Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms

Hilal Asi
Stanford University
asi@stanford.edu

John C. Duchi
Stanford University
jduchi@stanford.edu

Abstract

We study and provide instance-optimal algorithms in differential privacy by extending and approximating the inverse sensitivity mechanism. We provide two approximation frameworks, one which only requires knowledge of local sensitivities, and a gradient-based approximation for optimization problems, which are efficiently computable for a broad class of functions. We complement our analysis with instance-specific lower bounds for vector-valued functions, which demonstrate that our mechanisms are (nearly) instance-optimal under certain assumptions and that minimax lower bounds may not provide an accurate estimate of the hardness of a problem in general: our algorithms can significantly outperform minimax bounds for well behaved instances. Finally, we use our approximation framework to develop private mechanisms for unbounded-range mean estimation, principal component analysis, and linear regression. For PCA, our mechanisms give an efficient (pure) differentially private algorithm with near-optimal rates.

1 Introduction

We study the estimation of a function (statistic) of interest under differential privacy, where strong privacy protections usually decrease utility relative to non-private data analysis. In an effort to improve the utility of private algorithms, it is of utmost importance to design mechanisms that adapt to the hardness of the underlying data. Such mechanisms are of growing prevalence in the privacy literature, with prominent examples including the smooth sensitivity [25] and propose-test-release [12] frameworks.

To further investigate adaptivity to underlying instance, Asi and Duchi [4] recently study instance-optimal mechanisms—which, in a sense, achieve optimal utility for every possible data instance—in differentially private release of 1-dimensional quantities, moving beyond the more standard (worst case) minimax optimality. Inspired by classical statistical theory, Asi and Duchi develop local-minimax optimality and optimality against unbiased mechanisms, both of which aim to capture the hardness of the underlying data. By developing instance-specific lower bounds, they show that classical frameworks such as smooth sensitivity and propose-test-release may not be instance-optimal in general. To overcome this challenge, they investigate what they term the *inverse sensitivity mechanism*, showing it is instance-optimal for a wide range of functions.

Yet instance-optimality in private statistical estimation remains widely unexplored. First, the implementation of the inverse sensitivity mechanism requires a calculation of a particular sample distance (see Section 1.1.1), which may be intractable. Moreover, the current instance-optimality guarantees are not sharp for vector-valued functions. This is in part because the paper [4] tailors its instance-optimality notions for 1-dimensional functions by leveraging Stein's "hardest one-dimensional alternative" approach to lower bounds [31, 9], which gives tight lower bounds for 1-dimensional functions but fails to yield correct bounds in higher dimensions.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

To address these challenges, in this work we develop extensions and approximations to the inverse sensitivity mechanism with efficient implementations for a broad class of functions, which allows us to (for example) develop efficient algorithms for private PCA with near-optimal sample complexity. We also establish complementary instance-optimality results for vector-valued functions by proposing two approaches for instance-specific lower bounds: (i) a local minimax approach that measures the risk of an instance through the loss that an algorithm must incur on instances in a small neighborhood around it, and (ii) lower bounds against families of appropriately unbiased mechanisms, which includes many standard mechanisms. These instance-specific bounds suggest the limitations of more prevalent minimax (worst-case) bounds in privacy [19, 11]: they do not always give the correct limits on the performance of algorithms, and algorithms exist that achieve lower error on many instances.

1.1 Preliminaries

Given a function $f : \mathcal{X}^n \to \mathcal{T}$ and instance $\boldsymbol{x} \in \mathcal{X}^n$, we wish to design differentially private mechanisms that accurately estimates the value $f(\boldsymbol{x})$. We usually take $\mathcal{X}, \mathcal{T} \subset \mathbb{R}^d$ for a dimension d.

We begin by recalling the standard definition of differential privacy [16, 15]. We say that two instances $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ are *neighboring* if they differ in at most one example, that is, $d_{\text{ham}}(x, x') \le 1$.

Definition 1.1. *A randomized algorithm $M : \mathcal{X}^n \to \mathcal{T}$ is (ε, δ)-*differentially private *if for all neighboring datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ and all measurable $S \subseteq \mathcal{T}$,*

$$\mathbb{P}\left(M(\boldsymbol{x}) \in S\right) \le e^{\varepsilon}\mathbb{P}\left(M(\boldsymbol{x}') \in S\right) + \delta.$$

*If $\delta = 0$, then M is ε-*differentially private*.*

Given a loss function $L : \mathcal{T} \times \mathcal{T} \to \mathbb{R}_+$, we quantify the utility of a mechanism M on instance $\boldsymbol{x}$ through its expected loss $\mathbb{E}[L(M(\boldsymbol{x}), f(\boldsymbol{x}))]$. A mechanism is instance-optimal if it achieves the best utility for every instance. We formalize this through instance-specific lower bounds in Section 3.

For a function $f : \mathcal{X}^n \to \mathbb{R}$, the standard method to preserve privacy is the Laplace mechanism [16]. Defining the global sensitivity of f to be $\mathrm{GS}_f := \sup_{\boldsymbol{x},\boldsymbol{x}' : d_{\text{ham}}(\boldsymbol{x},\boldsymbol{x}') \le 1} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|$, it adds Laplace noise, $M_{\text{Lap}}(\boldsymbol{x}) := f(\boldsymbol{x}) + \frac{\mathrm{GS}_f}{\varepsilon}\mathsf{Lap}(1)$. This can be conservative, therefore Nissim et al. [25] consider the local sensitivity at instance $\boldsymbol{x}$ at hand $\mathrm{LS}_f(\boldsymbol{x}) := \sup_{\boldsymbol{x}' : d_{\text{ham}}(\boldsymbol{x},\boldsymbol{x}') \le 1} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|$. Directly using the local sensitivities may compromise privacy, hence the smooth sensitivity framework adds noise that is proportional to a smooth upper bound $S^\beta(\boldsymbol{x})$ on the local sensitivity, that is, $M_{\text{sm}}(\boldsymbol{x}) := f(\boldsymbol{x}) + \frac{2S^\beta(\boldsymbol{x})}{\varepsilon} Z$, where Z is sampled from an admissible noise distribution and $S^\beta(\boldsymbol{x})$ is the smooth sensitivity satisfying $\mathrm{LS}(\boldsymbol{x}) \le S^\beta(\boldsymbol{x})$ and $S^\beta(\boldsymbol{x}) \le e^\beta S^\beta(\boldsymbol{x}')$ for neighboring instances $\boldsymbol{x}, \boldsymbol{x}'$, and β is chosen appropriately to guarantee the desired privacy level.

1.1.1 The inverse sensitivity mechanism

Our work builds on the inverse sensitivity mechanism [4], which we review. Key to the mechanism is the path-length (inverse sensitivity), which, for a target t, measures how many users we must change in $\boldsymbol{x}$ to reach $\boldsymbol{x}'$ with a target value t:

$$\mathsf{len}_f(\boldsymbol{x}; t) := \inf_{\boldsymbol{x}'} \left\{ d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \mid f(\boldsymbol{x}') = t \right\}. \tag{1}$$

The basic inverse sensitivity mechanism then instantiates the exponential mechanism [24] with the path-length function (1), yielding the density

$$\pi_{M_{\text{inv}}(\boldsymbol{x})}(t) = \frac{e^{-\mathsf{len}_f(\boldsymbol{x};t)\varepsilon/2}}{\int_{\mathcal{T}} e^{-\mathsf{len}_f(\boldsymbol{x};s)\varepsilon/2} ds}. \tag{M.1}$$

A smoother variant of mechanism (M.1) is sometimes necessary to achieve instance-optimality, where one instead uses

$$\mathsf{len}_f^\rho(\boldsymbol{x}; t) = \inf_{s \in \mathcal{T} : \|s - t\| \le \rho} \mathsf{len}_f(\boldsymbol{x}; s),$$

with a smoothing parameter $\rho > 0$ [4]. Different variations of these mechanisms are instance-optimal for a range of real-valued functions. Yet while examples exist, it is often unclear how to compute the length (1).

Instance-specific bounds depend on the *modulus of continuity*, which (focusing in this work on a vector space $\mathcal{T}$ with norm $\|\cdot\|_p$) measures the sensitivity of a function when changing k users:

$$\omega_f^p(\boldsymbol{x}; k) = \sup_{\boldsymbol{x}' \in \mathcal{X}^n} \left\{ \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\|_p : d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le k \right\}. \tag{2}$$

Instance-specific lower bounds show that the risk we expect for an ε-differentially private algorithm on instance $\boldsymbol{x}$ is in general roughly $\omega_f(\boldsymbol{x}; 1/\varepsilon)$ for 1-dimensional functions (with $p = 1$) and loss $L(s, t) = |s - t|$ [4]. Unfortunately, this is not tight for d-dimensional functions.

Notation We denote samples using bold symbols $\boldsymbol{x} \in \mathcal{X}^n$ and individual examples using non-bold symbol $x \in \mathcal{X}$. We let $d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}')$ denote the Hamming distance of instance $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$. The local sensitivity of $f : \mathcal{X}^n \to \mathcal{T}$ at instance $\boldsymbol{x}$ is $\text{LS}_f^p(\boldsymbol{x}) = \omega_f^p(\boldsymbol{x}; 1)$, and the global sensitivity of f is $\text{GS}_f^p = \sup_{\boldsymbol{x} \in \mathcal{X}^n} \omega_f^p(\boldsymbol{x}; 1)$. To facilitate notation, we sometimes remove the superscript p if $p = 2$. We let $\text{diam}_p(\mathcal{T}) = \sup_{s,t \in \mathcal{T}} \|s - t\|_p$, and $\mathbb{B}_p^{d-1} = \{x \in \mathbb{R}^d : \|x\|_p \le 1\}$ denote the ℓ_p-ball.

1.2 Contributions

Approximate inverse sensitivity mechanisms We develop two approximation methods for the inverse sensitivity mechanism: (i) using local sensitivities in Section 2.1 and (ii) a gradient-based method for minimization problems in Section 2.2. These methods have efficient implementations for a wide range of problems and can outperform smooth sensitivity mechanisms for pure differential privacy. In contrast to Cauchy and Student's T distributions used in such instantiations [7]—which have infinite first and third moment respectively—our mechanisms add noise with bounded p'th moments for all finite p, resulting in improved high-probability bounds for utility analysis which is especially important for high dimensional functions as our examples demonstrate.

Instance-optimality and lower bounds We propose two notions of instance optimality for vector-valued functions and prove tight lower bounds for both notions in Section 3. Similarly to the 1-dimensional setting, our results give a characterization of the risk through the modulus of continuity. Combined with our instance-specific upper bounds, these bounds establish that approximate and exact inverse mechanisms are (nearly) instance-optimal for vector-valued functions under some assumptions.

Applications We study three problems that illustrate the methodological possibilities of the inverse sensitivity framework and its approximations in Section 4: mean estimation, PCA and linear regression. The utility improvements in these examples demonstrate the advantages of our mechanisms over standard frameworks and the importance of these notions of instance-optimality. Here we highlight the PCA example where smooth sensitivity algorithms require sample complexity (for dimension d and ignoring other parameters) $O(d^{3/2})$ [18], whereas our mechanisms require $O(d)$ samples, which is the optimal dependence on the dimension d according to PCA lower bounds [10, 21].

1.3 Related work

The most widely used frameworks for instance-dependent noise are smooth sensitivity [25] and propose-test-release [12]. The former adds noise that scaling with a smooth upper bound on the local sensitivity, and the latter adds noise scaling with a prespecified upper bound on the local sensitivity—whose validity the algorithm tests—in a neighborhood of the instance. Applications are numerous: Smith and Thakurta [30] develop an algorithm based on propose-test-release for high-dimensional regression problems, and Bun and Steinke [7] design noise distributions for smooth sensitivity and use them to estimate the mean of distributions with unbounded range. Other applications include principal component analysis [18], outlier analysis [26], and graph data [22, 32]. The inverse sensitivity framework is a distinct approach to instance-dependent noise that Asi and Duchi [4] investigate ([20, 29, 8] propose variants of the mechanism). Their results suggest that this framework, in contrast to smooth sensitivity and propose-test-release, is instance-optimal for a range of functions, and can have quadratically better sample complexity than smooth sensitivity mechanisms.

