Private Geometric Median in Nearly-Linear Time

Syamantak Kumar* Daogao Liu† Kevin Tian‡ Chutong Yang§

Abstract

Estimating the geometric median of a dataset is a robust counterpart to mean estimation, and is a fundamental problem in computational geometry. Recently, [HSU24] gave an (ϵ, δ)-differentially private algorithm obtaining an α-multiplicative approximation to the geometric median objective, $\frac{1}{n}\sum_{i\in[n]} \|\cdot - \mathbf{x}_i\|$, given a dataset $\mathcal{D} := \{\mathbf{x}_i\}_{i\in[n]} \subset \mathbb{R}^d$. Their algorithm requires $n \gtrsim \sqrt{d} \cdot \frac{1}{\alpha\epsilon}$ samples, which they prove is information-theoretically optimal. This result is surprising because its error scales with the *effective radius* of $\mathcal{D}$ (i.e., of a ball capturing most points), rather than the worst-case radius. We give an improved algorithm that obtains the same approximation quality, also using $n \gtrsim \sqrt{d} \cdot \frac{1}{\alpha\epsilon}$ samples, but in time $\widetilde{O}(nd + \frac{d}{\alpha^2})$. Our runtime is nearly-linear, plus the cost of the cheapest non-private first-order method due to [CLM+16]. To achieve our results, we use subsampling and geometric aggregation tools inspired by FriendlyCore [TCK+22] to speed up the "warm start" component of the [HSU24] algorithm, combined with a careful custom analysis of DP-SGD's sensitivity for the geometric median objective.

1 Introduction

The *geometric median* problem, also known as the Fermat-Weber problem, is one of the oldest problems in computational geometry. In this problem, we are given a dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i\in[n]} \subset \mathbb{R}^d$, and our goal is to find a point $\mathbf{x}_\star \in \mathbb{R}^d$ that minimizes the average Euclidean distance to points in the dataset:

$$\mathbf{x}_\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^d} f_{\mathcal{D}}(\mathbf{x}), \text{ where } f_{\mathcal{D}}(\mathbf{x}) := \frac{1}{n}\sum_{i\in[n]} \|\mathbf{x} - \mathbf{x}_i\|. \tag{1}$$

This problem has received widespread interest due to its applications in high-dimensional statistics. In particular, the geometric median of a dataset $\mathcal{D}$ enjoys robustness properties that the mean (i.e., $\frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i$, the minimizer of $\frac{1}{n}\sum_{i\in[n]} \|\mathbf{x} - \mathbf{x}_i\|^2$) does not. For example, it is known (cf. Lemma 24, [CLM+16]) that if greater than half of $\mathcal{D}$ lies within a distance r of some $\bar{\mathbf{x}} \in \mathbb{R}^d$, then the geometric median lies within $O(r)$ of $\bar{\mathbf{x}}$. Thus, the geometric median provides strong estimation guarantees even when $\mathcal{D}$ contains outliers. This is in contrast to simpler estimators such as the mean, which can be arbitrarily corrupted by a single outlier. As a result, studying the properties and computational aspects of the geometric median has a long history, see e.g., [Web29, LR91] for some famous examples.

In this paper, we provide improved algorithms for estimating (1) subject to (ϵ, δ)-differential privacy, the de facto notion of provable privacy in modern machine learning. Privately computing the geometric median naturally fits into a recent line of work on designing DP algorithms in the presence of outliers. To explain the challenge of such problems, the definition of DP implies that the privacy-preserving guarantee must hold for *worst-case* datasets. This stringent definition affords DP a variety of desirable properties, most notably *composition* of private mechanisms (cf. [DR14], Section 3.5). However, it also begets challenges: for example, estimating the empirical mean of $\mathcal{D}$ subject to (ϵ, δ)-DP necessarily results in error scaling $\propto R$, the diameter of the dataset (cf. Section 5, [BST14]). Moreover, the worst-case nature of DP is at odds with typical *average-case* machine learning settings, where most (or all) of $\mathcal{D}$ is drawn from a distribution that we wish to learn about. From an algorithm design standpoint, the question

*University of Texas at Austin, syamantak@utexas.edu
†Google Research, liudaogao@gmail.com
‡University of Texas at Austin, kjtian@cs.utexas.edu
§University of Texas at Austin, cyang98@utexas.edu

follows: how do we design methods that provide privacy guarantees for worst-case data, but also yield improved utility guarantees for (mostly) average-case data?

