On Privately Estimating a Single Parameter

Hilal Asi[1] John C. Duchi[2] Kunal Talwar[1]
[1]Apple [2]Stanford University

March 21, 2025

Abstract

We investigate differentially private estimators for individual parameters within larger parametric models. While generic private estimators exist, the estimators we provide repose on new local notions of estimand stability, and these notions allow procedures that provide private certificates of their own stability. By leveraging these private certificates, we provide computationally and statistical efficient mechanisms that release private statistics that are, at least asymptotically in the sample size, essentially unimprovable: they achieve instance optimal bounds. Additionally, we investigate the practicality of the algorithms both in simulated data and in real-world data from the American Community Survey and US Census, highlighting scenarios in which the new procedures are successful and identifying areas for future work.

1 Introduction

The challenges of privately estimating high-dimensional objects are myriad: dimension dependent costs make private estimation of even simple models notoriously challenging [19, 31, 12]; optimal methods require sophisticated algorithmic strategies and analyses [21, 2]; the practicality of the methods can be dubious [2, 18, 13]; until recently, we did not even have methods that could computationally efficiently estimate a mean vector with error commensurate with the covariance of the observed data [18, 11]. We instead take a complementary goal, developing methodology for estimating a single parameter in a parametric statistical model. While this may seem pedestrian—how hard could it be to estimate a single scalar?—such problems and questions of their efficiency motivate both substantial applied work, where estimating a (single scalar) causal treatment effect motivates hundreds of thousands of studies [25], as well as deep theoretical work delineating what functionals can and cannot be estimated [7, 34, 32]. Less prosaically, how can we expect any applied work to leverage the insights of differential privacy if we cannot even efficiently estimate a single parameter?

To set the stage, consider the classical M-estimation problem [33, 24]. For a population P on data points $z \in \mathcal{Z}$, we wish to estimate the minimizer of the population (expected) loss

$$L(\theta) \coloneqq P\ell_\theta = \int \ell_\theta(z) dP(z),$$

where $\ell_\theta(z)$ measures the loss of the parameter θ on observation z and we use the empirical process notation that $Pf = \int f(z)dP(z)$. Given a sample of n observations $(z_1, \ldots, z_n)$ and associated empirical distribution $P_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{z_i}$ placing a point mass on each z_i, classical M-estimators release $\theta(P_n) = \operatorname{argmin}_\theta\{P_n\ell_\theta\}$. We augment this slightly to incorporate ℓ_2-regularization around a point $\theta_0 \in \mathbb{R}^d$, considering private release of

$$\theta(P_n) \coloneqq \operatorname{argmin}\left\{P_n\ell_\theta + \frac{\lambda_{\rm reg}}{2}\left\|\theta - \theta_0\right\|_2^2\right\}, \tag{1}$$

where $\lambda_{\rm reg} \geq 0$. Taking as motivation estimating treatment effects or other individual scalars, we also carefully consider estimating linear functionals

$$u^T \theta(P_n)$$

of the parameter, where u is (without loss of generality) a unit vector.

We develop differentially private [23, 22] estimators for these tasks, adopting notation that is a bit different from standard formulations but more convenient for estimation problems. Let $\mathcal{P}_n$ denote the collection of probability measures supported on at most n points in $\mathcal{Z}$, where $P_n(\{z\}) \in \{0, 1/n, 2/n, \ldots, 1\}$; we can identify a sample $\{z_1, \ldots, z_n\}$ by its associated empirical distribution P_n. We say that two samples P_n, P_n' are *neighboring* if they differ in only a single observation, equivalently, that their variation distance satisfies

$$\left\|P_n - P_n'\right\|_{\rm TV} := \sup_A |P_n(A) - P_n'(A)| \leq \frac{1}{n}.$$

We develop mechanisms, meaning a randomized functions on $\mathcal{P}_n$, satisfying

Definition 1.1. *A mechanism M is (ε, δ)-*differentially private *if*

$$\mathbb{P}(M(P_n) \in A) \leq e^\varepsilon \mathbb{P}(M(P_n') \in A) + \delta$$

for all neighboring empirical distributions P_n, P_n' and all measurable sets A.

For M-estimators of the form (1), the most natural seeming approach is to add noise commensurate with the *sensitivity*, or the modulus of continuity, of the statistic of interest with respect to changes in the sample P_n. If we wish to release a statistic $\theta(P_n)$, then we consider the local modulus of continuity

$$\omega_\theta(P_n; k) := \sup \left\{ \left\|\theta(P_n) - \theta(P_n')\right\|_2 \mid n \left\|P_n - P_n'\right\|_{\rm TV} \leq k \right\} \tag{2}$$

of θ for the ℓ_2-distance at P_n, where the supremum is taken over samples $P_n' \in \mathcal{P}_n$ differing by at most k observations from P_n. Adding noise scaling as $\omega_\theta(P_n; 1)$ is, essentially, the best we could possibly hope to achieve in private estimation [2]. However, this local modulus is sensitive to the underying sample P_n, so naively using it cannot work, which motivates Nissim et al.'s smooth sensitivity [27].

Numerous other strategies for privately computing M-estimators (1) exist, and we touch briefly on a few here before turning to our own development. *Objective perturbation* strategies add a random linear term to the objective (1), and they appear to be among the most practical private estimators, though their adaptivity to particular problems is unclear [15, 29]. Other general approaches for convex M-estimation include (stochastic) gradient approaches, which perturb data within a gradient descent method [5, 6], enjoy some worst-case guarantees, but they also do not appear adaptive to local stability (2). Asi and Duchi [2] take a different approach and focus on low-dimensional quantities, introducing the *inverse sensitivity mechanism*. This mechanism is essentially instance optimal for releasing one-dimensional quantities, but appears challenging to compute except in certain special cases, as more sophisticated problems require high-dimensional integrals. Our investigation takes as a departing point insights from both Asi and Duchi [2] and Nissim et al. [27], but then carefully investigates particular stability properties inherent in M-estimators.

1.1 Heuristic development, motivating approach, and main results

To motivate all our following development, we begin with a quite heuristic derivation of the estimators we develop, and this overview allows us to highlight a few of the challenges along the way. The basic idea is straightforward: we would like to add noise commensurate with the local modulus (2). A Taylor approximation provides the starting point for our heuristic. Let P_n and P_n' be neighboring samples satisfying $\|P_n - P_n'\|_{\rm TV} \le \frac{1}{n}$, and let $\theta = \theta(P_n)$ and $\theta' = \theta(P_n')$ be the associated empirical minimizers. Then via a Taylor approximation, we should have

$$\begin{aligned} 0 = P_n' \dot{\ell}_{\theta'} + \lambda_{\rm reg}(\theta' - \theta_0) &= (P_n' - P_n)\dot{\ell}_{\theta'} + P_n \dot{\ell}_{\theta'} + \lambda_{\rm reg}(\theta' - \theta_0) \\ &= (P_n' - P_n)\dot{\ell}_{\theta'} + (P_n \ddot{\ell}_\theta + \lambda_{\rm reg} I + E)(\theta' - \theta), \end{aligned}$$

where $E = o(\|\theta' - \theta\|)$ is an error matrix and $P_n \dot{\ell}_\theta + \lambda_{\rm reg}(\theta - \theta_0) = 0$. Inverting, we obtain

$$\theta(P_n') - \theta(P_n) = (P_n \ddot{\ell}_\theta + \lambda_{\rm reg} I + E)^{-1}(P_n - P_n')\dot{\ell}_{\theta'} = (P_n \ddot{\ell}_\theta + \lambda_{\rm reg} I)^{-1}(P_n - P_n')\dot{\ell}_{\theta'} + o(1/n),$$

where we cavalierly assumed $P_n \ddot{\ell}_\theta + \lambda_{\rm reg} I + E$ is invertible and $(P_n - P_n')\dot{\ell}_{\theta'} = O(1/n)$.

Continuing with this heuristic motivation, we define the collection of possible gradients

$$\mathcal{G} := \left\{ \dot{\ell}_\theta(z) \mid z \in \mathcal{Z}, \theta \in \mathbb{R}^d \right\}.$$

Providing privacy relies on controlling the amount changing a single example can modify a parameter of interest; upon changing a single example, we have $\theta(P_n') \approx \theta(P_n) + \frac{1}{n}(P_n \ddot{\ell}_\theta + \lambda_{\rm reg} I)^{-1}(g_0 - g_1)$ for gradients $g_0, g_1 \in \mathcal{G}$. Thus, to within higher order error terms, the most the parameter $\theta(P_n)$ may change is the local *parameter* sensitivity

$$\Delta(P_n) := \frac{1}{n} \sup_{g_0, g_1 \in \mathcal{G}} \left\| (P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\rm reg} I)^{-1}(g_0 - g_1) \right\|_2. \tag{3a}$$

If instead we wish to estimate the linear functional $u^T \theta(P_n)$, then the *directional* sensitivity

$$\Delta(P_n, u) := \frac{1}{n} \sup_{g_0, g_1 \in \mathcal{G}} u^T (P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\rm reg} I)^{-1}(g_0 - g_1) \tag{3b}$$

bounds the change. That is, we should obtain the guarantees

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \le \Delta(P_n) \quad \text{and} \quad |u^T(\theta(P_n) - \theta(P_n'))| \le \Delta(P_n, u).$$

The sensitivies (3) asymptotically capture *exactly* the amount that $\theta(P_n)$ or $u^T\theta(P_n)$ may change in substituting a single example in P_n when n is large (that is, the local sensitivity or local modulus of continuity); for example, for robust regression (see Example 1 to come) with covariate vectors x drawn from any compact set, we have

$$\frac{\omega_\theta(P_n; 1)}{\Delta(P_n)} \to 1 \quad \text{and} \quad \frac{\sup\{|u^T(\theta(P_n) - \theta(P_n'))| \text{ s.t. } n\|P_n - P_n'\|_{\rm TV} \le 1\}}{\Delta(P_n, u)} \to 1$$

with probability 1 as $n \to \infty$ under i.i.d. sampling.

Making these heuristics rigorous will require privately certifying a lower and upper bounds on the minimal (respectively, maximal) eigenvalues

$$\lambda_{\min}(P_n) := \lambda_{\min}(P_n \ddot{\ell}_{\theta(P_n)}) \quad \text{and} \quad \lambda_{\max}(P_n) := \lambda_{\max}(P_n \ddot{\ell}_{\theta(P_n)}).$$

Our sharpest insights thus revolve around developing conditions that guarantee $P_n \ddot{\ell}_\theta$ itself provides explicit control over the error matrix E, making an elegant and practical use case for the theory of self-concordant functions [26, 9].

We can summarize our approach as follows: **(i)** Privately certify an estimate $\widehat{\lambda}$ of $\lambda_{\min}(P_n)$ that satisfies $\widehat{\lambda} \le \lambda_{\min}(P_n)$ with high probability. **(ii)** Given such an estimate, demonstrate that the sensitivities (3) or a proxy for them are stable to changes in P_n when $\lambda_{\min}(P_n)$ is far from 0, and that they bound (respectively) the local sensitivities $\|\theta(P_n) - \theta(P_n')\|_2$ and $|u^T\theta(P_n) - u^T\theta(P_n')|$. Then **(iii)** we release $\theta(P_n)$ or $u^T\theta(P_n)$ with additive Gaussian noise whose variance scales roughly with the sensitivies (3). To obtain (ε, δ)-differential privacy, we will essentially show the following: if we wish to release $\theta(P_n)$, (effectively) release

$$\widehat{\theta} \coloneqq \theta(P_n) + \mathsf{N}\left(0, O(1)\frac{\log \delta^{-1}}{\varepsilon^2}\Delta(P_n)^2\right), \tag{4a}$$

and if we wish to instead release the linear functional $u^T\theta(P_n)$, then we (effectively) release

$$\widehat{T} \coloneqq u^T\theta(P_n) + \mathsf{N}\left(0, O(1)\frac{\log \delta^{-1}}{\varepsilon^2}\Delta(P_n, u)^2\right). \tag{4b}$$

Combining each of these steps with appropriate privacy composition guarantees then gives a full procedure whose error scales roughly as

$$\|\widehat{\theta} - \theta(P_n)\|_2^2 = O(1)\frac{\log \delta^{-1}}{\varepsilon^2}\Delta(P_n)^2 \cdot d$$

or

$$|\widehat{T} - u^T\theta(P_n)| = O(1)\frac{\sqrt{\log \frac{1}{\delta}}}{\varepsilon}\Delta(P_n, u),$$

each holding with high probability. Each of these is unimprovable [12, 2]. To the extent that we can achieve these—which we make precise in the sequel—we obtain error scaling precisely with the local sensitivity (modulus of continuity) of the parameter of interest.

2 Preliminaries: loss classes and private mechanisms

We describe the classes of losses we study in problem (1) and provide privacy building blocks.

2.1 Loss classes of interest

Recalling problem (1), the smoothness and related properties of the loss function ℓ_θ will determine much of the difficulty of the problems we consider. We consider both general smooth losses and a more nuanced perspective tied to generalized linear models.

2.1.1 Generic smooth losses

The first class of losses we consider are Lipschitzean of up to second order. In particular, for each $z \in \mathcal{Z}$, we assume that $\theta \mapsto \ell_\theta(z)$ is G_0-Lipschitz, has G_1-Lipschitz gradient, and G_2-Lipschitz Hessian, all with respect to the ℓ_2-norm, meaning (respectively) that

$$\|\dot{\ell}_\theta\|_2 \le G_0, \quad \|\ddot{\ell}_\theta\|_{\mathrm{op}} \le G_1, \quad \|\ddot{\ell}_\theta - \ddot{\ell}_{\theta'}\|_{\mathrm{op}} \le G_2 \left\|\theta - \theta'\right\|_2,$$

where we leave the observation z implicit. For d-dimensional problems, we typically expect the scaling that $G_0 \lesssim \sqrt{d}$, $G_1 \lesssim d$, and $G_2 \lesssim d^{3/2}$.

We take two working examples, arising from typical applications in robust statistics and estimation in which the data is in pairs $z = (x, y) \in \mathbb{R}^d \times \mathcal{Y}$. Both are generalized linear model losses, where $\ell_\theta(z) = h(\langle \theta, x \rangle, y)$ for a function h convex in its first argument, so that

$$\dot{\ell}_\theta(x, y) = h'(\langle \theta, x \rangle, y)x, \quad \text{and} \quad \ddot{\ell}_\theta(x, y) = h''(\langle \theta, x \rangle, y)xx^T.$$

The Lipschitz constants then must scale with the ℓ_2-diameter of $x \in \mathcal{X}$, that is,

$$G_0 = \sup_{\theta, x \in \mathcal{X}, y \in \mathcal{Y}} \|\dot{\ell}_\theta(x, y)\|_2 = \left\|h'(\cdot, \cdot)\right\|_\infty \sup_{x \in \mathcal{X}} \|x\|_2$$

and similarly $G_1 = \|h''\|_\infty \sup_{x \in \mathcal{X}} \|x\|_2^2$ and $G_2 = \|h^{(3)}\|_\infty \sup_{x \in \mathcal{X}} \|x\|_2^3$.

Example 1 (Robust regression)**:** The standard approaches to robust regression [24] use either absolute error or Huber's robust loss, neither of which is $\mathcal{C}^2$—making private estimation quite challenging—so we consider a smoother variant. Consider

$$h(t) = \log(1 + e^t) + \log(1 + e^{-t}),$$

and for $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$ define the loss $\ell_\theta(y \mid x) = h(y - \langle x, \theta \rangle)$. We have $h'(t) = \frac{e^t - 1}{e^t + 1} \in (-1, 1)$, $0 < h''(t) = \frac{2e^t}{(e^t+1)^2} \le \frac{1}{2}$, while $h'''(t) = \frac{2e^t}{(e^t+1)^2} - \frac{4e^{2t}}{(e^t+1)^3} = \frac{2e^t - 2e^{2t}}{(e^t+1)^3} \in (-\frac{1}{5}, \frac{1}{5})$. Different settings of the domain $x \in \mathcal{X}$ yield different Lipschitz constants, assuming y may take on any real value. When $\mathcal{X} = [-1, 1]^d$, we thus obtain

$$G_0 = \sqrt{d}, \quad G_1 = \frac{d}{2}, \quad G_2 = \|h^{(3)}\|_\infty \sup_{x \in \mathcal{X}} \|x\|_2^3 \approx .19245 \cdot d^{3/2}.$$

Taking $\mathcal{X}$ to be the ℓ_2-ball of radius $r\sqrt{d}$ gives $G_0 = r\sqrt{d}$, $G_1 = \frac{r^2 d}{2}$, $G_2 < \frac{r^3 d^{3/2}}{5}$. $\diamond$

Example 2 (Binary logistic regression)**:** For binary logistic regression, we assume the data $(x, y) \in \mathbb{R}^d \times \{-1, 1\}$, and for

$$h(t) = \log(1 + e^{-t}) \quad \text{define} \quad \ell_\theta(x, y) = h(y\langle x, \theta \rangle).$$

For $\sigma(t) = \frac{1}{1+e^t}$ we have $h'(t) = -\sigma(t) \in (-1, 0)$, $0 < h''(t) = \sigma(t)(1 - \sigma(t)) \le \frac{1}{4}$, and $h'''(t) = \sigma(t)(1 - \sigma(t))(1 - 2\sigma(t)) \in (-.0963, .0963)$. So as in Example 1, if we assume that $x \in [-r, r]^d$, then we have

$$G_0 = r\sqrt{d}, \quad G_1 = \frac{r^2 d}{4}, \quad G_2 \le \frac{r^3 d^{3/2}}{10}$$

for binary logistic regression. $\diamond$

2.1.2 Quasi-self-concordant losses and generalized linear models

Combining the generalized linear model setting with some mild restrictions on the loss h allows us to obtain stronger results. To use our heuristic derivation in Section 1.1 to guarantee privacy, we need to provide *and* privately certify fairly precise control over the error matrix E. This suggests considering loss functions whose second derivatives appropriately bound themselves or for which second derivatives control the third derivatives, leading us to consider families of (approximately) self-concordant losses [9, 28].

Classical self-concordant functions [26, 9] satisfy

$$|f'''(t)| \le 2f''(t)^{3/2}$$

for all t. We will use variations on this classical condition, saying that a convex function $f : \mathbb{R} \to \mathbb{R}$ is *φ-quasi-self-concordant (q.s.c.)* if

$$f''(t)\left[1 - \varphi(|s|)\right]_+ \le f''(t+s) \le f''(t)(1 + \varphi(|s|)) \tag{5a}$$

for all $t, s \in \mathbb{R}$. In some cases, we only require the lower bound on the second derivative (5a), so we say that $f : \mathbb{R} \to \mathbb{R}$ is *φ-lower q.s.c.* if

$$f''(t)\left[1 - \varphi(|s|)\right]_+ \le f''(t+s) \tag{5b}$$

for all $t, s \in \mathbb{R}$. For a loss function $h : \mathbb{R} \times \mathcal{Y} \to \mathbb{R}$, we say that h (or the induced loss $\ell_\theta(x, y) = h(\langle \theta, x\rangle, y)$) is φ-quasi-self-concordant if $t \mapsto h(t, y)$ is for each y. Several important properties derive from these self concordance definitions, including that self-concordance implies the quasi-self-concordance condition (5a). Before returning to Examples 1 and 2, we thus collect a few properties of quasi-self-concordance and self-concordance.

Lemma 2.1 (Self-concordance properties)**.** *The following properties hold.*

(i) If for some $c < \infty$, the function f satisfies $|f'''(t)| \le cf''(t)$ for all t, then

$$e^{-c|s|}f''(t) \le f''(t+s) \le e^{c|s|}f''(t)$$

for all t, s, and so is φ-q.s.c. for $\varphi(s) = e^{cs} - 1$, or for $\varphi(s) = (e^c - 1)s$ for $s \le 1$ and $\varphi(s) = \infty$ otherwise.

(ii) Let f be self-concordant with $t \in \operatorname{dom} f$. Then

$$\frac{f''(t)}{(1 + |s|f''(t)^{1/2})^2} \le f''(t+s) \le \frac{f''(t)}{\left[1 - |s|f''(t)^{1/2}\right]_+^2},$$

where the lower bound holds when $t + s \in \operatorname{dom} f$.

(iii) If f is self-concordant, then it is φ-q.s.c. with $\varphi(s) = \left[1 - s \sup_t f''(t)^{1/2}\right]_+^{-2} - 1$.

The proofs of these results are standard; we include them in Appendix A.2 for completeness.

We now revisit our examples above in the context of quasi-self-concordance (5a).

Example 3 (Robust regression with log loss; Example 1 continued)**:** For robust regression with the log loss, recall that for $y \in \mathbb{R}$ we have $h(t, y) = \log(1 + e^{t-y}) + \log(1 + e^{y-t})$. Setting $\phi(t) = \log(1 + e^t) + \log(1 + e^{-t})$ yields

$$\frac{\phi'''(t)}{\phi''(t)} = \frac{e^t - e^{2t}}{e^t(e^t + 1)} = -\frac{e^{2t} - e^t}{e^{2t} + e^t} \in [-1, 1].$$

Leveraging Lemma 2.1.(i), we thus obtain

$$\begin{aligned} h''(t, y)\left[1 - |s|\right]_+ \le e^{-|s|}h''(t, y) &\le h''(t+s, y) \\ &\le e^{|s|}h''(t, y) \overset{(\star)}{\le} h''(t, y)\left(1 + (e-1)|s|\right), \end{aligned} \tag{6}$$

where inequality $(\star)$ holds for $|s| \le 1$. Robust regression with the log loss is φ-lower q.s.c. with $\varphi(s) = 1 - e^{-s}$ and is φ-q.s.c. with $\varphi(s) = (e^s - 1)$, as $e^{-s} \ge 1 - (e^s - 1)$ for $s \ge 0$. $\diamond$

Example 4 (Binary logistic regression; Example 2 continued)**:** For binary logistic regression, we have $h(t, y) = \log(1 + e^{-ty})$. Then for the sigmoid function $\sigma(t) = 1/(1 + e^t)$, the derivative calculations in Example 2 give

$$\frac{|h'''(t, y)|}{h''(t, y)} = |1 - 2\sigma(t)| \le 1,$$

so the bounds (6) hold as in Example 3; logistic regression is φ-q.s.c. with $\varphi(s) = (e^s - 1)$. $\diamond$

Ostrovskii and Bach [28] give additional examples of self-concordant loss functions satisfying the classical self-concordance definitions.

2.2 Composition, test-and-release mechanisms, and privacy building blocks

We record a few building block results on privacy that form the basis for our guarantees in the sequel. The first instantiates propose-test-release approaches [20] and composition of (ε, δ)-differentially private mechanisms. The application of the results we present will be to estimate a quantity approximating the local sensitivity of a statistic θ of interest, with a (private) certificate that the quantity upper bounds the local sensitivity $\omega_\theta(P_n; 1)$; we can then release the statistic with noise added commensurate to this bound, as in the motivation (4). We will use the shorthand

$$Z_0 \stackrel{d}{=}_{\varepsilon,\delta} Z_1$$

to mean that $\mathbb{P}(Z_0 \in A) \le e^\varepsilon \mathbb{P}(Z_1 \in A) + \delta$ for any measurable sets A.

2.2.1 Composition

We begin with the basic composition bound, which considers drawing a (private) random variable conditional on P_n, and then conditional on this value, releasing another statistic. To formalize this, assume we have a random variable $W \sim \mu(\cdot \mid P_n)$ taking values in a set $\mathcal{W}$ and a (randomized) mechanism M mapping $\mathcal{P}_n \times \mathcal{W} \to \mathcal{T}$ for a target set $\mathcal{T}$, where for each sample distribution P_n there exists a good set $G = G(P_n)$ for which

$$\mathbb{P}(M(P_n, w) \in A) \le e^\varepsilon \mathbb{P}(M(P_n', w) \in A) + \delta$$

for all $w \in G(P_n)$. We have the following minor extension of standard compositional guarantees [21] (which require that the good set is the full space, $G(P_n) = \mathcal{W}$).

Lemma 2.2. *Assume that W is $(\varepsilon_0, \delta_0)$-differentially private and $\mathbb{P}(W \in G(P_n) \mid P_n) \ge 1 - \gamma$ for all $P_n \in \mathcal{P}_n$. Then the composed pair*

$$(M(P_n, W), W)$$

is $(\varepsilon + \varepsilon_0, \delta + \delta_0 + \gamma)$ differentially private.

See Appendix A.1.1 for a proof of this lemma.

As a typical application of Lemma 2.2, we demonstrate a Gaussian mechanism. Letting $\Phi(\cdot)$ denote the standard normal c.d.f., define the (ε, δ)-variance

$$\sigma^2(\varepsilon, \delta) := \inf \left\{ \sigma^2 \mid \Phi(-\sigma\varepsilon - 1/2\sigma) + \Phi(-\sigma\varepsilon + 1/2\sigma) \le \delta \right\}. \tag{7}$$

As $\Phi(-t) \le e^{-t^2/2}$ for $t \ge 0$, it suffices to choose σ large enough that $-\sigma\varepsilon + \frac{1}{2\sigma} \le -\sqrt{2\log\frac{2}{\delta}}$, so solving the quadratic in σ guarantees

$$\sigma(\varepsilon, \delta) \le \sigma_{\text{naive}}(\varepsilon, \delta) := \frac{\sqrt{2\log\frac{2}{\delta}}}{2\varepsilon} + \frac{\sqrt{2\log\frac{2}{\delta} + 2\varepsilon}}{2\varepsilon}.$$

But as $\Phi(-t) \asymp \frac{1}{t}e^{-t^2/2}$ for t large, the formulation (7) is tighter.

The following lemma, which we prove for completeness in Appendix A.1.2, shows that this quantity is sufficient to guarantee privacy (see also [22]).

Lemma 2.3. *Let $\Phi(\cdot)$ denote the standard normal c.d.f., and let $\mu_0, \mu_1 \in \mathbb{R}^d$ and $\Delta^2 \ge \|\mu_0 - \mu_1\|_2^2$. Then $Z_i \sim \mathsf{N}(\mu_i, \Delta^2\sigma^2(\varepsilon, \delta)I_d)$, $i = 0, 1$, satisfy $Z_0 \overset{d}{=}_{\varepsilon,\delta} Z_1$.*

Now let $f : \mathcal{P}_n \to \mathbb{R}^d$ be a function of interest, and let W be an $(\varepsilon_0, \delta_0)$-differentially private estimate of the local modulus $\omega_f(P_n; 1)$ satisfying $\mathbb{P}(W \ge \omega_f(P_n; 1) \mid P_n) \ge 1 - \gamma$ for all sample distributions P_n. Define the mechanism

$$M(P_n) = f(P_n) + \mathsf{N}(0, W^2 \cdot \sigma^2(\varepsilon, \delta)I_d).$$

Then the following observation is an immediate consequences of Lemmas 2.2 and 2.3.

Observation 2.4. *The mechanism $M(P_n)$ above is $(\varepsilon + \varepsilon_0, \delta + \delta_0 + \gamma)$-differentially private.*

In our most sophisticated functional estimation problems, we will require a bit more subtlety in the closeness of Gaussian distributions; we defer such discussion until then.