2 Approximate inverse sensitivity mechanisms

Having described the difficulty of sampling from the inverse sensitivity mechanism in general, in this section we develop two approximation frameworks that are applicable for a broader range of functions while maintaining some of the instance-optimality guarantees of the exact mechanism. First, in Section 2.1 we describe a method that uses the local sensitivities to approximate the path-length, and in Section 2.2 we describe an approximation for the specific setting of empirical risk minimization.

2.1 Approximation using local sensitivities

The mechanisms we develop in this section first construct an approximation $\overline{\mathsf{len}}_f(\boldsymbol{x}; t)$ for the path-length, then apply the exponential mechanism for a base measure μ on $\mathcal{T}$ with this approximation

$$\pi_{M_{\mathrm{appr}}(\boldsymbol{x})}(t) = \frac{e^{-\overline{\mathsf{len}}_f(\boldsymbol{x};t)\varepsilon/2}}{\int_{\mathcal{T}} e^{-\overline{\mathsf{len}}_f(\boldsymbol{x};s)\varepsilon/2} d\mu(s)}. \tag{M.2}$$

Our main tool for calculating $\overline{\mathsf{len}}_f(\boldsymbol{x}; t)$ are the local sensitivities of instance $\boldsymbol{x}$ at distance ℓ

$$\mathrm{LS}^p_\ell(\boldsymbol{x}) = \sup_{\boldsymbol{x}': d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}') = \ell} \mathrm{LS}^p(\boldsymbol{x}').$$

This definition implies that changing k users can vary the function value by at most $\sum_{\ell=1}^k \mathrm{LS}^p_\ell(\boldsymbol{x})$. As a consequence, we have the lower bound $\mathsf{len}_f(\boldsymbol{x}; t) \geq \min\{k : \sum_{i=1}^k \mathrm{LS}^p_i(\boldsymbol{x}) \geq \|t - f(\boldsymbol{x})\|_p\}$. Unfortunately, directly using this lower bound may result in mechanisms that are not private. The following theorem shows how to construct suitable approximations that preserve privacy.

Theorem 1. *Let $f : \mathcal{X}^n \to \mathcal{T}$ and $R_\ell : \mathcal{X}^n \to \mathbb{R}$ satisfy $\mathrm{LS}^p(\boldsymbol{x}) \leq R_1(\boldsymbol{x})$ and $R_\ell(\boldsymbol{x}) \leq R_{\ell+1}(\boldsymbol{x}')$ for any neighboring instances $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$. Then, using the approximation*

$$\overline{\mathsf{len}}_f(\boldsymbol{x}; t) = \min\left\{ k : \sum_{i=1}^k R_i(\boldsymbol{x}) \geq \|t - f(\boldsymbol{x})\|_p \right\}, \tag{3}$$

mechanism (M.2) is ε-differentially private.

Algorithm 1 efficiently samples from the approximate inverse sensitivity mechanism for reasonable choices of p: the main bottleneck is step 3 but efficient algorithms exist for $p \in \{1, 2\}$ using truncated Gamma distributions [23].

Algorithm 1: Sampling from approximate inverse sensitivity

Input: $\boldsymbol{x} \in \mathcal{X}^n$, p, $\{R_i(\cdot)\}_{i=1}^n$

1 Denote $S_k = \{t : \sum_{\ell=1}^{k-1} R_\ell(\boldsymbol{x}) \leq \|t\|_p \leq \sum_{\ell=1}^k R_\ell(\boldsymbol{x})\}$;

2 Sample $k \sim K$ from $\mathbb{P}(K = k) \propto \mathrm{Vol}(S_k) e^{-k\varepsilon/2}$ for $1 \leq k \leq n$;

3 Sample $z \sim \mathsf{Uni}(S_k)$;

4 **return** $f(\boldsymbol{x}) + z$

Before proceeding to our utility analysis, we show that given the local sensitivities, we can always find an appropriate choice of R_i without calculating the smooth sensitivities.

Proposition 2.1. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$ and assume $\overline{\mathrm{LS}}(\boldsymbol{x})$ is such that $\mathrm{LS}(\boldsymbol{x}) \leq \overline{\mathrm{LS}}(\boldsymbol{x})$ for every $\boldsymbol{x}$. Then mechanism (M.2) using the approximation (3) with $R_\ell(\boldsymbol{x}) = \sup_{\boldsymbol{x}': d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}') \leq \ell} \overline{\mathrm{LS}}(\boldsymbol{x})$ is ε-differentially private.*

2.1.1 Utility guarantees for vector-valued functions

In this section, we provide utility guarantees for the exact and approximate inverse sensitivity mechanisms for vector-valued functions. Combined with our lower bounds of Section 3, this establishes (near) instance optimality of these methods. Our guarantees hold with high probability, in contrast to those of the smooth sensitivity framework which uses distributions with heavy tails. We also show that our approximations can outperform smooth Laplace for real-valued functions.

We begin by analyzing the utility of the exact and approximate inverse sensitivity mechanisms.

Theorem 2. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$, $\operatorname{diam}_2(f(\mathcal{X}^n)) \le D$, $r > 0$ and $1 \le K \le n$. Then the (smooth) inverse sensitivity mechanism (M.1) with $\rho = 1/n^r$ has*

$$\mathbb{P}\left(\|M_{\text{inv}}(\boldsymbol{x}) - f(\boldsymbol{x})\|_2 \ge \omega_f(\boldsymbol{x}; K) + 1/n^r\right) \le e^{-K\varepsilon/2}(n^r D)^d.$$

Moreover, if $R_i(\boldsymbol{x}) \le D$, the approximate mechanism (M.2) using (3) with $p = 2$ has

$$\mathbb{P}\left(\|M_{\text{appr}}(\boldsymbol{x}) - f(\boldsymbol{x})\|_2 \ge \sum_{i=1}^{K} R_i(\boldsymbol{x})\right) \le e^{-K\varepsilon/2+1}\left(nD/\sum_{i=1}^{1/\varepsilon} R_i(\boldsymbol{x})\right)^d.$$

We remark that using the smooth sensitivity framework to preserve pure differential privacy does not usually result in such high probability bounds due to using noise distributions with heavy tails such as Cauchy distribution [25]. Moreover, Theorem 2 implies that using $k \approx \frac{Cd\log n}{\varepsilon}$ for large constant C, with high probability the inverse sensitivity mechanism roughly has

$$\|M_{\text{inv}}(\boldsymbol{x}) - f(\boldsymbol{x})\|_2 \le O(\omega_f(\boldsymbol{x}; Cd\log n/\varepsilon)).$$

The approximate mechanism has similar loss whenever our approximate R_i are accurate such that $\sum_{i=1}^{K} R_i(\boldsymbol{x}) = O(\omega_f(\boldsymbol{x}; K))$. The lower bounds in Section 3 show this is (near) instance optimal.

We conclude this section with another choice of R_i that uses the smooth sensitivities instead. This guarantees that the approximate mechanism always outperforms the smooth Laplace mechanism [25].

Proposition 2.2. *Let $f : \mathcal{X}^n \to \mathbb{R}$, $\varepsilon = O(1)$ and $R_\ell(\boldsymbol{x}) = \sup_{\boldsymbol{x}' : d_{\text{ham}}(\boldsymbol{x},\boldsymbol{x}') \le \ell} \mathrm{S}^\beta(\boldsymbol{x}')$. Then mechanism (M.2) using (3) is ε-differentially private. If $p = 1$ and $\beta = \frac{\varepsilon}{8}$ then $\mathbb{E}\left[|M_{\text{appr}}(\boldsymbol{x}) - f(\boldsymbol{x})|\right] \le O(\frac{\mathrm{S}^\beta(\boldsymbol{x})}{\varepsilon})$.*

The smooth Laplace mechanism—which guarantees only approximate $(\varepsilon, \delta > 0)$-DP—has loss $O(\frac{\mathrm{S}^\beta(\boldsymbol{x})}{\varepsilon})$ with a much smaller $\beta = \frac{\varepsilon}{2\log 2/\delta}$, which can be $\log 1/\delta$ worse in some settings.

2.2 Gradient-based approximations for empirical risk minimization

In this section, we describe our second approximation which applies to empirical risk minimization problems. Given data points $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ and L-Lipschitz loss function $\ell(\theta; x_i)$ for $\theta \in \Theta$, we wish to solve the following minimization problem

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta\in\Theta} L_n(\theta; \boldsymbol{x}, \boldsymbol{y}) := \frac{1}{n}\sum_{i=1}^{n} \ell(\theta; x_i, y_i).$$

It is possible to calculate the path-length using gradients for robust regression [4]. Here, we use similar techniques to approximate the inverse sensitivity mechanism in general settings. As ℓ is L-Lipschitz, we need to change $\mathsf{len}(\boldsymbol{x}, \boldsymbol{y}; \theta) \ge \frac{n}{L}\|\nabla L_n(\theta; \boldsymbol{x}, \boldsymbol{y})\|_2$ users to make θ a minimizer with $\nabla L_n(\theta; \boldsymbol{x}', \boldsymbol{y}') = 0$. The *gradient mechanism* uses this approximation of len, resulting in the density

$$\pi_{\mathsf{Grad}}(\theta \mid \boldsymbol{x}, \boldsymbol{y}) \propto e^{-\frac{n\varepsilon}{2L}\|\nabla L_n(\theta; \boldsymbol{x}, \boldsymbol{y})\|_2}. \tag{4}$$

Sampling from this distribution can be hard in general, but we show an efficient implementation for linear regression in Section 4.3. For general twice differentiable functions, we propose an efficient heuristic of the gradient mechanism based on Taylor's expansion which gives $\nabla L_n(\theta; \boldsymbol{x}, \boldsymbol{y}) \approx \nabla^2 L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})(\theta - \hat{\theta}_n)$. Letting $\mathrm{GS}_{\mathsf{Hess}}$ denote the global sensitivity of $\|\nabla^2 L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})(\theta - \hat{\theta}_n)\|_2$, we define the following *Hessian-based mechanism* for $\theta \in \Theta$

$$\pi_{\mathsf{Hess}}(\theta \mid \boldsymbol{x}, \boldsymbol{y}) \propto e^{-\frac{\varepsilon}{2\mathrm{GS}_{\mathsf{Hess}}}\left\|\nabla^2 L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})(\theta - \hat{\theta}_n)\right\|_2}. \tag{5}$$

The main advantage of the Hessian mechanism (5) is that now we can design efficient and simple sampling procedures (see Section 4.3). It also provides an accurate approximation of the gradient mechanism with good utility whenever $\nabla^2 L_n$ is H-Lipschitz with small H.

The privacy of these mechanisms follow immediately from the privacy of the exponential mechanism. For utility, we start with the following lemma which upper bound $\mathrm{GS}_{\mathsf{Hess}}$.

Lemma 2.1. *Assume $\ell(\cdot; x_i, y_i)$ is L-Lipschitz, $\nabla^2 L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) \succeq \lambda I$, $\hat{\theta}_n \in \operatorname{int}\Theta$, and $\nabla^2 L_n(\cdot; \boldsymbol{x}, \boldsymbol{y})$ is H-Lipschitz. If $H \le \lambda = O(1)$ and $n \ge 4L \operatorname{diam}_2(\Theta) + 1$ then $\mathrm{GS}_{\mathsf{Hess}} \le O\left(\frac{L}{n}\right)$.*

We are now ready to analyze the utility of the Hessian-based mechanism.