Such questions have been successfully addressed for various statistical tasks in recent work, including parameter estimation [BD14, KV17, BKSW19, DFM+20, BDKU20, BGS+21, AL22, LKJO22, KDH23, BHS23], clustering [NRS07, NSV16, CKM+21, TCK+22], and more. However, existing approaches for estimating (1) (even non-privately) are based on iterative optimization methods, as the geometric median does not admit a simple, closed-form solution. Much of the DP optimization toolkit is exactly plagued by the aforementioned "worst-case sensitivity" issues, e.g., lower bounds for general stochastic optimization problems again scale with the domain size. This is troubling in the context of (1), because a major appeal of the geometric median is its robustness: its error should not be significantly affected by any small subset of the data. Privately estimating the geometric median thus poses an interesting technical challenge, beyond its potential appeal as a subroutine in downstream robust algorithms.

To explain the distinction between worst-case and average-case error rates in the context of (1), we introduce the following helpful notation: for all quantiles $\tau \in [0, 1]$, we let

$$r^{(\tau)} := \arg\min_{r \geq 0} \left\{ \sum_{i \in [n]} \mathbb{I}_{\|\mathbf{x}_i - \mathbf{x}_\star\| \leq r} \geq \tau n \right\}, \text{ where } \mathbf{x}_\star := \arg\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{n} \sum_{i \in [n]} \|\mathbf{x} - \mathbf{x}_i\|, \tag{2}$$

when $\mathcal{D} = \{\mathbf{x}_i\}_{i \in [n]} \subset \mathbb{R}^d$ is clear from context. In other words, $r^{(\tau)}$ is the smallest radius describing a ball around the geometric median $\mathbf{x}_\star$ containing at least τn points in $\mathcal{D}$. We also use R to denote an a priori overall domain size bound, where we are guaranteed that $\mathcal{D} \subset \mathbb{B}^d(R)$. Note that in general, it is possible for, e.g., $r^{(0.9)} \ll R$ if $\approx 10\%$ of $\mathcal{D}$ consists of outliers with atypical norms. Due to the robust nature of the geometric median (i.e., the aforementioned Lemma 34, [CLM+16]), a natural target is estimation error scaling with the "effective radius" $r^{(\tau)}$ for some quantile $\tau \in (0.5, 1)$. This is a much stronger guarantee than the error rates $\propto R$ that typical DP optimization methods give.

A simple argument based on Markov's inequality shows $r^{(\tau)} = O(f_{\mathcal{D}}(\mathbf{x}_\star))$ for all $\tau < 1$. Thus, in this introduction our goal will be to approximate the minimizer of (1) to additive error $\alpha f_{\mathcal{D}}(\mathbf{x}_\star)$ for some $\alpha \in (0, 1)$, i.e., to give *α-multiplicative error* guarantees on optimizing $f_{\mathcal{D}}$.[1] Again, datasets with outliers may have $f_{\mathcal{D}}(\mathbf{x}_\star) \ll R$, so this goal is beyond the reach of naïvely applying DP optimization methods.

In a recent exciting work, [HSU24] bypassed this obstacle and obtained such private multiplicative approximations to the geometric median, and with near-optimal sample complexity. Assuming that $\mathcal{D}$ has size $n \gtrsim \sqrt{d} \cdot \frac{1}{\alpha\epsilon}$,[2] [HSU24] gave two algorithms for estimating (1) to α-multiplicative error. They also proved a matching lower bound, showing that this many samples is information-theoretically necessary.[3] From both a theoretical and practical perspective, the main outstanding question left by [HSU24] is that of computational efficiency: in particular, the [HSU24] algorithms ran in time $\widetilde{O}(n^2 d + n^3 \epsilon^2)$ or $\widetilde{O}(n^2 d + n d^2 + d^{4.372})$. This leaves a significant gap between algorithms for privately solving (1), and their counterparts in the non-private setting, where [CLM+16] showed that (1) could be approximated to α-multiplicative error in nearly-linear time $\widetilde{O}(\min(nd, \frac{d}{\alpha^2}))$.

1.1 Our results

Our main contribution is a faster algorithm for privately approximating (1) to α-multiplicative error.