2.2.2 Test and release

Lemma 2.2 allows us to present variants of the test and release framework [20], which privately tests that a sample P_n is "good enough," then uses a separate mechanism that is private on "good" samples. Thus, consider two mechanisms: the first, M_0, computes a statistic $(\varepsilon_0, \delta_0)$-differentially privately. If $M_0(P_n)$ satisfies some condition, we execute $M_1(P_n)$. We make the following abstract assumption:

A.1. There is a statistic $\lambda : \mathcal{P}_n \to \Lambda$ and a (deterministic) good set $G \subset \Lambda$ such that if $\lambda(P_n) \in G$, then $M_1(P_n) \overset{d}{=}_{\varepsilon,\delta} M_1(P_n')$ for all P_n' neighboring P_n.

A.2. There is an acceptance set A such that if $\lambda(P_n) \notin G$, then $\mathbb{P}(M_0(P_n) \in A) \le \delta_0$.

Loosely, we have the probabilistic implication that whenever $M_0(P_n) \in A$, excepting an event with probability δ_0, the statistic $\lambda(P_n) \in G$. We instantiate the following test/release scheme:

Algorithm 1: The Test/Release Scheme

Require: $(\varepsilon_0, \delta_0)$ and (ε, δ)-differentially private mechanisms M_0 and M_1 satisfying Assumptions A.1 and A.2, along with associated acceptance set A.

i. Release $M_0(P_n)$.

ii. If $M_0(P_n) \in A$, then release $M_1(P_n)$. Otherwise, release $\perp$.

Let $M(P_n)$ be the final output of the procedure 1, Then in Appendix A.1.3, we provide a proof of the following guarantee that $M(P_n)$ is private.

Lemma 2.5. *The mechanism $M(P_n)$ is $(\varepsilon_0 + \varepsilon, e^{\varepsilon_0}\delta_0 + \delta)$-differentially private.*

3 Parameter release algorithms for GLMs

We present the two main algorithms that apply our ideas, first to release the full vector $\theta(P_n)$, and second to release individual linear functionals $u^T\theta(P_n)$. The latter is our main interest, but the former exhibits the same techniques. As we outline in Section 2, this consists of two phases: first, we privately release a (putative) lower bound $\widehat{\lambda}$ on $\lambda_{\min}(P_n)$, which is both accurate and differentially private. Given such an estimate $\widehat{\lambda}$, we can then use (recall the definition (3a)) that $\|\theta(P_n) - \theta(P_n')\|_2 \le (1 + o(1))\frac{2G_0}{\lambda_{\min}(P_n)+\lambda_{\reg}}$ to release an estimate $\widehat{\theta}$ with appropriate noise. We use this approach in Section 3.1; in the subsequent Section 3.2, we extend the ideas to release individual parameters. In Section 3.3, we discuss the implied dimension dependence and accuracy guarantees of the main results here, especially in relation to the underlying geometry of the data, with some commentary on optimality as well.

To develop the ideas, we focus on generalized linear model losses ℓ that are φ-quasi-self-concordant, meaning that $\ell_\theta(x, y) = h(x^T\theta, y)$, where h satisfies inequality (5). We make a few restrictions to allow concrete algorithms, tacitly assuming these throughout this section: we require that for a constant $\alpha \ge 0$ and $\rho \in (0, 1)$, the self-bounding functional φ satisfies

$$\varphi(t) \le \alpha t \quad \text{if} \quad t \le \frac{1-\rho}{\alpha}. \tag{8}$$

Recalling Examples 3 (robust regression) and 4 (binary logistic regression), the choice $\varphi(t) = (e^t - 1)$ guarantees inequality (8) holds whenever $t = \frac{1-\rho}{\alpha}$ satisfies $e^t - 1 \le 1 - \rho$; the choices $\alpha = 1.2332 \le 1.234$ and $\rho = \frac{1}{2}$ suffice.

3.1 Releasing full parameter vectors

We preview the stability guarantees we prove in Section 5. Let $0 \le \alpha < \infty$ and $\rho \in (0, 1)$ be the constants in the linear bound (8) on the self-concordance function φ. Define the condition

$$\lambda_{\min}(P_n) + \frac{1}{\rho}\lambda_{\reg} \ge \frac{4G_0\alpha\rad(\mathcal{X})}{\rho(1-\rho)n} + \frac{G_1}{\rho n}, \tag{C1}$$

which guarantees that $\lambda_{\min}(P_n) + \lambda_{\reg}$ is large enough to certify stability: as a consequence of Proposition 5.4 in Sec. 5.1.2, we have

Corollary 3.1. *Let condition (C1) hold. Define*

$$t(\lambda) := \frac{1 - \sqrt{1 - \frac{8\alpha\rad(\mathcal{X})G_0}{\lambda n}}}{2\alpha\rad(\mathcal{X})}. \tag{9}$$

Then for any neighboring samples P_n and P_n' and $0 \le \lambda \le \lambda_{\min}(P_n) + \lambda_{\reg}$, we have

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \le t(\lambda).$$

So any guarantee that the (regularized) minimal eigenvalue $\lambda_{\min}(P_n) + \lambda_{\reg} \ge \lambda$ implies a stability guarantee on the parameters via the parameter change bound $t(\lambda)$ equation (9) defines. The bound satisfies $t(\lambda) \le \frac{3G_0}{n\lambda}$ under Condition (C1) (see the discussion following Proposition 5.4), it is monotonically decreasing in λ, and satisfies the asymptotic $t(\lambda) = \frac{2G_0}{n\lambda}(1 + o(1))$ as $n \to \infty$. The bound in Corollary 3.1 is thus sharp, in that for large n it converges to local sensitivity (3a) whenever the gradients $\mathcal{G}$ are a scaled ℓ_2-ball.

To leverage Corollary 3.1 and the test/release framework (Alg. 1), we thus seek to privately release the minimal eigenvalue $\lambda_{\min}(P_n)$. For this, we develop a new family of techniques for releasing parameters whose stability one can control recursively. Deferring the full development to Section 6, let R be a "recursion" function satisfying the following: for some nonnegative quantity $\lambda(P_n)$, we have the bound

$$\lambda(P_n') \geq R(\lambda(P_n)) \quad \text{and} \quad |\lambda(P_n') - \lambda(P_n)| \leq \lambda(P_n) - R(\lambda(P_n)),$$

for all neighboring P_n, P_n', so that R lower bounds $\lambda(P_n')$ and upper bounds the change in the parameter of interest: it provides a (local) guarantee of stability of $\lambda(P_n)$. For the N-fold composition R^N of R, we can calculate the smallest N yielding $R^N(\lambda(P_n)) = 0$; this value N is stable with respect to the sample P_n. By releasing a privatized variant $\widehat{N}$ of N and inverting the recursion, we may then release a private version of the quantity $\lambda(P_n)$ of interest.

To work in the context of GLMs, define the recursion

$$R(\lambda) := \begin{cases} \lambda\left[1 - \varphi\left(\mathrm{rad}(\mathcal{X})t(\lambda + \lambda_{\mathrm{reg}})\right)\right]_+ - \frac{G_1}{n} & \text{if } \lambda \text{ satisfies (C1)} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Corollary 5.2 in Section 5.2 to come shows this recursion provides the stability guarantee that $R(\lambda_{\min}(P_n)) \leq \lambda_{\min}(P_n')$ for any P_n' neighboring P_n. The following algorithm instantiates our discussion and releases (with high-probability) a lower bound on $\lambda_{\min}(P_n)$.

Algorithm 2: Privately lower bounding $\lambda_{\min}(P_n)$ for quasi-self-concordant GLMs.

Require: A φ-quasi-self-concordant loss where φ locally satisfies the linear upper bound (8), privacy parameters $\varepsilon \geq 0$ and $\delta \in (0, 1)$

i. Set the recursion R as in (10).

ii. Set

$$\widehat{N} := \min\left\{N \in \mathbb{N} \mid R^N(\lambda_{\min}(P_n)) = 0\right\} + \frac{1}{\varepsilon}\mathsf{Lap}(1).$$

iii. Set $k(\varepsilon, \delta) = \frac{1}{\varepsilon}\log\frac{1}{2\delta}$, then return $\widehat{N}$ and

$$\widehat{\lambda} := \sup\left\{\lambda \geq 0 \mid R^{[\widehat{N} - k(\varepsilon,\delta)]_+}(\lambda) = 0\right\}.$$

Proposition 6.1 in Section 6.1 then implies the following privacy guarantee.

Corollary 3.2. *Let the loss ℓ be φ-q.s.c. for $\varphi(t) = (e^t - 1)$. Then Algorithm 2 is ε-differentially private, and with probability at least $1-\delta$, $\widehat{\lambda}$ satisfies $\widehat{\lambda} \leq \lambda_{\min}(P_n)$. Additionally, there exists a numerical constant $C < \infty$ such that if $C\frac{G_0\mathrm{rad}(\mathcal{X})}{n} \leq \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$,*

$$\widehat{\lambda} \geq \lambda_{\min}(P_n) - O(1)\frac{1}{\varepsilon}\log\frac{1}{\delta}\left[\frac{G_0\mathrm{rad}(\mathcal{X})}{n}\frac{\lambda_{\min}(P_n)}{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}} + \frac{G_1}{n}\right]$$

with the same probability.

See Section 6.2.2 for a proof of the corollary.

With Corollary 3.2 in hand, Corollary 3.1 coupled with the privacy composition results we enumerate in Section 2.2 (Lemma 2.3 and Observation 2.4), this guarantees that Algorithm 3 is private, as the next theorem captures.

Algorithm 3: Local output perturbation for releasing $\theta(P_n)$

Require: A φ-quasi-self-concordant GLM loss $h : \mathbb{R} \times \mathcal{Y} \to \mathbb{R}$ satisfying the self-bounding condition (5a), privacy parameters $\varepsilon \geq 0$ and $\delta \in (0, 1)$

i. Let $\widehat{\lambda}$ be the output of Algorithm 2

ii. If $\widehat{\lambda} + \lambda_{\text{reg}} = 0$, return $\perp$

iii. Otherwise, let $\sigma^2(\varepsilon, \delta)$ be the normal variance (7). Return

$$\widehat{\theta} \coloneqq \theta(P_n) + \mathsf{N}\left(0, \sigma^2(\varepsilon, \delta) \cdot t^2(\widehat{\lambda} + \lambda_{\text{reg}}) I_d\right).$$

Theorem 1. *The output $\widehat{\theta}$ of Alg. 3 is $(2\varepsilon, 2\delta)$-differentially private. Additionally, there exists a numerical constant $C < \infty$ such that if*

$$\lambda_{\min}(P_n) + \lambda_{\text{reg}} \geq C\left(\frac{1}{\varepsilon}\log\frac{1}{\delta} \cdot \left[\frac{G_1}{n} + \frac{G_0 \text{rad}(\mathcal{X})}{n}\right] + \frac{G_0 \text{rad}(\mathcal{X})}{n} + \frac{G_1}{n}\right),$$

then with probability at least $1 - \delta - \gamma$,

$$\left\|\widehat{\theta} - \theta(P_n)\right\|_2 \leq C\frac{G_0}{n(\lambda_{\min}(P_n) + \lambda_{\text{reg}})}\frac{1}{\varepsilon}\sqrt{\log\frac{1}{\delta}}\left[\sqrt{d} + \sqrt{\log\frac{1}{\gamma}}\right].$$

See Section 7.2 for the full proof.

3.2 Releasing individual model parameters

One of our main desiderata is to release a single coordinate of the vector $\theta(P_n)$, or, more generally, to release

$$u^T\theta(P_n)$$

for a unit vector u. The key is that for different problem geometries—relating to the gradient set $\mathcal{G} = \{\dot{\ell}_\theta(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y}, \theta \in \mathbb{R}^d\}$—the minimal and maximal eigenvalues $\lambda_{\min}(P_n)$ and $\lambda_{\max}(P_n)$ of $P_n\ddot{\ell}_{\theta(P_n)}$ certify bounds on the stability of $u^T\theta(P_n)$. The following corollary (Lemma 5.2 in the proof of Proposition 5.4) captures this for self-concordant losses (8).

Corollary 3.3. *For $u \in \mathbb{R}^d$ and $\lambda \geq 0$, let $\Delta(P_n, u)$ be the directional difference (3b) and $t(\lambda)$ be the parameter change bound (9). Define*

$$\gamma(\lambda) \coloneqq \alpha \cdot t(\lambda)\text{rad}_2(\mathcal{X})$$

and

$$\omega(u \mid P_n) \coloneqq \Delta(P_n, u) + \frac{2G_0}{n(\lambda_{\min}(P_n) + \lambda_{\text{reg}})} \cdot \frac{\gamma(\lambda_{\min}(P_n) + \lambda_{\text{reg}})}{1 - \gamma(\lambda_{\min}(P_n) + \lambda_{\text{reg}})}\|u\|_2. \tag{11}$$

Then

$$\left|u^T(\theta(P_n') - \theta(P_n))\right| \leq \omega(u \mid P_n).$$

Corollary 3.3 allows us to use the propose-test-release scheme to argue that releasing

$$u^T\theta(P_n) + \omega(u \mid P_n) \cdot Z$$

for a Gaussian Z with variance scaling as $\frac{1}{\varepsilon^2}\log\frac{1}{\delta}$ is private so long as we can privately certify that $\lambda_{\min}(P_n)$ is large enough and $\lambda_{\max}(P_n)$ is small enough, because these combined imply that the ratio $\omega(u \mid P_n)/\omega(u \mid P_n')$ is close to one whenever P_n and P_n' are neighboring.

3.2.1 Releasing the maximal eigenvalue

To address that we must certify that $\lambda_{\max}(P_n)$ is not too large, we adapt Algorithm 2 to release an approximation to $\lambda_{\max}(P_n)$, and follows here. Recalling the self-concordance function $\varphi(t) \le \alpha t$, for a fixed value $\widehat{\lambda}$, we define the increasing recursion

$$R_{\widehat{\lambda}}(\lambda) := \min\left\{\lambda \cdot \left(1 + \varphi(t(\widehat{\lambda}) \cdot \mathrm{rad}(\mathcal{X}))\right) + \frac{G_1}{n}, G_1\right\}. \tag{12}$$

Then via a derivation and justification completely parallel to that we have done for the lower eigenvalues, so that we find the smallest N such that $R^N(\lambda_{\max}(P_n)) \ge G_1$ (recalling that the Lipschitz constant G_1 of the gradients upper bounds $\lambda_{\max}(P_n)$), we obtain that the following algorithm is ε-differentially private.

Algorithm 4: A private upper bound on $\lambda_{\max}(P_n)$

Require: A φ-quasi-concordant loss where φ locally satisfies the linear upper bound (8), privacy parameters $\varepsilon \ge 0$ and $\delta \in (0,1)$, ε-differentially private estimate $\widehat{\lambda}_{\min}$ satisfying $\widehat{\lambda}_{\min} \le \lambda_{\min}(P_n)$ with probability at least $1-\delta$.

i. Set the recursion $R = R_{\widehat{\lambda}_{\min}+\lambda_{\mathrm{reg}}}$ as in (12).

ii. Set
$$\widehat{N} := \min\left\{N \in \mathbb{N} \mid R^N(\lambda_{\max}(P_n)) \ge G_1\right\} + \frac{1}{\varepsilon}\mathsf{Lap}(1).$$

iii. Set $k(\varepsilon,\delta) = \frac{1}{\varepsilon}\log\frac{1}{2\delta}$, then return $\widehat{N}$ and
$$\widehat{\lambda} = \min\left\{\inf\left\{\lambda \mid R^{N-k(\varepsilon,\delta)}(\lambda) \ge G_1\right\}, G_1\right\}.$$

Corollary 3.4. *Let the loss ℓ be φ-q.s.c. for $\varphi(t) = e^t - 1$. Let $\widehat{\lambda}_{\min}$ be the output of Algorithm 2 and $\widehat{\lambda}_{\max}$ be the output of Algorithm 4 with this input. Then the pair*

$$(\widehat{\lambda}_{\min}, \widehat{\lambda}_{\max})$$

is $(2\varepsilon,\delta)$-differentially private and satisfies $\widehat{\lambda}_{\min} \le \lambda_{\min}(P_n)$ and $\widehat{\lambda}_{\max} \ge \lambda_{\max}(P_n)$ with probability at least $1-2\delta$. Additionally, there exists a numerical constant $C < \infty$ such that if $C\frac{G_0\mathrm{rad}(\mathcal{X})}{n} \le \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$ then with the same probability

$$\widehat{\lambda}_{\max} \le \lambda_{\max}(P_n) + O(1)\frac{1}{\varepsilon}\log\frac{1}{\delta}\left[\frac{G_0\mathrm{rad}(\mathcal{X})}{n}\frac{\lambda_{\max}(P_n)}{\lambda_{\min}(P_n)+\lambda_{\mathrm{reg}}} + \frac{G_1}{n}\right].$$

See Section 6.2.3 for a proof of this corollary.

3.2.2 Releasing the linear functional

Now that Algorithms 2 and 4 demonstrate that accurately releasing minimal and maximal eigenvalues is possible, we can provide a test-release scheme that first checks whether one can certify that the local moduli of continuity $\omega(u \mid P_n)$ are similar for P_n' neighboring P_n, and then—assuming they are—releases a noisy version of $u^T\theta(P_n)$. The key are stability guarantees on the ratio $\omega(u \mid P_n)/\omega(u \mid P_n')$, which in turn imply that $\mathsf{N}(0, \omega(u \mid P_n))$ and

$\mathsf{N}(0, \omega(u \mid P_n'))$ are appropriately close distributions so that the release $u^T\theta(P_n) + \omega(u \mid P_n) \cdot Z$ is private. These rely on a series of constants, all implicitly dependent on the estimated minimal eigenvalue $\lambda_0 \approx \lambda_{\min}(P_n) + \lambda_{\reg}$ and maximal eigenvalue $\lambda_1 \approx \lambda_{\max}(P_n) + \lambda_{\reg}$, and that (to actually describe the algorithm) we define here:

$$\begin{aligned} t &:= t(\lambda_0), \quad r := \rad_2(\mathcal{X}), \quad \beta := \frac{\|h''\|_\infty}{[1-\alpha t]_+} \frac{r^2}{n\lambda_0}, \quad \gamma := \alpha r \cdot t, \quad \gamma' := \alpha r \cdot t(R(\lambda_0)) \\ s_1 &:= \frac{1}{[1-\alpha r t]_+} - 1, \quad s_2 := \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{[1-\alpha r t]_+}, \quad \kappa := \frac{\lambda_1}{\lambda_0}. \end{aligned} \tag{13}$$

Then Propositions 8.2 and 8.3 in Section 8.1 imply the following corollary.

Corollary 3.5. *For $2 \le p \le \infty$, let the gradient set $\mathcal{G}_p := \{g \in \mathbb{R}^d \mid \|g\|_p \le d^{\frac{1}{p}-\frac{1}{2}} G_0\}$ and R be the recursion (10). Then for $p = 2$,*

$$\left(1 + \kappa(s_1 + s_2 r) + \frac{\kappa\lambda_0}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}\right)^{-1} \le \frac{\omega(u \mid P_n)}{\omega(u \mid P_n')} \le \frac{1 + \kappa\frac{\gamma}{1-\gamma}}{1 - \kappa(s_1 + s_2 r)}.$$

For $p > 2$, let $d_p = d^{1-2/p}$. Then

$$\left(\left(1 + \sqrt{d_p} s_1 \kappa + \frac{2 s_2 d_p}{\lambda_0}\right) + \frac{\sqrt{d_p}\kappa\lambda_0}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}\right)^{-1} \le \frac{\omega(u \mid P_n)}{\omega(u \mid P_n')} \le \frac{1 + \sqrt{d_p}\kappa\frac{\gamma}{1-\gamma}}{1 - \sqrt{d_p} s_1 \kappa - 2 d_p s_2/\lambda_0}.$$

We shall see that $(\frac{\omega(u|P_n)}{\omega(u|P_n')})^2 - 1 \lesssim \varepsilon / \log\frac{1}{\delta}$ and $(\frac{\omega(u|P_n')}{\omega(u|P_n)})^2 - 1 \lesssim \varepsilon / \log\frac{1}{\delta}$ is enough to guarantee that releasing $u^T\theta(P_n) + \mathsf{N}(0, \sigma^2(\varepsilon,\delta) \cdot \omega(u \mid P_n)^2)$ is private. Let Φ^{-1} denote the standard inverse Gaussian cumulative distribution function, so that $\Phi^{-1}(1-\delta)^2 \le \log\frac{1}{\delta}$. Recalling the generalized linear models in Examples 1 and 2, where the radius of the covariate vectors $x \in \mathcal{X}$ governs smoothness properties, we consider two cases. In the case that gradients belong to the $p = 2$-norm ball, we check that

$$\max\left\{\frac{1 + \kappa\frac{\gamma}{1+\gamma}}{1 - \kappa(s_1 + s_2 r)}, 1 + \kappa(s_1 + s_2 r) + \frac{\lambda_1}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}\right\}^2 - 1 \le \frac{2\varepsilon}{1 + \Phi^{-1}(1-\delta/2)^2}. \tag{14a}$$

For $p > 2$, let $d_p = d^{1-\frac{2}{p}}$ and check that

$$\begin{aligned} \max\left\{\frac{1 + \sqrt{d_p}\kappa\frac{\gamma}{1+\gamma}}{1 - \sqrt{d_p}\kappa s_1 - 2 d_p s_2/\lambda_0}, 1 + \sqrt{d_p}\kappa s_1 + \frac{2 d_p s_2}{\lambda_0} + \frac{\sqrt{d_p}\lambda_1}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}\right\}^2 - 1 \\ \le \frac{2\varepsilon}{1 + \Phi^{-1}(1-\delta/2)^2}. \end{aligned} \tag{14b}$$

With these definitions, Algorithm 5 then privately releases a version of $u^T\theta(P_n)$.

Algorithm 5: Releasing a one-dimensional statistic

Require: A φ-quasi-concordant loss where φ locally satisfies the linear upper bound (8), privacy parameters $\varepsilon \geq 0$ and $\delta \in (0,1)$

i. Let $\widehat{\lambda}_{\min}$ and $\widehat{\lambda}_{\max}$ be the outputs of Algorithms 2 and 4, respectively

ii. If ε fails to satisfy the appropriate inequality (14) return $T = \perp$.

iii. So long as the pair $\widehat{\lambda}_{\min}$ and λ_{reg} satisfy condition (C1), return

$$T := u^T \theta(P_n) + \mathsf{N}\left(0, \sigma^2(\varepsilon, \delta) \cdot \omega(u \mid P_n)^2\right)$$

Otherwise return $T = \perp$.

Then as a corollary of the main results in Section 8 (see Section 8.1.1 for the proof), we have the following result.

Corollary 3.6. *Let $\varepsilon \geq 0$ and $\delta \in (0,1)$. Then T is $(3\varepsilon, (1 + e^{\varepsilon} + e^{2\varepsilon})\delta)$-differentially private.*

Unpacking Corollary 3.6, we see the following: as soon as we can guarantee the conditions (14) hold, we can then release the actual statistic $u^T\theta(P_n)$ with noise scaling precisely as the (slightly enlarged) local modulus of continuity (11).

In passing, we note a slight but practically important improvement that we employ in our experiments. Because Algorithm 5 performs three private operations: two using Laplace mechanisms and the last with a Gaussian, we can use privacy loss random variables and explicit calculations to programatically achieve sharper privacy bounds than that in the corollary [1].

3.3 Dimension dependence and commentary

While the algorithms we have presented are relatively straightforward to implement, we still must discuss the dimension and sample-size scaling they require and accuracy guarantees they provide, especially in the context of the necessary dimension-dependent penalties privacy enforces [31, 4, 12]. Let us focus on the scaling that arises in Corollary 3.6 for large sample sizes n (see also Proposition 8.3 to come). For simplicity, we assume the self-bounding constants (8) satisfy $\alpha = O(1)$ (e.g., as in Examples 3 and 4), and that the q.s.c. function h has $O(1)$-Lipschitz zeroth, first, and second derivatives. Then the problem scaling all boils down to the ℓ_2 radius $r = \text{rad}_2(\mathcal{X})$ of the covariates $\mathcal{X}$, so that $G_i = O(1) \cdot r^{1+i}$. In this case, Condition (C1) becomes that $\lambda_{\min}(P_n) + \lambda_{\text{reg}} \gtrsim \frac{r^2}{n}$, while letting $\lambda = \lambda_{\min}(P_n)$ be shorthand for the smallest eigenvalue, the parameter change (9) satisfies

$$t(\lambda) = \frac{2G_0}{\lambda n}(1 + o(1)) \asymp \frac{r}{\lambda n}$$

Substituting $r = \text{rad}_2(\mathcal{X})$ and the above values into the constants (13), we obtain $s_1 \asymp \frac{1}{1 - r^2/\lambda n} - 1 \asymp \frac{r^2}{\lambda n}$, $s_2 \asymp \frac{1}{n}$, and $\gamma \asymp \frac{r^2}{\lambda n}$. Finally, we specialize a bit to the case that $\mathcal{X}$ is contained in a scaled ℓ_p-ball for some $2 \leq p \leq \infty$. In this case, condtion (14b) essentially subsumes condition (14a). Letting $d_p = d^{1 - \frac{2}{p}}$ and $\kappa = \frac{\lambda_{\max}(P_n) + \lambda_{\text{reg}}}{\lambda_{\min}(P_n) + \lambda_{\text{reg}}}$, condition (14b) thus (for large n) becomes equivalent to

$$\kappa \frac{\sqrt{d_p} r^2}{\lambda n} \lesssim \frac{\varepsilon}{\log \frac{1}{\delta}}.$$

Once this scaling holds, then Algorithm 5 releases $T = u^T\theta(P_n) + \omega(u \mid P_n) \cdot \sigma(\varepsilon, \delta)\mathsf{N}(0,1)$, where we recall the definition (11) of $\omega(u \mid P_n) = \Delta(P_n, u) + O(1)\frac{d^{3/2}}{n^2}$, the optimal scaling. We summarize this with the following theorem.

Theorem 2. *Let the losses ℓ_θ satisfy the smoothness conditions of Algorithm 5, and assume that the covariate domain $\mathcal{X}$ has ℓ_2-radius $r = \mathrm{rad}_2(\mathcal{X})$. Then the output T of Algorithm 5 is $(3\varepsilon, (1 + e^\varepsilon + e^{2\varepsilon})\delta)$-differentially private. Let the notation above hold. Then additionally, there is a numerical constant $C < \infty$ such that if*

$$C \cdot \kappa \frac{\sqrt{d_p} r^2}{(\lambda_{\min}(P_n) + \lambda_{\reg})n} \le \frac{\varepsilon}{\log \frac{1}{\delta}},$$

then with probability at least $1 - \delta - \gamma$,

$$\left|T - u^T\theta(P_n)\right| \le C \frac{1}{n} \cdot \sup_{x \in \mathcal{X}} \left|u^T (P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\reg} I)^{-1} x\right| \cdot \frac{1}{\varepsilon} \sqrt{\log \frac{1}{\delta}} \sqrt{\log \frac{1}{\gamma}}.$$

So we see a somewhat interesting behavior, which we believe is worth investigating, though leave to future work: once the sample size is large enough, then Algorithm (5) releases $u^T\theta(P_n)$ with noise scaling exactly (up to the higher-order term) as the local modulus of continuity $\Delta(P_n, u)$. But until we have sufficient sample size to dominate the dimension, the presented procedures are likely impractical. There appear to be two condition-number-like quantities: the actual condition number $\kappa = \frac{\lambda_{\max}(P_n) + \lambda_{\reg}}{\lambda_{\min}(P_n) + \lambda_{\reg}}$, and one relating the scale of the covariates $\mathcal{X}$ to the curvature $\lambda_{\min}$ of the problem, with a dimension-dependent penalty.