Proposition 2.3. *Let the set of instances $(x_i, y_i)_{i=1}^n$ satisfy the assumptions of Lemma 2.1. Then the Hessian mechanism (5) is ε-DP. If $\inf_{\theta \in \operatorname{bd}\Theta} \|\theta - \hat{\theta}_n\|_2 \ge \Omega(\frac{dL^2}{n^2\varepsilon^2} \operatorname{tr}(\nabla^2 L(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})^{-2}))$, then*

$$\mathbb{E}_{\theta \sim \pi_{\mathsf{Hess}}(\cdot|\boldsymbol{x},\boldsymbol{y})}\left[\|\theta - \hat{\theta}_n\|_2^2\right] \le O\left(\frac{dL^2 \operatorname{tr}(\nabla^2 L(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})^{-2})}{n^2\varepsilon^2}\right).$$

When $H = 0$, the gradient (4) and Hessian mechanisms (5) are identical and the gradient mechanism has the same utility. In Section 4.3, we use these mechanisms for solving regression problems and show the significant advantages of instance-specific bounds over standard minimax bounds.

Finally, we remark that our gradient-based approximations of the inverse sensitivity mechanism are closely related to the K-norm mechanism [27] where the authors use gradient norms as a score function for the exponential mechanism. However, their work only provides asymptotic utility analyses without finite-sample guarantees, and they propose an approximate implementation of their mechanisms using an MCMC procedure without providing privacy guarantees for the implementation.

3 Instance-specific lower bounds for vector-valued functions

Given a function $f : \mathcal{X}^n \to \mathbb{R}^d$, in this section we prove instance-specific lower bounds on the loss that any private mechanism must incur. Unfortunately, the instance-specific notions in [4] were tailored for 1-dimensional functions, hence do not result in satisfactory lower bounds in our setting. To this end, we propose extensions that result in tight bounds. The first notion gives lower bounds by restricting to families of appropriately unbiased mechanisms. The second is a local-minimax approach that measures the performance in a small neighborhood around a given instance.

We begin with our optimality notion for unbiased mechanisms, which we define now.

Definition 3.1. *We say that a randomized algorithm M is $\|\cdot\|$-unbiased if for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ and $\rho > 0$,*

$$\mathbb{P}(\|M(\boldsymbol{x}) - f(\boldsymbol{x})\| \le \rho) \ge \mathbb{P}(\|M(\boldsymbol{x}) - f(\boldsymbol{x}')\| \le \rho).$$

Definition 3.1 says that when applying an unbiased mechanism M on instance $\boldsymbol{x}$, the output is more likely to be in a ball around the correct value $f(\boldsymbol{x})$ rather than $f(\boldsymbol{x}')$ for some other instance $\boldsymbol{x}'$. Anderson's theorem [2] implies that the Laplace mechanism, Gaussian mechanism, their smooth sensitivity instantiations, the approximate inverse sensitivity mechanism, and any instantiation of the exponential mechanism with a concave score function are $\|\cdot\|$-unbiased.

Our lower bounds require a growth condition on the set of values at distance at most k, $W_f(\boldsymbol{x}; k) = \{f(\boldsymbol{x}') : d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le k\}$. We have the following instance-specific lower bound for unbiased mechanisms.

Theorem 3. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$ and assume $W_f(\boldsymbol{x}; k) \supseteq c \cdot \omega_f(\boldsymbol{x}; k) \cdot \mathbb{B}_2^{d-1}$ for $c > 0$. If M is ε-DP,*

$$\sup_{\boldsymbol{x} \in \mathcal{X}^n} \mathbb{E}\left[\|M(\boldsymbol{x}) - f(\boldsymbol{x})\|_2\right] \ge \frac{c}{8} \sup_{\boldsymbol{x} \in \mathcal{X}^n} \max_{1 \le k \le n} e^{-k\varepsilon/d} \omega_f(\boldsymbol{x}; k).$$

Moreover, if M is $\|\cdot\|_2$-unbiased, then for any $\boldsymbol{x} \in \mathcal{X}^n$,

$$\mathbb{E}\left[\|M(\boldsymbol{x}) - f(\boldsymbol{x})\|_2\right] \ge \frac{c}{8} \max_{1 \le k \le n} e^{-2k\varepsilon/d} \omega_f(\boldsymbol{x}; k).$$

Theorem 3 suggests that worst-case lower bounds may be too pessimistic: while the minimax risk is roughly $\sup_{\boldsymbol{x} \in \mathcal{X}^n} \omega_f(\boldsymbol{x}; d/\varepsilon)$, we may hope to achieve a better risk for instance $\boldsymbol{x}$, that is, $\omega_f(\boldsymbol{x}; d/\varepsilon)$.

Now we define the local-minimax risk for an instance $\boldsymbol{x}$ following similar ideas in statistical theory [cf. 33, Ch. 8]. Let $\mathcal{M}_\varepsilon$ be the family of ε-differentially private mechanisms. For a radius r, we define the local-minimax risk of $\boldsymbol{x}$ to be the worst-case risk in a small neighborhood around $\boldsymbol{x}$, that is,

$$\mathcal{R}(\boldsymbol{x}; r) := \inf_{M \in \mathcal{M}_\varepsilon} \sup_{\boldsymbol{x}' : d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le r} E\left[\|M(\boldsymbol{x}') - f(\boldsymbol{x}')\|_2\right]. \tag{6}$$

The choice of r in this definition is important to exclude trivial mechanisms such as $M(\boldsymbol{x}') = f(\boldsymbol{x})$; briefly, we choose the smallest radius that excludes such mechanisms, which is in this case $r = \Theta(d/\varepsilon)$ (see Appendix C.2 for more details about this definition).

We have the following lower bound for the local-minimax risk.

Theorem 4. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$ and assume $W_f(\boldsymbol{x}; k) \supseteq c \cdot \omega_f(\boldsymbol{x}; k) \cdot \mathbb{B}_2^{d-1}$ for $c > 0$. Then for any $\boldsymbol{x} \in \mathcal{X}^n$, $\mathcal{R}(\boldsymbol{x}; d/\varepsilon) \geq \Omega\left(\omega_f(\boldsymbol{x}; d/\varepsilon)\right)$.*

Similarly to our lower bounds for unbiased mechanisms, Theorem 4 shows that any mechanism must incur local-minimax risk roughly $\omega_f(\boldsymbol{x}; d/\varepsilon)$ for instance $\boldsymbol{x}$. The upper bounds of Theorem 2 show that the exact inverse mechanism achieves this loss for every instance up to logarithmic factors, as well as the approximate version if the approximations R_i are accurate.

4 Applications

We investigate three examples that demonstrate different advantages and applications of the exact, approximate, and gradient inverse sensitivity mechanisms. Our examples include (i) mean estimation with unbounded range, (ii) principal component analysis and (iii) linear regression, and show that our techniques yield private algorithms with better noise distributions resulting in improved utility, which in some cases can significantly outperform existing minimax-optimal algorithms.

4.1 Unbounded-range mean estimation

Given $x_i \overset{\text{iid}}{\sim} P$ with unbounded range, our goal is to privately estimate the mean $\mu = \mathbb{E}_{x\sim P}[x]$. The difficulty here is that the empirical mean has infinite global and even local sensitivity, leading Bun and Steinke [7] to use the trimmed mean which calculates the mean after removing the smallest and largest m samples. Letting $x_{(1)} \leq \cdots \leq x_{(n)}$ denote the order statistics, the trimmed mean is

$$\mathsf{trim}_m(\boldsymbol{x}) = \frac{x_{(m+1)} + x_{(m+2)} + \cdots + x_{(n-m)}}{n - 2m}. \tag{7}$$

This is useful as it leads to small local sensitivity under distributional assumptions. Bun and Steinke [7] use the smooth sensitivity to estimate the trimmed mean, resulting in strong utility but only with the weaker concentrated differential privacy [14, 6]. To preserve pure differential privacy, they use Student's T distribution which has infinite third moments and consequently heavy tails.

We use the exact inverse mechanism to estimate the mean with strong utility. This algorithm has finite p'th moment for any finite p, therefore yields tight confidence intervals. We assume $\mu \in [a, b]$ and let $[c]_{[a,b]}$ denote projection to $[a, b]$. The following lemma enables exact calculation of the path-length.

Lemma 4.1. *Let $f(\boldsymbol{x}) = [\mathsf{trim}_m(\boldsymbol{x})]_{[a,b]}$. Then, for any $t \in [a, b]$, if $t \geq \mathsf{trim}_m(\boldsymbol{x})$, we have $\mathsf{len}_f(\boldsymbol{x}; t) = \min\{k : k \leq m, t - f(\boldsymbol{x}) \leq \frac{1}{n-2m}\sum_{i=1}^{k}(x_{(n-m+i)} - x_{(m+i)})\} \cup \{m + 1\}$.*

The calculation for $t < \mathsf{trim}_m(\boldsymbol{x})$ is similar and we present it in Appendix D. Using Lemma 4.1, we can efficiently sample (Algorithm 4 in Appendix D.1 which runs in $O(n \log n)$ time) from the inverse sensitivity mechanism. To analyze the performance of this algorithm, we assume P is σ-subgaussian while noting that these results can be extended to other settings in [7]. The following proposition upper bounds the error of our algorithm, which resembles the bounds that the algorithms of [7] achieve with the weaker concentrated differential privacy.

Proposition 4.1. *Let $a, b \in \mathbb{R}$ and $x_i \overset{\text{iid}}{\sim} P$ where P is σ-subgaussian with mean $\mu \in [a, b]$. If P is symmetric about its mean and $n \geq \frac{12\log(n(b-a)/\sigma^2)}{\varepsilon}$, the inverse sensitivity mechanism (Algorithm 4) with $\rho = \frac{\sigma^2}{n^2}$ is ε-differentially private and has*

$$\mathbb{E}[(\hat{x} - \mu)^2] \leq \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2} \cdot O\left(\frac{\log((b-a)/\sigma)}{\varepsilon} + \frac{\log n}{\varepsilon^2}\right).$$

4.2 Principal component analysis

In this section, we apply our approximations to calculate a rank k approximation of a matrix. Given $x_1, \ldots, x_n \in \mathbb{B}_2^{d-1}$ with covariance $\Sigma(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n x_i x_i^T$, we wish to find $V \in \mathbb{R}^{d\times k}$ that solves

$$\hat{V}(\boldsymbol{x}) = \operatorname*{argmax}_{V:V^TV=I_k} F(V) \coloneqq \operatorname{tr}(V^T\Sigma(\boldsymbol{x})V) \tag{8}$$

Gonen and Gilad-Bachrach [18] design ε-DP algorithms based on the smooth sensitivity framework with suboptimal error of roughly $\frac{d^{3/2}}{n\mathsf{GAP}(\boldsymbol{x})\varepsilon}$. We show that our algorithms achieve a near-optimal rate $\frac{d}{n\mathsf{GAP}(\boldsymbol{x})\varepsilon}$. Though there exist algorithms that achieve this rate using the exponential mechanism [10, 21], these algorithms require sampling from complex distributions and the only implementation with theoretical runtime analysis requires $O(d^6)$ time. In contrast, given the eigenvectors, our algorithm (Algorithm 2) returns a private version of the leading eigenvector in time $O(n+d)$ with high probability.

We only consider $k = 1$ as extensions to larger k are straightforward using QR factorization [18]. Our algorithm builds on techniques from [18] and the approximate inverse mechanism. It requires a non-private PCA algorithm $\mathcal{A}_1$ that calculates the first eigenvector $\hat{v} \in \mathbb{R}^d$ (which maximizes (8)) and the gap between the two largest eigenvalues $\mathsf{GAP}(\boldsymbol{x}) \coloneqq \lambda_1(\Sigma(\boldsymbol{x})) - \lambda_2(\Sigma(\boldsymbol{x}))$. Then it randomly flips the sign of $\hat{v}$ as $-\hat{v}$ is also a solution, and adds noise using the approximate inverse sensitivity. Algorithm 2 describes our private PCA procedure. Given the output of the non-private PCA algorithm, the main computational difficulty in Algorithm 2 is step 3 which requires sampling from the noise distribution of Algorithm 1. Using the rejection-sampling algorithms of Laud et al. [23] for sampling from truncated Gamma distributions (which has constant success probability in our setting), we can efficiently sample from Algorithm 1 in time $O(n+d)$ with high probability.

