\frac{w_{S'}(B_{\tau,S''})}{w_{S'}(B_{\tau,S'})} \\
&\leq e^{\varepsilon/2}\frac{w_{S'}(B_{\tau+\Delta,S})}{w_{S'}(B_{\tau,S'})} \leq e^{\varepsilon/2}(1+2\delta) \, ,
\end{aligned}
$$

where the second inequality follows from the fact that $w_{S''}(A) \leq e^{\varepsilon/6}w_{S'}(A)$ for any set $A \subset B_{\tau,S'} \cup B_{\tau,S''} \subseteq B_{\tau+(k^*+3)\Delta,S}$ and the third inequality follows from the fact that $B_{\tau,S''} \subseteq B_{\tau+\Delta,S'}$. From the assumption on the weights, it follows that $w_{S'}(B_{\tau+\Delta,S'})/w_{S'}(B_{\tau,S'}) \leq w_{S'}(B_{\tau+\Delta,S'})/w_{S'}(B_{\tau-\Delta,S'}) \leq 1/(1-\delta) \leq 1 + 2\delta$ for $\delta < 1/2$. Similarly,

$$
\begin{aligned}
\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \notin B_{\tau,S''}) &\leq \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S')}}(\hat{\theta} \notin B_{\tau-\Delta,S'}) \\
&\leq 1 - \frac{w_{S'}(B_{\tau-\Delta,S'})}{w_{S'}(B_{\tau,S'})} \leq 1 - \frac{w_{S'}(B_{\tau-\Delta,S'})}{w_{S'}(B_{\tau+\Delta,S'})} \leq \delta \, .
\end{aligned}
$$

Putting these together, we get $\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S')}}(\hat{\theta} \in E) \leq e^{\varepsilon/2} \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S'')}}(\hat{\theta} \in E) + 4e^{\varepsilon/2}\delta$.

Next, we show the other side of the $(\varepsilon/2, 4e^{\varepsilon/2}\delta)$-DP condition: $\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S')}}(\hat{\theta} \in E) \leq e^{\varepsilon/2} \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S)}}(\hat{\theta} \in E) + 4e^{2\varepsilon}\delta$ for all $S' \sim S$. We need to show an upper bound on the ratio:

$$
\begin{aligned}
\frac{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S')}}(\hat{\theta} \in E \cap B)}{\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S)}}(\hat{\theta} \in E \cap B)} &\leq e^{\varepsilon/4} \frac{w_S(B_{\tau,S})}{w_{S'}(B_{\tau,S'})} \\
&\leq e^{\varepsilon/2} \frac{w_S(B_{\tau,S})}{w_S(B_{\tau,S'})} \\
&\leq e^{\varepsilon/2} \frac{w_S(B_{\tau,S})}{w_S(B_{\tau-\Delta,S})} \leq (1 + 2\delta)e^{\varepsilon/2} \ ,
\end{aligned}
$$

For the probability outside $B_{\tau,S'}$,

$$
\begin{aligned}
\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S'')}}(\hat{\theta} \notin B_{\tau,S'}) &\leq \mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon, \Delta, \tau, S'')}}(\hat{\theta} \in B_{\tau+\Delta,S'} \setminus B_{\tau,S'}) \\
&\leq \frac{w_{S''}(B_{\tau+\Delta,S'} \setminus B_{\tau,S'})}{w_{S''}(B_{\tau,S''})} \\
&\leq e^{\varepsilon/2} \frac{w_{S'}(B_{\tau+\Delta,S'} \setminus B_{\tau,S'})}{w_{S'}(B_{\tau,S''})} \\
&\leq e^{\varepsilon/2} \frac{w_{S'}(B_{\tau+\Delta,S'}) - w_{S'}(B_{\tau,S'})}{w_{S'}(B_{\tau-\Delta,S'})} \\
&\leq e^{\varepsilon/2}(1 + 2\delta - 1) = 2e^{\varepsilon/2}\delta \ .
\end{aligned}
$$