Reifying the theorem by considering the cases that $\mathcal{X}$ is an ℓ_2-ball of radius $\sqrt{d}$ or the hypercube $\{-1, 1\}^d$, we see that once the sample size is large enough—i.e., it satisfies $\kappa d \ll n$ in the former case and $\kappa d^{3/2} \ll n$ in the latter—we achieve optimal private estimation.[1] It would be interesting to understand if this dimensional scaling is fundamental in some sense. In statistical problems in the "high-dimensional asymptotic" regime that $d/n \to c$ for a constant $0 < c < \infty$ (or even $d^2/n \to 0$), certain interesting functionals are estimable, such as the mean-squared error of a predictor [16]. With privacy, these questions appear to be subtle.

4 Experiments

We complement our theoretical work with experimental results that compare the proposed algorithm to existing algorithms for privately estimating a single parameter. We consider two settings. In the first, we evaluate the procedures here, along with alternative private algorithms, on a simulated robust regression dataset (as in Example 1), where we create a synthetic dataset to allow us to test different aspects of the algorithms here and their relationship with others. In the second, we consider the Folktables dataset [17], which consists of datasets derived from the US Census. In each experiment, we consider five procedures:

1. Localized output perturbation (the methods in this paper). When estimating the entire parameter $\theta(P_n)$, this corresponds to Alg. 3, while estimating the linear functional $u^T\theta(P_n)$ corresponds to Algorithm 5.

[1] We note in passing that these scalings appear to be unimprovable using our analyses, though plausibly a much cleaner treatment is possible.

2. A non-private and idealized variant of the procedure above, where we release the parameter of interest with noise scaling as its local sensitivity, that is,

$$\theta(P_n) + \mathsf{N}(0, \Delta^2(P_n) \cdot \sigma^2(\varepsilon, \delta) I_d) \quad \text{or} \quad u^\top \theta(P_n) + \mathsf{N}(0, \Delta^2(P_n, u) \cdot \sigma^2(\varepsilon, \delta) I_d),$$

where $\Delta(P_n)$ and $\Delta(P_n, u)$ are the sensitivies (3) and $\sigma^2(\varepsilon, \delta)$ is the private variance (7).

3. Differentially private stochastic gradient descent (DP-SGD) [5, 6], where we have non-privately searched for hyper-parameters to select batch sizes and total iterations.

4. Naive output perturbation, which for $\theta_{\lambda_{\rm reg}}(P_n) := \operatorname{argmin}_\theta \{L_n(\theta) + \frac{\lambda_{\rm reg}}{2} \|\theta\|_2^2\}$ releases

$$\theta(P_n) + \mathsf{N}\left(0, \frac{4G_0^2}{n^2 \lambda_{\rm reg}^2} \cdot \sigma^2(\varepsilon, \delta) I_d\right),$$

as $\|\theta_{\lambda_{\rm reg}}(P_n) - \theta_{\lambda_{\rm reg}}(P_n')\|_2 \le \frac{2G_0}{n\lambda_{\rm reg}}$ is trivially stable. We set $\lambda_{\rm reg} = 10^{-2}$.

5. Objective perturbation [15] with optimized parameter settings [29]. For linear models (logistic or robust regression) as in this paper, this releases

$$\widehat{\theta}(P_n) = \operatorname*{argmin}_\theta \left\{ L_n(\theta) + W^T \theta + \frac{\lambda_{\rm reg}}{2} \|\theta\|_2^2 \right\}, \tag{15}$$

where $\lambda_{\rm reg} = \frac{4G_1}{n\varepsilon}$ and $W \sim \mathsf{N}(0, \sigma^2)$ for $\sigma^2 = \frac{2{\rm rad}(\mathcal{X})}{n\varepsilon}\sqrt{2\log\frac{4}{\delta}} + \sqrt{2\varepsilon + 2\log\frac{4}{\delta}}$. These choices guarantee (ε, δ)-differential privacy.

4.1 Robust Regression (synthetic experiments)

In our simulated data, we experiment with robust regression. To generate data for the experiments, we fix a sample size n and dimension d, then generate $\theta^\star \sim r\mathsf{Uni}(\mathbb{S}^{d-1})$, varying the radius $r = \|\theta^\star\|_2$. We sample $x_i \stackrel{\rm iid}{\sim} \mathsf{Uni}[-1, 1]^d$, and draw $y_i = \langle \theta^\star, x_i \rangle + z_i$ for $z_i \stackrel{\rm iid}{\sim} \sigma \cdot \mathsf{Lap}(1)$, $i = 1, \ldots, n$. With this setting, we consider either estimating $\theta_1^\star$, the first coordinate of $\theta^\star$, or the vector $\theta^\star$. Each of our reported results corresponds to average results over 25 such experiments, where we provide (approximate) 95% confidence intervals of ± 2 standard errors. We consider four distinct experimental settings: (1) releasing eigenvalues, (2) error in releasing the functional $e_1^T \theta(P_n)$ versus sample size n, and (3) and (4) evaluating error versus privacy ε in releasing $e_1^T \theta(P_n)$ or the full vector $\theta(P_n)$.

Because Algorithm 5 outputs $\perp$ when it cannot certify $\lambda_{\min}(P_n) > 0$, our first experiment investigates the relative error $|\widehat{\lambda} - \lambda_{\min}(P_n)|/\lambda_{\min}(P_n)$ in the eigenvalue Algorithm 2 releases. We show results in Figure 1, varying the sample size ratio n/d and for dimensions $d = 5, 10, 20$ and fixing $\varepsilon = 1$, $\delta = 10^{-6}$. This figure makes clear that, while the procedure eventually achieves quite small error, the sample sizes necessary may be quite large for most tasks—hence, in the sequel, our focus on census data with relatively small numbers of covariates. We will return to this point in the discussion, when we suggest future work.

For our second set of experiments, we fix $\varepsilon = 4$ to investigate the scaling of errors with the sample size n for estimating $e_1^T \theta(P_n)$, the first coordinate of $\theta(P_n)$. Based on our theoretical results and Fig. 1, we expect that Algorithm 5 should exhibit a type of thresholding behavior: when n is too small, we cannot certify that $\lambda_{\min}(P_n)$ is large enough to guarantee stability, and so must release statistics with substantial noise. When n is large enough, we expect to achieve error near that of the non-private procedure adding noise scaling exactly as the

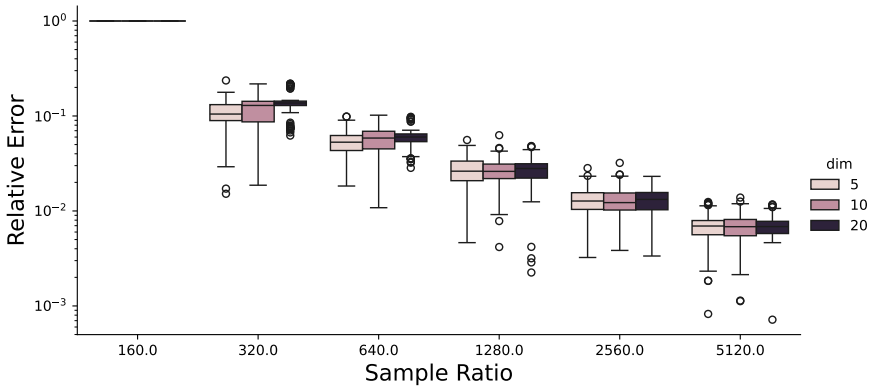

Figure 1. Relative error $\frac{|\widehat{\lambda}-\lambda_{\min}(P_n)|}{\lambda_{\min}(P_n)}$ in eigenvalue estimate versus sampling ratio $r = n/d$ for dimensions $d \in \{5, 10, 20\}$ on simulated robust regression. For ratios $r \geq 160$, Alg. 2 releases $\widehat{\lambda} = 0$.

local sensitivity (see item 2 above). While prior work suggests objective perturbation (15) should be a competitive and easy-to-use algorithm [15, 29], that it regularizes its parameter θ around 0 suggests that there should be a gap between its performance and the methods here as $\|\theta^\star\|_2$ grows. Figures 2 and 3 show the results of these experiments for different dimensions d; the results are consistent with our expectations: the non-private method has (by far) the best performance, while Algorithm 5 eventually achieves errors near the "best possible" non-private release.

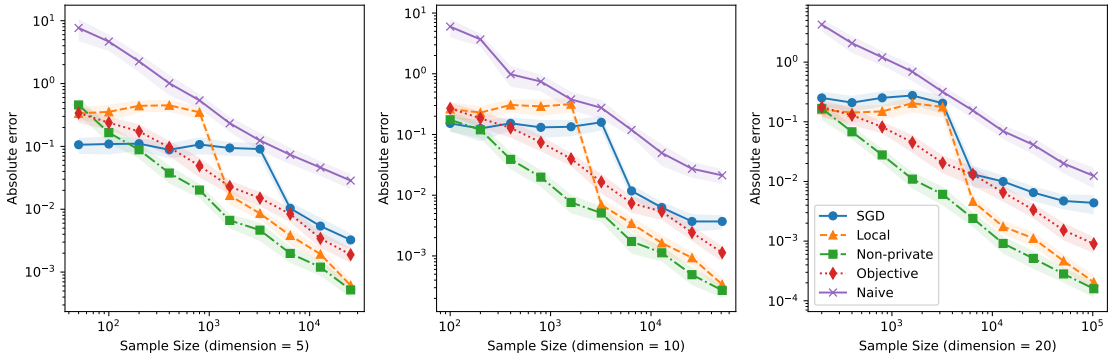

Figure 2. Error $|\theta_1^\star - \widehat{\theta}_1(P_n)|$ in the first-coordinate of the target $\theta(P_n)$ versus sample size for varying dimensions $d = 5, 10, 20$ in a robust regression experiment, where $\|\theta^\star\|_2 = 1$. The method Local (orange triangle) is Algorithm 5; SGD is DPSGD, non-private is the non-private idealized version of the methods here (item 2), objective is objective perturbation (15), and naive is the naive output perturbation estimator. Objective perturbation and the methods here exhibit the best performance, with Alg. 5 exhibiting a noticeable improvement at sufficiently large sample size.

Finally, Figures 4 and 5 investigate the error $\widehat{\theta} - \theta(P_n)$ for single coordinates (Fig. 4) and the full parameter vector $\theta(P_n)$, respectively, as ε increases, for fixed sample size $n = 10^5$ and dimension $d = 10$. The plots are consistent with our observations and expectations to this point: differentially private SGD and objective perturbation both exhibit reasonable performance, but are worse than the local release Algorithm 5 for estimating a single coordinate. On the other hand, objective perturbation is very competitive when $\theta^\star$ is small, but has some degradation as the norm of $\theta^\star$ increases (plots (b) in each figure).

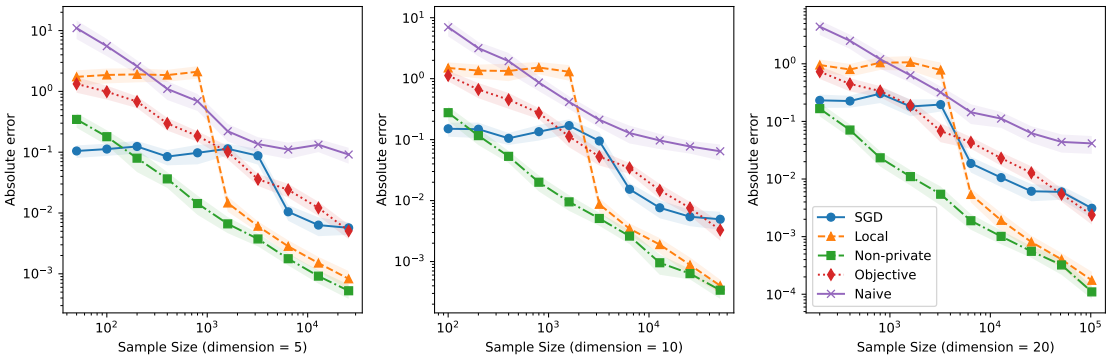

Figure 3. Identical to Fig. 2, except that $\|\theta^\star\|_2 = 5$. Note that the gap in performance between objective perturbation and Alg. 5 is larger than in the case that $\|\theta^\star\|_2 = 1$.

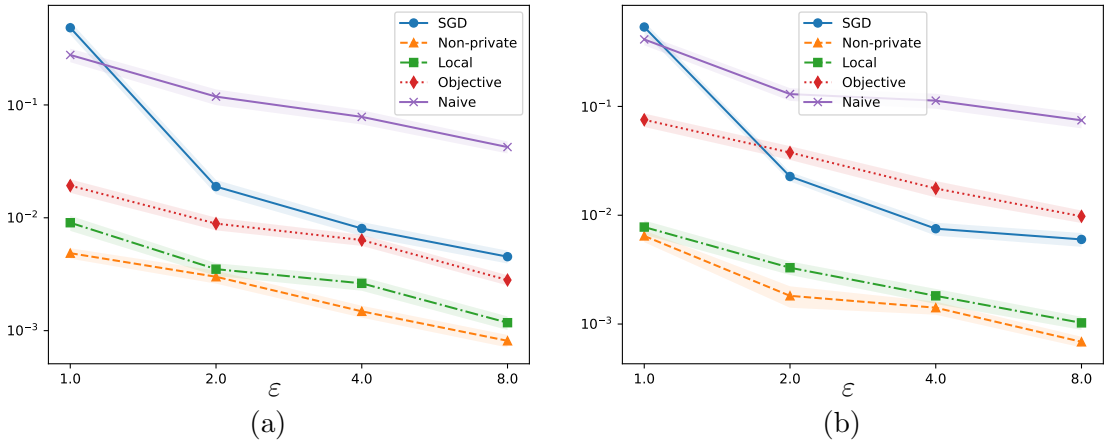

Figure 4. Error $|\widehat{\theta}_j - \theta^\star_j|$ as a function of the privacy parameter ε for simulated robust regression estimating a single (random) coordinate j. (a) Small norm $\|\theta^\star\|_2 = 1$ (b) Larger norm $\|\theta^\star\|_2 = 6$. Dimension $d = 10$ in both and sample size $n = 10^5$.

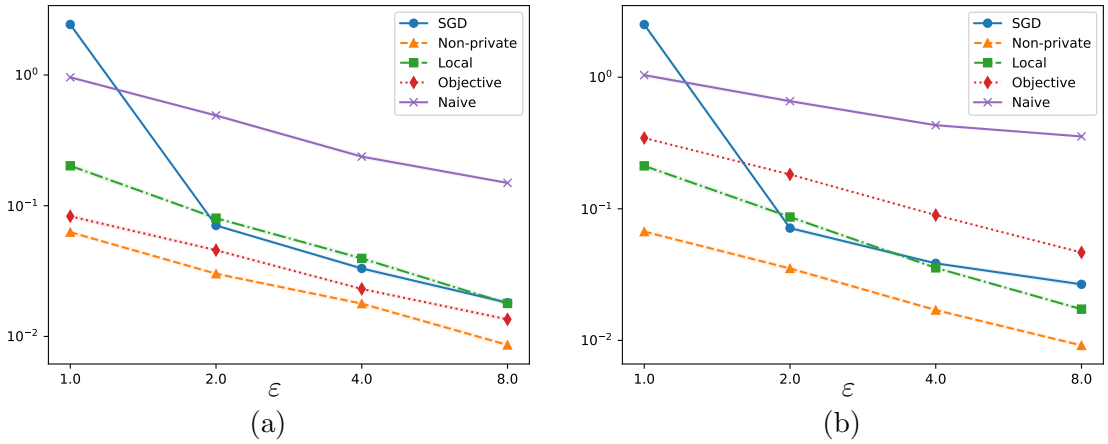

Figure 5. Error $\|\theta - \theta^\star\|_2$ as a function of the privacy parameter ε for simulated robust regression estimating entire parameter $\theta^\star \in \mathbb{R}^d$, where $d = 10$, and sample size $n = 10^5$. (a) Small norm $\|\theta^\star\|_2 = 1$ (b) Larger norm $\|\theta^\star\|_2 = 6$.

4.2 Logistic Regression (folktables)

We also investigate the performance of the methods we develop on an income prediction with data from the American Community Survey (ACS), part of the US Census, as implemented in the FolkTables datasets [17]. We use state-level data drawn from the 2018 edition of the survey, fitting logistic regression predictors of income, where $Y = 1$ if the income of an individual is above \$40,000 and $Y = -1$ otherwise. As features we take the following covariates: an indicator of working age (between 18 and 60 years old); hours-per week of work over the past year (normalized to the interval $[-1, 1]$); schooling level from -1 (no grade school) to 1 (graduate degree); an indicator of whether an individual is white; a 1-hot encoding of occupation mapped to 8 distinct areas[2]; an indicator of marital status; an indicator of sex; and a 1-hot encoding of whether an individual is employed in a private corporation, government, self-employed, or has unknown employment. Including an intercept term and eliminating linear dependence in the features, this yields $d = 17$-dimensional problem data, with covariate vectors $x_i \in [-1, 1]^d$.

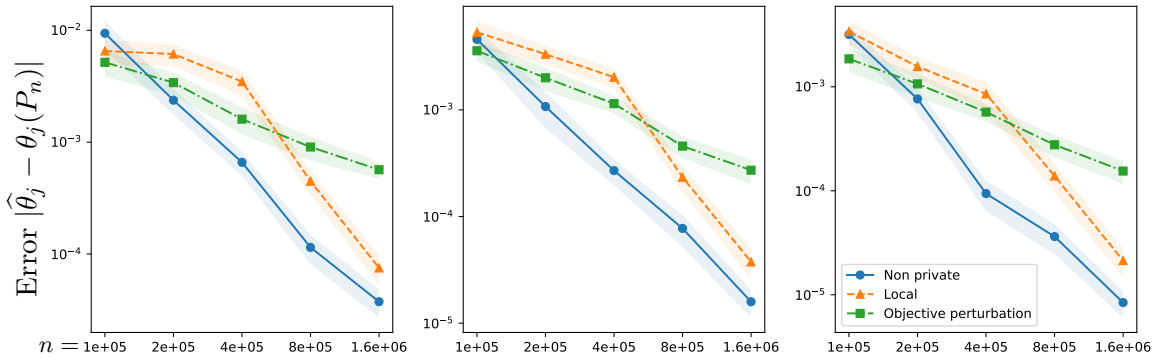

Figure 6. Estimation error versus sample size n for the parameter θ_j corresponding to the `SEX` indicator in predicting income levels using California survey data with approximate 95% confidence intervals based on 25 experiments. Left: privacy level $\varepsilon = 2$. Middle: privacy level $\varepsilon = 4$. Right: privacy level $\varepsilon = 8$. The true parameter $\theta_j(P_n) \approx -.27$. The "Non-private" estimator is the non-private idealized version (item 2), objective perturbation corresponds to (15), and Local is Algorithm 5.

We present results for experiments on data from California and Michigan; results on other states are similar. For each set of experiments, we treat the available data for the state as the population, then draw a sample (with replacement) of size n to give an empirical distribution P_n, investigating estimators on the sample P_n. We modify Algorithm 5 slightly, so that if the estimated minimal eigenvalue $\widehat{\lambda} = 0$, the method instead uses objective perturbation with privacy parameter $\varepsilon/2$ (as we use half of the privacy budget ε to estimate $\widehat{\lambda}$). On the data from California, the Hessian $\nabla^2 L(\theta^\star)$ has condition number 300 ± 4, while the Michigan data yields condition number 450 ± 6, making the problems moderately poorly conditioned (these condition numbers were large enough that even tuned DP-SGD had error $\geq 10^{-1}$, so we do not include its results in the experiments). Based on our results in simulation, we expect two main results in our experiments: first, for small sample sizes, we expect objective perturbation to outperform the private local modulus procedures (Alg. 5), but there ought to be a transition

[2]These correspond to top-level occupation codes from the 2018 census and are "Management, Business, and Financial", "Computer, Engineering, and Science", "Education, Legal, Community Service, Arts, and Media", "Healthcare Practitioners and Technical", "Service", "Sales and Office", "Natural Resources, Construction, and Maintenance", and the union of the categories "Production, Transportation, and Material Moving" and "Military Specific" occupations

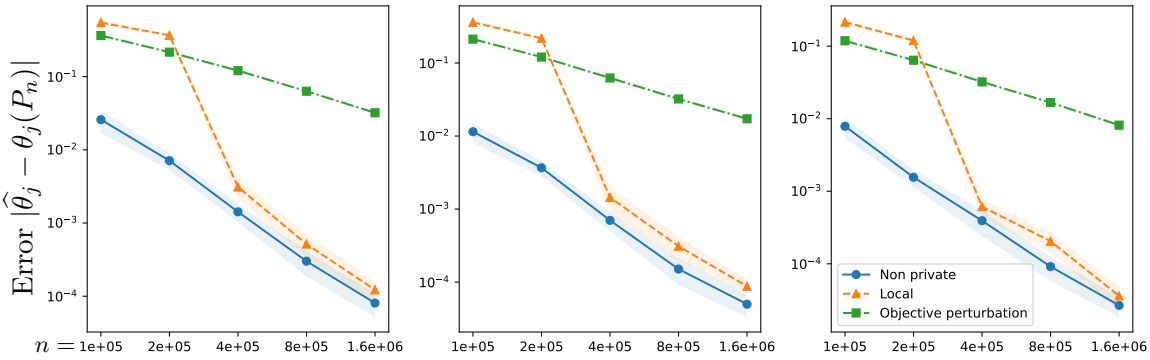

Figure 7. Estimation error versus sample size n for the parameter θ_j corresponding to amount of schooling in predicting income levels using Michigan survey data with approximate 95% confidence intervals based on 25 experiments. Left: privacy level $\varepsilon = 2$. Middle: privacy level $\varepsilon = 4$. Right: privacy level $\varepsilon = 8$. The true parameter $\theta_j(P_n) \approx 3.0$. The legend matches that in Fig. 6.

once the sample size n is large enough that Alg. 2 can effectively estimate $\lambda_{\min}(P_n)$. Second, when the parameter $\theta_j(P_n)$ of interest is large, we expect to see a larger gap in performance between objective perturbation and the local noise addition procedures we develop here.

In Figures 6 and 7, we see results roughly consistent with these expectations. Both figures exhibit the transition somewhere in the neighborhood of $n = 4 \cdot 10^5$ datapoints, where the error in using Algorithm 5 drops substantially, reflecting that it is possible to certify stability of the local modulus of continuity $\sup_{x \in \mathcal{X}} u^T \nabla^2 L_n(\theta(P_n))^{-1} x$. The plots also make clear that there remains a substantial gap between methods that can explicitly leverage the local modulus of continuity of the estimand of interest and those that cannot.

5 Parameter and eigenvalue stability guarantees

The building blocks out of which all of our theoretical results follow are stability analyses that demonstrate that if the empirical minimizer of a (smooth enough) loss $P_n \ell_\theta$ has Hessian with minimal eigenvalue $\lambda > 0$, then (i) the empirical minimizers associated with neighboring samples P_n' are close, and (ii) the Hessians at these empirical minimizers have minimal eigenvalues $\lambda' \geq \lambda - o(1)$, where the $o(1)$ term depends in somewhat nontrivial ways on P_n and P_n'. Accordingly, in this section, we present several results in this vein. The first set, in Section 5, gives quantitative bounds on the stability of empirical minimizers for both general smooth losses (Sec. 5.1.1) and quasi-self-concordant generalized linear model losses (Sec. 5.1.2). We then build off of these stability results to provide Hessian and associated eigenvalue perturbation bounds in Section 5.2. Throughout this section we let

$$\theta(P_n) = \operatorname*{argmin}_{\theta \in \Theta} P_n \ell_\theta + \frac{\lambda_{\text{reg}}}{2} \|\theta - \theta_0\|_2^2 \quad \text{and} \quad \lambda_{\min}(P_n) = \lambda_{\min}\left(P_n \ddot{\ell}_{\theta(P_n)}\right),$$

and P_n' denotes the empirical distribution of a sample satisfying $n \|P_n - P_n'\|_{\text{TV}} \leq 1$.

5.1 Stability bounds for the full parameter

We collect the main bounds on the deviation $\|\theta(P_n) - \theta(P_n')\|_2$, making the heuristic development in Sec. 1.1 rigorous and giving the appropriate numerical constants necessary to implement our associated private algorithms. We defer proofs of the results to Section 5.3.

5.1.1 Generic smooth losses

When the losses have G_i-Lipschitz continuous ith derivative for $i = 0, 1, 2$, we have the following two propositions; the first gives a "basic" guarantee, while the second sharpens it by a particular recursive bound. In each, we require that $\lambda_{\min}(P_n)$ is large enough (highlighting the importance of privately certifying lower bounds on $\lambda_{\min}(P_n)$); we thus require the condition

$$\lambda_{\min}(P_n) + \lambda_{\reg} \geq \max\left\{\frac{3G_1}{n}, \sqrt{\frac{12G_0G_2}{n}}\right\}. \tag{C2}$$

As a brief remark, we can compare Condition (C2) to Condition (C1). Let us assume a "typical" scenario, where the Lipschitz constants exhibit the scalings $G_i \propto d^{(i+1)/2}$ (recall Examples 1 and 2), and we expect that $\lambda_{\min}(P_n)$ should be roughly of constant order. (Think of classical linear regression, where $P_n\ddot{\ell}_\theta = \frac{1}{n}\sum_{i=1}^n x_i x_i^T$, so that if $x_i \overset{\text{iid}}{\sim} \mathsf{Uni}(\{\pm 1\}^d)$ then $P_n\ddot{\ell}_\theta \approx I_d$.) Then Condition (C2) requires a sample size scaling at least as $n \gtrsim d^2$ or that $\lambda_{\reg} \gtrsim d/\sqrt{n}$, while Condition (C1) requires only that $n \gtrsim d$ or $\lambda_{\reg} \gtrsim \frac{d}{n}$, a quadratic difference in required sample size.

Proposition 5.1. *Let Condition (C2) hold. Then for any empirical distribution P_n' with $\|P_n - P_n'\|_{\TV} \leq 1/n$, the minimizer $\theta(P_n')$ exists and satisfies*

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \leq \frac{12G_0}{n(\lambda_{\min}(P_n) + \lambda_{\reg})}.$$

See Section 5.3.1 for a proof.

When the problem is unconstrained (so that $\Theta = \mathbb{R}^d$), we can provide a sharper recursive bound. Proposition 5.1 guarantees the existence of a solution $\theta(P_n')$ for all P_n' neighboring P_n whenever Condition (C2) holds, and so we can perform a Taylor expansion to yield sharper guarantees. (We defer the proof to Section 5.3.2).