Algorithm 2: Private PCA using approximate inverse sensitivity

Input: $\boldsymbol{x}$

1 Calculate $\hat{v} = \mathcal{A}_1(\boldsymbol{x})$, $\mathsf{GAP} = \mathsf{GAP}(\boldsymbol{x})$;
2 Set $\overline{v} = B\hat{v}$ for $B \sim \mathsf{Uni}\{-1, +1\}$;
3 Sample z from (M.2) using Algorithm 1 with $p = 2$, $R_i = \min(C_{\mathsf{pca}}/(n\mathsf{GAP} - 2k), \sqrt{2})$;
4 **return** $v_{\mathsf{out}} = \frac{\overline{v}+z}{\|\overline{v}+z\|_2}$;

Following [18], we define the local sensitivity ($k = 1$) while taking into consideration the vector sign

$$\mathrm{LS}(\boldsymbol{x}) = \sup_{\boldsymbol{x}':d_{\mathrm{ham}}(\boldsymbol{x},\boldsymbol{x}')\le 1} \min(\|\hat{V}(\boldsymbol{x}) - \hat{V}(\boldsymbol{x}')\|_2, \|\hat{V}(\boldsymbol{x}) + \hat{V}(\boldsymbol{x}')\|_2).$$

We build on the following key lemma that bounds the local sensitivity.

Lemma 4.2 ([18], Theorem 5, Lemma 11)**.** *If* $\mathsf{GAP}(\boldsymbol{x}) > 0$ *then there is a universal constant* $C_{\mathsf{pca}} < \infty$ *such that* $\mathrm{LS}(\boldsymbol{x}) \le \min(\frac{C_{\mathsf{pca}}}{n\mathsf{GAP}(\boldsymbol{x})}, \sqrt{2})$*. Moreover,* $|\mathsf{GAP}(\boldsymbol{x}) - \mathsf{GAP}(\boldsymbol{x}')| \le 2d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}')/n$*.*

Using this bound and the guarantees of our approximate mechanism, we get the following proposition.

Proposition 4.2. *Assume* $n \ge 1/C_{\mathsf{pca}}$*,* $\beta > 0$ *and* $\Omega(\frac{d}{\mathsf{GAP}(\boldsymbol{x})\varepsilon}) \le \frac{n}{\log n/\beta}$*. Algorithm 2 is* ε*-differentially private and with probability* $1-\beta$*,*

$$|F(v_{\mathsf{out}}) - F(\hat{v})| \le O\left(\frac{d\log n/\beta}{n\mathsf{GAP}(\boldsymbol{x})\varepsilon} + \frac{1}{n^4}\right).$$

4.3 Linear regression

For our final example, we investigate the setting of linear regression where we have data points $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$. Our goal here is to find $\theta \in \Theta$ that minimizes

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta\in\Theta} L_n(\theta; \boldsymbol{x}, \boldsymbol{y}) \coloneqq \frac{1}{2n}\sum_{i=1}^n (\langle\theta, x_i\rangle - y_i)^2.$$

We let $X \in \mathbb{R}^{n\times d}$ has x_i in the i'th row, $\Sigma_n = \frac{1}{n}X^TX$, and $\boldsymbol{y} \in \mathbb{R}^d$ denote the vector of y_i.

The gradient mechanism (4) have an efficient implementation in this setting. Assuming $\Sigma_n \succ 0$, we get that $\nabla L_n(\theta) = \Sigma_n(\theta - \bar{\theta})$ where $\bar{\theta} = \frac{1}{n}\Sigma_n^{-1} X^T \boldsymbol{y}$. The gradient mechanism has density

$$\pi(\bar{\theta} + \Delta \mid \boldsymbol{x}) \propto e^{-\frac{n\varepsilon}{2L}\|\Sigma_n \Delta\|_2},$$

for $\bar{\theta} + \Delta \in \Theta$ (and Lipschitz constant L) which Algorithm 3 samples from (see Appendix F.1). The main difficulty in Algorithm 3 is calculating the non-private estimator $\bar{\theta}$ in step 1 while the remaining steps (for private noise addition) only require sampling from simple distributions and matrix-vector products.

Algorithm 3: Gradient mechanism for linear regression

Input: $(x_i, y_i)_{i=1}^n$

1 Calculate $\Sigma_n = \frac{1}{n} X^T X$, $\bar{\theta} = \frac{1}{n}\Sigma_n^{-1} X^T \boldsymbol{y}$;

2 Sample $R \sim \mathsf{Gamma}(d, 1)$, $U \sim \mathsf{Uni}(\mathbb{S}^{d-1})$;

3 Set $\theta_{\mathsf{out}} = \bar{\theta} + \frac{2L}{n\varepsilon}\Sigma_n^{-1} \cdot R \cdot U$;

4 **if** $\theta_{\mathsf{out}} \notin \Theta$ **then** go to 3 ;

5 **return** θ_{out};

The following proposition states the utility and privacy guarantees of Algorithm 3.

Proposition 4.3. *For the set of instances $(x_i, y_i)_{i=1}^n$ with $\Sigma_n(\boldsymbol{x}) \succ 0$ and Lipschitz constant L, Algorithm 3 is ε-differentially private. Moreover, if $\inf_{\theta \in \mathrm{bd}\,\Theta} \|\theta - \hat{\theta}_n\|_2 \geq \Omega\left(\frac{dL^2}{n^2\varepsilon^2} \operatorname{tr}\left(\Sigma_n(\boldsymbol{x})^{-2}\right)\right)$,*

$$\mathbb{E}\left[L_n(\theta_{\mathsf{out}}; \boldsymbol{x}, \boldsymbol{y}) - L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})\right] \leq O\left(\frac{dL^2 \operatorname{tr}\left(\Sigma_n(\boldsymbol{x})^{-1}\right)}{n^2\varepsilon^2}\right).$$

To appreciate the instance-specific upper bounds of Proposition 4.3, recall that existing private algorithms for empirical risk minimization of L-Lipschitz and λ-strongly convex functions achieve excess loss $\mathbb{E}[L_n(\theta) - L_n(\hat{\theta}_n)] = O(\frac{d^2L^2}{n^2\varepsilon^2\lambda})$ which is minimax optimal in some regimes [5]. In contrast, for natural instances where $\Sigma_n(\boldsymbol{x})^{-1}$ has polynomially decaying eigenvalues $\lambda_j = j^{-\alpha}$ for $\alpha \in (0, 1]$, Proposition 4.3 implies that Algorithm 3 achieves excess loss $\widetilde{O}(\frac{d^{2-\alpha}L^2}{n^2\varepsilon^2})$ which can offer up to $\widetilde{O}(d)$ improvement. Finally, we note that Wang [34]—which focuses on approximate (ε, δ)-DP—develops private algorithms for linear regression that exhibit good adaptivity to the difficulty of the underlying instance. There exists an extensive prior work on private linear regression and—as this is not the main focus of our work—we refer the reader to [34, 28] for a survey of results.

Comparison with the smooth sensitivity framework We conclude the paper with a short comparison of the smooth sensitivity framework and inverse sensitivity mechanisms. While smooth sensitivity mechanisms may not be instance-optimal in many settings, Asi and Duchi [4] show that the inverse sensitivity mechanism is (nearly) instance-optimal for most well-behaved functions and can offer quadratic improvement in sample complexity over smooth sensitivity mechanisms in certain settings. The inverse sensitivity mechanism also outperforms smooth Laplace uniformly for every instance for natural families of sample-monotone functions (see Section 4.3 in [4]). As our development in this paper shows, the approximate versions of the inverse sensitivity mechanism still enjoy similar advantages over smooth mechanisms. Proposition 2.2 shows that—for certain choices of approximations—the approximate inverse sensitivity mechanisms uniformly outperform the smooth Laplace mechanism for every instance. Moreover, the smooth sensitivity framework requires adding noise with heavy-tailed distributions and unbounded moments (such as Cauchy) to preserve ε-differential privacy, in contrast to the approximate inverse sensitivity mechanisms which (depending on the approximation and inverse sensitivity) has noise with exponentially decaying tails, resulting in better high-probability bounds and confidence intervals. The PCA example clearly demonstrates these advantages where the approximate inverse sensitivity mechanism enjoys a factor of $\sqrt{d}$ improvement in sample complexity over smooth sensitivity mechanisms.

Broader Impact

The substantial growth in data collection and analysis and the increasing awareness for privacy concerns has led to a growing body of work on privacy risks in both academic [16] and industrial settings [17, 3]. Differential privacy [16] has emerged as the standard method for preserving privacy and has enjoyed several applications including in statistical estimation [11], machine learning [5], and game theory [24].

Unfortunately, it is usually challenging to develop private algorithms that achieve satisfactory utility [11]. Therefore, while differential privacy has been successfully deployed in several industrial companies, most applications instantiate a large privacy parameter ε to achieve acceptable utility, potentially compromising the privacy of users [1].

However, the standard approach in differential privacy to measure the performance of an algorithm is through its (worst case) minimax risk [11]. This—as our theory demonstrates—may be too pessimistic in general and may not capture the correct trade-off between privacy and utility for natural data that arises in real-life. An instance-specific understanding of this trade-off can therefore result in significant improvements in both utility and privacy.

We hope that this work—and instance-optimality in differential privacy in general [4]—can lead to a better understanding of the privacy-utility trade-off of private algorithms for the underlying data at hand. By exploiting the average-case nature of data in real life, we believe that the instance-optimal algorithms we develop can achieve satisfying utility with significantly stronger privacy protections for users.

Funding Transparency Statement

Funding in direct support of this work: NSF CAREER CCF-1553086, ONR YIP N00014-19-2288, Sloan Foundation, NSF HDR 1934578 (Stanford Data Science Collaboratory), and Stanford DAWN Consortium.

References

[1] M. Abadi, U. Erlingsson, I. Goodfellow, H. B. McMahan, N. Papernot, I. Mironov, K. Talwar, and L. Zhang. On the protection of private information in machine learning systems: Two recent approaches. In *30th IEEE Computer Security Foundations Symposium (CSF)*, pages 1–6. IEEE, 2017.

[2] T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955.

[3] Apple Differential Privacy Team. Learning with privacy at scale, 2017. Available at https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

[4] H. Asi and J. Duchi. Near instance-optimality in differential privacy. *arXiv:2005.10630 [cs.CR]*, 2020.

[5] R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *55th Annual Symposium on Foundations of Computer Science*, pages 464–473, 2014.

[6] M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference (TCC)*, pages 635–658, 2016.

[7] M. Bun and T. Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. In *Advances in Neural Information Processing Systems 32*, pages 181–191, 2019.

[8] M. Bun, G. Kamath, T. Steinke, and Z. S. Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems 32*, pages 156–167, 2019.

[9] T. Cai and M. Low. A framework for estimating convex functions. *Statistica Sinica*, 25:423–456, 2015.

[10] K. Chaudhuri, A. Sarwate, and K. Sinha. Near-optimal algorithms for differentially-private principal components. In *Advances in Neural Information Processing Systems 25*, 2012. URL http://arxiv.org/abs/1207.2812.

[11] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation (with discussion). *Journal of the American Statistical Association*, 113(521):182–215, 2018.

[12] C. Dwork and J. Lei. Differential privacy and robust statistics. In *Proceedings of the Forty-First Annual ACM Symposium on the Theory of Computing*, pages 371–380, 2009.

[13] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3 & 4):211–407, 2014.

[14] C. Dwork and G. Rothblum. Concentrated differential privacy. *arXiv:1603.01887 [cs.DS]*, 2016.

[15] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology (EUROCRYPT 2006)*, 2006.

[16] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference*, pages 265–284, 2006.

[17] U. Erlingsson, V. Pihur, and A. Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS)*, 2014.

[18] A. Gonen and R. Gilad-Bachrach. Smooth sensitivity based approach for differentially private PCA. In *Algorithmic Learning Theory*, pages 438–450, 2018.