where the first inequality follows from $B_{\tau,S''} \subseteq B_{\tau+\Delta,S'}$, the second inequality follows from $(B_{\tau+\Delta,S'} \setminus B_{\tau,S'}) \cap B_{\tau,S''} \subseteq B_{\tau+\Delta,S'} \setminus B_{\tau,S'}$, the third inequality follows from the fact that $B_{\tau,S''} \subseteq B_{\tau+\Delta,S'}$ and the local sensitivity assumption, and the last inequality follows from the weight assumption and $B_{\tau-\Delta,S'} \subseteq B_{\tau,S'}$. ∎

The next lemma identifies the range of the threshold $k^* = O(\tau/\Delta)$ that ensures safety.

Lemma 50 *Under the assumption (b), if there exists a $g > 0$ such that $\tau - \Delta(k^* + g + 1) > 0$ and*

$$
\frac{\text{Vol}(B_{\tau+\Delta(k^*+1),S})}{\text{Vol}(B_{\tau-\Delta(k^*+g+1),S})} e^{\frac{-\varepsilon g}{4}} \leq \frac{1}{8} e^{-\varepsilon/2}\delta \ , \tag{79}
$$

then $S' \in \mathsf{SAFE}_{(\varepsilon/2, \delta/2, \tau)}$ for all S' within Hamming distance k^* from S.

Proof Consider S' at Hamming distance k away from S. From Lemma 49 it suffices to show that $w_{S'}(B_{\tau-\Delta,S'})/w_{S'}(B_{\tau+\Delta,S'}) \geq 1 - \delta'$ for $\delta' = (1/8)e^{-\varepsilon/2}\delta$, which is equivalent to

$$
w_{S'}(B_{\tau+\Delta,S'} \setminus B_{\tau-\Delta,S'})/w_{S'}(B_{\tau+\Delta,S'}) \leq \delta' \ .
$$

The denominator is lower bounded by

$$
\begin{aligned}
w_{S'}(B_{\tau+\Delta,S'}) \ \geq \ w_{S'}(B_{\tau-\Delta(1+g),S'}) \ &\geq \ \mathrm{Vol}(B_{\tau-\Delta(1+g),S'})e^{-\varepsilon(\tau-\Delta(1+g))/(4\Delta)} \\
&\geq \ \mathrm{Vol}(B_{\tau-\Delta(1+g+k),S})e^{-\varepsilon(\tau-\Delta(1+g))/(4\Delta)} \ ,
\end{aligned}
$$

where the last inequality uses the local sensitivity (the assumption (b)). The numerator is upper bounded by

$$
w_{S'}(B_{\tau+\Delta,S'} \setminus B_{\tau-\Delta,S'}) \ \leq \ w_{S'}(B_{\tau+(k+1)\Delta,S} \setminus B_{\tau-\Delta,S'}) \ \leq \ \mathrm{Vol}(B_{\tau+(k+1)\Delta,S})e^{-\varepsilon(\tau-\Delta)/(4\Delta)} \ ,
$$

where the first inequality uses the local sensitivity. Together, it follows that

$$
\frac{w_{S'}(B_{\tau+\Delta,S'} \setminus B_{\tau-\Delta,S'})}{w_{S'}(B_{\tau+\Delta,S'})} \ \leq \ \frac{\mathrm{Vol}(B_{\tau+(k+1)\Delta,S})e^{-\varepsilon(\tau-\Delta)/(4\Delta)}}{\mathrm{Vol}(B_{\tau-\Delta(1+g+k),S})e^{-\varepsilon(\tau-\Delta(1+g))/(4\Delta)}} \ \leq \ \delta' \ = \ \frac{1}{8}e^{\varepsilon/2}\delta \ ,
$$

as $e^{-\varepsilon(\tau-\Delta)/(4\Delta)}/e^{-\varepsilon(\tau-\Delta(1+g))/(4\Delta)} = e^{-\varepsilon g/4}$, which implies safety. ∎

We next show that $k^* = O((1/\varepsilon)\log(1/(\delta\zeta)))$ is sufficient to ensure a large enough safety margin of $m_\tau - k^* = \Omega((1/\varepsilon)\log(1/\zeta))$.