Proposition 5.2. *Let the conditions of Proposition 5.1 hold, but assume $\Theta = \mathbb{R}^d$. Then for all P_n' neighboring P_n, we have*

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \leq \frac{1}{2G_2}\left[\lambda_{\min}(P_n) + \lambda_{\reg} - \sqrt{(\lambda_{\min}(P_n) + \lambda_{\reg})^2 - \frac{8G_0G_2}{n}}\right].$$

Proposition 5.2 is always sharper than Proposition 5.1 and is (for n large) asymptotically tight. Indeed, letting $\lambda_{\reg} = 0$ for simplicity and assuming n is large, a Taylor expansion of $\sqrt{a^2 + \delta} = a + \delta/2a + O(\delta^2)$ gives

$$\frac{1}{2G_2}\left[\lambda_{\min}(P_n) - \sqrt{\lambda_{\min}^2(P_n) - \frac{8G_0G_2}{n}}\right] = \frac{2G_0}{\lambda_{\min}(P_n)n} + O(n^{-2}),$$

which is essentially as sharp as we could expect in these generic settings; recall the definition (3b) of the (asymptotic) local sensitivity.

5.1.2 Quasi-self-concordant GLMs

Given more conditions on the losses at play, we can obtain sharper stability guarantees for $\theta(P_n)$; we provide a few of these here. Recall the definitions (5) of quasi-self-concordance (q.s.c.), so that we consider generalized linear model (GLM) losses of the form

$$\ell_\theta(x, y) = h(\langle\theta, x\rangle, y).$$

Notably, for any such GLM-type loss, we have

$$\dot{\ell}_\theta(x,y) = h'(\langle\theta,x\rangle,y)x \quad \text{and} \quad \ddot{\ell}_\theta(x,y) = h''(\langle\theta,x\rangle,y)xx^T.$$

Thus, if the radius $\mathrm{rad}(\mathcal{X}) \coloneqq \sup_{x\in\mathcal{X}} \|x\|_2$ is finite, for any unit vector v and $t \geq 0$ we have the key semidefinite lower bound

$$\ddot{\ell}_{\theta+tv} \succeq [1 - \varphi(t \cdot \mathrm{rad}(\mathcal{X}))]_+ \ddot{\ell}_\theta \tag{16}$$

for any parameter θ. The self-bounding inequality (16) is the key that allows us more precise control on the error matrices in the heuristic derivation of stability in Section 1.1.

We begin with a proposition that applies to any lower q.s.c. loss with $\varphi(t) = \alpha t$; as Examples 3 and 4 show, this applies with $\alpha = 1$ for robust regression with the log loss and binary logistic regression. (See Section 5.3.3 for a proof.)

Proposition 5.3. *Define* $\mathrm{rad}(\mathcal{X}) = \sup_{x\in\mathcal{X}} \|x\|_2$ *and the loss* ℓ *be lower q.s.c. with* $\varphi(t) \leq \alpha t$. *Let* $\rho \in (0,1)$, *and define*

$$\mathsf{d}(P_n) \coloneqq \frac{4G_0}{n} \frac{1}{\rho\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}} - G_1/n}.$$

Then $\|\theta(P_n) - \theta(P_n')\|_2 \leq \mathsf{d}(P_n)$ *so long as* $\mathsf{d}(P_n) \leq \frac{1-\rho}{\alpha \mathrm{rad}(\mathcal{X})}$.

As in Section 5.1.1, we can leverage Proposition 5.3 to obtain sharper guarantees by iterating its implied bound on $\|\theta(P_n) - \theta(P_n')\|_2$ in a more careful Taylor expansion of $P_n'\dot{\ell}_{\theta(P_n')}$ around $P_n\dot{\ell}_{\theta(P_n)}$. We assume Condition (C1) that $\lambda_{\min}(P_n) + \frac{1}{\rho}\lambda_{\mathrm{reg}} \geq \frac{4G_0\alpha\mathrm{rad}(\mathcal{X})}{\rho(1-\rho)n} + \frac{G_1}{\rho n}$. Fixing $\theta = \theta(P_n)$, we define the shorthands

$$\tau_1 \coloneqq \sup_{P_n',\theta'} \left\| (P_n\ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1}(P_n - P_n')\dot{\ell}_{\theta'} \right\|_2 \leq \frac{2G_0}{n(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}})} \quad \text{and}$$

$$\tau_2 \coloneqq \sup_{P_n',\theta'} \left\| (P_n - P_n')\dot{\ell}_{\theta'} \right\|_2 \leq \frac{2G_0}{n}.$$

We prove the following guarantee in Section 5.3.4.

Proposition 5.4. *Let Condition* (C1) *hold for a given* $\rho \in (0,1)$ *and let* ℓ *be be* φ*-q.s.c.* (5a), *where* φ *satisfies inequality* (8) *that* $\varphi(t) \leq \alpha t$ *for* $0 \leq t \leq \frac{1-\rho}{\alpha}$. *Then*

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \leq \frac{1 - \sqrt{1 - \frac{4\alpha\mathrm{rad}(\mathcal{X})\tau_2}{\lambda_{\min}(P_n)+\lambda_{\mathrm{reg}}}}}{2\alpha\mathrm{rad}(\mathcal{X})}.$$

Recalling the parameter $t(\lambda)$ from its definition (9), Proposition 5.4 shows that

$$\left\|\theta(P_n) - \theta(P_n')\right\|_2 \leq t(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}) = \frac{1 - \sqrt{1 - \frac{8\alpha\mathrm{rad}(\mathcal{X})G_0}{(\lambda_{\min}(P_n)+\lambda_{\mathrm{reg}})n}}}{2\alpha\mathrm{rad}(\mathcal{X})}.$$

A few more careful Taylor approximations show that Proposition 5.4 sharpens Proposition 5.3. Recognize that $\sqrt{1-\epsilon} \geq 1 - \frac{\epsilon}{2} - \frac{\epsilon^2}{2}$ for $\epsilon \in [0,1]$, so that for $\lambda = \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$, the right hand side of the bound in the proposition satisfies

$$t(\lambda) \leq \frac{2G_0}{n\lambda} + \frac{16\alpha\mathrm{rad}(\mathcal{X})G_0^2}{n^2\lambda^2} \overset{\text{(C1)}}{\leq} \frac{2G_0}{n(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}})} + \frac{4\rho(1-\rho)G_0}{n(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}})},$$

where the second inequality holds under Condition (C1),

5.2 Stability bounds for minimal eigenvalues

With the parameter stability bounds in the preceding section, we can obtain corollaries about the perturbation stability of minimal (and maximal) eigenvalues of the empirical Hessian matrix $P_n \ddot{\ell}_\theta$. As these are all relatively quick, we give them sequentially and include proofs.

Corollary 5.1. *Let the conditions of Proposition 5.2 hold. Then for all P_n' neighboring P_n,*

$$\frac{\lambda_{\min}(P_n) + \lambda_{\reg}}{2} \left[1 + \sqrt{1 - \frac{8G_0G_2}{n(\lambda_{\min}(P_n) + \lambda_{\reg})^2}}\right] - \frac{G_1}{n}$$
$$\le \lambda_{\min}(P_n') + \lambda_{\reg} \le \frac{\lambda_{\min}(P_n) + \lambda_{\reg}}{2} \left[3 - \sqrt{1 - \frac{8G_0G_2}{n(\lambda_{\min}(P_n) + \lambda_{\reg})^2}}\right] + \frac{G_1}{n}.$$

Proof Let $\theta = \theta(P_n)$ and $\theta' = \theta(P_n')$ for shorthand. We prove the lower bound; the upper bound is completely similar. Then we have the semidefinite ordering inequalities

$$P_n' \ddot{\ell}_{\theta'} = P_n \ddot{\ell}_{\theta'} + (P_n' - P_n)\ddot{\ell}_{\theta'} \succeq P_n \ddot{\ell}_\theta - G_2 \left\|\theta - \theta'\right\|_2 - \frac{G_1}{n} I,$$

because of the assumptions that $\ddot{\ell}_\theta$ is G_2-Lipschitz and that $\theta \mapsto \dot{\ell}_\theta$ is G_1-Lipschitz, so that $\left\|\ddot{\ell}_\theta\right\|_{\op} \le G_1$ for any parameter θ. Substituting the bound of Proposition 5.2 for $\|\theta(P_n) - \theta(P_n')\|_2$ then gives the corollary. □

The cleaner behavior of Hessians for self-concordant GLMs allows sharper recursive guarantees. Recall the parameter change quantity (9) of Section 3, that is,

$$t(\lambda) = \frac{1 - \sqrt{1 - \frac{8\alpha \rad(\mathcal{X})G_0}{\lambda n}}}{2\alpha \rad(\mathcal{X})}.$$

Corollary 5.2. *Let the conditions of Proposition 5.4 hold, so that ℓ is φ-q.s.c., and let $\lambda_j(P_n) = \lambda_j(P_n \ddot{\ell}_{\theta(P_n)})$ denote the jth eigenvalue of $P_n \ddot{\ell}_{\theta(P_n)}$. Then for all P_n' neighboring P_n,*

$$\left|\lambda_j(P_n') - \lambda_j(P_n)\right| \le \lambda_j(P_n)\varphi\big(t(\lambda_{\min}(P_n) + \lambda_{\reg}) \cdot \rad(\mathcal{X})\big) + \frac{G_1}{n}.$$

Proof Let $\theta' = \theta(P_n')$ and $\theta = \theta(P_n)$ for shorthand as usual. Then we have

$$P_n' \ddot{\ell}_{\theta'} = P_n \ddot{\ell}_{\theta'} + (P_n' - P_n)\ddot{\ell}_{\theta'}.$$

As $\ddot{\ell}_v(z) \succeq 0$ for all z, v and $\left\|\ddot{\ell}_v\right\|_{\op} \le G_1$, we thus have

$$-\frac{G_1}{n} I + P_n \ddot{\ell}_{\theta'} \preceq P_n' \ddot{\ell}_{\theta'} \preceq P_n \ddot{\ell}_{\theta'} + \frac{G_1}{n} I.$$

Let $t = \|\theta - \theta'\|_2$ and $r = \rad(\mathcal{X})$ for shorthand. Then for the upper bound, note that $\ddot{\ell}_{\theta'} \preceq \ddot{\ell}_\theta(1 + \varphi(tr))$, so that $P_n' \ddot{\ell}_{\theta'} \preceq P_n \ddot{\ell}_\theta(1 + \varphi(tr)) + \frac{G_1}{n} I$. A similar derivation gives $P_n' \ddot{\ell}_{\theta'} \succeq P_n \ddot{\ell}_\theta \left[1 - \varphi(tr)\right]_+ - \frac{G_1}{n} I$. Applying Weyl's inequalitites then gives

$$\left[1 - \varphi(tr)\right]_+ \lambda_j(P_n) - \frac{G_1}{n} \le \lambda_j(P_n') \le (1 + \varphi(tr))\lambda_j(P_n) + \frac{G_1}{n},$$

and rearranging this yields the corollary. □

Unpacking Corollary 5.2, consider the "standard" scalings in which $\mathrm{rad}(\mathcal{X}) \lesssim \sqrt{d}$ and $G_0 \lesssim \sqrt{d}$, for example, in the context of Examples 3 and 4. Then taking $\varphi(t) = e^t - 1 \approx t$ and assuming $n \gg d$, we have $t(\lambda) = \frac{2G_0}{n\lambda} + O(n^{-2})$ and Corollary 5.2 unpacks to the bound that (roughly)

$$|\lambda_j(P_n') - \lambda_j(P_n)| \le \frac{\lambda_j(P_n)}{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}} \frac{2G_0 \mathrm{rad}(\mathcal{X})}{n} + \frac{G_1}{n} + O(n^{-2}) \lesssim \frac{\lambda_j(P_n)}{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}} \frac{d}{n}.$$

So we have the guarantees that

$$|\lambda_{\min}(P_n) - \lambda_{\min}(P_n')| \lesssim \frac{d}{n} \quad \text{and} \quad |\lambda_{\max}(P_n) - \lambda_{\max}(P_n')| \lesssim \frac{\lambda_{\max}(P_n) + \lambda_{\mathrm{reg}}}{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}} \frac{d}{n},$$

and the eigenvalues are (quite) stable to perturbations. In passing, we note that a first-order expansion of $\lambda_{\max}(P_n')$ suggests these bounds are likely hard to improve.

5.3 Proofs of parameter stability

We collect our omitted proofs from this section.

5.3.1 Proof of Proposition 5.1

Let $\theta = \theta(P_n)$ for shorthand and $\mathsf{R}(\theta) = \frac{\lambda_{\mathrm{reg}}}{2}\|\theta - \theta_0\|_2^2$ be the regularization. Fixing an arbitrary unit vector v, let $\theta' = \theta + tv \in \Theta$ and $t \ge 0$ be any value. By the first-order conditions for optimality of convex optimization we have $(P_n \dot{\ell}_\theta + \nabla \mathsf{R}(\theta))^T v \ge 0$ for any such setting. Then our various Lipschitz continuity assumptions give

$$\begin{aligned}
P_n' \ell_{\theta'} + \mathsf{R}(\theta') &\ge P_n' \ell_\theta + R(\theta) + t(P_n' \dot{\ell}_\theta + \nabla \mathsf{R}(\theta))^T v + \frac{t^2}{2}(v^T P_n' \ddot{\ell}_\theta v + \lambda_{\mathrm{reg}}) - \frac{G_2}{6} t^3 \\
&= P_n' \ell_\theta + \mathsf{R}(\theta) + t \underbrace{(P_n \dot{\ell}_\theta + \nabla \mathsf{R}(\theta))^T v}_{\ge 0} + \frac{t^2}{2}(v^T P_n \ddot{\ell}_\theta v + \lambda_{\mathrm{reg}}) - \frac{G_2}{6} t^3 + (P_n' - P_n)\left[t \ddot{\ell}_\theta^T v + \frac{t^2}{2} v^T \ddot{\ell}_\theta v\right] \\
&\ge P_n' \ell_\theta + \mathsf{R}(\theta) + \frac{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}}{2} t^2 - \frac{G_2}{6} t^3 - \frac{2G_0}{n} t - \frac{G_1}{2n} t^2.
\end{aligned}$$

That is, if $\theta' = \theta + tv$ we have for $\lambda = \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$ that

$$P_n' \ell_{\theta'} + \mathsf{R}(\theta') \ge P_n' \ell_\theta + \mathsf{R}(\theta) + \frac{\lambda}{2} t^2 - t\left[\frac{2G_0}{n} + \frac{G_1}{2n} t + \frac{G_2}{6} t^2\right]. \tag{17}$$

Because $t \mapsto P_n' \ell_{\theta + tv} + \mathsf{R}(\theta + tv)$ is convex, if for some $t_0 > 0$ the sum of the final two terms on the right hand side of inequality (17) is positive, then we evidently have $P_n' \ell_{\theta+tv} + \mathsf{R}(\theta + tv) > P_n' \ell_\theta + \mathsf{R}(\theta)$ for all $t \ge t_0$. We now proceed to find a fairly gross upper bound on this critical radius t_0. By condition (C2), n is large enough that $\frac{G_1}{2n} \le \frac{\lambda}{6}$. Then for any $t \le \frac{\lambda}{G_2}$, we have $\frac{G_2}{6} t^3 \le \frac{\lambda}{6} t^2$, and under these conditions inequality (17) implies

$$P_n' \ell_{\theta+tv} + \mathsf{R}(\theta + tv) \ge P_n' \ell_\theta + \mathsf{R}(\theta) + \frac{\lambda}{3} t^2 - \frac{G_1}{2n} t^2 - \frac{2G_0}{n} t \ge P_n' \ell_\theta + \mathsf{R}(\theta) + \frac{\lambda}{6} t^2 - \frac{2G_0}{n} t.$$

Then if $t > t_0 = \frac{12G_0}{n\lambda}$, we have $P_n' \ell_{\theta+tv} + \mathsf{R}(\theta + tv) > P_n \ell_\theta + \mathsf{R}(\theta)$, and it is possible to find such a t so long as $\frac{\lambda}{G_2} > \frac{12G_0}{n\lambda}$, that is, $\lambda > \sqrt{12G_0G_2/n}$, which holds per (C2). Restating this, whenever Condition (C2) holds, we necessarily have $\|\theta(P_n) - \theta(P_n')\|_2 \le \frac{12G_0}{n\lambda}$ as Proposition 5.1 requires.

5.3.2 Proof of Proposition 5.2

Proposition 5.1 guarantees the existence of a solution $\theta(P_n')$ minimizing $P_n'\ell_v$ in v; letting $\theta' = \theta(P_n')$ and $\theta = \theta(P_n)$ for shorthand, and $\mathsf{R}(\theta) = \frac{\lambda_{\rm reg}}{2}\|\theta - \theta_0\|_2^2$ be the ℓ_2-regularization, we therefore can perform a series of Taylor approximations to obtain

$$\begin{aligned} 0 = P_n'\dot\ell_{\theta'} + \nabla\mathsf{R}(\theta') &= P_n\dot\ell_{\theta'} + \nabla\mathsf{R}(\theta') + (P_n' - P_n)\dot\ell_{\theta'} \\ &= P_n\dot\ell_\theta + \nabla\mathsf{R}(\theta) + (P_n\ddot\ell_\theta + \lambda_{\rm reg}I + E)(\theta' - \theta) + (P_n' - P_n)\dot\ell_{\theta'}, \end{aligned}$$

where the error matrix E satisfies $\|E\|_{\rm op} \le G_2\|\theta - \theta'\|_2$ and we have used $\nabla\mathsf{R}(\theta) - \nabla\mathsf{R}(\theta') = \lambda_{\rm reg}(\theta - \theta')$. Under Condition (C2), we know that $P_n\ddot\ell_\theta + \lambda_{\rm reg}I + E$ is invertible (even more, $\lambda_{\min}(P_n) = \lambda_{\min}(P_n\ddot\ell_\theta) + \lambda_{\rm reg}I > \|E\|_{\rm op}$). Thus we obtain

$$\theta - \theta' = (P_n\ddot\ell_\theta + \lambda_{\rm reg}I + E)^{-1}(P_n - P_n')\dot\ell_{\theta'}.$$

Defining the shorthand $H = P_n\ddot\ell_\theta + \lambda_{\rm reg}I$ for the Hessian, because $\|E\|_{\rm op} < \lambda_{\min}(P_n)$, we have $(H+E)^{-1} = H^{-1} + \sum_{i=1}^\infty (-1)^i(H^{-1}E)^iH^{-1}$, which in turn satisfies

$$\begin{aligned} \left\|\sum_{i=1}^\infty (-1)^i(H^{-1}E)^iH^{-1}\right\|_{\rm op} &\le \frac{\|E\|_{\rm op}}{(\lambda_{\min}(P_n) + \lambda_{\rm reg})^2}\frac{1}{1 - \|E\|_{\rm op}/(\lambda_{\min}(P_n) + \lambda_{\rm reg})} \\ &= \frac{\|E\|_{\rm op}}{\lambda_{\min}(P_n) + \lambda_{\rm reg}}\frac{1}{\lambda_{\min}(P_n) + \lambda_{\rm reg} - \|E\|_{\rm op}}. \end{aligned}$$

Substituting above and using the shorthand $\lambda = \lambda_{\min}(P_n) + \lambda_{\rm reg}$, we obtain

$$\|\theta - \theta'\|_2 \le \left\|H^{-1}(P_n - P_n')\dot\ell_{\theta'}\right\|_2 + \frac{\|E\|_{\rm op}}{\lambda}\frac{1}{\lambda - \|E\|_{\rm op}}\left\|(P_n - P_n')\dot\ell_{\theta'}\right\|_2$$

As $\|E\|_{\rm op} \le G_2\|\theta - \theta'\|_2$ and $\|(P_n - P_n')\dot\ell_{\theta'}\|_2 \le \frac{2G_0}{n}$, if we let $t = \|\theta - \theta'\|_2$, then t necessarily satisfies

$$t \le \frac{2G_0}{n\lambda} + \frac{2G_0G_2t}{n\lambda(\lambda - G_2t)} \quad \text{or} \quad \lambda t - G_2t^2 \le \frac{2G_0}{n\lambda} + \frac{2G_0}{n\lambda}G_2t.$$

Solving the implied quadratic yields

$$\|\theta(P_n) - \theta(P_n')\|_2 = t \le \frac{\lambda - \sqrt{\lambda^2 - \frac{8G_0G_2}{n}}}{2G_2}.$$

5.3.3 Proof of Proposition 5.3

As usual, we let P_n and P_n' be neighboring datasets, $\mathsf{R}(\theta) = \frac{\lambda_{\rm reg}}{2}\|\theta - \theta_0\|_2^2$, and let $\theta = \theta(P_n) = \operatorname{argmin}_\theta P_n\ell_\theta + \mathsf{R}(\theta)$. Then for any vector v, we have

$$\ddot\ell_{\theta+v}(x,y) = h''(\langle\theta + v, x\rangle, y)xx^T \succeq h''(\langle\theta, x\rangle, y)xx^T\left[1 - \alpha_h|\langle v, x\rangle|\right]_+.$$

Rewriting this in with the typical shorthand of suppressing x and y and the dependence of α_h on h, we have $\ddot\ell_{\theta+v} \succeq \ddot\ell_\theta\left[1 - \alpha|\langle v, x\rangle|\right]_+$. Thus for any v, there is some $s \in [0,1]$ for which we have

$$\begin{aligned} P_n'\ell_{\theta+v} + \mathsf{R}(\theta + v) &= P_n'\ell_\theta + \mathsf{R}(\theta) + (P_n'\dot\ell_\theta + \nabla\mathsf{R}(\theta))^Tv + \frac{1}{2}v^T(P_n'\ddot\ell_{\theta+sv} + \lambda_{\rm reg}I)v \\ &\ge P_n'\ell_\theta + \mathsf{R}(\theta) + (P_n' - P_n)\dot\ell_\theta^Tv + \frac{1}{2}v^T(P_n\ddot\ell_{\theta+sv} + \lambda_{\rm reg}I)v - \frac{G_1}{2n}\|v\|_2^2 \end{aligned}$$

for some $s \in [0,1]$, where the inequality follows because $\|\ddot{\ell}\|_{\rm op} \le G_1$ and $P_n\dot{\ell}_\theta + \nabla \mathsf{R}(\theta) = 0$. Using the assumption (8) on the losses and its consequence (16), we then have

$$\begin{aligned}
&P_n'\ell_{\theta+v} + \mathsf{R}(\theta+v) \\
&\ge P_n'\ell_\theta + \mathsf{R}(\theta) + (P_n' - P_n)\dot{\ell}_\theta^T v + \frac{1}{2}v^T(P_n\ddot{\ell}_\theta\left[1 - \alpha|\langle v, X\rangle|\right]_+ + \lambda_{\rm reg} I)v - \frac{G_1}{2n}\|v\|_2^2 \qquad (18)\\
&\ge P_n'\ell_\theta + \mathsf{R}(\theta) - \frac{2G_0}{n}\|v\|_2 + \left(\lambda_{\min}(P_n)(1 - \alpha\|v\|_2 \operatorname{rad}(\mathcal{X})) + \lambda_{\rm reg} I - \frac{G_1}{n}\right)\frac{\|v\|_2^2}{2}.
\end{aligned}$$

Let $t = \|v\|_2$ for shorthand. Then by convexity, if

$$-\frac{2G_0}{n}t + \left(\lambda_{\min}(P_n)(1 - \alpha t \cdot \operatorname{rad}(\mathcal{X})) + \lambda_{\rm reg} - \frac{G_1}{n}\right)\frac{t^2}{2} > 0,$$

then we necessarily have $P_n'\ell_{\theta+v} + \mathsf{R}(\theta+v) > P_n'\ell_\theta + \mathsf{R}(\theta)$, and moreover, $P_n'\ell_{\theta+u} + \mathsf{R}(\theta+u) > P_n'\ell_\theta + \mathsf{R}(\theta)$ whenever $\|u\|_2 > t$, so that $\|\theta(P_n) - \theta(P_n')\|_2 \le t$.

Notably, whenever $t\alpha\operatorname{rad}(\mathcal{X}) \le 1 - \rho$, it suffices to find a t satisfying

$$-\frac{2G_0}{n}t + \left(\lambda_{\min}(P_n)\rho + \lambda_{\rm reg} - \frac{G_1}{n}\right)\frac{t^2}{2} = 0 \quad \text{and} \quad t \le \frac{1-\rho}{\alpha\operatorname{rad}(\mathcal{X})}.$$

This occurs whenever $t = \frac{4G_0}{n}\frac{1}{\rho\lambda_{\min}(P_n) + \lambda_{\rm reg} - G_1/n} < \frac{1-\rho}{\alpha\operatorname{rad}(\mathcal{X})}$, which is the claim of the proposition.

5.3.4 Proof of Proposition 5.4

The proof of the proposition requires some manipulations of Hessian error terms, so we provide a matrix inequality to address them.

Lemma 5.1. *Let $A \succ 0$ satisfy $-\delta A \preceq E \preceq \delta A$. Then for each $k \in \mathbb{N}$, there exists a symmetric D satisfying $-\frac{\delta^{k+1}}{1-\delta} \preceq D \preceq \frac{\delta^{k+1}}{1-\delta}$ and for which*

$$(A+E)^{-1} = A^{-1} + \sum_{i=1}^{k}(-1)^i(A^{-1}E)^iA^{-1} + A^{-1/2}DA^{-1/2}.$$

Proof We have $(A^{-1}E)^iA^{-1} = A^{-1/2}(A^{-1/2}EA^{-1/2})^iA^{-1/2}$, and $-\delta I \preceq A^{-1/2}EA^{-1/2} \preceq \delta I$ by assumption. Thus $\|A^{-1/2}EA^{-1/2}\|_{\rm op}^i \le \delta^i$. In particular, we can therefore perform the standard matrix inverse expansion that

$$(A+E)^{-1} = A^{-1} + \sum_{i=1}^{\infty}(-1)^i(A^{-1}E)^iA^{-1} = A^{-1} + A^{-1/2}\sum_{i=1}^{\infty}(-1)^i(A^{-1/2}EA^{-1/2})^iA^{-1/2}.$$

Now note that

$$\left\|\sum_{i=k+1}^{\infty}(-1)^i(A^{-1/2}EA^{-1/2})^i\right\|_{\rm op} \le \sum_{i=k+1}^{\infty}\left\|A^{-1/2}EA^{-1/2}\right\|_{\rm op}^i \le \sum_{i=k+1}^{\infty}\delta^i = \frac{\delta^{k+1}}{1-\delta}.$$

Letting $D = \sum_{i=k+1}^{\infty}(-1)^i(A^{-1/2}EA^{-1/2})^i$ completes the proof. □

With Lemma 5.1, we can perform the manipulations of the gradient conditions for optimality of $\theta(P_n)$ with the necessary Hessian perturbations.