Theorem 1. *Let $\mathcal{D} = \{\mathbf{x}_i\}_{i \in [n]} \subset \mathbb{B}^d(R)$ for $R > 0$, $0 < r \leq r^{(0.9)}$, and $(\alpha, \epsilon, \delta) \in [0, 1]^3$. There is an (ϵ, δ)-DP algorithm that returns $\hat{\mathbf{x}}$ such that with probability $\geq 1 - \delta$, $f_{\mathcal{D}}(\hat{\mathbf{x}}) \leq (1 + \alpha) f_{\mathcal{D}}(\mathbf{x}_\star)$, assuming $n \gtrsim \frac{\sqrt{d}}{\alpha\epsilon}$. The algorithm runs in time $\widetilde{O}(nd + \frac{d}{\alpha^2})$.*

To briefly explain Theorem 1's statement, it uses a priori knowledge of $0 < r < R$ such that R upper bounds the domain size of $\mathcal{D}$, and r lower bounds the "effective radius" $r^{(0.9)}$. However, its runtime only

[1] Our results, as well as those of [HSU24], in fact give stronger additive error bounds of $\alpha r^{(\tau)}$ for any fixed $\tau \in (0.5, 1)$.

[2] In this introduction only, we use $\widetilde{O}, \lesssim, \gtrsim$ to hide polylogarithmic factors in problem parameters, i.e., d, $\frac{1}{\alpha}$, $\frac{1}{\epsilon}$, $\frac{1}{\delta}$, and $\frac{R}{r}$, where $\mathcal{D} \subseteq \mathbb{B}^d(R)$ and $r \leq r^{(0.9)}$.

[3] Intuitively, we require $\alpha \approx d^{-1/2}$ to obtain nontrivial mean estimation when $\mathcal{D}$ consists of i.i.d. Gaussian data (as a typical radius is $\approx \sqrt{d}$), matching known sample complexity lower bounds of $\approx \frac{d}{\epsilon}$ for Gaussian mean estimation [KLSU19].

depends polylogarithmically on the aspect ratio $\frac{R}{r}$, rather than polynomially (as naïve DP optimization methods would); we also remark that our sample complexity is independent of $\frac{R}{r}$.

The runtime of Theorem 1 is nearly-linear in the regime $n \gtrsim \frac{1}{\alpha^2}$ (e.g., if $\sqrt{d} \cdot \frac{1}{\epsilon} \gtrsim \frac{1}{\alpha}$), but more generally it does incur an additive overhead of $\frac{d}{\alpha^2}$. This overhead matches the fastest non-private first-order method for approximating (1) to α-multiplicative error, due to [CLM+16]. We note that [CLM+16] also gave a custom second-order interior-point method, that non-privately solves (1) in time $\widetilde{O}(nd)$, i.e., with polylogarithmic dependence on $\frac{1}{\alpha}$. We leave removing this additive runtime term in the DP setting, or proving this is impossible in concrete query models, as a challenging question for future work.

Our algorithm follows a roadmap given by [HSU24], who split their algorithm into two phases: an initial "warm start" phase that computes an $O(1)$-multiplicative approximation of the geometric median, and a secondary "boosting" phase that uses iterative optimization methods to improve the warm start to an α-multiplicative approximation. The role of the warm start is to improve the domain size of the boosting phase to scale with the effective radius. However, both the warm start and the boosting phases of [HSU24] required superlinear $\approx n^2 d$ time. Our improvement to the warm start phase of the [HSU24] is quite simple, and may be of independent interest, so we provide a self-contained statement here.

Theorem 2. *Let $\mathcal{D} = \{\mathbf{x}_i\}_{i \in [n]} \subset \mathbb{B}^d(R)$ for $R > 0$, $0 < r \le r^{(0.9)}$, and $(\epsilon, \delta) \in [0,1]^2$. There is an (ϵ, δ)-DP algorithm that returns $\hat{\mathbf{x}}$ such that with probability $\ge 1 - \delta$, $f_{\mathcal{D}}(\hat{\mathbf{x}}) = O(f_{\mathcal{D}}(\mathbf{x}_\star))$, assuming $n \gtrsim \frac{\sqrt{d}}{\epsilon}$. The algorithm runs in time $\widetilde{O}(nd)$.*

1.2 Our techniques

As discussed previously, our algorithm employs a similar framework as [HSU24]. It is convenient to further split the warm start phase of the algorithm into two parts: finding an estimate $\hat{r}$ of the effective radius of $\mathcal{D}$, and finding an approximate centerpoint at distance $O(\hat{r})$ from the geometric median $\mathbf{x}_\star$.