Lemma 51 *Under the assumptions (a), (b), and (c) of Theorem 15, for $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$, if $d_H(S',S) \leq (2/\varepsilon)\log(4/(\zeta\delta))$ then $S' \in \mathsf{SAFE}_{(\varepsilon/2,\delta/2,\tau)}$.*

Proof Applying Lemma 50 with $k^* = (2/\varepsilon)\log(4/(\delta\zeta))$ and $g = (1/(8\Delta))\tau$, we require

$$
\frac{\mathrm{Vol}(B_{\tau+\Delta(k^*+1),S})}{\mathrm{Vol}(B_{(7/8)\tau-\Delta(k^*+1),S})}e^{\frac{-\varepsilon\tau}{32\Delta}} \ \leq \ \frac{1}{8}e^{-\varepsilon/2}\delta \ .
$$

From the assumption (a), it is sufficient to have

$$
\exp\left\{c_2 p - \frac{\tau\varepsilon}{32\Delta}\right\} \ \leq \ \frac{1}{8}e^{-\varepsilon/2}\delta \ .
$$

For $\Delta \leq (\tau\varepsilon)/(32(c_2 p + (\varepsilon/2) + \log(8/\delta)))$, which follows from the assumption (c), this is satisfied. ∎

F.2. Proof of Theorem 15

We first show that we pass the safety test with high probability. Define the error event E as the event that we output $\perp$ in the TEST step. From Lemma 51, we have $m_\tau > (2/\varepsilon)\log(4/(\delta\zeta))$ under the assumptions (a), (b), and (c). This implies that

$$
\mathbb{P}(E) \ = \ \mathbb{P}\big(m_\tau + \mathrm{Lap}(2/\varepsilon) < (2/\varepsilon)\log(2/\delta)\big) \ \leq \ \frac{\zeta}{2} \ .
$$

We next show that resilience implies good utility (once safety test has passed). We want the exponential mechanism to output an accurate $\hat{\theta}$ near θ with high probability, i.e., $\mathbb{P}_{\hat{\theta} \sim r_{(\varepsilon,\Delta,\tau,S)}}(D_\phi(\hat{\theta},\theta) \geq$

$c_0\rho) \leq \zeta/2$. We omit the subscript in the probability for brevity, and it is assumed that randomness is in the sampling of the exponential mechanism. We want to bound by $\zeta/2$ the failure probability:

$$
\begin{aligned}
\mathbb{P}\big(D_\phi(\hat{\theta}, \theta) \geq c_0\rho \big) &\leq \frac{\mathbb{P}\big(D_\phi(\hat{\theta}, \theta) \geq c_0\rho \big)}{\mathbb{P}\big(D_\phi(\hat{\theta}, \theta) \leq c_1\rho_1 \big)} \\
&\leq \frac{\mathrm{Vol}(B_{\tau,S})}{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_1\rho\})} \frac{\max_{\hat{\theta}:D_\phi(\hat{\theta},\theta)\geq c_0\rho} \mathbb{P}(\hat{\theta})}{\min_{\hat{\theta}:D_\phi(\hat{\theta},\theta)\leq c_1\rho_1} \mathbb{P}(\hat{\theta})} ,
\end{aligned}
$$

as long as $\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_0\rho\} \subseteq B_{\tau,S}$ (otherwise we are under-estimating the volume), which follows from the assumption (d); $D_S(\hat{\theta}) \leq (D_\phi(\hat{\theta}, \theta) + c_1\rho) \leq (c_0 + c_1)\rho = \tau$.