Lemma 5.2. *Let P_n, P_n' be neighboring samples and the conditions of Proposition 5.4 hold. Then $\theta = \theta(P_n)$ and $\theta' = \theta(P_n')$ exist, and $\gamma := \alpha \left\|\theta - \theta'\right\|_2 \operatorname{rad}(\mathcal{X}) < 1$. Additionally, there is a symmetric matrix D satisfying $-\frac{\gamma}{1-\gamma} \preceq D \preceq \frac{\gamma}{1-\gamma}$ for which*

$$\theta' - \theta = (P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1}(P_n - P_n')\dot{\ell}_{\theta'} + (P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1/2} D (P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1/2}(P_n - P_n')\dot{\ell}_{\theta'}.$$

Proof Let $r = \operatorname{rad}(\mathcal{X})$ and $\lambda = \lambda_{\min}(P_n)$ for shorthand. By Proposition 5.3, for any $\rho \in (0,1)$ such that $\rho\lambda + \lambda_{\mathrm{reg}} \geq \frac{4\alpha G_0 r}{(1-\rho)n} + \frac{G_1}{n}$, we have $\left\|\theta(P_n) - \theta(P_n')\right\|_2 \leq \frac{4G_0}{n} \frac{1}{\lambda\rho + \lambda_{\mathrm{reg}} - G_1/n} \leq \frac{1-\rho}{\alpha r}$. The solution $\theta' = \theta(P_n')$ then exists, and so a Taylor expansion gives

$$\begin{aligned} 0 = P_n' \dot{\ell}_{\theta'} + \nabla \mathsf{R}(\theta') &= P_n \dot{\ell}_{\theta'} + (P_n' - P_n)\dot{\ell}_{\theta'} + \nabla \mathsf{R}(\theta') \\ &= P_n \dot{\ell}_\theta + \nabla \mathsf{R}(\theta) + (P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I + E)(\theta' - \theta) + (P_n' - P_n)\dot{\ell}_{\theta'}, \end{aligned}$$

where the error matrix E satisfies

$$-\alpha P_n \ddot{\ell}_\theta |(\theta - \theta')^T X| \preceq E \preceq \alpha P_n \ddot{\ell}_\theta |(\theta - \theta')^T X|$$

by assumption, as $|(\theta - \theta')^T x| \leq (1-\rho)/\alpha$, and so the self-bounding conditions apply.

Define the Hessian $H = P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I$ for shorthand. So long as $\sup_{x \in \mathcal{X}} (\theta - \theta')^T x < 1/\alpha$, for which it is sufficient that $\left\|\theta - \theta'\right\|_2 r < 1/\alpha$, $H + E$ is invertible, because in this case $-H \prec E \prec H$. With the choice $\gamma = \alpha \left\|\theta - \theta'\right\|_2 \operatorname{rad}(\mathcal{X})$, we have $\gamma \leq 1 - \rho < 1$, where $\rho \in (0,1)$ is as in Condition (C1). Lemma 5.1 therefore gives that

$$(H + E)^{-1} = H^{-1} + H^{-1/2} D H^{-1/2}$$

for a symmetric matrix D satisfying $-\gamma/(1-\gamma) \preceq D \preceq \gamma/(1-\gamma)$, and rewriting the Taylor expansion above then gives

$$\begin{aligned} (\theta' - \theta) &= (H + E)^{-1}(P_n - P_n')\dot{\ell}_{\theta'} \\ &= H^{-1}(P_n - P_n')\dot{\ell}_{\theta'} + H^{-1/2} D H^{-1/2}(P_n - P_n')\dot{\ell}_{\theta'}, \end{aligned}$$

as desired. □

Taking norms of both sides in Lemma 5.2 yields the inequality

$$\left\|\theta - \theta'\right\|_2 \leq \left\|(P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1}(P_n - P_n')\dot{\ell}_{\theta'}\right\|_2 + \frac{\left\|D\right\|_{\mathrm{op}}}{\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}} \left\|(P_n - P_n')\dot{\ell}_{\theta'}\right\|_2. \tag{19}$$

Let $t = \left\|\theta(P_n) - \theta(P_n')\right\|_2$ for shorthand; we will find bounds on t so that the bound (19) holds. Substituting in the earlier bounds on the error matrix D, we have $\left\|D\right\|_{\mathrm{op}} \leq \alpha t r/(1 - \alpha t r)$, and we see that t necessarily satisfies

$$t \leq \left\|(P_n \ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I)^{-1}(P_n - P_n')\dot{\ell}_{\theta'}\right\|_2 + \frac{\alpha r t}{(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}})(1 - \alpha r t)} \left\|(P_n - P_n')\dot{\ell}_{\theta'}\right\|_2.$$

As we note above, Condition (C1) is sufficient to guarantee that $1 - \alpha r t \geq \rho > 0$, and so for $\tau_2 = \left\|(P_n - P_n')\dot{\ell}_{\theta'}\right\|_2$ and $\lambda = \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$, so the preceding display implies that

$$t \leq \frac{\tau_2}{\lambda} + \frac{\alpha r t}{1 - \alpha r t} \frac{\tau_2}{\lambda} \quad \text{i.e.} \quad 0 \leq \alpha r t^2 - t + \frac{\tau_2}{\lambda}.$$

Solving the quadratic, this implies

$$t \leq \frac{1 - \sqrt{1 - 4\frac{\alpha r \tau_2}{\lambda}}}{2\alpha r}.$$

6 Releasing private quantities via recursive bounds

As we outline in Section 2, the first stage in our algorithms is to privately release a lower bound on $\lambda_{\min}(P_n)$. In this section, we develop the tools to do so by introducing new algorithms for privately releasing statistics whose values on neighboring samples can be bounded recursively, that is, by functionals of the statistic itself. The prototypical example on which we focus is the minimal eigenvalue $\lambda_{\min}(P_n)$, which satisfies a number of recursive bounds. From Corollary 5.1, for example, with $\lambda_{\reg} = 0$, we see that

$$\lambda_{\min}(P_n') \geq \frac{\lambda_{\min}(P_n)}{2}\left[1 + \sqrt{1 - \frac{8G_0G_2}{n\lambda_{\min}^2(P_n)}}\right] - \frac{G_1}{n},$$

so long as Condition (C2) holds, while Corollary 5.2 gives a sharper guarantee for generalized linear models with quasi-self-concordance.

To develop the mechanisms, we leverage Asi and Duchi's approximate inverse sensitivity mechanism [3], which releases a real-valued statistic $f(P_n)$ with (ε, δ)-differential privacy and high accuracy. We first recapitulate their mechanism, then show how to apply it to eigenvalues in Section 6.2. Recall the modulus of continuity (2) for a function f acting on finitely supported measures,

$$\omega_f(P_n; k) := \sup_{P_n' \in \mathcal{P}_n} \left\{|f(P_n) - f(P_n')| \text{ s.t. } n\left\|P_n - P_n'\right\|_{\TV} \leq k\right\}.$$

The mechanism requires a set of upper bounding functions $U_i : \mathcal{P}_n \to \mathbb{R}_+$, $i = 1, \ldots, n$, acting on the sample measures $\mathcal{P}_n$ and satisfying

$$\omega_f(P_n; 1) \leq U_1(P_n)$$

and the local upper bounding condition that

$$U_k(P_n) \leq U_{k+1}(P_n') \text{ for all } k \in [n] \text{ and } P_n, P_n' \text{ with } \left\|P_n - P_n'\right\|_{\TV} \leq \frac{1}{n}. \tag{20}$$

The upper inverse modulus of continuity

$$\overline{\mathsf{len}}_f(P_n; t) := \min\left\{k \in \mathbb{N} \mid \sum_{i=1}^{k} U_i(P_n) \geq |t - f(P_n)|\right\} \tag{21}$$

then defines the approximate inverse sensitivity mechanism

$$\mathbb{P}(M(P_n) \in A) = \frac{\int_A e^{-\varepsilon\overline{\mathsf{len}}_f(P_n;t)/2} d\mu(t)}{\int_{\mathcal{T}} e^{-\varepsilon\overline{\mathsf{len}}_f(P_n;t)/2} d\mu(t)}. \tag{M.A}$$

Asi and Duchi [3, Thm. 1] show that the mechanism is differentially private:

Corollary 6.1. *The length function (21) satisfies* $|\overline{\mathsf{len}}_f(P_n; t) - \overline{\mathsf{len}}_f(P_n'; t)| \leq n\left\|P_n - P_n'\right\|_{\TV}$, *and hence the mechanism (M.A) is ε-differentially private.*

Using the mechanism (M.A), we show how to release one-dimensional quantities whose stability is governed by the quantity itself, after which (in Sec. 6.2) we show how this applies more concretely to releasing minimal (and maximal) eigenvalues.

6.1 Private one-dimensional statistics via recursive bounds

Let $C \subset \mathbb{R}$ be a closed convex set (typically, this will be $[0, \infty)$ or an interval $[a, b]$). A mapping $R : \mathbb{R} \to \mathbb{R}$ is an *accelerating decreasing recursion* on C if for all $\lambda \le \lambda' \in C$, we have $R(\lambda) \le \lambda$ and the acceleration condition

$$\lambda - R(\lambda) \ge \lambda' - R(\lambda') \quad \text{whenever} \quad R(\lambda) > \inf C, \tag{22}$$

so that the recursion $\lambda \mapsto R(\lambda)$ accelerates toward the lower limit $\inf C$. If R is differentiable, then condition (22) is equivalent to

$$R'(\lambda) \ge 1 \quad \text{whenever} \quad R(\lambda) > \inf C.$$

(To see this, set $\lambda' = \lambda + \delta$ and take $\delta \downarrow 0$.) Additionally, if R is accelerating and H_ρ is the hard-thresholding operator $H_\rho(t) = t1\{t \ge \rho\}$, then $H_\rho \circ R$ is also an accelerating decreasing recursion by inspection. Previewing our applications to eigenvalues, examples include the linear mapping $R(\lambda) = [\lambda - a]_+$, or the mapping $R(\lambda) = \lambda - a/\lambda - b$ with $a, b \ge 0$, which are both accelerating over $\lambda \in \mathbb{R}_+$. Define the k-fold composition

$$R^k \coloneqq \underbrace{R \circ \cdots \circ R}_{k \text{ times}}$$

Assume we have a statistic $\lambda : \mathcal{P}_n \to C$ satisfying the one-step recursive guarantee that $\lambda(P_n') \ge R(\lambda(P_n))$ whenever P_n', P_n are neighboring. Define the upper bound sequence

$$U_k(P_n) \coloneqq \begin{cases} R^{k-1}(\lambda(P_n)) - R^k(\lambda(P_n)) & \text{if } R^k(\lambda(P_n)) > \inf C \\ +\infty & \text{otherwise.} \end{cases} \tag{23}$$

We claim the following lemma, which shows that this sequence upper bounds the local modulus of continuity as required in the definition (21).

Lemma 6.1. *Let C be closed convex, $\lambda : \mathcal{P}_n \to C$, and $R : C \to \mathbb{R}$ be an accelerating decreasing recursion. Assume that λ satisfies the bounds*

$$\lambda(P_n) - R(\lambda(P_n)) \ge \lambda(P_n') - \lambda(P_n) \ge R(\lambda(P_n)) - \lambda(P_n)$$

for all neighboring $P_n, P_n' \in \mathcal{P}_n$. Then the mapping (23) satisfies

$$U_1(P_n) \ge \omega_\lambda(P_n; 1) \quad \textit{and} \quad U_k(P_n) \le U_{k+1}(P_n') \textit{ for all } k \in \mathbb{N}.$$

Proof Fix $\lambda_0 = \lambda(P_n)$, and let $\lambda_0' = \lambda(P_n')$ for some P_n, P_n' with $\|P_n - P_n'\|_{\mathrm{TV}} \le 1/n$. We have the recursions

$$\lambda_{k+1} \coloneqq R(\lambda_k) \quad \text{and} \quad \lambda_{k+1}' = R(\lambda_k'),$$

and define $u_k \coloneqq U_k(P_n) = \lambda(P_n) - R_k(\lambda(P_n))$ and $u_k' \coloneqq U_k(P_n') = \lambda(P_n') - R_k(\lambda(P_n'))$ for shorthand. The first claim that $u_1 \ge \omega_\lambda(P_n; 1)$ is immediate, as $u_1 = R(\lambda(P_n)) - \lambda(P_n) \ge |\lambda(P_n') - \lambda(P_n)|$. So we need only show that $u_k \le u_{k+1}'$ for all k, which we do via induction, demonstrating both that $u_k \le u_{k+1}'$ and that $\lambda_k \ge \lambda_{k+1}'$ for all k.

Base case. The case $k = 0$ is that $\lambda_0 \ge \lambda_1'$, which is equivalent to the claim that

$$\lambda_0 \ge \lambda_1' = R(\lambda_0'), \quad \text{i.e.} \quad \lambda_0 - \lambda_0' \ge R(\lambda_0') - \lambda_0',$$

which is immediate by the assumed bounds on $\lambda(\cdot)$.

Induction. Assume that the inequalities $u_i \le u'_{i+1}$ and $\lambda_i \ge \lambda'_{i+1}$ hold for all $i < k$. We wish to show they hold for $i = k$. The monotonicity of the recursive mapping guarantees that $\lambda_k = R(\lambda_{k-1}) \ge R(\lambda'_k) = \lambda'_{k+1}$ by the assumption that $\lambda_{k-1} \ge \lambda'_k$. If $\lambda_k = \inf C$ then $\lambda'_{k+1} = \inf C$, and so $u_k = u'_{k+1} = +\infty$. Otherwise, we have $\lambda_k > \inf C$ and then

$$u_k = \lambda_{k-1} - \lambda_k = \lambda_{k-1} - R(\lambda_{k-1}) \stackrel{(i)}{\le} \lambda'_k - R(\lambda'_{k-1}) \stackrel{(ii)}{\le} u'_{k+1},$$

where inequality (i) is the acceleration condition (22) and (ii) is an equality unless $R(\lambda'_{k-1}) \le \inf C$, in which case $u'_{k+1} = +\infty$. This gives the induction and the lemma. □

Inverting R^k provides a clean approach to releasing lower bounds on $\lambda(P_n)$. Define the inverse

$$(R^k)^{-1}(\gamma) := \sup\left\{\lambda \ge \inf C \mid R^k(\lambda) \le \gamma\right\}.$$

If we have a high probability guarantee on a (random) N that $R^N(\lambda(P_n)) > \inf C$, then the monotonicity of R guarantees that $\lambda(P_n) \ge (R^N)^{-1}(\inf C)$, leading to the following algorithm.

Algorithm 6: A private lower bound on $\lambda(P_n)$

Require: Privacy parameters $\varepsilon \ge 0$ and $\delta \in (0, 1)$ and an accelerating decreasing recursion $R : C \to \mathbb{R}$ satisfying $|\lambda(P_n) - \lambda(P'_n)| \le \lambda(P_n) - R(\lambda(P_n))$ for all neighboring empirical distributions P_n, P'_n.

i. Set
$$\widehat{N} := \min\left\{N \in \mathbb{N} \mid R^N(\lambda(P_n)) \le \inf C\right\} + \frac{1}{\varepsilon}\mathsf{Lap}(1).$$

ii. Set $k(\varepsilon, \delta) = \frac{1}{\varepsilon}\log\frac{1}{2\delta}$, then return $\widehat{N}$ and
$$\widehat{\lambda} = \left(R^{\widehat{N} - k(\varepsilon,\delta)}\right)^{-1}(\inf C).$$

The discussion above and that if $W \sim \mathsf{Lap}(1)$, we have $\mathbb{P}(W \ge \log\frac{1}{2\delta}) = \frac{1}{2}\int_{\log\frac{1}{2\delta}} e^{-t}dt = \delta$, then immediately yield the following proposition.

Proposition 6.1. *Algorithm 6 is ε-differentially private, and $\lambda(P_n) \ge \widehat{\lambda}$ with probability at least $1 - \delta$.*

Proof By the construction of the upper bound mapping (23), we have

$$\overline{\mathsf{len}}_\lambda(P_n; \inf C) = \min\{N \in \mathbb{N} \mid R^N(\lambda(P_n)) \le \inf C\}.$$

Thus for $W \sim \mathsf{Lap}(1)$, we have $\widehat{N} \stackrel{d}{=} \overline{\mathsf{len}}_\lambda(P_n; \inf C) + \frac{1}{\varepsilon}W$, and $\widehat{N}$ is ε-differentially private (and so is $\widehat{\lambda}$ by post-processing). To obtain that $\mathbb{P}(\lambda(P_n) \ge \widehat{\lambda}) \ge 1 - \delta$, note that $\mathbb{P}(\widehat{N} < \overline{\mathsf{len}}_\lambda(P_n; \inf C) + k(\varepsilon, \delta)) = \mathbb{P}(W < \log\frac{1}{2\delta}) = 1 - \delta$. On the event $\widehat{N} < \overline{\mathsf{len}}_\lambda(P_n; \inf C) + k(\varepsilon, \delta)$, we know that $R^{\widehat{N} - k(\varepsilon,\delta)}(\lambda(P_n)) > \inf C$, and so monotonicity of R guarantees $\lambda(P_n) \ge \widehat{\lambda}$. □

6.2 Releasing eigenvalues for M-estimation problems

We finish this section by giving explicit algorithms for releasing minimal eigenvalues for M-estimation problems (1) as well as proving Corollaries 3.2 and 3.4. Algorithm 6 guarantees privacy, by Proposition 6.1, so if we can demonstrate a recursion R for $\lambda_{\min}(P_n)$ satisfying

$$|\lambda_{\min}(P_n) - \lambda_{\min}(P_n')| \le \lambda_{\min}(P_n) - R(\lambda_{\min}(P_n)),$$

then we may simply apply Algorithm 6.

We begin with a generic lemma, which gives two somewhat more sophisticated recursions, based (respectively) on Corollaries 5.1 and 5.2. (See Appendix A.3.1 for a proof.)

Lemma 6.2. *Let $a, b, c \ge 0$ and $\lambda_0 \ge 0$. Then the functions*

$$R(\lambda) = \frac{\lambda}{2}\left[1 + \sqrt{1 - \frac{a}{\lambda^2}}\right] - b \quad or \quad R(\lambda) = \lambda\left[2 - \exp\left(b\left(1 - \sqrt{1 - \frac{a}{\lambda + \lambda_0}}\right)\right)\right] - c$$

are accelerating decreasing recursions for, respectively, $\lambda > a$ and $\lambda + \lambda_0 > a$.

6.2.1 Releasing the minimal eigenvalue for a general smooth loss

We now revisit Corollary 5.1, which applies when all we know are that the losses ℓ have Lipschitz derivatives. In this case, we have the following corollary.

Corollary 6.2. *Let the loss ℓ have G_i-Lipschitz continuous ith derivative for $i = 0, 1, 2$. Define the recursion*

$$R(\lambda) := \max\left\{\frac{\lambda}{2}\left[1 + \sqrt{1 - \frac{8G_0G_2}{n\lambda^2}}\right] - \frac{G_1}{n}, \lambda_{\rm reg}\right\},$$

where $\sqrt{x} = -\infty$ for $x \le 0$. Then Algorithm 6 applied with this recursion releases an ε-differentially private $\widehat{\lambda}$. With probability at least $1 - \delta$, $\widehat{\lambda}$ satisfies both $\widehat{\lambda} \le \lambda_{\min}(P_n) + \lambda_{\rm reg}$ and

$$\widehat{\lambda} \ge \lambda_{\min}(P_n) + \lambda_{\rm reg} - O(1)\frac{1}{\varepsilon}\log\frac{1}{\delta}\left[\frac{G_0G_2}{(\lambda_{\min}(P_n) + \lambda_{\rm reg})n} + \frac{G_1}{n}\right].$$

Proof The first claim of the corollary is immediate by combining Proposition 6.1, Lemma 6.2 that R is accelerating, and the deviation bounds in Corollary 5.1.

The guarantees on the relationship between $\widehat{\lambda}$ and $\lambda_{\min}(P_n)$ require more work. Let $\lambda^\star = \lambda_{\min}(P_n) + \lambda_{\rm reg}$ for shorthand. That $\widehat{\lambda} \le \lambda^\star$ with probability $1 - \delta$ is immediate by definition of $\widehat{N}$. To obtain the lower bound $\widehat{\lambda} \ge \lambda^\star - O(\frac{1}{n\varepsilon})$, introduce the shorthands $a = \frac{8G_0G_2}{n}$ and $b = \frac{G_1}{n}$, where we assume $\max\{\sqrt{a}, b\} \ll \lambda^\star$. (Otherwise, the guarantee is vacuous.) Recall the definition $(R^k)^{-1}(0) = \inf\{\lambda \mid R^k(\lambda) = \lambda_{\rm reg}\}$, and let N be the smallest value necessary to obtain $R^N(\lambda^\star) = \lambda_{\rm reg}$, so $R^{N-1}(\lambda^\star) > 0$. Consider a single iteration of the recursion

$$\lambda \mapsto R(\lambda) = \frac{\lambda}{2}\left(1 + \sqrt{1 - \frac{a}{\lambda^2}}\right) - b = \frac{\lambda}{2}\left(2 - \frac{a}{\lambda^2} + O(a^2/\lambda^4)\right) - b = \lambda - \frac{a}{\lambda} - b - O\left(\frac{a^2}{\lambda^3}\right).$$

Then for some (numerical) constant c the recursion $R^N(\lambda^\star - \frac{a}{\lambda^\star} - b - c\frac{a^2}{\lambda^{\star 3}}) = \lambda_{\rm reg}$, so that $(R^N)^{-1}(\lambda_{\rm reg}) \ge \lambda^\star - \frac{a}{\lambda^\star} - b - O(\frac{a^2}{\lambda^{\star 3}})$. Applying k steps of the recursion with the above linearization, we obtain

$$R^k(\lambda^\star) = \lambda^\star - k\left(\frac{a}{\lambda^\star} + b\right) - O\left(k\frac{a^2}{\lambda^{\star 3}}\right).$$

For $k = k(\varepsilon, \delta) = \frac{1}{\varepsilon} \log \frac{1}{2\delta}$ we have $\widehat{N} \geq N - k(\varepsilon, \delta)$ with probability at least $1 - \delta$, so recognizing that $a^2/\lambda^{\star 2} \ll a/\lambda^\star$ gives

$$(R^{\widehat{N}})^{-1}(\lambda_{\mathrm{reg}}) \geq R^{k+1}\left((R^N)^{-1}(\lambda_{\mathrm{reg}})\right) \geq \lambda^\star - O(1)k\left(\frac{a}{\lambda^\star} + b\right) + O\left(k\frac{a^2}{\lambda^{\star 3}}\right).$$

Substituting for a and b gives the corollary once we recognize that it is vacuous whenever $ka/\lambda^\star \gtrsim \lambda^\star$. □

6.2.2 Proof of Corollary 3.2

We can revisit Corollary 5.2 to apply to (quasi) self-concordant losses. Define

$$a = \frac{4G_0\alpha\mathrm{rad}(\mathcal{X})}{n}, \quad b = \frac{1}{2\alpha}, \quad c = \frac{G_1}{n}.$$

Then the defined recursion satisfies

$$R(\lambda) = \lambda\left(2 - \exp\left(b\left(1 - \sqrt{1 - \frac{a}{\lambda + \lambda_{\mathrm{reg}}}}\right)\right)\right) - c$$

as in Lemma 6.2 (so long as λ satisfies Condition (C1), and hard-thresholding to 0 otherwise) so that it is an accelerating and decreasing recursion. Corollary 5.2 shows that R bounds the changes in $\lambda_{\min}(P_n)$ to $\lambda_{\min}(P_n')$. Proposition 6.1 thus gives the differential privacy.

For the claimed lower bound on $\widehat{\lambda}$, we consider the behavior of R for λ near $\lambda_{\min}(P_n)$. Let $r = \mathrm{rad}_2(\mathcal{X})$ for shorthand. Under the assumption that $C\frac{G_0 r}{n} \leq \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$ for a suitably large numerical constant C, we have

$$b\left(1 - \sqrt{1 - \frac{a}{\lambda + \lambda_{\mathrm{reg}}}}\right) = b\left(\frac{a}{2(\lambda + \lambda_{\mathrm{reg}})} + O(a^2/(\lambda + \lambda_{\mathrm{reg}})^2)\right) = \frac{2G_0 r}{n(\lambda + \lambda_{\mathrm{reg}})} + O\left(\frac{G_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})^2}\right)$$

assuming that α and ρ are numerical constants. Ignoring the higher order terms and using that $e^t = 1 + t + O(t^2)$, we thus obtain

$$\begin{aligned} R(\lambda) &= \lambda\left(1 - \frac{2G_0 r}{n(\lambda + \lambda_{\mathrm{reg}})} - O\left(\frac{G_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})^2}\right)\right) - \frac{G_1}{n} \\ &= \lambda - \frac{2G_0 r}{n}\frac{\lambda}{\lambda + \lambda_{\mathrm{reg}}} - \frac{G_1}{n} - O\left(\frac{G_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})}\right). \end{aligned}$$

Following the same strategy as that in the proof of Corollary 6.2, we see that k steps of this linearization yields

$$R^k(\lambda) = \lambda - k\left(\frac{2G_0 r}{n}\frac{\lambda}{\lambda + \lambda_{\mathrm{reg}}} - \frac{G_1}{n}\right) - O\left(\frac{kG_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})}\right).$$

Then if N is the smallest value necessary to obtain $R^N(\lambda) = 0$ for $\lambda = \lambda_{\min}(P_n)$, we have $R^{N-1}(\lambda) > 0$, and $(R^N)^{-1}(0) \geq \lambda - \frac{2G_0 r}{n}\frac{\lambda}{\lambda + \lambda_{\mathrm{reg}}} - \frac{G_1}{n} - O(\frac{G_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})})$. Setting $k = k(\varepsilon, \delta) = \frac{1}{\varepsilon}\log\frac{1}{2\delta}$, we have $\widehat{N} \geq N - k(\varepsilon, \delta)$ with probability at least $1 - \delta$, and as $\frac{G_0^2 r^2}{n^2(\lambda + \lambda_{\mathrm{reg}})} \lesssim \frac{G_0 r}{n}$ under the settings of the corollary, we have

$$(R^{\widehat{N}})^{-1}(0) \geq \lambda - O(1)k(\varepsilon, \delta)\left(\frac{G_0 r}{n}\frac{\lambda}{\lambda + \lambda_{\mathrm{reg}}} + \frac{G_1}{n}\right)$$

as desired.

6.2.3 Proof of Corollary 3.4

We first recognize that given any $\lambda \le \lambda_{\min}(P_n) + \lambda_{\text{reg}}$, Corollary 5.2 guarantees that

$$\lambda_{\max}(P_n') \le \lambda_{\max}(P_n)\left(1 + \varphi(t(\lambda) \cdot \mathrm{rad}(\mathcal{X}))\right) + \frac{G_1}{n}.$$

So the bounds required for recursive algorithms to provide privacy hold. For the actual privacy guarantee, we rely on the composition guarantee of Lemma 2.2, and privacy follows from Proposition 6.1 as in the proof of Corollary 3.2.