Radius estimation. Our radius estimation algorithm is almost identical to that in [HSU24], Section 2.1, which uses the sparse vector technique (cf. Theorem 3.23, [DR14]) to detect the first time an estimate $\hat{r}$ is such that most points have $\ge \frac{3}{4}$ of $\mathcal{D}$ at a distance of $\approx \hat{r}$. The estimate $\hat{r}$ is geometrically updated over a grid of size $O(\log(\frac{R}{r}))$. Naïvely implemented, this strategy takes $\gtrsim n^2 d$ time due to the need for pairwise distance comparisons; even if dimesionality reduction techniques are used, this step appears to require $\Omega(n^2)$ time. We make a simple observation that a random sample of $\approx \log(\frac{1}{\delta})$ points from $\mathcal{D}$ is enough to determine whether a given point has $\gg \beta$ neighbors, or $\ll \gamma$, for appropriate (constant) quantile thresholds β, γ, which is enough to obtain an $\widetilde{O}(nd)$ runtime.

Centerpoint estimation. Our centerpoint estimation step departs from [HSU24], Section 2.2, who analyzed a custom variant of DP gradient descent with geometrically-decaying step sizes. We make the simple observation that directly applying the FriendlyCore algorithm of [HSU24] yields the same result. However, the standard implementation of FriendlyCore again requires $\Omega(n^2)$ time to estimate weights for each data point. We again show that FriendlyCore can be sped up to run in $\widetilde{O}(nd)$ time (independently of $\frac{R}{r}$) via weights estimated through subsampling. Our privacy proof of this subsampled variant is subtle, and based on an argument that couples our algorithm to an idealized algorithm that never fails to be private. We use this to account for the privacy loss due to the failure of our subsampling, i.e., if the estimates are inaccurate. We note that the [HSU24] algorithm for this step already ran in nearly-linear $\approx nd \log(\frac{R}{r})$ time, so we obtain an asymptotic improvement only if $\frac{R}{r}$ is large.

Boosting. The most technically novel part of our algorithm is in the boosting phase, which takes as input a radius and centerpoint estimate from the previous steps, and outputs an α-multiplicative approximation to (1). Like [HSU24], we use iterative optimization methods to implement this phase. However, a major bottleneck to a faster algorithm is the lack of a nearly-linear time DP solver for non-smooth empirical risk minimization (ERM) problems. Indeed, such $\widetilde{O}(1)$-pass optimizers are known only when the objective is convex and sufficiently smooth [FKT20], or $n \gtrsim d^2$ samples are taken [CJJ+23]. This is an issue, because while computing the geometric median (1) is a convex ERM problem, it is non-smooth, and nontrivial multiplicative guarantees are possible even with $n \approx \sqrt{d}$ samples.

We give a custom analysis of DP-SGD, specifically catered to the (non-smooth) ERM objective (1). Our main contribution is a tighter sensitivity analysis of DP-SGD's iterates, leveraging the structure of the geometric median. To motivate this observation, consider coupled algorithms with iterates $\mathbf{z}$, $\mathbf{z}'$, both taking gradient steps with respect to the subsampled function $\|\cdot - \mathbf{x}_i\|$ for some dataset element $\mathbf{x}_i \in \mathcal{D}$. A simple calculation shows these gradients are unit vectors $\mathbf{u}$, $\mathbf{u}'$, in the directions of $\mathbf{z} - \mathbf{x}_i$ and $\mathbf{z}' - \mathbf{x}_i$ respectively. It is not hard to formalize that updating $\mathbf{z} \leftarrow \mathbf{z} - \eta\mathbf{u}$ and $\mathbf{z}' \leftarrow \mathbf{z}' - \eta\mathbf{u}'$ is always contractive, unless $\mathbf{z}, \mathbf{z}'$ were both already very close to $\mathbf{x}_i$ (and hence, each other) to begin with. We use this structural result to inductively control DP-SGD's sensitivity, which lets us leverage a prior reduction from private optimization to stable optimization [FKT20].