Similarly, since $\hat{\theta} \in B_{\tau,S}$ implies $D_\phi(\hat{\theta}, \theta) \leq \tau + c_1\rho = (c_0 + 2c_1)\rho$, the volume ratio is bounded by

$$
\frac{\mathrm{Vol}(B_{\tau,S})}{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_1\rho\})} \leq \frac{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq (c_0 + 2c_1)\rho\})}{\mathrm{Vol}(\{\hat{\theta} : D_\phi(\hat{\theta}, \theta) \leq c_1\rho\})} \leq e^{c_2 p} ,
$$

under the assumption (a). The probability ratio can be bounded similarly. From the assumption (d), we have

$$
\frac{\max_{\hat{\theta}:D_\phi(\hat{\theta},\theta)\geq c_0\rho} \mathbb{P}(\hat{\theta})}{\min_{\hat{\theta}:D_\phi(\hat{\theta},\theta)\leq c_1\rho} \mathbb{P}(\hat{\theta})} \leq \exp\left\{ -\frac{\varepsilon}{4\Delta}\big((c_0 - c_1) - (2c_1)\big)\rho \right\} \leq \exp\left\{ -\frac{\varepsilon(c_0 - 3c_1)\rho}{4\Delta} \right\} .
$$

When $e^{c_2 p - (\varepsilon(c_0 - 3c_1)\rho/(4\Delta))} \leq \zeta/2$, we have the desired bound. This is guaranteed with our assumption (c).

Appendix G. Auxiliary lemmas

Lemma 52 *For any symmetric $\Sigma \succ 0$ and vector $u \in \mathbb{R}^d$,*

$$
\max_{v:\|v\|=1} \frac{\langle v, u \rangle}{v^\top \Sigma v} = \left\| \Sigma^{-1/2} u \right\| . \tag{80}
$$

Proof This follows analogously from the proof of Lemma 7. ∎

Lemma 53 *Let $\Sigma, A \in \mathbb{R}^{d \times d}$ be a symmetric matrix. If $-c\mathbf{I}_{d\times d} \preceq \Sigma^{-1/2} A \Sigma^{-1/2} - \mathbf{I}_{d\times d} \preceq c\mathbf{I}_{d\times d}$ for some $c > 0$, then we have for any $u \in \mathbb{R}^d$,*

$$
\|\Sigma^{-1/2}(A - \Sigma)u\| \leq c\|\Sigma^{1/2}u\| . \tag{81}
$$

Proof Using the fact that $-\mathbf{I}_{d\times d} \preceq M \preceq \mathbf{I}_{d\times d}$ implies $-\mathbf{I}_{d\times d} \preceq M^2 \preceq \mathbf{I}_{d\times d}$, for any symmetric matrix M, we know

$$
-c^2\mathbf{I}_{d\times d} \preceq \Sigma^{-1/2}(A - \Sigma)\Sigma^{-1}(A - \Sigma)\Sigma^{-1/2} \preceq c^2\mathbf{I}_{d\times d} , \tag{82}
$$

which implies that

$$-c^2\Sigma \preceq (A - \Sigma)\Sigma^{-1}(A - \Sigma) \preceq c^2\Sigma . \tag{83}$$

Thus, we know

$$\|\Sigma^{-1/2}(A - \Sigma)u\|^2 = u^\top(A - \Sigma)\Sigma^{-1}(A - \Sigma)u \leq c^2 u^\top \Sigma u = c^2\|\Sigma^{1/2}u\|^2 . \tag{84}$$