[19] M. Hardt and K. Talwar. On the geometry of differential privacy. In *Proceedings of the Forty-Second Annual ACM Symposium on the Theory of Computing*, pages 705–714, 2010. URL http://arxiv.org/abs/0907.3754.

[20] A. Johnson and V. Shmatikov. Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1079–1087, 2013.

[21] M. Kapralov and K. Talwar. On differentially private low rank approximation. In *Proceedings of the Twenty-Fourth ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1395–1414, 2013.

[22] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In A. Sahai, editor, *Theory of Cryptography*, volume 7785 of *Lecture Notes in Computer Science*, pages 457–476. Springer, 2013.

[23] P. W. Laud, P. Damien, and T. S. Shively. Sampling some truncated distributions via rejection algorithms. *Communications in Statistics: Simulation and Computation*, 39(6):1111–1121, 2010.

[24] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *48th Annual Symposium on Foundations of Computer Science*, 2007.

[25] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on the Theory of Computing*, 2007.

[26] R. Okada, K. Fukuchi, and J. Sakuma. Differentially private analysis of outliers. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 458–473. Springer, 2015.

[27] M. Reimherr and J. Awan. KNG: The K-norm gradient mechanism. In *Advances in Neural Information Processing Systems 32*, pages 10208–10219, 2019.

[28] O. Sheffet. Differentially private ordinary least squares. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3105–3114, 2017.

[29] O. Sheffet. Homework 1 for *Differential Privacy: Privacy-preserving Data Analysis*. University of Alberta course CMPUT651, 2018. URL http://webdocs.cs.ualberta.ca/~osheffet/HW1W18.pdf.

[30] A. Smith and A. Thakurta. Differentially private feature selection via stability arguments, and the robustness of the Lasso. In *Proceedings of the Twenty Sixth Annual Conference on Computational Learning Theory*, pages 819–850, 2013. URL http://proceedings.mlr.press/v30/Guha13.html.

[31] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 197–206, 1956.

[32] J. Ullman and A. Sealfon. Efficiently estimating Erdos-Renyi graphs with node differential privacy. In *Advances in Neural Information Processing Systems 32*, pages 3765–3775, 2019.

[33] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[34] Y. X. Wang. Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, pages 93–103, 2018.

Appendix

A Proofs of Section 2.1

A.1 Proof of Theorem 1

Using the privacy guarantees of the exponential mechanism [24, Theorem 6], it is enough to prove that $\overline{\mathsf{len}}_f(\boldsymbol{x}; t)$ is 1-Lipschitz. Let $\boldsymbol{x}, \boldsymbol{x}'$ be two neighboring datasets and let $t \in \mathcal{T}$. Assume w.l.o.g. that $\ell = \overline{\mathsf{len}}_f(\boldsymbol{x}; t) \leq \overline{\mathsf{len}}_f(\boldsymbol{x}'; t)$. We need to prove that $\overline{\mathsf{len}}_f(\boldsymbol{x}'; t) \leq \ell + 1$. From the definition of $\overline{\mathsf{len}}$ in Equation (1), we have that $\sum_{i=1}^{\ell} R_i(\boldsymbol{x}) \geq \|t - f(\boldsymbol{x}')\|_p$, so the claim follows from conditions on $R_\ell(\cdot)$ since

$$\|t - f(\boldsymbol{x}')\|_p \leq \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\|_p + \|t - f(\boldsymbol{x})\|_p \leq R_1(\boldsymbol{x}') + \sum_{i=1}^{\ell} R_i(\boldsymbol{x}) \leq \sum_{i=1}^{\ell+1} R_i(\boldsymbol{x}').$$

A.2 Proof of Proposition 2.1

To prove the claim about privacy, it is enough to show that R_ℓ satisfy the conditions of Lemma 1. We clearly have $\mathrm{LS}(\boldsymbol{x}) \leq \overline{\mathrm{LS}}(\boldsymbol{x}) \leq R_1(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}^n$. Moreover, we have that $R_\ell(\boldsymbol{x}) \leq R_{\ell+1}(\boldsymbol{x}')$ for all neighboring datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ since

$$R_\ell(\boldsymbol{x}) = \sup_{\boldsymbol{x}_1 : d_{\mathrm{ham}}(\boldsymbol{x}, \boldsymbol{x}_1) \leq \ell} \overline{\mathrm{LS}}(\boldsymbol{x}_1) \leq \sup_{\boldsymbol{x}_1 : d_{\mathrm{ham}}(\boldsymbol{x}', \boldsymbol{x}_1) \leq \ell+1} \overline{\mathrm{LS}}(\boldsymbol{x}_1) = R_{\ell+1}(\boldsymbol{x}').$$

A.3 Proof of Theorem 2

We begin with the exact inverse sensitivity mechanism. Let $C^k = \{t : \mathsf{len}^\rho_f(\boldsymbol{x}; t) = k\}$. The definition of len^ρ implies that $\mathsf{len}^\rho_f(\boldsymbol{x}; t) = 0$ for t such that $\|t - f(\boldsymbol{x})\|_2 \leq \rho$ and that $\mathsf{len}^\rho_f(\boldsymbol{x}; t) \geq K$ for any t such that $\|t - f(\boldsymbol{x})\|_2 \geq \omega_f(\boldsymbol{x}; K) + \rho$. Thus we have that

$$\begin{aligned}
P\left(\|M_{\mathrm{inv}}(\boldsymbol{x}) - f(\boldsymbol{x})\|_2 \geq \omega_f(\boldsymbol{x}; K) + \rho\right) &\leq \sum_{k=K}^{n} \mathbb{P}(M_{\mathrm{inv}}(\boldsymbol{x}) \in C^k) \\
&\leq \frac{e^{-K\varepsilon/2} \sum_{k=K}^{n} \int_{s \in C^k} ds}{\int_{s \in \mathcal{T}} e^{-\mathsf{len}^\rho_f(\boldsymbol{x}; t)\varepsilon/2} ds} \\
&\leq e^{-K\varepsilon/2} \frac{\mathrm{Vol}\{t : \|t\|_2 \leq D\}}{\mathrm{Vol}\{t : \|t\|_2 \leq \rho\}} \\
&\leq e^{-K\varepsilon/2} (D/\rho)^d.
\end{aligned}$$

This gives the first part of the claim.

Now we prove the bounds for the approximate mechanism. First, we notice that the noise added by the approximate mechanism (M.2) satisfies

$$z(\boldsymbol{x}) \coloneqq M(\boldsymbol{x}) - f(\boldsymbol{x}),$$

where $\mathbb{P}(z(\boldsymbol{x}) = z) \propto e^{-k\varepsilon/2}$ for $z \in B^k = \{z \in \mathbb{R}^d : \sum_{i=1}^{k-1} S_i \le \|z\|_2 < \sum_{i=1}^{k} S_i\}$. Noting that $\|z(\boldsymbol{x})\|_2 \ge \sum_{i=1}^{K} R_i(\boldsymbol{x})$ implies that $z(\boldsymbol{x}) \in B^k$ for $k \ge K$, we get that

$$\begin{aligned}
\mathbb{P}\left(\|M_{\text{appr}}(\boldsymbol{x}) - f(\boldsymbol{x})\|_2 \ge \sum_{i=1}^{K} R_i(\boldsymbol{x})\right) &= \mathbb{P}\left(\|z(\boldsymbol{x})\|_2 \ge \sum_{i=1}^{K} R_i(\boldsymbol{x})\right) \\
&\le \sum_{k=K}^{n} \mathbb{P}(z(\boldsymbol{x}) \in B^k) \\
&\le e^{-K\varepsilon/2} \frac{\text{Vol}\{t : \|t\|_2 \le nD\}}{e^{-1}\text{Vol}\{t : \|t\|_2 \le \sum_{i=1}^{1/\varepsilon} R_i(\boldsymbol{x})\}} \\
&\le e^{-K\varepsilon/2+1} \left(\frac{nD}{\sum_{i=1}^{1/\varepsilon} R_i(\boldsymbol{x})}\right)^d,
\end{aligned}$$

where the last inequality follows using that the ratio of the volumes of two ℓ_p-balls with radii r_1 and r_2 is $(r_1/r_2)^d$.

A.4 Proof of Proposition 2.2

The claim about privacy follows from identical arguments to the proof of Proposition 2.1. We now prove the claim about utility. We remove $\boldsymbol{x}$ to simplify notation. First, we have that $1 \le \frac{R_i}{R_1} \le e^{i\beta}$ from the definition of smooth sensitivity. Thus we have

$$\begin{aligned}
\mathbb{E}\left[|M(\boldsymbol{x}) - f(\boldsymbol{x})|\right] &= \frac{\sum_{i=1}^{n} e^{-i\varepsilon/2} R_i \sum_{j=1}^{i} R_j}{\sum_{i=1}^{n} e^{-i\varepsilon/2} R_i} \\
&\le \frac{\sum_{i=1}^{n} e^{-i\varepsilon/2} e^{i\beta} R_1^2 \sum_{j=1}^{i} e^{j\beta}}{\sum_{i=1}^{n} e^{-i\varepsilon/2} R_1} \\
&\le R_1 \frac{\sum_{i=1}^{n} e^{-i\varepsilon/2} e^{i\beta} \frac{e^{i\beta}}{e^{\beta}-1}}{\sum_{i=1}^{n} e^{-i\varepsilon/2}} \\
&= \frac{R_1}{e^{\varepsilon/8}-1} \frac{\sum_{i=1}^{n} e^{-i\varepsilon/4}}{\sum_{i=1}^{n} e^{-i\varepsilon/2}} \\
&= \frac{R_1}{e^{\varepsilon/8}-1} \frac{e^{\varepsilon/2}-1}{e^{\varepsilon/4}-1} = O\left(\frac{\text{S}^{\beta}(\boldsymbol{x})}{\varepsilon}\right),
\end{aligned}$$

where the last equality follows since $\varepsilon = O(1)$.

B Proofs of Section 2.2

B.1 Proof of Lemma 2.1

We begin with some notation. Let $(\boldsymbol{x}, \boldsymbol{y})$ and $(\boldsymbol{x}', \boldsymbol{y}')$ be two neighboring instances, and denote their minimizers by $\hat{\theta}_n$ and $\hat{\theta}'_n$, respectively. We let $D = \text{diam}_2(\Theta)$.

The following lemma bounds the distance between these minimizers.

Lemma B.1. *Under the assumptions of Proposition 2.1,*

$$\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \le \frac{2L}{\lambda n}.$$

To prove Lemma B.1, we first prove the following weaker version.