Our result is the first we are aware of that obtains a nearly-linear runtime for DP-SGD on a structured non-smooth problem. We were inspired by [ALT24], who also gave faster runtimes for (smooth) DP optimization problems with outliers under further assumptions on the objective. We hope that our work motivates future DP optimization methods that harness problem structure for improved rates.

1.3 Related work

Differentially private convex optimization. Differentially private convex optimization has been studied extensively for over a decade [CM08, KST12, BST14, KJ16, BFGT20, FKT20, BGN21, GLL22, GLL+23] and inspired the influential DP-SGD algorithm widely adopted in deep learning [ACG+16]. In the classic setting, where functions are assumed to be Lipschitz and defined over a convex domain of diameter R, optimal rates have been achieved with linear dependence on R [BFTGT19]. Recent years have seen significant advancements in optimizing the gradient complexity of DP stochastic convex optimization [FKT20, AFKT21, KLL21, ZTC22, CJJ+23, CCGT24]. Despite these efforts, a nearly-linear gradient complexity has only been established for sufficiently smooth functions [FKT20, ZTC22, CCGT24] and for non-smooth functions [CJJ+23] when the condition $\sqrt{n} \gtrsim d$ is satisfied.

Differential privacy with average-case data. Adapting noise to the inherent properties of data, rather than catering to worst-case scenarios, is critical for making differential privacy practical in real-world applications. Several important approaches have emerged in this direction: smooth sensitivity frameworks [NRS07] that refine local sensitivity to make it private; instance optimality techniques [AD20] that provide tailored guarantees for specific datasets; methods with improved performance under distributional assumptions such as sub-Gaussian or heavy-tailed i.i.d. data [CWZ21, AL23, ALT24]; and data-dependent sensitivity computations that adapt during algorithm execution [ATMR21]. These approaches collectively represent the frontier in balancing privacy and utility beyond worst-case analyses. We view our work as another contribution towards this broader program.

1.4 Full version

This is an extended abstract meant for presentation at TPDP 2025. A full version of this paper, complete with all proofs and an empirical evaluation, will be posted on arXiv by the date of the workshop.

References

[ACG+16] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[AD20] Hilal Asi and John C Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. *Advances in neural information processing systems*, 33:14106–14117, 2020.

[AFKT21] Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: Optimal rates in l1 geometry. In *International Conference on Machine Learning*, pages 393–403. PMLR, 2021.

[AL22] Hassan Ashtiani and Christopher Liaw. Private and polynomial time algorithms for learning gaussians and beyond. In *Conference on Learning Theory*, pages 1075–1076. PMLR, 2022.

[AL23] Hilal Asi and Daogao Liu. User-level differentially private stochastic convex optimization: Efficient algorithms with optimal rates. *arXiv preprint arXiv:2311.03797*, 2023.

[ALT24] Hilal Asi, Daogao Liu, and Kevin Tian. Private stochastic convex optimization with heavy tails: Near-optimality from simple reductions. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024*, 2024.

[ATMR21] Galen Andrew, Om Thakkar, Brendan McMahan, and Swaroop Ramaswamy. Differentially private learning with adaptive clipping. *Advances in Neural Information Processing Systems*, 34:17455–17466, 2021.

[BD14] Rina Foygel Barber and John C Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*, 2014.

[BDKU20] Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan R. Ullman. Coinpress: Practical private mean and covariance estimation. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

[BFGT20] Raef Bassily, Vitaly Feldman, Cristóbal Guzmán, and Kunal Talwar. Stability of stochastic gradient descent on nonsmooth convex losses. *Advances in Neural Information Processing Systems*, 33:4381–4391, 2020.

[BFTGT19] Raef Bassily, Vitaly Feldman, Kunal Talwar, and Abhradeep Guha Thakurta. Private stochastic convex optimization with optimal rates. *Advances in neural information processing systems*, 32, 2019.

[BGN21] Raef Bassily, Cristóbal Guzmán, and Anupama Nandi. Non-euclidean differentially private stochastic convex optimization. In *Conference on Learning Theory*, pages 474–499. PMLR, 2021.

[BGS+21] Gavin Brown, Marco Gaboardi, Adam Smith, Jonathan Ullman, and Lydia Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. *Advances in neural information processing systems*, 34:7950–7964, 2021.