∎

Appendix H. Existing lower bounds

Theorem 54 (Lower bound on DP Gaussian mean estimation with known covariance (Kamath et al., 2019, Lemma 6.7)) *Let $\hat{\mu} : \mathbb{R}^{n \times d} \to [-R\sigma, R\sigma]^d$ be an (ε, δ)-differentially private estimator (with $\delta \leq \sqrt{d}/(48\sqrt{2}Rn\sqrt{\log(48\sqrt{2}Rn/\sqrt{d})}))$ such that for every Gaussian distribution $P = \mathcal{N}(\mu, \sigma^2 \mathbf{I}_{d \times d})$ (for $-R\sigma \leq \mu_j \leq R\sigma$ where $j \in [d]$) and*

$$\mathbb{E}_{S \sim P^n}\left[\|\hat{\mu}(S) - \mu\|^2\right] \leq \alpha^2 \leq \frac{d\sigma^2 R^2}{6} , \tag{85}$$

then $n \geq \frac{d\sigma}{24\alpha\varepsilon}$.

Theorem 55 (Lower bound on DP covariance bounded mean estimation (Kamath et al., 2020, Theorem 6.1)) *Suppose $\hat{\mu}$ is an $(\varepsilon, 0)$-DP estimator such that, for every product distribution $P \in \mathbb{R}^d$ such that $\mathbb{E}[P] = \mu$, $\sup_{v:\|v\|=1} \mathbb{E}_{x \sim P}[\langle v, x - \mu\rangle^2] \leq 1$ and*

$$\mathbb{E}_{S \sim P^n}\left[\|\hat{\mu}(S) - \mu\|^2\right] \leq \alpha^2 . \tag{86}$$

Then $n = \Omega\left(d/(\varepsilon\alpha^2)\right)$

Theorem 56 (Lower bound of the error rate for hypercontractive linear regression with independent noise(Bakshi and Prasad, 2021, Theorem 6.1)) *Consider linear model $y = \langle \beta, x \rangle + \eta$, where optimal hyperplane β is used to generate data, and the noise η is independent of the samples x. Then there exists two distribution D_1 and D_2 over $\mathbb{R}^2 \times \mathbb{R}$ such that the marginal distribution over $\mathbb{R}^2$ has covariance Σ and is (κ_k, k)-hypercontractive yet $\|\Sigma^{1/2}(\beta_1 - \beta_2)\| = \Omega(\sqrt{\kappa_k}\gamma\alpha^{1-1/k})$, where β_1 and β_2 are the optimal hyperplanes for D_1 and D_2 respectively, $\gamma < 1/\alpha^{1/k}$ and the noise η is uniform over $[-\gamma, \gamma]$.*

Theorem 57 (Lower bound of the error rate for hypercontractive linear regression with dependent noise(Bakshi and Prasad, 2021, Theorem 6.2)) *There exists two distributions D_1, D_2 over $\mathbb{R}^2 \times \mathbb{R}$ such that the marginal distribution over $\mathbb{R}^2$ has covariance Σ and is κ_k, k-hypercontractive yet $\|\Sigma^{1/2}(\beta_1 - \beta_2)\| = \Omega(\sqrt{\kappa_k}\gamma\alpha^{1-2/k})$, where β_1 and β_2 are least square solutions for D_1 and D_2, respectively, $\gamma < 1/\alpha^{1/k}$ and the noise is a function of the marginal distribution of $\mathbb{R}^2$,*

Theorem 58 (**Lower bound of Gaussian DP covariance estimation** (Kamath et al., 2019, **Lemma 6.11**))
Let $\widehat{\Sigma} : \mathbb{R}^{n \times d} \to \Theta$ be an $(\varepsilon, 0)$-DP estimator (where Θ is the space of all $d \times d$ PSD matriaces), and for every $\mathcal{N}(0, \Sigma)$ over $\mathbb{R}^d$ such that $1/2\mathbf{I}_{d \times d} \preceq \Sigma \preceq 3/2\mathbf{I}_{d \times d}$,

$$\mathbb{E}_{S \sim \mathcal{N}(0, \Sigma)^n} \left[\|\widehat{\Sigma}(S) - \Sigma\|_F^2 \right] \leq \frac{\alpha^2}{64} \, , \tag{87}$$

then $n \geq \Omega(d^2/(\varepsilon\alpha))$.