Lemma B.2. *Assume $L_n(\theta; \boldsymbol{x}, \boldsymbol{y})$ is λ-strongly convex and L-Lipschitz. Then*

$$\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \le \frac{L}{\lambda n}.$$

Proof Since L_n is λ-strongly convex and L-Lipschitz, we have

$$\begin{aligned}\lambda \left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2^2 &\leq \langle \nabla L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) - \nabla L_n(\hat{\theta}'_n; \boldsymbol{x}, \boldsymbol{y}), \hat{\theta}_n - \hat{\theta}'_n \rangle \\ &\leq \left\|\nabla L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) - \nabla L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})\right\|_2 \left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2.\end{aligned}$$

The claim now follows since $\nabla L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) = 0$ and $\left\|\nabla L_n(\hat{\theta}'_n; \boldsymbol{x}, \boldsymbol{y})\right\|_2 \leq \frac{L}{n}$. □

Proof [of Lemma B.1] First, we have that $\nabla L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) = 0$ and $\left\|\nabla L_n(\hat{\theta}'_n; \boldsymbol{x}, \boldsymbol{y})\right\|_2 \leq \frac{L}{n}$ since $\ell(\cdot; x_i)$ is L-Lipschitz. We split to cases whether $\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \leq \frac{\lambda}{2H}$. First, if $\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \leq \frac{\lambda}{2H}$ then we know that the function $L_n(\theta; \boldsymbol{x}, \boldsymbol{y})$ is $\lambda/2$-strongly convex on the set $A = \{\theta : \left\|\hat{\theta}_n - \theta\right\|_2 \leq \frac{\lambda}{2H}\}$. We have $\hat{\theta}'_n \in A$, and therefore Lemma B.2 implies that $\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \leq \frac{2L}{\lambda n}$.
Now assume that $\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 > \frac{\lambda}{2H}$ and we get a contradiction. Indeed let $\theta_t = (1-t)\hat{\theta}_n + t\hat{\theta}'_n$. For any $0 \leq t \leq 1$ such that $\left\|\hat{\theta}_n - \theta_t\right\|_2 \leq \frac{\lambda}{2H}$, we have that

$$\begin{aligned}\lambda \left\|\hat{\theta}_n - \theta_t\right\|_2^2 &\leq \langle \nabla L_n(\theta_t; \boldsymbol{x}, \boldsymbol{y}), \theta_t - \hat{\theta}_n \rangle \\ &\overset{(i)}{\leq} \langle \nabla L_n(\theta_1; \boldsymbol{x}, \boldsymbol{y}), \theta_1 - \hat{\theta}_n \rangle \\ &\leq \frac{L \left\|\hat{\theta}'_n - \hat{\theta}_n\right\|_2}{n},\end{aligned}$$

where the third inequality follows from Cauchy-Schwartz inequality since $\theta_1 = \hat{\theta}'_n$ and (i) follows from a monotonicity argument which we explain presently. This implies that $\left\|\hat{\theta}'_n - \hat{\theta}_n\right\|_2 \geq \frac{n\lambda^2}{4LH^2}$ which is a contradiction.
Let us now explain why inequality (i) holds. First we denote $u = \hat{\theta}'_n - \hat{\theta}_n$ and we notice that $\theta_t = \hat{\theta}_n + tu$. Define $g(t) = L_n(\theta_t; x)$ which is convex in t. As g is convex with minimizer at $t^\star = 0$, we have $g'(0) = 0$ and $0 \leq g'(t) \leq g'(s)$ for $0 \leq t \leq s$. Therefore we have that $0 \leq \langle \nabla L_n(\theta_t; \boldsymbol{x}, \boldsymbol{y}), u \rangle \leq \langle \nabla L_n(\theta_s; \boldsymbol{x}, \boldsymbol{y}), u \rangle$. Inequality (i) now follows since

$$\begin{aligned}\langle \nabla L_n(\theta_t; \boldsymbol{x}, \boldsymbol{y}), \theta_t - \hat{\theta}_n \rangle &= t\langle \nabla L_n(\theta_t; \boldsymbol{x}, \boldsymbol{y}), u \rangle \\ &\leq s\langle \nabla L_n(\theta_s; \boldsymbol{x}, \boldsymbol{y}), u \rangle \\ &= \langle \nabla L_n(\theta_s; \boldsymbol{x}, \boldsymbol{y}), \theta_s - \hat{\theta}_n \rangle.\end{aligned}$$

□

Now we are ready to prove Proposition 2.1. First, Lemma B.1 implies that

$$\left\|\hat{\theta}_n - \hat{\theta}'_n\right\|_2 \leq \frac{2L}{\lambda n}.$$

Therefore as $\nabla^2 L_n(\cdot; \boldsymbol{x}, \boldsymbol{y})$ is H-Lipschitz, we get

$$\left\|\nabla^2 L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) - \nabla^2 L_n(\hat{\theta}'_n; \boldsymbol{x}, \boldsymbol{y})\right\|_2 \leq \frac{2LH}{\lambda n}.$$

As a consequence we have

$$\begin{aligned}
&\left|\left\|\nabla^2 L_n(\hat{\theta}_n;\boldsymbol{x},\boldsymbol{y})(\theta-\hat{\theta}_n)\right\|_2 - \left\|\nabla^2 L_n(\hat{\theta}'_n;\boldsymbol{x}',\boldsymbol{y}')(\theta-\hat{\theta}'_n)\right\|_2\right| \\
&= \left|\left\|\nabla^2 L_n(\hat{\theta}_n;\boldsymbol{x},\boldsymbol{y})(\theta-\hat{\theta}_n)\right\|_2 - \left\|\nabla^2 L_n(\hat{\theta}'_n;\boldsymbol{x}',\boldsymbol{y}')(\theta-\hat{\theta}_n) + \nabla^2 L_n(\hat{\theta}'_n;\boldsymbol{x}',\boldsymbol{y}')(\hat{\theta}_n-\hat{\theta}'_n)\right\|_2\right| \\
&\le \left\|\nabla^2 L_n(\hat{\theta}_n;\boldsymbol{x},\boldsymbol{y}) - \nabla^2 L_n(\hat{\theta}'_n;\boldsymbol{x},\boldsymbol{y})\right\|_2 \left\|\theta-\hat{\theta}_n\right\|_2 + \left\|\nabla^2 L_n(\hat{\theta}'_n;\boldsymbol{x}',\boldsymbol{y}')(\hat{\theta}_n-\hat{\theta}'_n)\right\|_2 \\
&\le \frac{2DLH}{\lambda n} + \left\|\nabla L_n(\hat{\theta}_n;\boldsymbol{x}',\boldsymbol{y}')\right\|_2 + O(H\left\|\hat{\theta}_n-\hat{\theta}'_n\right\|_2^2) \\
&\le \frac{2DLH}{\lambda n} + \frac{L}{n} + O\left(H\left(\frac{2LH}{\lambda n}\right)^2\right).
\end{aligned}$$

B.2 Proof of Proposition 2.3

The claim about privacy is immediate from the exponential mechanism. Let us now argue about the claim for utility. To simplify notation, we let $\Sigma = \nabla^2 L(\hat{\theta}_n;\boldsymbol{x},\boldsymbol{y})$. We later show (see Section F.1) that to sample from the distribution (5), one can sample $R \sim \mathsf{Gamma}(d,1)$ and $U \sim \mathsf{Uni}(\mathbb{S}^{d-1})$ and then set $\theta = \hat{\theta}_n + Z$ where $Z = \frac{2\mathrm{GS}_{\mathsf{Hess}}}{\varepsilon}\Sigma^{-1}\cdot R\cdot U$, and finally we accept θ if $\theta\in\Theta$, otherwise we repeat the process, It is easy to show that for Z we have

$$\mathbb{E}\left[\|Z\|_2^2\right] = \frac{4\mathbb{E}[R^2]\mathrm{GS}_{\mathsf{Hess}}^2}{\varepsilon^2}\mathbb{E}\left[\left\|\Sigma^{-1}U\right\|_2^2\right] \le \frac{Cd\mathrm{GS}_{\mathsf{Hess}}^2}{\varepsilon^2}\operatorname{tr}\left(\Sigma^{-2}\right),$$

for a universal constant C. But we need to upper bound $\mathbb{E}\left[\|Z\|_2^2 \mid \hat{\theta}_n + Z\in\Theta\right]$ as this is the error of the mechanism. To finish the proof, we now prove that for every random variable W,

$$\mathbb{E}\left[\|W\|_2^2 \mid \hat{\theta}_n + Z\in\Theta\right] \le 2\mathbb{E}\left[\|W\|_2^2\right]. \tag{9}$$

To this end, we let $\rho^2 = \mathbb{E}\left[\|Z\|_2^2\right]$ and define three disjoint sets, $S_1 = \{Z : \|Z\|_2 \le 2\rho\}$, $S_2 = \{Z : \hat{\theta}_n + Z\in\Theta, Z\notin S_1\}$, and $S_3 = \mathbb{R}^d\setminus(S_1\cup S_2)$. Clearly these sets are disjoint and the assumptions of the Proposition imply that $S_1\subseteq\Theta$ and therefore $\Theta = S_1\cup S_2$. Using conditional expectation and denoting $p_i = \mathbb{P}(Z\in S_i)$, we have that

$$\mathbb{E}\left[\|W\|_2^2\right] = \sum_{i=1}^3 p_i\mathbb{E}\left[\|W\|_2^2 \mid Z\in S_i\right].$$

Noting that $p_1\ge 1/2$ by Markov inequality, we now get that

$$\begin{aligned}
\mathbb{E}\left[\|W\|_2^2 \mid \hat{\theta}_n + Z\in\Theta\right] &= \frac{p_1}{p_1+p_2}\mathbb{E}\left[\|W\|_2^2 \mid Z\in S_1\right] + \frac{p_2}{p_1+p_2}\mathbb{E}\left[\|W\|_2^2 \mid Z\in S_2\right] \\
&\le 2p_1\mathbb{E}\left[\|W\|_2^2 \mid Z\in S_1\right] + 2p_2\mathbb{E}\left[\|W\|_2^2 \mid Z\in S_2\right] \\
&\le 2\mathbb{E}\left[\|W\|_2^2\right].
\end{aligned}$$

The claim follows.

C Proofs of Section 3 (lower bounds)

C.1 Proofs of Theorem 3

We start with the lower bound for unbiased mechanisms. Fix $\boldsymbol{x}\in\mathcal{X}^n$ and assume towards a contradiction that $\mathbb{E}[\|M(\boldsymbol{x}) - f(\boldsymbol{x})\|_2] \le \frac{1}{8}\beta_k$ where $\beta_k = c\cdot e^{-2k\varepsilon/d}\omega_f(\boldsymbol{x};k)$. The definition of $W_f(\boldsymbol{x};k)$ implies that there exists a $\beta_k/4$ packing, namely S, of $W_f(\boldsymbol{x};k)$ of size at least $m_{\beta_k} \ge (\frac{4c\omega_f(\boldsymbol{x};k)}{\beta_k})^d \ge 4^d e^{2k\varepsilon}$. The definition of $W_f(\boldsymbol{x};k)$ implies that there is an instance $\boldsymbol{x}'$ such

that $d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le k$ and $f(\boldsymbol{x}') = t$ for every $t \in S$, hence we have a set A of size m_{β_k} of datasets $\boldsymbol{x}'$ such that $f(\boldsymbol{x}') \in S$. For every $\boldsymbol{x}' \in A$, we define

$$B_{\boldsymbol{x}'} = \{y : \|y - f(\boldsymbol{x}')\|_2 \le \beta_k/4\}.$$

We now have that:

$$\mathbb{P}(M(\boldsymbol{x}) \in B_{\boldsymbol{x}'}) \overset{(i)}{\ge} \mathbb{P}(M(\boldsymbol{x}') \in B_{\boldsymbol{x}'})e^{-k\varepsilon} \overset{(ii)}{\ge} \mathbb{P}(M(\boldsymbol{x}') \in B_{\boldsymbol{x}})e^{-k\varepsilon}$$
$$\overset{(iii)}{\ge} \mathbb{P}(M(\boldsymbol{x}) \in B_{\boldsymbol{x}})e^{-2k\varepsilon} \overset{(iv)}{\ge} \frac{e^{-2k\varepsilon}}{2},$$

where (i) and (iii) follow from the definition of differential privacy, (ii) follows since M is $\|\cdot\|_2$-unbiased, and (iv) follows from Markov inequality. As the sets $B_{\boldsymbol{x}'}$ are disjoint for $\boldsymbol{x}' \in A$, we have a contradiction

$$1 \ge \sum_{\boldsymbol{x}' \in A} \mathbb{P}(M(\boldsymbol{x}) \in B_{\boldsymbol{x}'}) \ge \frac{1}{2} m_{\beta_k} e^{-2k\varepsilon} \ge \frac{4^d}{2}.$$

To prove the first part of the claim (i.e., the minimax lower bound), we use similar ideas while starting from the assumption that for every $\boldsymbol{x}$ we have $\mathbb{E}[\|M(\boldsymbol{x}) - f(\boldsymbol{x})\|_2] \le \frac{1}{8}\beta_k$ where $\beta_k = c \cdot \sup_{\boldsymbol{x}'} e^{-k\varepsilon/d}\omega_f(\boldsymbol{x}'; k)$. We again define a packing A (now with $m_{\beta_k} \ge 4^d e^{k\varepsilon}$) and we get using Markov inequality and the definition of differential privacy that $\mathbb{P}(M(\boldsymbol{x}) \in B_{\boldsymbol{x}'}) \ge \mathbb{P}(M(\boldsymbol{x}') \in B_{\boldsymbol{x}'})e^{-k\varepsilon} \ge \frac{e^{-k\varepsilon}}{2}$. This gives a contradiction similarly to our argument above.