The proof of accuracy is also similar to that of Corollary 3.2. Let $r = \mathrm{rad}_2(\mathcal{X})$ as before and $\widehat{\lambda} = \widehat{\lambda}_{\min}(P_n) + \lambda_{\text{reg}}$. Using the assumption that $\frac{G_0 r}{n(\lambda_{\min}(P_n)+\lambda_{\text{reg}})} \le 1/C$ for a suitably large numerical constant C, the output $\widehat{\lambda}_{\min}$ of Algorithm 3 satisfies $\widehat{\lambda}_{\min} \ge \lambda_{\min}(P_n) - O(1)k(\varepsilon,\delta)(\frac{G_0 r}{n} + \frac{G_1}{n}) \gtrsim \lambda_{\min}(P_n)$ with probability at least $1-\delta$. So on this event, we obtain

$$\begin{aligned} R(\lambda) &= \lambda \exp\left(t(\widehat{\lambda})r\right) + \frac{G_1}{n} \\ &= \lambda + \frac{2G_0 r}{n} \cdot \frac{\lambda}{\lambda_{\min}(P_n) + \lambda_{\text{reg}}} + \frac{G_1}{n} + O(1)\frac{G_0^2 r^2}{n^2(\lambda_{\min}(P_n) + \lambda_{\text{reg}})} \cdot \frac{\lambda}{\lambda_{\min}(P_n) + \lambda_{\text{reg}}}, \end{aligned}$$

where we have used Corollary 3.2 so that $t(\widehat{\lambda}) = \frac{2G_0}{n(\lambda_{\min}(P_n)+\lambda_{\text{reg}})}(1+o(1))$, that $e^t = 1+t+O(t^2)$ for t small. By assumption $\frac{G_0 r}{n(\lambda_{\min}(P_n)+\lambda_{\text{reg}})} \le C^{-1}$, we obtain

$$R(\lambda) \le \lambda + O(1)\frac{G_0 r}{n} \cdot \frac{\lambda}{\lambda_{\min}(P_n) + \lambda_{\text{reg}}} + \frac{G_1}{n}.$$

As in the proof of Corollary 3.2 (*mutatis mutandis*), if N is the smallest value such that $R^N(\lambda_{\max}(P_n)) = G_1$, we have $R^{N-1}(\lambda_{\max}(P_n)) < G_1$ and $(R^N)^{-1}(G_1) \le \lambda_{\max}(P_n) + O(1)\frac{G_0 r}{n} \cdot \frac{\lambda_{\max}(P_n)}{\lambda_{\min}(P_n)+\lambda_{\text{reg}}} + \frac{G_1}{n}$. Iterating this $k = k(\varepsilon,\delta)$ times from $\lambda_{\max}(P_n)$ yields

$$\inf\left\{\lambda \mid R^{N-k(\varepsilon,\delta)}(\lambda) \ge G_1\right\} \le \lambda_{\max}(P_n) + O(1)k(\varepsilon,\delta)\frac{G_0 r}{n}\frac{\lambda_{\max}(P_n)}{\lambda_{\min}(P_n) + \lambda_{\text{reg}}} + k(\varepsilon,\delta)\frac{G_1}{n}.$$

7 Private algorithms for parameter release

When we wish to release a full parameter vector $\theta(P_n)$, we focus on the more basic composition approaches from Section 2.2. Letting

$$\omega_\theta(P_n; 1) := \sup\left\{\left\|\theta(P_n) - \theta(P_n')\right\|_2 \mid n\left\|P_n - P_n'\right\|_{\text{TV}} \le 1\right\}$$

be the modulus of continuity of $\theta(P_n)$ for the ℓ_2-norm with respect to changing a single example (the local sensitivity), Observation 2.4 shows that if a private random variable W satisfies $W \ge \omega_\theta(P_n; 1)$ with high probability, then

$$\theta(P_n) + \mathsf{N}(0, W^2 \cdot \sigma^2(\varepsilon,\delta) I_d)$$

is differentially private. We apply this insight to the two main cases we consider: generic smooth losses and for quasi-self-concordant (q.s.c.) generalized linear models (GLMs). For both, we focus on the unregularized case that the parameter set $\Theta = \mathbb{R}^d$.

7.1 General smooth losses without regularization

Focusing on generic smooth losses, Proposition 5.2 shows that

$$\omega_\theta(P_n; 1) \le \frac{1}{2G_2}\left[\lambda_{\min}(P_n) + \lambda_{\rm reg} - \sqrt{(\lambda_{\min}(P_n) + \lambda_{\rm reg})^2 - \frac{8G_0G_2}{n}}\right]$$

so long as $\lambda_{\min}(P_n) + \lambda_{\rm reg} \ge \max\{3G_1/n, \sqrt{12G_0G_2/n}\}$, as in Condition (C2). Thus, the following algorithm is differentially private and releases an approximation to $\theta(P_n)$.

Algorithm 7: Parameter release for generic smooth losses

Require: privacy level (ε, δ) and Lipschitz constants G_i, $i = 0, 1, 2$, of loss ℓ

i. Let $\widehat{\lambda}$ be the output of Alg. 6 with privacy $(\varepsilon/2, \delta/2)$, statistic $\lambda(P_n) = \lambda_{\min}(P_n) + \lambda_{\rm reg}$, and recursion

$$R(\lambda) = \max\left\{\frac{\lambda}{2}\left(1 + \sqrt{1 - \frac{8G_0G_2}{n\lambda^2}}\right) - \frac{G_1}{n}, \lambda_{\rm reg}\right\}.$$

ii. If $\widehat{\lambda}$ satisfies Condition (C2), set

$$W := \frac{1}{2G_2}\left[\widehat{\lambda} - \sqrt{\widehat{\lambda}^2 - \frac{8G_0G_2}{n}}\right]$$

and return

$$\widehat{\theta} = \theta(P_n) + \mathsf{N}\left(0, W^2\sigma^2\left(\frac{\varepsilon}{2}, \frac{\delta}{2}\right) \cdot I_d\right).$$

By combining the pieces of our results together, we obtain the following proposition on the accuracy and privacy of Algorithm 7.

Proposition 7.1. *The output $\widehat{\theta}$ of Alg. 7 is (ε, δ)-differentially private. Additionally, there exists a numerical constant $C < \infty$ such that if*

$$\lambda_{\min}(P_n) \ge C \max\left\{\frac{G_1}{n\varepsilon}\log\frac{1}{\delta}, \sqrt{\frac{G_0G_2}{n\varepsilon}\log\frac{1}{\delta}}\right\}$$

then with probability at least $1 - \delta - \gamma$,

$$\left\|\theta(P_n) - \widehat{\theta}\right\|_2 \le C\frac{G_0}{n\varepsilon\lambda_{\min}(P_n)}\sqrt{\log\frac{1}{\delta}}\left[\sqrt{d} + \sqrt{\log\frac{1}{\gamma}}\right].$$

Proof The privacy guarantee is nearly immediate via Corollary 6.2, which gives that $\lambda_{\min}(P_n) + \lambda_{\rm reg} \ge \widehat{\lambda}$ with probability at least $1 - \delta/2$ and $\widehat{\lambda}$ is $\varepsilon/2$-differentially private. Then Observation 2.4 guarantees that $\widehat{\theta}$ is (ε, δ)-differentially private.

Corollary 6.2 guarantees $\widehat{\lambda} \ge \lambda_{\min}(P_n) - O(1)(\frac{G_0G_2}{\varepsilon\lambda_{\min}(P_n)n} + \frac{G_1}{n\varepsilon})\log\frac{1}{\delta}$ with probability at least $1 - \delta$. On this event, so long as the lower bound on $\lambda_{\min}(P_n)$ in the statement of the proposition holds (for suitably large numerical constant C), the random variable

$$W \lesssim \frac{1}{2G_2}\left[\lambda_{\min}(P_n) - \sqrt{\lambda_{\min}(P_n)^2 - \frac{G_0G_2}{n}}\right] \lesssim \frac{G_0}{n\lambda_{\min}(P_n)}$$

by a Taylor approximation of $\sqrt{1-\gamma} = 1 - \gamma/2 + O(\gamma^2)$, valid for γ small. Noting that for any $0 < \gamma < 1$, a random Gaussian $Z \sim \mathsf{N}(0, \sigma^2 I)$ satisfies $\|Z\|_2 \le \sigma(\sqrt{d} + O(1)\sqrt{\log(1/\gamma)})$ with probability at least $1 - \gamma$ (cf. [35, Thm. 3.1.1]), then because $\sigma^2(\varepsilon, \delta) \lesssim \varepsilon^{-1} \log \frac{1}{\delta}$, we have the proposition. $\square$

7.2 Quasi-self-concordant GLMs and the proof of Theorem 1

For q.s.c. GLMs, Proposition 5.4 shows that so long as $\lambda_{\min}(P_n)$ and $\lambda_{\reg}$ satisfy inequality (C1), then

$$\omega_\theta(P_n; 1) \le t(\lambda_{\min}(P_n) + \lambda_{\reg}),$$

where Eq. (9) defines the parameter change constant $t(\lambda) = \frac{2G_0}{n\lambda}(1 + o(1))$. In this case, by leveraging Corollary 3.2, we can prove the claimed deviation guarantee on $\widehat{\theta}$ relative to $\theta(P_n)$. The privacy guarantee of the theorem follows by combining Observation 2.4 with Proposition 6.1.

For the accuracy guarantee, Corollary 3.2 shows that under the conditions on $\lambda_{\min}(P_n)$ and $\lambda_{\reg}$ in the statement of Theorem 1, we have $\widehat{\lambda} \ge \lambda_{\min}(P_n) - O(\varepsilon^{-1} \log \frac{1}{\delta}) \max\{\frac{G_1}{n}, \frac{G_0 \rad(\mathcal{X})}{n}\}$ with probability at least $1 - \delta$, and a Taylor approximation yields that the parameter change quantity (9) satisfies

$$t(\widehat{\lambda} + \lambda_{\reg}) \lesssim \frac{G_0}{n(\widehat{\lambda} + \lambda_{\reg})} \lesssim \frac{G_0}{n(\lambda_{\min}(P_n) + \lambda_{\reg})}$$

on this event. Setting $t = t(\widehat{\lambda} + \lambda_{\reg})$ and $\sigma^2 = \sigma^2(\varepsilon, \delta)$, the quantity $\widehat{\theta} = \theta(P_n) + \mathsf{N}(0, t^2\sigma^2 I_d)$ satisfies $\|\widehat{\theta} - \theta(P_n)\|_2 \lesssim t\sigma\sqrt{d}(1 + \sqrt{\log \frac{1}{\gamma}})$ with probability at least $1 - \gamma$.

7.3 Dimension and accuracy scaling

As in Section 3.3, let us briefly discuss the scaling of the accuracy with dimension in Proposition 7.1 and Theorem 1 and when these scalings apply. We focus on the case of a "typical" generalized linear modeling scenario, where we have a loss of the form $\ell_\theta(x, y) = h(y - \langle x, \theta\rangle)$ or $\ell_\theta(x, y) = h(y\langle x, \theta\rangle)$, as in the robust regression or (binary) logistic regression Examples 1 and 2, where the covariate vectors $x \in [-1, 1]^d$ and h has Lipschitz zeroth, first, and second derivatives. Then we have the scalings

$$G_0 \asymp \sqrt{d}, \quad G_1 \asymp d, \quad G_2 \asymp d^{3/2}, \quad \rad(\mathcal{X}) \asymp \sqrt{d}.$$

In both cases, if either of Algorithms 3 or 7 releases an estimate $\widehat{\theta}$,

$$\|\widehat{\theta} - \theta(P_n)\|_2 \lesssim \frac{G_0}{n\lambda_{\min}(P_n)} \cdot \frac{\sqrt{d \log \frac{1}{\delta}}}{\varepsilon} \tag{24}$$

with high probability by Proposition 7.1 and Theorem 1. As in our discussion in the introduction, for n large, at $\theta = \theta(P_n)$ the local modulus (2) has scaling

$$\omega_\theta(P_n; 1) \asymp \sup_{x \in \mathcal{X}, y} \frac{1}{n} \left\| (P_n \ddot{\ell}_\theta + \lambda_{\reg} I)^{-1} \dot{\ell}_\theta(x, y) \right\|_2 \overset{(\star)}{\le} \frac{1}{n} \frac{G_0}{\lambda_{\min}(P_n)},$$

where inequality ($\star$) holds with (approximate) equality when the Hessian $P_n\ddot{\ell}_\theta$ is near a scaled identity matrix or $\mathcal{X}$ is a scaled ℓ_2-ball. By Cai et al.'s score attack [12], the additional scaling with $\sqrt{d}/\varepsilon$ is unavoidable, making the accuracy of these algorithms unimprovable in a worst-case sense, though they adapt to the particular (local) strong convexity of the problem.

At the grossest level, then, the main difference between the algorithms is when they may actually release parameters, as the accuracy guarantees (24) they provide are indistinguishable. The basic Algorithm 7 states that as soon as

$$\lambda_{\min}(P_n) + \lambda_{\reg} \gg \max\left\{\frac{d}{n\varepsilon}, \frac{d}{\sqrt{n\varepsilon}}\right\},$$

so that $n \gg d^2$, the algorithm applies, while Algorithm 3 requires the weaker condition that

$$\lambda_{\min}(P_n) + \lambda_{\reg} \gg \frac{d}{n\varepsilon},$$

so that $n \gg d$. (In both cases, we ignore the logarithmic scaling with $\frac{1}{\delta}$.) Such a requirement is, at least in the worst case, unavoidable under differential privacy.

8 Releasing linear functionals of the parameter

By combining the algorithms we have developed for releasing minimal eigenvalues in Section 6.2, the privacy guarantees of the propose-test-release framework in Section 2.2.2, and the stability bounds in Section 5, we can finally return to one of our original motivations: releasing a single coordinate of the vector $\theta(P_n)$, or, more generally, releasing

$$u^T\theta(P_n)$$

for a unit vector u. To develop the methodology, we will require a few more sophisticated deviation bounds on the parameter θ and functionals of θ. Note from Lemma 5.2 in the proof of Proposition 5.4 that under the conditions of the proposition, for

$$\gamma = \gamma(P_n) \coloneqq \alpha \cdot t(\lambda_{\min}(P_n) + \lambda_{\reg})\mathrm{rad}(\mathcal{X}) < 1,$$

there exists a symmetric D with $\|\!|D|\!\|_{\op} \le \frac{\gamma}{1-\gamma}$ such that the Hessian $H \coloneqq P_n\ddot{\ell}_\theta + \lambda_{\reg} I$ satisfies

$$\theta(P_n') - \theta(P_n) = H^{-1}(P_n - P_n')\dot{\ell}_{\theta'} + H^{-1/2}DH^{-1/2}(P_n - P_n')\dot{\ell}_{\theta'},$$

where we use $\theta' = \theta(P_n')$ and $\theta = \theta(P_n)$. Then for a any ℓ_2-unit vector u, inequality (11) holds:

$$\left|u^T(\theta(P_n') - \theta(P_n))\right| \le \omega(u \mid P_n) \coloneqq \Delta(P_n, u) + \frac{2G_0}{n(\lambda_{\min}(P_n) + \lambda_{\reg})} \cdot \frac{\gamma(P_n)}{1 - \gamma(P_n)}$$

where we recall the directional sensitivity (3b)

$$\Delta(P_n, u) = \frac{1}{n} \sup_{g_0, g_1 \in \mathcal{G}} u^T(P_n\ddot{\ell}_{\theta(P_n)} + \lambda_{\reg} I)^{-1}(g_0 - g_1).$$

We use the propose-test-release scheme to argue that releasing

$$u^T\theta(P_n) + \omega(u \mid P_n) \cdot Z$$

for a Gaussian Z with variance scaling as $\frac{1}{\varepsilon^2}\log\frac{1}{\delta}$ is private so long as we can privately certify that $\lambda_{\min}(P_n)$ is large enough and $\lambda_{\max}(P_n)$ is small enough.

8.1 Propose-test-release for the local modulus of continuity

The approach to the (somewhat) naive release above introduces subtleties, however, because neighboring samples P_n, P_n' may have different directional sensitivies Δ, so that even if $u^T\theta(P_n) - u^T\theta(P_n')$ is small, the magnitude of the noise added may leak information. We therefore adopt an approach building out of literature on private mean estimation algorithms that adapt to the covariance of the underlying data [8, 10, 11, 18]. Thus, we control the ratio

$$\frac{\omega(u \mid P_n)}{\omega(u \mid P_n')} = \frac{\Delta(P_n, u) + \frac{2G_0}{n(\lambda_{\min}(P_n)+\lambda_{\reg})} \frac{\gamma(P_n)}{1-\gamma(P_n)}}{\Delta(P_n', u) + \frac{2G_0}{n(\lambda_{\min}(P_n')+\lambda_{\reg})} \frac{\gamma(P_n')}{1-\gamma(P_n')}}. \tag{25}$$

We can control this ratio as soon as we have (high probability) lower bounds $\widehat{\lambda}_0 \le \lambda_{\min}(P_n)$ and $\widehat{\lambda}_1 \ge \lambda_{\max}(P_n)$. To that end, assume there exists a ratio bounding term $\mathtt{r}(u, \lambda)$ for $\lambda = (\lambda_0, \lambda_1)$ such that whenever $0 \le \lambda_0 \le \lambda_{\min}(P_n)$ and $\lambda_{\max}(P_n) \le \lambda_1$, we have

$$\frac{1}{1 + \mathtt{r}^2(u, \lambda)} \le \frac{\omega(u \mid P_n)^2}{\omega(u \mid P_n')^2} \le 1 + \mathtt{r}^2(u, \lambda) \;\; \text{for all neighboring } P_n, P_n'. \tag{26}$$

Once we have such a guarantee, then so long as $\varepsilon \ge \frac{1}{2}(1 + \Phi^{-1}(1-\delta/2)^2)\mathtt{r}^2(u, \widehat{\lambda})$, we release

$$T := u^T\theta(P_n) + \mathsf{N}\left(0, \sigma^2(\varepsilon, \delta) \cdot \omega(u \mid P_n)^2\right) \tag{27}$$

and $T = \perp$ otherwise. The following result, based on the test-release framework (Algorithm 1), guarantees privacy.

Proposition 8.1. *Let $\varepsilon \ge \frac{1}{2}(1 + \Phi^{-1}(1-\delta/2)^2)\mathtt{r}^2(u, \widehat{\lambda})$ and $\delta > 0$. Then T is $(3\varepsilon, (1 + e^{\varepsilon} + e^{2\varepsilon})\delta)$-differentially private.*

Because the proposition is more or less a consequence of the propose-test-release scheme, we prove it in Appendix A.1.4.

We provide two main results that allow us to apply Proposition 8.1. First, we address the case in which the gradient set is a scaled ℓ_2-ball, and in the second, when it is a scaled ℓ_∞-ball. In either case, we must specify several constants to allow (private) certification that the ratio (26) is bounded. Assume that the loss $\ell_\theta(x, y) = h(\langle \theta, x\rangle, y)$ where h satisfies the self-concordance guarantees (8) with self-bounding parameter α satisfying $\varphi(t) \le \alpha t$. For covariate domain $\mathcal{X}$, let $r = \rad_2(\mathcal{X})$ be the ℓ_2-radius of the data, and recall the definition (9) of $t(\lambda)$, which guarantees that $\|\theta(P_n) - \theta(P_n')\|_2 \le t(\lambda_{\min}(P_n) + \lambda_{\reg})$ under Condition (C1). Recall additionally the recursion R defined by the cases (10).

Now, we control the ratio (26). Fix λ_0 and λ_1 to be any positive values (in the sequel, we take them to estimate the minimal and maximal eigenvalues $\lambda_{\min}(P_n) + \lambda_{\reg}$ and $\lambda_{\max}(P_n) + \lambda_{\reg}$). Recall the definitions (13) of the constants

$$t := t(\lambda_0), \quad r := \rad_2(\mathcal{X}), \quad \beta := \frac{\|h''\|_\infty}{[1 - \alpha t]_+} \frac{r^2}{n\lambda_0}, \quad \gamma := \alpha r \cdot t, \quad \gamma' := \alpha r \cdot t(R(\lambda_0))$$
$$s_1 := \frac{1}{[1 - \alpha r t]_+} - 1, \quad s_2 := \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{[1 - \alpha r t]_+}, \quad \kappa := \frac{\lambda_1}{\lambda_0}.$$

We then have the following guarantees.

Proposition 8.2. *Let the preceding conditions hold, assume that* $\lambda_0 \le \lambda_{\min}(P_n) + \lambda_{\rm reg}$ *and* $\lambda_{\max}(P_n) + \lambda_{\rm reg} \le \lambda_1$*, that Condition (C1) holds, and define* $\kappa = \frac{\lambda_1}{\lambda_0}$*. Let the gradient set*

$$\mathcal{G} = \left\{ g \in \mathbb{R}^d \mid \|g\|_2 \le G_0 \right\}.$$

Then

$$1 \le \frac{\omega(u \mid P_n)}{\Delta(P_n, u)} \le 1 + \kappa \frac{\gamma}{1-\gamma}$$

and

$$1 - \kappa(s_1 + s_2 r) \le \frac{\omega(u \mid P_n')}{\Delta(P_n, u)} \le 1 + \kappa(s_1 + s_2 r) + \kappa \frac{\lambda_0}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}.$$

Proposition 8.3. *Let the conditions of Proposition 8.2 hold, but define* $d_p = d^{1-2/p}$ *and let the gradient set*

$$\mathcal{G} = \left\{ g \in \mathbb{R}^d \mid \|g\|_p \le G_0 \right\}.$$

Then

$$1 \le \frac{\omega(u \mid P_n)}{\Delta(P_n, u)} \le 1 + \sqrt{d_p}\kappa \frac{\gamma}{1-\gamma}$$

and

$$1 - \sqrt{d_p} s_1 \kappa - 2\frac{2 d_p s_2}{\lambda_0} \le \frac{\omega(u \mid P_n')}{\Delta(P_n, u)} \le 1 + \sqrt{d_p}\kappa s_1 + \frac{2 s_2 d_p}{\lambda_0} + \frac{\sqrt{d_p}\kappa\lambda_0}{R(\lambda_0)} \frac{\gamma'}{1-\gamma'}.$$

Propositions 8.2 and 8.3 rely on fairly careful control over non-symmetric quadratic forms. After giving some commentary on the results, we build up to them over the remainder of this section, beginning in Section 8.2, which addresses the similarity of Hessians for neighboring samples P_n and P_n', with the proofs of the propositions following in Sections 8.3 and 8.4.

8.1.1 Proof of Corollary 3.6

To obtain privacy using these results, we apply Proposition 8.1. We need to guarantee that the ratio of the (local) moduli of continuity satisfy the appropriate bounds on the ratio $\mathtt{r}(u, \lambda)$ in inequality (26). The inequalities (14) guarantee that

$$\frac{1}{1+\mathtt{r}^2} \le \left(\frac{\omega(u \mid P_n)}{\omega(u \mid P_n')} \right)^2 \le 1 + \mathtt{r}^2 \quad \text{for some} \quad \mathtt{r}^2 \le \frac{2\varepsilon}{1 + \Phi^{-1}(1-\delta/2)^2}.$$

Then Corollary 3.6 follows as an immediate corollary to Propositions 8.2 and 8.3, coupled with Proposition 8.1.

8.2 Self-similar matrices and Hessians

Proposition 8.1 makes it clear that what is essential to releasing a statistic accurately is to provide sufficient bounds on the ratio (26). This turns out to be a fairly subtle question, and we develop a few tools to bound ratios of matrix-vector products here to address the issue. We provide proofs of the results in Section A.4, which require a few auxiliary results as well. Abstractly—treating H_0 and H_1 as the Hessians $P_n \ddot{\ell}_{\theta(P_n)}$ and $P_n' \ddot{\ell}_{\theta(P_n')}$, respectively—we will control ratios of

$$\sup_{v \in \mathcal{V}} u^T H_0^{-1} v \quad \text{to} \quad \sup_{v \in \mathcal{V}} u^T H_1^{-1} v,$$

where $\mathcal{V}$ is a symmetric convex body. In evaluating the ratios of these quantities, we consider matrices H_0 and H_1 that we term (s_1, s_2)-*self-similar relative to* $\mathcal{X}$, meaning that H_0 and H_1 satisfy

$$H_1^{-1} = H_0^{-1} + E_1 + E_2, \tag{28}$$

where the error matrices E_1 and E_2 satisfy that there exist vectors $x_0, x_1 \in \mathcal{X}$ such that

$$-s_1 H_0^{-1} \preceq E_1 \preceq s_1 H_0^{-1} \quad \text{and} \quad -s_2 H_0^{-1} x_0 x_0^T H_0^{-1} \preceq E_2 \preceq s_2 H_0^{-1} x_1 x_1^T H_0^{-1}.$$

The key to applying these similarity results is that the Hessians $H_0 = P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\text{reg}} I$ and $H_1 = P_n' \ddot{\ell}_{\theta(P_n')} + \lambda_{\text{reg}} I$ are self-similar.

Lemma 8.1. *Assume that* $\|\theta - \theta'\|_2 \le t$ *and* $\operatorname{rad}(\mathcal{X}) \le r$. *Define* $\beta = \frac{\|h''\|_\infty}{1-\alpha r t} \frac{1}{\lambda_{\min}(P_n \ddot{\ell}_\theta) + \lambda_{\text{reg}}} \frac{r^2}{n}$. *If* $\beta < 1$, *there there exist* $x_0, x_1 \in \mathcal{X}$ *such that*

$$\begin{aligned} &\frac{1}{1+\alpha r t} H_0^{-1} - \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{(1+\alpha r t)^2} H_0^{-1} x_0 x_0^T H_0^{-1} \\ &\qquad \preceq H_1^{-1} \preceq \frac{1}{1-\alpha r t} H_0^{-1} + \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{(1-\alpha r t)^2} H_0^{-1} x_1 x_1^T H_0^{-1}. \end{aligned}$$

See Section 8.5 for a proof of Lemma 8.1. Rewriting the result in a more modular form, Lemma 8.1 shows the following:

Lemma 8.2. *Let* $H_0 = P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\text{reg}} I$ *and* $H_1 = P_n' \ddot{\ell}_{\theta(P_n')} + \lambda_{\text{reg}} I$. *Assume the bounds of Lemma 8.1 that* $\|\theta(P_n) - \theta(P_n')\|_2 \le t$. *Then* H_0 *and* H_1 *are* (s_1, s_2)*-self-similar with*

$$s_1 = \frac{1}{1-\alpha r t} - 1 \quad \text{and} \quad s_2 = \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{(1-\alpha r t)^2}.$$

We specialize these results to bound the ratios of $\omega(u \mid P_n)/\omega(u \mid P_n')$ when the $\mathcal{V}$ is a norm ball. We begin by capturing the case in which $\mathcal{V}$ is an ℓ_2 ball, so that $\sup_{v \in \mathcal{V}} u^T H_0^{-1} v = \left\|H_0^{-1} u\right\|_2$. (Scalings of the ℓ_2-ball follow trivially.)

Lemma 8.3. *Let* H_0 *and* H_1 *be* (s_1, s_2)*-self-similar* (28) *relative to* $\mathcal{X}$. *Then*

$$\left\|H_1^{-1} u\right\|_2 \le \left(1 + \frac{s_1}{2}\left(1 + \frac{\lambda_{\max}(H_0)}{\lambda_{\min}(H_0)}\right) + s_2 \sup_{x \in \mathcal{X}} \lambda_{\max}(H_0) \left\|H_0^{-1} x\right\|_2^2\right) \left\|H_0^{-1} u\right\|_2.$$

Similarly,

$$\left\|H_1^{-1} u\right\|_2 \ge \left(1 - \frac{s_1}{2}\left(1 + \frac{\lambda_{\max}(H_0)}{\lambda_{\min}(H_0)}\right) - s_2 \sup_{x \in \mathcal{X}} \lambda_{\max}(H_0) \left\|H_0^{-1} x\right\|_2^2\right) \left\|H_0^{-1} u\right\|_2.$$

See Section A.4.2 for the proof of the lemma.