[BHS23] Gavin Brown, Samuel B. Hopkins, and Adam D. Smith. Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. In *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023*, volume 195 of *Proceedings of Machine Learning Research*, pages 5578–5579. PMLR, 2023.

[BKSW19] Mark Bun, Gautam Kamath, Thomas Steinke, and Steven Z Wu. Private hypothesis selection. *Advances in Neural Information Processing Systems*, 32, 2019.

[BST14] Raef Bassily, Adam D. Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014*, pages 464–473. IEEE Computer Society, 2014.

[CCGT24] Christopher A Choquette-Choo, Arun Ganesh, and Abhradeep Thakurta. Optimal rates for $o(1)$-smooth dp-sco with a single epoch and large batches. *arXiv preprint arXiv:2406.02716*, 2024.

[CJJ+23] Yair Carmon, Arun Jambulapati, Yujia Jin, Yin Tat Lee, Daogao Liu, Aaron Sidford, and Kevin Tian. Resqueing parallel and private stochastic convex optimization. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*, pages 2031–2058. IEEE, 2023.

[CKM+21] Edith Cohen, Haim Kaplan, Yishay Mansour, Uri Stemmer, and Eliad Tsfadia. Differentially-private clustering of easy instances. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pages 2049–2059. PMLR, 2021.

[CLM+16] Michael B Cohen, Yin Tat Lee, Gary Miller, Jakub Pachocki, and Aaron Sidford. Geometric median in nearly linear time. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 9–21, 2016.

[CM08] Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. *Advances in neural information processing systems*, 21, 2008.

[CWZ21] T Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

[DFM+20] Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[FKT20] Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: optimal rates in linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, pages 439–449. ACM, 2020.

[GLL22] Sivakanth Gopi, Yin Tat Lee, and Daogao Liu. Private convex optimization via exponential mechanism. In *Conference on Learning Theory*, pages 1948–1989. PMLR, 2022.

[GLL+23] Sivakanth Gopi, Yin Tat Lee, Daogao Liu, Ruoqi Shen, and Kevin Tian. Private convex optimization in general norms. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5068–5089. SIAM, 2023.

[HSU24] Mahdi Haghifam, Thomas Steinke, and Jonathan Ullman. Private geometric median. *Advances in Neural Information Processing Systems*, 37:46254–46293, 2024.

[KDH23] Rohith Kuditipudi, John Duchi, and Saminul Haque. A pretty fast algorithm for adaptive private mean estimation. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2511–2551. PMLR, 2023.

[KJ16] Shiva Prasad Kasiviswanathan and Hongxia Jin. Efficient private empirical risk minimization for high-dimensional learning. In *International Conference on Machine Learning*, pages 488–497. PMLR, 2016.

[KLL21] Janardhan Kulkarni, Yin Tat Lee, and Daogao Liu. Private non-smooth erm and sco in subquadratic steps. *Advances in Neural Information Processing Systems*, 34:4053–4064, 2021.

[KLSU19] Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan R. Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory, COLT 2019, 25-28 June 2019*, volume 99 of *Proceedings of Machine Learning Research*, pages 1853–1902. PMLR, 2019.

[KST12] Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *Conference on Learning Theory*, pages 25–1. JMLR Workshop and Conference Proceedings, 2012.

[KV17] Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

[LKJO22] Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. DP-PCA: statistically optimal and differentially private PCA. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022*, 2022.

[LR91] Hendrik P. Lopuhaa and Peter J. Rousseuw. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 19(1):229–248, 1991.

[NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 75–84. ACM, 2007.

[NSV16] Kobbi Nissim, Uri Stemmer, and Salil P. Vadhan. Locating a small cluster privately. In Tova Milo and Wang-Chiew Tan, editors, *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016*, pages 413–427. ACM, 2016.

[TCK+22] Eliad Tsfadia, Edith Cohen, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Friendly-core: Practical differentially private aggregation. In *International Conference on Machine Learning*, pages 21828–21863. PMLR, 2022.

[Web29] Alfred Weber. *Theory of the Location of Industries*. University of Chicago Press, 1929.

[ZTC22] Qinzi Zhang, Hoang Tran, and Ashok Cutkosky. Differentially private online-to-batch for smooth losses. In *NeurIPS*, 2022.