C.2 Local-minimax lower bound and proof of Theorem 4

First, we start by explaining why $r = \Omega(d/\varepsilon)$ is necessary to exclude trivial mechanisms in the local minimax definition (6). Assume that we choose $r \ll d/\varepsilon$. Then for an instance $\boldsymbol{x} \in \mathcal{X}^n$, consider the trivial constant mechanism that sets $M_{\mathsf{triv}}(\boldsymbol{x}') = f(\boldsymbol{x})$ for every $\boldsymbol{x}' \in \mathcal{X}^n$. Clearly this mechanism is ε-differentially private and its local-minimax risk for $\boldsymbol{x}$ is

$$\sup_{\boldsymbol{x}' : d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le r} E\left[\|M_{\mathsf{triv}}(\boldsymbol{x}') - f(\boldsymbol{x}')\|_2\right] = \sup_{\boldsymbol{x}' : d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le r} \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\|_2 = \omega_f(\boldsymbol{x}; r).$$

Our lower bounds of Lemma C.1 on the local minimax risk with radius r show this is the optimal risk—up to constant factors that do not depend on r—for $\boldsymbol{x}$ whenever $r \le d/\varepsilon$. Therefore we need to pick a larger value of r such that the optimal mechanism is not the trivial constant mechanism. Picking $r = C \cdot d/\varepsilon$, our lower bounds are roughly $\sup_{1 \le i \le C} e^{-i}\omega_f(\boldsymbol{x}; id/\varepsilon)$ (which is usually maximized at $i = 1$ resulting in $\omega_f(\boldsymbol{x}; d/\varepsilon)$) and so the trivial mechanism does not achieve this for C large enough.

Moreover, when $r \ll d/\varepsilon$, no single mechanism M can be instance-optimal according to this definition: if there exists $M^\star$ such that $\mathbb{E}[\|M^\star(\boldsymbol{x}) - f(\boldsymbol{x})\|_2] \le O(1)\mathcal{R}(\boldsymbol{x}; r) \le O(1)\omega_f(\boldsymbol{x}; r)$ for every $\boldsymbol{x} \in \mathcal{X}^n$, then we get that

$$\sup_{\boldsymbol{x} \in \mathcal{X}^n} \mathbb{E}[\|M^\star(\boldsymbol{x}) - f(\boldsymbol{x})\|_2] \le O(1) \sup_{\boldsymbol{x} \in \mathcal{X}^n} \omega_f(\boldsymbol{x}; r) \ll \sup_{\boldsymbol{x} \in \mathcal{X}^n} \omega_f(\boldsymbol{x}; d/\varepsilon),$$

which contradict the minimax lower bounds of Theorem 3.

Theorem 4 follows from the following lemma by setting $r = d/\varepsilon$.

Lemma C.1. *Let the assumptions of Theorem 4 hold. Then for any $r \ge 1$, $\mathcal{R}(\boldsymbol{x}; r) \ge \frac{c}{8} \sup_{k \le r} e^{-k\varepsilon/d}\omega_f(\boldsymbol{x}; k)$.*

Proof The proof follows similar arguments to those we had in the proof of Theorem 3. Assume toward a contradiction that $\mathcal{R}(\boldsymbol{x}; r) \le \frac{1}{8}\beta_k$ where $\beta_k = c \cdot e^{-k\varepsilon/d}\omega_f(\boldsymbol{x}; k)$ for $k \le r$. This implies that there exists a mechanism M such that for every $\boldsymbol{x}'$ such that $d_{\text{ham}}(\boldsymbol{x}, \boldsymbol{x}') \le r$, we have $\mathbb{E}[\|M(\boldsymbol{x}') - f(\boldsymbol{x}')\|_2] \le \frac{1}{8}\beta_k$. Repeating the arguments of the proof of Theorem 3 with $W_f(\boldsymbol{x}; k)$ for $k \le r$ proves the claim. □

D Proofs and further details of Section 4.1 (mean estimation)

In this section we provide proofs for the claims in Section 4.1 and give our algorithm. We begin by giving the full version of Lemma 4.1, which we prove in Appendix D.2.

Lemma (Full version of Lemma 4.1)**.** *Let* $f(\boldsymbol{x}) = [\mathsf{trim}_m(\boldsymbol{x})]_{[a,b]}$. *Then, for any* $t \in [a,b]$, *if* $t \geq \mathsf{trim}_m(\boldsymbol{x})$

$$\mathsf{len}_f(\boldsymbol{x};t) = \min\{k : k \leq m, |t - f(\boldsymbol{x})| \leq \frac{1}{n-2m}\sum_{i=1}^{k}(x_{(n-m+i)} - x_{(m+i)})\} \cup \{m+1\}.$$

Moreover, if $t \leq \mathsf{trim}_m(\boldsymbol{x})$

$$\mathsf{len}_f(\boldsymbol{x};t) = \min\{k : k \leq m, |t - f(\boldsymbol{x})| \leq \frac{1}{n-2m}\sum_{i=1}^{k}(x_{(n-m+1-i)} - x_{(m+1-i)})\} \cup \{m+1\}.$$

D.1 Sampling from the inverse sensitivity mechanism

We describe an algorithm for sampling from the inverse sensitivity mechanism with $\rho > 0$ for the mean estimation problem of Section 4.1. Our goal is to sample from

$$\pi_{M_{\mathrm{inv}}(\boldsymbol{x})}(t) = \frac{e^{-\mathsf{len}^{\rho}(\boldsymbol{x};t)\varepsilon/2}}{\int_{\mathcal{T}} e^{-\mathsf{len}^{\rho}(\boldsymbol{x};s)\varepsilon/2} ds}.$$

Algorithm 4 shows how to sample from this distribution using Lemma 4.1. To see this, note that Lemma 4.1 implies that S_k in Algorithm 4 is exactly the set $\{t : \mathsf{len}^{\rho}(\boldsymbol{x};t) = k\}$. And so all values $t \in S_k$ have the same probability. Moreover, the probability of sampling a value from the set S_k is $\mathrm{Vol}(S_k)e^{-k\varepsilon/2}$ using the definition of the mechanism.

Algorithm 4: Inverse sensitivity for mean estimation

Input: $\boldsymbol{x} \in \mathbb{R}^n$, m, ρ, a, b

Calculate $\hat{x}_t = [\mathsf{trim}_m(\boldsymbol{x})]_{[a,b]}$;

Calculate $u_k = \min\left(\rho + \frac{1}{n-2m}\sum_{i=1}^{k}(x_{(n-m+i)} - x_{(m+i)}), b - \hat{x}_t\right)$ for $0 \leq k \leq m$;

Calculate $\ell_k = \min\left(\rho + \frac{1}{n-2m}\sum_{i=1}^{k}(x_{(n-m+1-i)} - x_{(m+1-i)}), \hat{x}_t - a\right)$ for $0 \leq k \leq m$;

Set $u_{m+1} = b - \hat{x}_t$ and $\ell_{m+1} = \hat{x}_t - a$;

Set $S_k = [-\ell_{k+1}, -\ell_k] \cup [u_k, u_{k+1}]$;

Sample $k \propto \mathrm{Vol}(S_k)e^{-k\varepsilon/2}$;

Sample $z \sim \mathsf{Uni}(S_k)$;

return $\hat{x} = \hat{x}_t + z$

D.2 Proof of Lemma 4.1

We only prove the case $t \geq \mathsf{trim}_m(\boldsymbol{x})$ as the other one is similar. In this case, to make the value of the trimmed mean t by changing k values, we must change the value of the k smallest samples $x_{(1)}, \cdots, x_{(k)}$ and set their value to ∞. Denote the resulting sample by $\boldsymbol{x}'$. The trimmed mean $\boldsymbol{x}'$ is

$$\mathsf{trim}_m(\boldsymbol{x}') = \frac{x'_{(m+1)} + x'_{(m+2)} + \cdots + x'_{(n-m)}}{n-2m}.$$

We split to two cases. First, if $k \geq m+1$, then we get that $x'_{n-m} = \infty$ and therefore $\mathsf{trim}_m(\boldsymbol{x}') = \infty$. This means that for any $t > \mathsf{trim}_m(\boldsymbol{x})$ we can set suitable new values to $x_{(1)}, \cdots, x_{(k)}$ (instead of ∞) such that $\mathsf{trim}_m(\boldsymbol{x}') = t$. Now assume $k \leq m$. In this case, we get that

$$\mathsf{trim}_m(\boldsymbol{x}') = \frac{x_{(m+k+1)} + x_{(m+k+2)} + \cdots + x_{(n+k-m)}}{n-2m}.$$

The claim follows as we have

$$\mathsf{trim}_m(\boldsymbol{x}') - \mathsf{trim}_m(\boldsymbol{x}) = \frac{\sum_{i=1}^{k}(x_{(n-m+i)} - x_{(m+i)})}{n-2m}.$$

D.3 Proof of Proposition 4.1

The privacy guarantees of Algorithm 4 follow from Proposition 3.2 in [4].

To prove the claim about utility, we use the following result from [7] which upper bounds the error of the trimmed mean estimator.

Lemma D.1 (Bun and Steinke [7], Proposition 10)**.** *Let $x_i \overset{\text{iid}}{\sim} P$ where P has mean μ and variance σ^2. Then*

$$\mathbb{E}[(\mathsf{trim}_m(\boldsymbol{x}) - \mu)^2] \le \frac{n(1+\sqrt{8}m)}{(n-2m)^2}\sigma^2 = O\left(\frac{m}{n}\right)\sigma^2.$$

Moreover, if P is symmetric about its mean then

$$\mathbb{E}[(\mathsf{trim}_m(\boldsymbol{x}) - \mu)^2] \le \left(1 + O\left(\frac{m}{n}\right)\right)\frac{\sigma^2}{n}.$$

We have that

$$\begin{aligned}\mathbb{E}[(\hat{x}-\mu)^2] &\le \mathbb{E}[([\mathsf{trim}_m(\boldsymbol{x})]_{[a,b]} - \mu)^2] + \mathbb{E}[z^2] \\ &\le \mathbb{E}[(\mathsf{trim}_m(\boldsymbol{x}) - \mu)^2] + \mathbb{E}[z^2],\end{aligned}$$

where the second inequality follows since $\mu \in [a,b]$, and so a projection to $[a,b]$ cannot increase error. Thus, given the bound of Lemma D.1, now we only need to upper bound $\mathbb{E}[z^2]$.

We begin with the following lemma for a fixed $\boldsymbol{x}$.