We can also consider more general sets. Let the set $\mathcal{V}$ be a symmetric convex body as before. Define the maximal ℓ_p-inscribed- and ℓ_p-radii by

$$\operatorname{rad}_p(\mathcal{V}) := \sup\{\|v\|_p \mid v \in \mathcal{V}\} \quad \text{and} \quad \operatorname{ins}_p(\mathcal{V}) := \sup\{t \mid t\mathbb{B}_p^d \subset \mathcal{V}\},$$

so that the ratio $\operatorname{rad}_p(\mathcal{V})/\operatorname{ins}_p(\mathcal{V})$ gives a type of condition number for $\mathcal{V}$. For example, $\mathcal{V} = [-1, 1]^d$ has $\operatorname{rad}_2(\mathcal{V}) = \sqrt{d}$ and $\operatorname{ins}_2(\mathcal{V}) = 1$. For a matrix A we define the $\mathcal{V}$-relative conditioning

$$\kappa(A, \mathcal{V}) := \frac{\sup_{v \in \mathcal{V}} v^T A v / \|v\|_2}{\inf_{\|u\|_2 = 1} \sup_{v \in \mathcal{V}} u^T A v}.$$

A quick calculation gives the following bound on this condition number.

Lemma 8.4. *The $\mathcal{V}$-relative condition number satisfies*

$$\kappa(A, \mathcal{V}) \le \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(A).$$

Additionally, for $\mathcal{V} = \mathbb{B}_2^d$, $\kappa(A, \mathcal{V}) = \kappa(A)$, and for $\mathcal{V} = [-1, 1]^d$, for any condition number $\kappa \ge 1$ there exist matrices A with $\kappa(A) = \kappa$ and

$$\frac{1}{\sqrt{2}} \cdot \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(A) \le \kappa(A, \mathcal{V}) \le \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(A).$$

See Section A.4.3 for a proof. With this lemma, we can provide an analogue of Lemma 8.3 when $\mathcal{V}$ is not an ℓ_2-ball.

Lemma 8.5. *Let H_0 and H_1 be (s_1, s_2)-self-similar (28), and let u be a unit vector. Let $\mathcal{V} = c\mathcal{X} = \{cx \mid x \in \mathcal{X}\}$, where $c > 0$ is a fixed constant. Then for any unit vector u,*

$$\sup_{v \in \mathcal{V}} u^T H_1^{-1} v \le \left(1 + s_1 \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(H_0) + s_2 \frac{2\mathrm{rad}_2^2(\mathcal{V})}{c^2 \lambda_{\min}(H_0)}\right) \sup_{v \in \mathcal{V}} u^T H_0^{-1} v$$

and

$$\sup_{v \in \mathcal{V}} u^T H_1^{-1} v \ge \left(1 - s_1 \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(H_0) - s_2 \frac{2\mathrm{rad}_2^2(\mathcal{V})}{c^2 \lambda_{\min}(H_0)}\right) \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

In the case that $\mathcal{V} = [-1, 1]^d$ and $H_0 = I$, these results are sharp for any standard basis vector $u \in \{e_1, \ldots, e_d\}$, in that there exists H_1 satisfying self-similarity and

$$\sup_{v \in \mathcal{V}} u^T H_1^{-1} v \ge \left(1 + s_1\sqrt{d} + s_2 d\right) = \left(1 + s_1 \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(H_0) + s_2 \frac{\mathrm{rad}_2^2(\mathcal{V})}{\lambda_{\min}(H_0)}\right) \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

See Section A.4.4 for the proof.

8.3 Proof of Proposition 8.2

Observe first that, by Lemma 8.2, if we have any quantity $\lambda_0 \le \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$ then the Hessians $H_0 = P_n \ddot{\ell}_{\theta(P_n)} + \lambda_{\mathrm{reg}} I$ and $H_1 = P_n' \ddot{\ell}_{\theta(P_n')} + \lambda_{\mathrm{reg}} I$ are (s_1, s_2)-self-similar, where for $t = t(\lambda_0)$, we recall our parameter definitions

$$s_1 := \frac{1}{[1 - \alpha r t]_+} - 1, \quad s_2 := \frac{1}{n(1-\beta)} \frac{\|h''\|_\infty}{[1 - \alpha r t]_+^2}, \quad \text{and} \quad \beta = \frac{\|h''\|_\infty}{[1 - \alpha r t]_+} \frac{r^2}{n\lambda_0},$$

Recalling the definition $\gamma(P_n) = \alpha t(\lambda_{\min}(P_n) + \lambda_{\mathrm{reg}})\mathrm{rad}(\mathcal{X})$, because $t(\lambda)$ is decreasing in λ, once we have the lower bound $\lambda_0 \le \lambda_{\min}(P_n) + \lambda_{\mathrm{reg}}$ we obtain an upper bound $\gamma = \alpha t r \ge \gamma(P_n)$.

We first provide an upper bound on $\omega(u \mid P_n) = \Delta(P_n, u) + \frac{2G_0}{n\lambda_0}\frac{\gamma}{1-\gamma}$ and a lower bound on $\omega(u \mid P_n') \ge \Delta(P_n', u)$. We may use the bounds from Lemmas 8.3 and 8.5 to control $\Delta(P_n', u)$. So long as $\lambda_1 \ge \lambda_{\max}(P_n) + \lambda_{\mathrm{reg}}$, the condition number estimate $\kappa = \frac{\lambda_1}{\lambda_0} \ge \kappa(H_0)$. Recalling the gradient set $\mathcal{G} = \{g \in \mathbb{R}^d \mid \|g\|_2 \le G_0\}$, Lemma 8.3 coupled with the recognition that

$$\Delta(P_n, u) = \frac{2}{n} \sup_{g \in \mathcal{G}_2} u^T H_0^{-1} g = \frac{2G_0}{n} \left\|H_0^{-1} u\right\|_2$$

implies

$$\Delta(P_n', u) \geq (1 - s_1\kappa - s_2 r\kappa)\, \Delta(P_n, u).$$

For the upper bound on $\omega(u \mid P_n)$, note that $\left\|H_0^{-1}u\right\|_2 \geq \lambda_1^{-1}$ for any unit vector u, so that $\frac{2G_0}{n\lambda_0} \leq \kappa\Delta(P_n, u)$.

To obtain the lower bound on $\omega(u \mid P_n)$ and upper bound on $\omega(u \mid P_n')$, note that $\gamma(P_n') = \alpha t(\lambda_{\min}(P_n'))\mathrm{rad}(\mathcal{X}) \leq \alpha t(R(\lambda_0))r =: \gamma'$. So proceeding as above, we obtain trivially that $\omega(u \mid P_n) \geq \Delta(P_n, u)$, and

$$\omega(u \mid P_n') \leq \Delta(P_n', u) + \frac{2G_0}{nR(\lambda_0)}\frac{\gamma'}{1-\gamma'} \leq \Delta(P_n, u)(1 + \kappa(s_1 + s_2 r)) + \frac{2G_0}{nR(\lambda_0)}\frac{\gamma'}{1-\gamma'}$$

by Lemma 8.3 again. Now again use that $\frac{2G_0}{n\lambda_0} \leq \kappa\Delta(P_n, u)$.

8.4 Proof of Proposition 8.3

In the case of ℓ_p, $p > 2$-bounded gradients, we must control the error terms somewhat differently than we did in the proof of Proposition 8.2. We still have that $H_0 = P_n\ddot{\ell}_{\theta(P_n)} + \lambda_{\mathrm{reg}}I$ and $H_1 = P_n'\ddot{\ell}_{\theta(P_n')} + \lambda_{\mathrm{reg}}I$ are s_1, s_2-similar, with the same definitions of the constants as in Proposition 8.2.

For $\mathcal{V} = \{v \in \mathbb{R}^d \mid \|v\|_p \leq 1\}$, we have relative condition bound $\kappa(A, \mathcal{V}) \leq \sqrt{d_p}\kappa(A)$, which is (nearly) as sharp as possible by Lemma 8.4. We therefore have via Lemma 8.5 that

$$\Delta(P_n', u) = \frac{2}{n}\sup_{g \in \mathcal{G}_p} u^T H_1^{-1} g \geq \left(1 - s_1\sqrt{d_p}\kappa(H_0) - s_2\frac{2d_p}{\lambda_{\min}(H_0)}\right)\Delta(P_n, u).$$

Using that $\Delta(P_n, u) = \frac{2G_0}{\sqrt{d_p}n}\left\|H_0^{-1}u\right\|_q \geq \frac{2G_0}{\sqrt{d_p}n\lambda_1}$, where $q = \frac{p}{p-1} < 2$ is conjugate to p, we rearrange to obtain $\frac{2G_0}{n\lambda_0} \leq \sqrt{d_p}\kappa\Delta(P_n, u)$, so that

$$\omega(u \mid P_n) \leq \Delta(P_n, u)\left(1 + \sqrt{d_p}\kappa\frac{\gamma}{1-\gamma}\right).$$

To obtain the converse bounds, note that applying Lemma 8.5 gives

$$\Delta(P_n', u) \leq \left(1 + s_1\sqrt{d_p}\kappa(H_0) + s_2\frac{2d_p}{\lambda_{\min}(H_0)}\right)\Delta(P_n, u).$$

and similar calculations thus yield

$$\begin{aligned}\omega(u \mid P_n') &\leq \Delta(P_n', u) + \frac{2G_0}{nR(\lambda_0)}\frac{\gamma'}{1-\gamma'} \\ &\leq \Delta(P_n, u)\left(1 + \sqrt{d_p}s_1\kappa + \frac{2s_2 d_p}{\lambda_0}\right) + \frac{\sqrt{d_p}\kappa\lambda_0}{R(\lambda_0)}\frac{\gamma'}{1-\gamma'}\Delta(P_n, u)\end{aligned}$$

where we used again that $\frac{2G_0}{n\lambda_0} \leq \sqrt{d_p}\kappa\Delta(P_n, u)$.

8.5 Proof of Lemma 8.1

Recalling our notation that $\|\theta - \theta'\|_2 \le t$ and $\mathrm{rad}(\mathcal{X}) \le r$, we have

$$P_n'\ddot{\ell}_{\theta'} = P_n\ddot{\ell}_{\theta'} + (P_n' - P_n)\ddot{\ell}_{\theta'} = P_n\ddot{\ell}_\theta + P_n(\ddot{\ell}_{\theta'} - \ddot{\ell}_\theta) + (P_n' - P_n)\ddot{\ell}_{\theta'}.$$

Let $E_1 = P_n\ddot{\ell}_{\theta'} - P_n\ddot{\ell}_\theta$, so that $-\alpha r t P_n\ddot{\ell}_\theta \preceq E_1 \preceq \alpha r t P_n\ddot{\ell}_\theta$. Let $E_2 = (P_n' - P_n)\ddot{\ell}_{\theta'}$, so leveraging that $\ddot{\ell}_\theta = h''(\langle\theta, x\rangle, y)xx^T$, there exist x_0, x_1, y_0, y_1 such that

$$n(P_n' - P_n)\ddot{\ell}_{\theta'} = nE_2 = \ddot{\ell}_{\theta'}(x_0, y_0) - \ddot{\ell}_{\theta'}(x_1, y_1) = h''(\langle\theta', x_0\rangle, y_0)x_0x_0^T - h''(\langle\theta', x_1\rangle, y_1)x_1x_1^T,$$

that is, there exist x_0, x_1 and $\|h''\|_\infty < \infty$ such that $-x_1x_1^T \|h''\|_\infty \preceq nE_2 \preceq x_0x_0^T \|h''\|_\infty$. Define

$$H_0 = P_n\ddot{\ell}_\theta + \lambda_{\mathrm{reg}} I, \quad \text{and} \quad H_1 = P_n'\ddot{\ell}_{\theta'} + \lambda_{\mathrm{reg}} I.$$

Then using the operator monotonicity properties of the matrix inverse, we obtain that for some $x \in \mathcal{X}$,

$$\begin{aligned} H_1^{-1} &\preceq \left(P_n\ddot{\ell}_\theta(1 - \alpha r t) + (1 - \alpha r t)\lambda_{\mathrm{reg}} I - n^{-1}\left\|h''\right\|_\infty xx^T\right)^{-1} \\ &= \frac{1}{1 - \alpha r t}H_0^{-1} + \frac{\|h''\|_\infty}{n(1 - \alpha r t)^2(1 - \frac{\|h''\|_\infty}{(1-\alpha r t)n}x^T H_0^{-1}x)}H_0^{-1}xx^TH_0^{-1} \end{aligned}$$

by the Sherman-Morrison inversion formula. Similarly, there exists $x \in \mathcal{X}$ such that

$$\begin{aligned} H_1^{-1} &\succeq \left(P_n\ddot{\ell}_\theta(1 + \alpha r t) + (1 + \alpha r t)\lambda_{\mathrm{reg}} I + n^{-1}\left\|h''\right\|_\infty xx^T\right)^{-1} \\ &= \frac{1}{1 + \alpha r t}H_0^{-1} - \frac{\|h''\|_\infty}{n(1 + \alpha r t)^2(1 + \frac{\|h''\|_\infty}{n(1+\alpha r t)}x^T H_0^{-1}x)}H_0^{-1}xx^TH_0^{-1}. \end{aligned}$$

Defining $\beta = \frac{\|h''\|_\infty}{1-\alpha r t}\frac{r^2}{(\lambda_{\min}(P_n\ddot{\ell}_\theta) + \lambda_{\mathrm{reg}})n}$, we thus obtain

$$\begin{aligned} &\frac{1}{1 + \alpha r t}H_0^{-1} - \frac{\|h''\|_\infty}{n(1 + \alpha t r)^2(1 - \beta)}H_0^{-1}xx^TH_0^{-1} \\ &\quad \preceq H_1^{-1} \preceq \frac{1}{1 - \alpha r t}H_0^{-1} + \frac{\|h''\|_\infty}{n(1 - \alpha t r)^2(1 - \beta)}H_0^{-1}xx^TH_0^{-1} \end{aligned}$$

as desired.

9 Discussion

The original motivation for this paper was in service to a hypothesis that we entertain, which is that to improve adoption of privacy-preserving procedures in sciences will require effective and practical methods. We admit that we have, perhaps, strayed from a simple set of procedures via detours through some nontrivial mathematical machinery, which still leaves us unable to easily test our hypothesis. In spite of this, our experimental results are promising: for large sample sizes, Algorithm 5 nearly achieves optimal performance, to within small numerical constant factors.

Nonetheless, there are several avenues for future work, which we hope that others will tackles. First, as we discuss in Section 3.3, even for well-conditioned problems, it appears

that Algorithm 5 becomes most effective when $n \gtrsim d^{3/2}$, at which point it more or less releases $u^T\theta(P_n) + \mathsf{N}(0, \Delta^2(P_n, u))$, which is optimal scaling. Identifying the precise dimension dependence at which this "local modulus"-dependent release is possible will be interesting. One plausible avenue here would be to develop procedures that rely not on the minimal eigenvalue $\lambda_{\min}(P_n)$, which governs the worst-case gross behavior of $\|\theta(P_n) - \theta(P_n')\|_2$ but instead on a more nuanced quantity relating directly to the differences $u^T\theta(P_n) - u^T\theta(P_n')$. Corollaries 3.3 and 5.2 both rely on this global bound in the change of $\theta(P_n)$ rather than the particular directionality that $\theta(P_n') \approx \theta(P_n) - (P_n\ddot{\ell}_\theta)^{-1}(P_n - P_n')\dot{\ell}_\theta$, so that more careful tracking there could allow better dimension-dependence. Of course, our approaches here may be simply mis-directed, and a more direct attempt to implement Asi and Duchi's [2] inverse sensitivity, which is instance optimal, may be more sensible. Regardless, we hope that continued interest in practicable procedures for private estimation continues.

A Technical appendices

A.1 Proofs of basic privacy building blocks

In this appendix, we collect the proofs of the privacy building blocks in Section 2.2.

A.1.1 Proof of Lemma 2.2

We wish to show that for any $A \subset \mathcal{T} \times \mathcal{W}$ and neighboring sample P_n', we have

$$\mathbb{P}((M(P_n, W), W) \in A) \le e^{\varepsilon+\varepsilon_0}\mathbb{P}((M(P_n', W'), W') \in A) + \delta_0 + \delta + \gamma, \tag{29}$$

where $W' \sim \mu(\cdot \mid P_n')$ is the mechanism W on input P_n'. Define the slices $A_w = \{t \mid (t, w) \in A\}$ and projection $A^{\mathcal{W}} = \{w \mid \text{there exist } (t, w) \in A\}$, which are measurable as A is [30, Ch. 12.4]. By standard conditional probability and (dis)integration arguments [14], we have

$$\begin{aligned}
\mathbb{P}(M(P_n, W) \in A) &= \int_{A^{\mathcal{W}}} \mathbb{P}(M(P_n, w) \in A_w) d\mu(w \mid P_n) \\
&\overset{(i)}{\le} \int_{G(P_n) \cap A^{\mathcal{W}}} \mathbb{P}(M(P_n, w) \in A_w) d\mu(w \mid P_n) + \mathbb{P}(W \notin G(P_n) \mid P_n) \\
&\overset{(ii)}{\le} \int_{G(P_n) \cap A^{\mathcal{W}}} \min\left\{(e^\varepsilon \mathbb{P}(M(P_n', w) \in A_w) + \delta), 1\right\} d\mu(w \mid P_n) + \gamma \\
&\overset{(iii)}{\le} \int_{A^{\mathcal{W}}} \min\left\{e^\varepsilon \mathbb{P}(M(P_n', w) \in A_w), 1\right\} d\mu(w \mid P_n) + \delta + \gamma,
\end{aligned}$$

where inequality (i) follows because $\int f d\mu \le 1$ whenever $0 \le f \le 1$, inequality (ii) by the assumptions that $M(P_n, w)$ is (ε, δ)-differentially private when $w \in G(P_n)$ and that $\mathbb{P}(W \notin G(P_n) \mid P_n) \le \gamma$, and inequality (iii) follows because $\min\{a + b, 1\} \le \min\{a, 1\} + b$ whenever $a, b \ge 0$.

For shorthand define the (measurable) function $f(w) := \min\{e^\varepsilon \mathbb{P}(M(P_n', w) \in A_w), 1\}$, noting that $0 \le f \le 1$. Then by the definition of the integral $\int f d\mu$ as a supremum over simple functions $0 \le \varphi \le f$ (e.g. [30, Ch. 11.3]), we obtain $\int f(w) d\mu(w \mid P_n) \le e^{\varepsilon_0} \int f(w) d\mu(w \mid$

$P_n') + \delta_0$ by the assumption that W is $(\varepsilon_0, \delta_0)$-DP. Substituting above gives

$$\begin{aligned}\mathbb{P}(M(P_n, W) \in A) &\le e^{\varepsilon_0} \int_{A^{\mathcal{W}}} \min\{e^{\varepsilon} \mathbb{P}(M(P_n', w) \in A_w), 1\} d\mu(w \mid P_n') + \gamma + \delta_0 + \delta \\ &\le e^{\varepsilon_0 + \varepsilon} \int_{A^{\mathcal{W}}} \mathbb{P}(M(P_n', w) \in A_w) d\mu(w \mid P_n') + \gamma + \delta_0 + \delta \\ &= e^{\varepsilon_0 + \varepsilon} \mathbb{P}((M(P_n', W'), W') \in A) + \gamma + \delta_0 + \delta,\end{aligned}$$

which is inequality (29).

A.1.2 Proof of Lemma 2.3

Without loss of generality by translation, we assume $\mu_0 = 0$ and let $\mu = \mu_1$. Let $p_i(z) = \frac{1}{2\sigma^2} \exp(-\frac{1}{2\sigma^2} \|z - \mu_i\|_2^2)$ for $i = 0, 1$, and define the log likelihood ratio $\ell(z) = \log \frac{p_0(z)}{p_1(z)} = \frac{1}{2\sigma^2}(\|\mu\|_2^2 + 2\langle \mu, z \rangle)$. Then we have $Z_0 \stackrel{d}{=}_{\varepsilon,\delta} Z_1$ if $\mathbb{P}(|\ell(Z)| \ge \varepsilon) \le \delta$ when $Z \sim \mathsf{N}(0, \sigma^2 I)$. A bit of linear algebra and the rotational invariance of the Gaussian distribution shows that if $W \sim \mathsf{N}(0, 1)$, then

$$\begin{aligned}\mathbb{P}(|\ell(Z)| \ge \varepsilon) &= \mathbb{P}\left(\left|\frac{\|\mu\|_2^2}{2\sigma^2} + \frac{\|\mu\|_2}{\sigma} W\right| \ge \varepsilon\right) \\ &= \mathbb{P}\left(W \ge \frac{\sigma}{\|\mu\|_2}\left(\varepsilon - \frac{\|\mu\|_2^2}{2\sigma^2}\right)\right) + \mathbb{P}\left(W \le -\frac{\sigma}{\|\mu\|_2}\left(\varepsilon + \frac{\|\mu\|_2^2}{2\sigma^2}\right)\right).\end{aligned}$$

The homogeneity of $\sigma / \|\mu\|_2$ gives the result.

A.1.3 Proof of Lemma 2.5

For any set B not including $\perp$, we have that $\{M(P_n) \in B\} = \{M_0(P_n) \in A, M_1(P_n) \in B\}$. Consider two cases: in the first, we have $\lambda(P_n) \in G$. Then $M_1(P_n) \stackrel{d}{=}_{\varepsilon,\delta} M_1(P_n')$, and so standard (ε, δ)-composition gives

$$\begin{aligned}\mathbb{P}(M(P_n) \in B) &= \mathbb{P}(M_0(P_n) \in A, M_1(P_n) \in B) \\ &\le e^{\varepsilon_0 + \varepsilon} \mathbb{P}(M_0(P_n') \in A, M_1(P_n') \in B) + \delta_0 + \delta.\end{aligned}$$

In the second, $\lambda(P_n) \notin G$. Then $\mathbb{P}(M_0(P_n) \in A) \le \delta_0$, and by $(\varepsilon_0, \delta_0)$-differential privacy, we have $\mathbb{P}(M_0(P_n') \in A) \le e^{\varepsilon_0} \delta_0$, so that

$$\begin{aligned}\mathbb{P}(M(P_n) \in B) &= \mathbb{P}(M_0(P_n) \in A, M_1(P_n) \in B) \le \mathbb{P}(M_0(P_n) \in A) \le \delta_0 \\ \mathbb{P}(M(P_n') \in B) &\le \mathbb{P}(M_0(P_n') \in A) \le e^{\varepsilon_0} \delta_0,\end{aligned}$$

and so combining the two guarantees gives

$$\mathbb{P}(M(P_n) \in B) \le e^{\varepsilon_0 + \varepsilon} \mathbb{P}(M(P_n') \in B) + e^{\varepsilon_0} \delta_0 + \delta.$$

Lastly, we consider the case that B may contain $\perp$. Note that

$$\mathbb{P}(M(P_n) = \perp) = \mathbb{P}(M_0(P_n) \notin A) \le e^{\varepsilon_0} \mathbb{P}(M_0(P_n') \notin A) + \delta_0 = e^{\varepsilon_0} \mathbb{P}(M(P_n') = \perp) + \delta_0.$$

Combining this display with the preceding derivation gives the result.

A.1.4 Proof of Proposition 8.1

Recapitulating a few results on Gaussian closeness, we say random variables X and Y satisfy

$$X \stackrel{d}{=}_{\varepsilon,\delta} Y \quad \text{if} \quad \mathbb{P}(X \in A) \le e^{\varepsilon}\mathbb{P}(Y \in A) + \delta \;\text{ and }\; \mathbb{P}(Y \in A) \le e^{\varepsilon}\mathbb{P}(X \in A) + \delta$$

for all measurable A. The following lemma gives sufficient conditions for closeness of Gaussian distributions, where we recall the nuclear norm $\|A\|_* = \sum_i \sigma_i(A)$, Mahalanobis norm $\|v\|_\Sigma^2 = v^T\Sigma^{-1}v$, and use the distance-like function on positive definite matrices

$$d_{\rm pd}(A, B) = \max\left\{\left\|A^{-1/2}(B - A)A^{-1/2}\right\|_*, \left\|B^{-1/2}(A - B)B^{-1/2}\right\|_*\right\}.$$

We also recall $\sigma^2(\varepsilon, \delta)$, the variance (7) necessary for Gaussians to provide (ε, δ)-privacy.

Lemma A.1. *Let $\varepsilon, \delta > 0$, and let $X \sim \mathsf{N}(\mu_1, \Sigma_1)$ and $Y \sim \mathsf{N}(\mu_2, \Sigma_2)$. Then X and Y satisfy $X \stackrel{d}{=}_{\varepsilon,\delta} Y$ in the following cases.*

i. If $\Sigma_1 = \Sigma_2 = \sigma^2\Sigma$, where $\sigma \ge \sigma(\varepsilon, \delta)\|\mu_1 - \mu_2\|_\Sigma$.

ii. If $\mu_1 = \mu_2$ and $\varepsilon \ge 6d_{\rm pd}(\Sigma_1, \Sigma_2)\log\frac{2}{\delta}$.

iii. In one dimension if $\Sigma_1 = \sigma_1^2$, $\Sigma_2 = \sigma_2^2$, and $\mu_1 = \mu_2 = \mu \in \mathbb{R}$, then $X \stackrel{d}{=}_{\varepsilon,\delta} Y$ if $\varepsilon \ge \frac{1}{2}(1 + \Phi^{-1}(1 - \delta/2)^2)d_{\rm pd}(\sigma_1^2, \sigma_2^2)$.

Proof The first claim is a trivial modification of Lemma 2.3. For the second, see, e.g., [18, Lemmas 2.5–2.6]. For the last (in one dimension), w.l.o.g. let $\mu = 0$, let p_1, p_2 denote the densities of X and Y, and consider $X \sim \mathsf{N}(0, \sigma_1^2)$; it suffices to show that $|\log\frac{p_1(X)}{p_2(X)}| \le \varepsilon$ with probability at least $1 - \delta$. To that end, note that

$$\begin{aligned}\left|\log\frac{p_1(x)}{p_2(x)}\right| &= \left|\frac{1}{2}\log\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}x^2\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right| \\ &\le \frac{1}{2}\max\left\{\log\frac{\sigma_2^2}{\sigma_1^2}, \log\frac{\sigma_1^2}{\sigma_2^2}\right\} + \frac{1}{2}\frac{x^2}{\sigma_1^2}\left|\frac{\sigma_1^2}{\sigma_2^2} - 1\right| \\ &\le \frac{1}{2}\max\left\{\frac{\sigma_2^2}{\sigma_1^2} - 1, \frac{\sigma_1^2}{\sigma_2^2} - 1\right\} + \frac{1}{2}\frac{x^2}{\sigma_1^2}\left|\frac{\sigma_1^2}{\sigma_2^2} - 1\right|,\end{aligned}$$

where we use that $\log t = \log(1 + t - 1) \le t - 1$ for all $t \ge 0$. As $X/\sigma_1 \sim \mathsf{N}(0, 1)$, it becomes sufficient to upper bound $\frac{1}{2}d_{\rm pd}(\sigma_1^2, \sigma_2^2) + \frac{1}{2}Z^2 d_{\rm pd}(\sigma_1^2, \sigma_2^2)$ for $Z \sim \mathsf{N}(0, 1)$. But of course, we have $|Z| \le \Phi^{-1}(1 - \delta/2)$ with probability at least $1 - \delta$, giving the result. □

We leverage part iii of Lemma A.1 to prove the proposition once the ratio of the directional modulus (11) quantities $\omega(u \mid P_n)$ and $\omega(u \mid P_n')$ is bounded. Let σ^2 be as in the statement of the proposition, and define random variables

$$\begin{aligned}Z_0 \sim \mathsf{N}(u^T\theta(P_n), \sigma^2 \cdot \omega(u \mid P_n)^2), \quad Z_1 \sim \mathsf{N}(u^T\theta(P_n'), \sigma^2 \cdot \omega(u \mid P_n)^2), \\ Z_2 \sim \mathsf{N}(u^T\theta(P_n'), \sigma^2 \cdot \omega(u \mid P_n')^2).\end{aligned}$$

We know that because $\widehat{\lambda} \le \lambda_{\min}(P_n)$, we have $|u^T\theta(P_n) - u^T\theta(P_n')| \le \omega(u \mid P_n)$, and so

$$Z_0 \stackrel{d}{=}_{\varepsilon,\delta} Z_1$$

by Lemma A.1. By assumption on the ratio (26), we have $d_{\rm pd}(\omega(u \mid P_n)^2, \omega(u \mid P_n')^2) \le \mathtt{r}^2(u, \widehat{\lambda})$, and applying Lemma A.1.iii,

$$Z_1 \stackrel{d}{=}_{\varepsilon,\delta} Z_2$$

so long as $\varepsilon \ge \frac{1}{2}(1 + \Phi^{-1}(1-\delta/2)^2)\mathtt{r}^2(u, \widehat{\lambda})$. By standard composition guarantees, we thus have $Z_0 \stackrel{d}{=}_{2\varepsilon, \delta + e^\varepsilon \delta} Z_2$. Applying Lemma 2.5 gives the proposition.

A.2 Proof of Lemma 2.1

We prove each statement in the lemma in turn.

(i) Define $g(t) = \log f''(t)$. Then $g'(t) = \frac{f'''(t)}{f''(t)}$, so that $|g(t+s) - g(t)| = |\int_t^{t+s} g'(u)du| \le c|s|$, giving the inequality. The choices of φ are immediate.

(ii) This is standard [9, Eq. (9.46)]. Without loss of generality, let $t = 0$. Self-concordance is equivalent to the statement that

$$\left|\frac{d}{ds}\left(f''(s)^{-1/2}\right)\right| \le 1, \quad \text{as} \quad \frac{d}{ds}\left(f''(s)^{-1/2}\right) = \frac{1}{2}\frac{f'''(s)}{f''(s)^{3/2}},$$

and the latter quantity has magnitude at most 1 if and only if f is self-concordant. Assuming w.l.o.g. that $s \ge 0$, then integrating from 0 to s then yields

$$-s \le \frac{1}{\sqrt{h''(s)}} - \frac{1}{\sqrt{h''(0)}} \le s.$$

Solving the upper and lower bounds gives claim (ii).

(iii) This is immediate from the upper bound of part (ii).

A.3 Proofs about recursions

A.3.1 Proof of Lemma 6.2

That each satisfies $R(\lambda) \le \lambda$ is immediate by convexity: we have $\sqrt{1-\delta} \le 1 - \delta/2$ and $\exp(\delta) \ge 1 + \delta$, respectively.

For $R(\lambda) = \frac{\lambda}{2} + \frac{1}{2}\sqrt{\lambda^2 - a} - b$, we observe that $R'(\lambda) = \frac{1}{2} + \frac{\lambda}{2\sqrt{\lambda^2 - a}} \ge 1$ for all $\lambda \ge a$.

Now we show that for any $\lambda_0 \ge 0$,

$$R(\lambda) := \lambda\left[2 - \exp\left(b\left(1 - \sqrt{1 - \frac{a}{\lambda + \lambda_0}}\right)\right)\right] - c$$

is an accelerating recursion, for which it suffices to show that $R'(\lambda) \ge 1$ for all $\lambda + \lambda_0 \ge a$. Taking derivatives and using that $\frac{\partial}{\partial\lambda} b(1 - \sqrt{1 - a/(\lambda+\lambda_0)}) = -\frac{ab}{2(\lambda+\lambda_0)^2\sqrt{1-a/(\lambda+\lambda_0)}}$, we have

$$\begin{aligned}
&R'(\lambda) \\
&= 2 - \exp\left(b(1 - \sqrt{1 - a/(\lambda+\lambda_0)})\right) + \exp\left(b(1 - \sqrt{1 - a/(\lambda+\lambda_0)})\right)\frac{ba}{2(\lambda+\lambda_0)\sqrt{1-a/(\lambda+\lambda_0)}} \\
&= 2 + \exp\left(b(1 - \sqrt{1 - a/(\lambda+\lambda_0)})\right)\left[\frac{ba}{2(\lambda+\lambda_0)\sqrt{1-a/(\lambda+\lambda_0)}} - 1\right].
\end{aligned}$$

Let $\delta = \frac{a}{\lambda+\lambda_0} < 1$ (as $\lambda + \lambda_0 > a$). Then $R'(\lambda) \geq 1$ if and only if

$$\frac{b\delta}{2\sqrt{1-\delta}} - 1 \geq -\exp\left(-b(1-\sqrt{1-\delta})\right) \quad \text{if and only if}$$
$$\exp\left(-b(1-\sqrt{1-\delta})\right) + \frac{b\delta}{2\sqrt{1-\delta}} \geq 1.$$

At $\delta = 0$ this inequality trivially holds. Define $f(\delta) = \exp(-b + b\sqrt{1-\delta})) + \frac{b\delta}{2\sqrt{1-\delta}}$. Then

$$f'(\delta) = \exp\left(-b(1-\sqrt{1-\delta})\right)\left(\frac{-b}{2\sqrt{1-\delta}}\right) + \frac{b}{2\sqrt{1-\delta}} + \frac{b\delta}{4(1-\delta)^{3/2}} > \frac{b\delta}{4(1-\delta)^{3/2}} \geq 0,$$

so $f(\delta) \geq 1$ for all $\delta = \frac{a}{\lambda} \in [0,1]$, and $R'(\lambda) \geq 1$.

A.4 Proofs about matrix ratios

We consider a few technical results that form useful building blocks for many of our results. Throughout, we let $\mathbf{S}^d = \{A \in \mathbb{R}^{d\times d} \mid A = A^T\}$ denote the symmetric matrices. The basic form of results in this section is as follows: for a (symmetric, positive definite) matrix X belonging to a set $\mathcal{C}$ of PSD matrices, we wish to provide bounds on quantities of the form

$$\sup_{X\in\mathcal{C}} \langle u, Xv\rangle \quad \text{and} \quad \sup_{X\in\mathcal{C}} \|Xu\| \tag{30}$$

where u and v are given vectors. For a symmetric matrix $A \in \mathbf{S}^d$, we let $[A]_+$ be its Euclidean projection onto the positive semidefinite matrices, so that if $A = U\Lambda U^T$ with $\Lambda = \operatorname{diag}(\lambda)$, then $[A]_+ = U[\Lambda]_+ U^T = U\operatorname{diag}([\lambda]_+)U^T$. Similarly, we let $[A]_- = -[-A]_+$ be the Euclidean projection of A onto the negative semidefinite matrices, or its negative semidefinite part, so that $A = [A]_+ - [-A]_+ = [A]_+ + [A]_-$.

A.4.1 Suprema of matrix inner products with semidefinite box constraints

Our main focus is on situations where the set $\mathcal{C}$ is of the form $\mathcal{C} = \{X \in \mathbf{S}^d \mid A \preceq X \preceq B\}$. In this case, the following lemma provides guidance in the solution of the first problem in (30). In the lemma, we say that a matrix X is invariant in the eigenspaces of Y and Z if for the spectral decompositions $Y = U\Lambda U^T$ and $Z = VDV^T$, where we include only the nonzero eigenvalues in Λ and D, we have $XUU^T = UU^TXUU^T$ and $XVV^T = VV^TXVV^T$.

Lemma A.2. *Let $A \preceq B$ be symmetric matrices and $C \in \mathbb{R}^{d\times d}$. Then*

$$\sup_{A\preceq X\preceq B} \operatorname{tr}(XC) = \inf\left\{\frac{1}{2}\langle B, C_+\rangle - \frac{1}{2}\langle A, C_-\rangle \mid C_+ \succeq 0, C_- \succeq 0, \frac{1}{2}(C + C^T) = C_+ - C_-\right\}.$$

Additionally,

$$\sup_{A\preceq X\preceq B} \operatorname{tr}(XC) \leq \frac{1}{2}\langle B, \left[C + C^T\right]_+\rangle + \frac{1}{2}\langle A, \left[C + C^T\right]_-\rangle,$$

and equality holds if A and B are invariant in the eigenspaces of $\left[C + C^T\right]_+$ and $\left[C + C^T\right]_-$.

Proof Without loss of generality assume that $C = C^T$, because $\operatorname{tr}(XC) = \operatorname{tr}(XC^T)$ for X symmetric; otherwise we simply replace C with its symmetrization $\frac{1}{2}(C + C^T)$. Introduce Lagrange multipliers $Y, Z \succeq 0$ for the constraints $A \preceq X$ and $X \preceq B$, respectively. Then

$$\mathcal{L}(X, Y, Z) = \langle X, C\rangle + \langle Y, X - A\rangle + \langle Z, B - X\rangle$$

satisfies

$$\sup_{X \in \mathbf{S}^d} \mathcal{L}(X, Y, Z) = \begin{cases} \langle B, Z \rangle - \langle A, Y \rangle & \text{if } C = Z - Y \\ +\infty & \text{otherwise.} \end{cases}$$

As $Y \succeq 0$, $Z \succeq 0$, this gives associated dual problem

$$\begin{array}{ll} \text{minimize} & \langle B, Z \rangle - \langle A, Y \rangle \\ \text{subject to} & Z \succeq 0, Y \succeq 0, C = Z - Y. \end{array} \tag{31}$$

There exist $Z, Y \succ 0$ satisfying the constraints of the dual—so that Slater's condition holds—and strong duality obtains for the problem (31). The first claim of the lemma follows.

Certainly the choices $Z = [C]_+$ and $Y = -[C]_-$ are feasible for the dual, giving the second claim of the lemma, though they may be suboptimal. For the special case attaining equality, let $C = U \Lambda U^T = C_+ - C_-$, where $C_+ \succeq 0$ and $C_- \succeq 0$ decompose C into its positive and negative eigenvalues. Let U_+ and U_- be the eigenvectors associated with the positive and negative eigenvalues of C, and let $\Pi_+ = U_+ U_+^T$ and $\Pi_- = U_- U_-^T$ be the associated projection matrices, and let Π_0 be the orthogonal projector to the null space of $\Pi_+ + \Pi_-$. If A and B are invariant in these eigenspaces, in that $A\Pi_+ = \Pi_+ A \Pi_+$ and $A\Pi_- = \Pi_- A \Pi_-$ (and similarly for B), then the eigenspace invariances imply that

$$X = \Pi_+ B \Pi_+ + \frac{1}{2} \Pi_0 (A + B) \Pi_0 + \Pi_- A \Pi_-$$

satisfies $A \preceq X \preceq B$ because we can write $A = \Pi_+ A \Pi_+ + \Pi_0 A \Pi_0 + \Pi_- A \Pi_-$, and similarly for B. Notably, the choices $Z = [C]_+$ and $Y = -[C]_-$ are feasible in the dual (31), and

$$\langle C, X \rangle = \langle \Pi_+ B \Pi_+, C \rangle + \langle \Pi_- A \Pi_-, C \rangle = \langle B, [C]_+ \rangle + \langle A, [C]_- \rangle = \langle B, Z \rangle - \langle A, Y \rangle,$$

showing equality in the dual. □

Lemma A.3. *Let $A \preceq B$ be symmetric and u, v be vectors. Then*

$$\sup_{A \preceq X \preceq B} u^T X v \leq \frac{1}{4} \left[\langle B, (u+v)(u+v)^T \rangle - \langle A, (u-v)(u-v)^T \rangle \right].$$

Proof Note that

$$\frac{1}{2}(uv^T + vu^T) = \frac{1}{4} \left[(u+v)(u+v)^T - (u-v)(u-v)^T \right],$$

a difference of the positive semidefinite matrices $(u+v)(u+v)^T$ and $(u-v)(u-v)^T$. Apply Lemma A.2. □

As a consequence of Lemma A.3, we have the following inequality.

Lemma A.4. *Let $A \preceq B$ be symmetric and u, v be vectors. Define $u_0 = u / \|u\|_2$ and $v_0 = v / \|v\|_2$. Then*

$$\sup_{A \preceq X \preceq B} u^T X v \leq \|u\|_2 \|v\|_2 \frac{1}{4} \left[\langle B, (u_0 + v_0)(u_0 + v_0)^T \rangle - \langle A, (u_0 - v_0)(u_0 - v_0)^T \right].$$

If $u_0 + v_0$ and $u_0 - v_0$ are eigenvectors of A and B, then equality holds.

Proof By homogeneity we have

$$u^T X v = \|u\|_2 \|v\|_2 (u/\|u\|_2)^T X (v/\|v\|_2) = \|u\|_2 \|v\|_2 u_0^T X v_0.$$

Now apply Lemma A.3, but note that as $\|u_0\|_2 = \|v_0\|_2 = 1$, we have $(u_0 + v_0)^T(u_0 - v_0) = 1 - 1 = 0$, so that Lemma A.2 gives the result as $u_0 v_0^T + v_0 u_0^T$ has eigedecomposition $\frac{1}{2}(u_0 + v_0)(u_0 + v_0)^T - \frac{1}{2}(u_0 - v_0)(u_0 - v_0)^T$.

The equality claim similarly follows from Lemma A.2. □

A.4.2 Proof of Lemma 8.3

We have

$$\begin{aligned}\left\|H_1^{-1}u\right\|_2 &\le \left\|H_0^{-1}u\right\|_2 + \|E_1 u\|_2 + \|E_2 u\|_2 \\ &\le \left\|H_0^{-1}u\right\|_2 + \|\!|E_1 H_0|\!\|_{\mathrm{op}} \left\|H_0^{-1}u\right\|_2 + \|\!|E_2 H_0|\!\|_{\mathrm{op}} \left\|H_0^{-1}u\right\|_2.\end{aligned}$$

We control the two error terms $\|\!|E_1 H_0|\!\|_{\mathrm{op}}$ and $\|\!|E_2 H_0|\!\|_{\mathrm{op}}$ in turn. Let w and v be unit vectors. Then normalizing by $\|H_0 v\|_2$, Lemma A.4 gives

$$\sup_{-s_1 H_0^{-1} \preceq E_1 \preceq s_1 H_0^{-1}} \frac{w^T E_1 H_0 v}{\|H_0 v\|_2} \le \frac{s_1}{2}\left[\langle w, H_0^{-1} w\rangle + \frac{1}{\|H_0 v\|_2^2}\langle v, H_0 v\rangle\right],$$

so that

$$\sup_{\|w\|_2=1, \|v\|_2=1} w^T E_1 H_0 v \le \sup_{\|w\|_2=\|v\|_2=1} \frac{\|H_0 v\|_2 s_1}{2}\left[\langle w, H_0^{-1} w\rangle + \frac{1}{\|H_0 v\|_2^2}\langle v, H_0 v\rangle\right].$$

Note that $\|H_0 v\|_2 \le \|\!|H_0|\!\|_{\mathrm{op}}$ and $\langle H_0 v, v\rangle / \|H_0 v\|_2 \le 1$, so taking a supremum over w yields

$$\sup_{\|w\|_2=1, \|v\|_2=1} w^T E_1 H_0 v = \|\!|E_1 H_0|\!\|_{\mathrm{op}} \le \frac{s_1}{2}\left(\|\!|H_0^{-1}|\!\|_{\mathrm{op}} \|\!|H_0|\!\|_{\mathrm{op}} + 1\right) = \frac{s_1}{2}\left(1 + \frac{\lambda_{\max}(H_0)}{\lambda_{\min}(H_0)}\right).$$

For the term involving $\|\!|E_2 H_0|\!\|_{\mathrm{op}}$, the naive bound

$$\sup_{-\Delta_0 \preceq E \preceq \Delta_1} \|\!|E H_0|\!\|_{\mathrm{op}} \le \max\{\|\!|\Delta_1|\!\|_{\mathrm{op}}, \|\!|\Delta_0|\!\|_{\mathrm{op}}\} \|\!|H_0|\!\|_{\mathrm{op}} \le \sup_{x \in \mathcal{X}} \left\|H_0^{-1}x\right\|_2^2 \lambda_{\max}(H_0)$$

suffices.

The proof of the lower bound is, *mutatis mutandis*, identical.

A.4.3 Proof of Lemma 8.4

For any unit vector u, we have $\sup_{v \in \mathcal{V}} u^T A v \ge \mathrm{ins}_2(\mathcal{V}) \|Au\|_2 \ge \lambda_{\min}(A)$. Simultaneously we have $\sup_{v \in \mathcal{V}} v^T A v / \|v\|_2 \le \sup_{v \in \mathcal{V}} \|Av\|_2 \le \mathrm{rad}_2(\mathcal{V}) \lambda_{\max}(A)$. This yields the claimed upper bound.

For the equalities and near equalities, note that if $\mathcal{V} = \mathbb{B}_2^d$, then $\sup_{v \in \mathcal{V}} v^T A v = \lambda_{\max}(A)$, while $\inf_{\|u\|_2=1} \sup_{v \in \mathcal{V}} u^T A v = \inf_u \|Au\|_2 = \lambda_{\min}(A)$. When $\mathcal{V} = [-1, 1]^d$, take the matrix $A = \frac{1}{d}\mathbf{1}\mathbf{1}^T + \lambda(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T)$. Then A has one eigenvalue 1 associated to the eigenvector $\mathbf{1}/\sqrt{d}$, and the rest are all λ, so that $\kappa(A) = 1/\lambda$. Then $v^T A v = \frac{1-\lambda}{d}\langle \mathbf{1}, v\rangle^2 + \lambda \|v\|_2^2$, and the supremum is achieved by $v = \mathbf{1}$, yielding $\sup_v v^T A v / \|v\|_2 = \sqrt{d}$. Now, take $u = (e_1 - e_2)/\sqrt{2}$. Then $\sup_{v \in \mathcal{V}} u^T A v = \|Au\|_1 = \lambda \|u\|_1 = \lambda\sqrt{2}$. Thus we have $\kappa(A, [-1, 1]^d) \ge \sqrt{d}/(\lambda\sqrt{2}) = \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})}\kappa(A)/\sqrt{2}$.

A.4.4 Proof of Lemma 8.5

Fix $v \in \mathcal{V}$. We control each of the error terms in the expansion

$$u^T H_1^{-1} v = u^T H_0^{-1} v + u^T E_1 v + u^T E_2 v.$$

For the first, we use Lemma A.4: for $v_0 = v/\|v\|_2$, we have

$$\begin{aligned}\sup_{-s_1 H_0^{-1} \preceq E_1 \preceq s_1 H_0^{-1}} u^T E_1 v &\leq \frac{\|v\|_2}{4}\left[\langle s_1 H_0^{-1}, (u+v_0)(u+v_0)^T + (u-v_0)(u-v_0)^T\rangle\right] \\ &= \frac{s_1 \|v\|_2}{2}\langle H_0^{-1}, uu^T + v_0 v_0^T\rangle.\end{aligned}$$

By Lemma 8.4, we have

$$v_0^T H_0^{-1} v \leq \kappa(H_0^{-1}, \mathcal{V}) \cdot \sup_{v \in \mathcal{V}} u^T H_0^{-1} v \leq \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})} \kappa(H_0) \cdot \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

Noting that the set $\{v/\mathrm{ins}_2(\mathcal{V}) \mid v \in \mathcal{V}\} \supset \mathbb{B}_2^d$, we have

$$u^T H_0^{-1} u \|v\|_2 \leq \|v\|_2 \cdot \sup_{v \in \mathcal{V}} u^T H_0^{-1} v/\mathrm{ins}_2(\mathcal{V}) \leq \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})} \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

Combining the preceding inequalities then gives that

$$u^T E_1 v \leq s_1 \frac{\mathrm{rad}_2(\mathcal{V})}{\mathrm{ins}_2(\mathcal{V})} \left(\frac{\kappa(H_0)+1}{2}\right) \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

We now control the second error term. Noting that $x_0 x_0^T + x_1 x_1^T \succeq x_0 x_0^T$ and $x_0 x_0^T + x_1 x_1^T \succeq x_0 x_0^T$ we have (again applying Lemma A.4)

$$\begin{aligned}\sup_{E_2} u^T E_2 v &\leq \frac{s_2 \|v\|_2}{2}\langle H_0^{-1}(x_0 x_0^T + x_1 x_1^T) H_0^{-1}, uu^T + v_0 v_0^T\rangle \\ &\leq s_2 \|v\|_2 \left[\sup_{x \in \mathcal{X}} \langle H_0^{-1} x x^T H_0^{-1}, uu^T\rangle + \sup_{x \in \mathcal{X}} \langle H_0^{-1} x x^T H_0^{-1}, v_0 v_0^T\rangle\right] \\ &= s_2 \|v\|_2 \left[\sup_{x \in \mathcal{X}} (u^T H_0^{-1} x)^2 + \sup_{x \in \mathcal{X}} (v_0^T H_0^{-1} x)^2\right]\end{aligned}$$

Now we use that $\mathcal{V}$ coincides with a scaled multiple of $\mathcal{X}$ to obtain

$$\sup_{\|u\|_2 = 1} \sup_{x \in \mathcal{X}} u^T H_0^{-1} x = c^{-1} \sup_{\|u\|_2 = 1} \sup_{v \in \mathcal{V}} u^T H_0^{-1} v \leq \frac{\mathrm{rad}_2(\mathcal{V})}{c \lambda_{\min}(H_0)},$$

and thus

$$u^T E_2 v \leq s_2 \cdot \frac{2\mathrm{rad}_2^2(\mathcal{V})}{c^2 \lambda_{\min}(H_0)} \sup_{v \in \mathcal{V}} u^T H_0^{-1} v.$$

Combining the inequalities gives the lemma. The lower bound calculation is completely similar.

To see the sharpness conditions, take $H_0 = I$ and $\mathcal{V} = [-1, 1]^d$. Then the constraint $-sI \preceq E \preceq sI$ is equivalent to the constraint that $\|\!|E|\!\|_{\rm op} \le s$, and so using Lemma A.4 with $v_0 = v/\left\|v\right\|_2$, we have

$$\sup_{\|\!|E|\!\|_{\rm op} \le s} u^T E v = \frac{s\left\|v\right\|_2}{2} \langle I, uu^T + v_0 v_0^T \rangle = s\left\|v\right\|_2,$$

because $u - v_0$ and $u + v_0$ are certainly eigenvectors of I and $\langle I, v_0 v_0^T \rangle = \langle I, uu^T \rangle = 1$ for ℓ_2-unit vectors u, v_0. Note also that $\mathrm{rad}_2(\mathcal{V})/\mathrm{ins}_2(\mathcal{V}) = \sup_{v \in \mathcal{V}} \left\|v\right\|_2 = \sqrt{d}$. Taking u to be any standard basis vector and $v = \mathbf{1}$ then yields $u^T H_0^{-1} v = u^T v = 1 = \sup_{v \in \mathcal{V}} u^T v$. In particular, we can choose E_1 such that

$$u^T(H_0^{-1} + E_1)v = u^T H_0^{-1} v + s_1 \left\|v\right\|_2 = (u^T v)(1 + s_1\sqrt{d}) = \sup_{v \in \mathcal{V}} (u^T v)(1 + s_1\sqrt{d}).$$

To control the second error term involving $u^T E_2 v$ take $x = v = \mathbf{1}$ and u to be any vector with nonnegative entries, so that $u^T E_2 v = u^T \mathbf{1} d = d \cdot \sup_{\left\|v\right\|_\infty \le 1} u^T v$. Thus, we have exhibited error matrices E_1 and E_2, when $H_0 = I$, with $-s_1 H_0^{-1} \preceq E_1 \preceq s_1 H_0^{-1}$ and $-s_2 xx^T \preceq E_2 \preceq s_2 xx^T$ such that

$$\sup_{v \in \mathcal{V}} u^T H_1^{-1} v \ge \left(1 + s_1\sqrt{d} + s_2 d\right) \sup_{v \in \mathcal{V}} u^T H_0^{-1} v$$

whenever u is a standard basis vector. As $\mathrm{rad}_2^2(\mathcal{V}) = d$, $\mathrm{ins}_2(\mathcal{V}) = 1$, and $\kappa(H_0) = 1$ in this case, the proof is complete.

References

[1] Sivakanth gopi and yin tat lee and lukas wutschitz. In *Advances in Neural Information Processing Systems 34*, 2021.

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

[3] H. Asi and J. C. Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. In *Advances in Neural Information Processing Systems 33*, 2020.

[4] R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv:1412.4451 [math.ST]*, 2014.

[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] R. Bassily, V. Feldman, K. Talwar, and A. Thakurta. Private stochastic convex optimization with optimal rates. In *Advances in Neural Information Processing Systems 32*, 2019.

[7] P. Bickel, C. A. J. Klaassen, Y. Ritov, and J. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer Verlag, 1998.

[8] S. Biswas, Y. Dong, G. Kamath, and J. Ullman. CoinPress: Practical private mean and covariance estimation. In *Advances in Neural Information Processing Systems 33*, 2020.

[9] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[10] G. Brown, M. Gaboardi, A. Smith, J. Ullman, and L. Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. *arXiv:2106.13329 [cs.LG]*, 2021.

[11] G. Brown, S. B. Hopkins, and A. Smith. Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. In *Proceedings of the Thirty Sixth Annual Conference on Computational Learning Theory*, 2023.

[12] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy: optimal rates of convergence for parameter estimation with differential privacy. *Annals of Statistics*, 49(5):2825–2850, 2021.

[13] K. Chadha, J. Duchi, and R. Kuditipudi. Resampling methods for private statistical inference. *arXiv:2402.07131 [stat.ML]*, 2024.

[14] J. Chang and D. Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3): 287–317, 1997.

[15] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

[16] L. H. Dicker. Variance estimation in high-dimensional linear models. *Biometrika*, 101 (2):269–284.

[17] F. Ding, M. Hardt, J. Miller, and L. Schmidt. Retiring Adult: New datasets for fair machine learning. *Advances in Neural Information Processing Systems 34*, 2021.

[18] J. Duchi, S. Haque, and R. Kuditipudi. A fast algorithm for adaptive private mean estimation. In *Proceedings of the Thirty Sixth Annual Conference on Computational Learning Theory*, 2023. URL https://arXiv.org/abs/2301.07078.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. John Wiley and Sons, second edition, 2009.

[25] G. Imbens and D. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[26] Y. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, 1994.

[27] 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.

[28] D. Ostrovskii and F. Bach. Finite-sample analysis of M-estimators using self-concordance. *Electronic Journal of Statistics*, 15:326–391, 2021.

[29] R. Redberg, A. Koskela, and Y.-X. Wang. Improving the privacy and practicality of objective perturbation for differentially private linear learners. In *Advances in Neural Information Processing Systems 36*, 2023.

[30] H. Royden. *Real Analysis*. Pearson, third edition, 1988.

[31] T. Steinke and J. Ullman. Between pure and approximate differential privacy. In *Proceedings of the Twenty Eighth Annual Conference on Computational Learning Theory*, 2015.

[32] M. van der Laan and S. Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, 2011.

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

[34] A. W. van der Vaart. Semiparametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour)*, pages 331–457. Springer, 2002.

[35] R. Vershynin. *High Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2019.