Lemma D.2. *Let $\boldsymbol{x} \in \mathbb{R}^n$ and $K \le m$. Let $L(\boldsymbol{x}) = \max_{1\le k\le K}(u_{k+1} - u_k, \ell_{k+1} - \ell_k)$. Then*

$$\mathbb{E}[z^2] \le O\left(\frac{L(\boldsymbol{x})^2}{\varepsilon^2}\right) + e^{-K\varepsilon/2}\frac{m(b-a)^2}{\rho}.$$

Proof Let $v_k = u_k + \ell_k$. Then we have that $v_{k+1} - v_k \le 2L$ for $k \le K$ and $v_0 \ge \rho$, hence we get

$$\begin{aligned}\mathbb{E}[z^2] &\le \frac{\sum_{k=0}^{m+1} e^{-k\varepsilon/2} v_k^2 (v_k - v_{k-1})}{\sum_{k=0}^{m+1} e^{-k\varepsilon/2}(v_k - v_{k-1})} \\ &\le \frac{\sum_{k=0}^{K} e^{-k\varepsilon/2} v_k^2 (v_k - v_{k-1})}{\sum_{k=0}^{m+1} e^{-k\varepsilon/2}(v_k - v_{k-1})} + e^{-K\varepsilon/2}\frac{m(b-a)^2}{\rho} \\ &\overset{(i)}{\le} O\left(\frac{L^2}{\varepsilon^2}\right) + e^{-K\varepsilon/2}\frac{m(b-a)^2}{\rho}.\end{aligned}$$

Inequality (i) follows from similar arguments to the proof of Proposition 4.3 in [4]: let $T \le K$ be the smallest such that $v_T \ge \frac{L}{\varepsilon}$. If no such T exists, then inequality (i) clearly holds. Note that $v_T \le \frac{L}{\varepsilon} + L \le O(\frac{L}{\varepsilon})$. Thus we get

$$\frac{\sum_{k=1}^{K} e^{-k\varepsilon/2} v_k^2 (v_k - v_{k-1})}{\sum_{k=1}^{m+1} e^{-k\varepsilon/2}(v_k - v_{k-1})} \le v_T + \frac{\sum_{k=T}^{K} 2e^{-k\varepsilon/2} k^2 L^3}{e^{-T\varepsilon/2} L/\varepsilon} = O\left(\frac{L^2}{\varepsilon^2}\right).$$

□

Now we are ready to finish the proof of Proposition 4.1. We notice that for any $\boldsymbol{x}$

$$L(\boldsymbol{x}) \le \frac{x_{(n)} - x_{(1)}}{n-m} \le \frac{2\max_{1\le i\le n}|x_i|}{n-m}.$$

Therefore using standard bound on the expectation of maximum of subgaussian variables (Lemma 45 in full version [7])

$$\mathbb{E}[L(\boldsymbol{x})^2] \le \frac{4\mathbb{E}[\max_{1\le i\le n} x_i^2]}{(n-m)^2} \le \frac{8\sigma^2 \log n}{(n-m)^2}.$$

Overall we have that using $K = m$ in Lemma D.2 implies that for a constant C,

$$\mathbb{E}[(\hat{x} - \mu)^2] \leq \mathbb{E}[(\mathsf{trim}_m(\boldsymbol{x}) - \mu)^2] + \frac{C\sigma^2 \log n}{(n-m)^2\varepsilon^2} + e^{-m\varepsilon/2}\frac{m(b-a)^2}{\rho}.$$

As $\rho = \frac{\sigma^2}{n^2}$ and $m \leq n$, setting $m = \frac{12\log(n(b-a)/\sigma^2)}{\varepsilon}$ and using Lemma D.1 proves the claim.

E Proofs of Section 4.2 (PCA)

Here we prove Proposition 4.2. First, we prove the claim about privacy then we proceed to show the utility analysis.

E.1 Proof of Proposition 4.2 (privacy)

We only need to prove that $w(\boldsymbol{x}) = \overline{v} + z$ is ε-DP as the claim for Algorithm 2 then follows since post-processing preserves privacy [13, Proposition 2.1]. To this end, first we note that Lemma 4.2 implies that the choice of $R_i(\boldsymbol{x})$ in Proposition 4.2 satisfies the conditions of Theorem 1. Now assume we have two neighboring instances $\boldsymbol{x}, \boldsymbol{x}'$ with leading eigenvectors $u_1, u_2 = -u_1$ and $u_1', u_2' = -u_1'$ respectively and assume without loss of generality $\|u_1 - u_1'\|_2 \leq \mathrm{LS}(\boldsymbol{x})$. We get that $w(\boldsymbol{x}) = w_1(\boldsymbol{x}) = u_1 + z$ with probability $1/2$ and $w(\boldsymbol{x}) = w_2(\boldsymbol{x}) = u_2 + z$ otherwise. Similarly we have $w_1(\boldsymbol{x}')$ and $w_2(\boldsymbol{x}')$. Theorem 1 now implies that the densities of $w_1(\boldsymbol{x})$ and $w_1(\boldsymbol{x}')$ are ε-DP (i.e., $\frac{\pi_{w_1(\boldsymbol{x})}(t)}{\pi_{w_1(\boldsymbol{x}')}(t)} \leq e^{\varepsilon}$) and similarly for the densities of $w_2(\boldsymbol{x})$ and $w_2(\boldsymbol{x}')$, therefore by quasi convexity we get that $w(\boldsymbol{x})$ is ε-DP.

E.2 Proof of Proposition 4.2 (utility)

To facilitate notation, we drop $\boldsymbol{x}$ from our analysis. First, we bound the norm of the noise z that the algorithm adds. We claim that there exists a universal constant $C_1 > 0$ such that the noise z in step 4 of Algorithm 2 has with probability $1 - \beta$

$$\|z\|_2 \leq C_1\left(\frac{d\log n/\beta}{n\mathsf{GAP}(\boldsymbol{x})\varepsilon} + \frac{1}{n^4}\right).$$

Deferring the proof of this, we can now complete the proof of the claim. We assume that n is large enough so that $\|z\|_2 \leq 1/2$. Notice that in our setting ($k = 1$) we have that $F(v) = v^T\Sigma v = \left\|\Sigma^{1/2}v\right\|_2^2$. Therefore denoting $\lambda = \frac{1}{\|\overline{v}+z\|_2}$ we get that

$$\begin{aligned}
|F(v_{\mathsf{out}}) - F(\overline{v})| &\leq |F(v_{\mathsf{out}}) - F(\overline{v}+z)| + |F(\overline{v}+z) - F(\overline{v})| \\
&= |\lambda^2 - 1|F(\overline{v}+z) + \left|\left\|\Sigma^{1/2}(\overline{v}+z)\right\|_2^2 - \left\|\Sigma^{1/2}\overline{v}\right\|_2^2\right| \\
&\leq |\lambda^2 - 1|\,\|\overline{v}+z\|_2 + \left\|\Sigma^{1/2}z\right\|_2^2\left(\left\|\Sigma^{1/2}(\overline{v}+z)\right\|_2^2 + \left\|\Sigma^{1/2}\overline{v}\right\|_2^2\right) \\
&\leq 2|\lambda^2 - 1| + 3\,\|z\|_2^2,
\end{aligned}$$

where we use the fact that $\|z\|_2 \leq 1/2$, $\|\overline{v}\|_2 = 1$, and $\|x_i\|_2 \leq 1$ so that $\left\|\Sigma^{1/2}u\right\|_2 \leq \|u\|_2$ for every u. Now we only need to upper bound $|\lambda^2 - 1|$. As $\lambda \leq 2$, we have

$$|\lambda^2 - 1| \leq 3|\lambda - 1| = 3\frac{|1 - \|\overline{v}+z\|_2|}{\|\overline{v}+z\|_2} \leq 6\,\|z\|_2.$$

Therefore overall we have that

$$|F(v_{\mathsf{out}}) - F(\hat{v})| \leq 15\,\|z\|_2,$$

which proves the claim.

Now we return to prove the claim about the norm of z. We use Theorem 2 with $K = \frac{c_1 d\log n/\beta}{\varepsilon}$ to get

$$\mathbb{P}\left(\|z\|_2 \geq \sum_{i=1}^{K} R_i(\boldsymbol{x})\right) \leq \frac{e^{-K\varepsilon/2}}{e}\left(\frac{n\sqrt{2}}{\sum_{i=1}^{1/\varepsilon} R_i(\boldsymbol{x})}\right)^d.$$

Assuming $K \le n\mathsf{GAP}(\boldsymbol{x})/4$ as we can take n large enough in the assumption of the proposition, we get that $R_i(\boldsymbol{x}) \le \frac{2C_{\mathsf{pca}}}{n\mathsf{GAP}(\boldsymbol{x})}$ for $i \le K$ and therefore $\sum_{i=1}^{K} R_i(\boldsymbol{x}) = O(\frac{C_{\mathsf{pca}} d \log n/\beta}{n\mathsf{GAP}(\boldsymbol{x})\varepsilon})$. Since $R_i(\boldsymbol{x}) \ge \frac{C_{\mathsf{pca}}}{n\mathsf{GAP}(\boldsymbol{x})}$, setting c_1 large enough we get that

$$\mathbb{P}\left(\|z\|_2 \ge \sum_{i=1}^{K} R_i(\boldsymbol{x})\right) \le \frac{e^{-K\varepsilon/2}}{e}\left(\frac{n^2\mathsf{GAP}(\boldsymbol{x})\sqrt{2}}{C_{\mathsf{pca}}}\right)^d \le \beta.$$

F Proofs of Section 4.3 (linear regression)

F.1 Sampling from the gradient mechanism

In this section, we show how Algorithm 3 is basically sampling from the distribution of the gradient mechanism. To this end, we show how to sample a vector $t \in \mathbb{R}^d$ with density $\pi(t) = \exp(-\|At\|)$ for a matrix $A \succ 0$. The change of variables $u = At$ and then using rotational symmetry gives that

$$\begin{aligned}\int \pi(t)dt = \frac{1}{\det(A)}\int \exp(-\|u\|)du &= \frac{1}{\det(A)}\int_0^\infty \exp(-r)\mathrm{Vol}_{d-1}(r\mathbb{S}^{d-1})dr \\ &= \frac{1}{\det(A)}\frac{d\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}\int_0^\infty r^{d-1}e^{-r}dr = \frac{d\pi^{d/2}\Gamma(d)}{\det(A)\Gamma(\frac{d}{2}+1)}.\end{aligned}$$

In particular, to sample T with the density $\pi(t) = \exp(-\|At\|)$, we draw $R \sim \mathsf{Gamma}(d,1)$, then $U \sim \mathsf{Uni}(\mathbb{S}^{d-1})$, and set $T = RA^{-1}U$.

Recall that the gradient mechanism has $\pi(t) = \exp(-\|At\|)$ only for $t \in S$ for some set $S \subset \mathbb{R}^d$ and $\pi(t) = 0$ otherwise. Therefore we apply rejection sampling until we get $t \in S$. This shows that Algorithm 3 is sampling from the gradient mechanism.

F.2 Proof of Proposition 4.3

Similarly to the proof of Proposition 2.3, and letting $Z = \frac{2L}{n\varepsilon}\Sigma_n^{-1} \cdot R \cdot U$ for $R \sim \mathsf{Gamma}(d,1)$, and $U \sim \mathsf{Uni}(\mathbb{S}^{d-1})$, we note that Algorithm 3 sets $\theta_{\mathsf{out}} = \bar{\theta} + Z$ and accepts it if $\theta_{\mathsf{out}} \in \Theta$. We also have

$$L_n(\theta_{\mathsf{out}}; \boldsymbol{x}, \boldsymbol{y}) - L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y}) = (\theta_{\mathsf{out}} - \hat{\theta}_n)^T\Sigma_n(\theta_{\mathsf{out}} - \hat{\theta}_n).$$

Since $\hat{\theta}_n \in \mathrm{int}(\Theta)$, we get that $\bar{\theta} = \hat{\theta}_n$ and thus the excess loss of the algorithm is

$$\mathbb{E}\left[L_n(\theta_{\mathsf{out}}; \boldsymbol{x}, \boldsymbol{y}) - L_n(\hat{\theta}_n; \boldsymbol{x}, \boldsymbol{y})\right] \le \mathbb{E}\left[Z^T\Sigma_n Z \mid \hat{\theta}_n + Z \in \Theta\right].$$

Using inequality (9) in the proof of Proposition 2.3, the claim follows since

$$\mathbb{E}\left[Z^T\Sigma_n Z \mid \bar{\theta} + Z \in \Theta\right] \le 2\mathbb{E}\left[Z^T\Sigma_n Z\right] = \frac{8\mathbb{E}[R^2]L^2}{n^2\varepsilon^2}\mathbb{E}\left[U\Sigma_n^{-1}U\right] = O\left(\frac{dL^2}{n^2\varepsilon^2}\,\mathrm{tr}\left(\Sigma_n^{-1}\right)\right).$$