4 PURE DP

In this section, we give a lower bound for ϵ-(pure) differentially private algorithms for minimizing unconstrained convex Lipschitz loss function $L(\theta; \mathcal{D})$. In the construction of lower bounds for constrained DP-ERM in the seminar paper Bassily et al. (2014), they chose linear function $\ell(\theta; d) = \langle \theta, d \rangle$ as the objective function, which isn't applicable in the unconstrained setting because it could decrease to negative infinity. Instead, we extend the linear loss in unit ball to the whole $\mathbb{R}^p$ while preserving its Lipschitzness and convexity.

4.1 EXTENSION OF LINEAR LOSS

We begin with the following lemma from Cobzas and Mustata (1978) which gives a Lipschitz extension of any convex Lipschitz function over some convex set to the whole domain $\mathbb{R}^p$.

Lemma 4.1 (Theorem 1 in Cobzas and Mustata (1978))**.** *Let f by an η-Lipschitz, convex function defined on a convex bounded set $\mathcal{C} \in \mathbb{R}^p$. Then there exists an efficiently computable, η-Lipschitz, convex function $\tilde{f}$ defined on $\mathbb{R}^p$ such that it's equal to f restricted on $\mathcal{C}$. The explicit construction is the following. Define auxiliary function $g_y(x)$:*

$$g_y(x) = f(y) + \eta ||x - y||_2, y \in \mathcal{C}, x \in \mathbb{R}^p \tag{5}$$

then $\tilde{f}$ is defined as $\tilde{f}(x) = \min_{y \in \mathcal{C}} g_y(x)$.

We use such extension to define our loss function in the unconstrained case. Namely, we define

$$\ell(\theta; d) = \min_{||y||_2 \leq 1} -\langle y, d \rangle + ||\theta - y||_2 \tag{6}$$

which is convex, 1-Lipschitz and equal to $-\langle \theta, d \rangle$ when $||\theta||_2 \leq 1$ according to Lemma 4.1. Specifically, it's easy to verify that $\ell(\theta; 0) = \max\{0, ||\theta||_2 - 1\}$, and when $||d||_2 = 1$

$$\ell(\theta; d) \geq \min_{||y||_2 \leq 1} -\langle y, d \rangle \geq -1, \tag{7}$$

where the equation holds if and only if $\theta = d$.

For any dataset $\mathcal{D} = \{d_1, ..., d_n\}$, we define $L(\theta; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^n \ell(\theta; d_i)$. The structure of the proof is similar to that in Bassily et al. (2014), while technical details are quite different as we need to handle a non-linear objective function.

4.2 LOWER BOUND

To prove the lower bound we need the following lemma from Bassily et al. (2014). The proof is similar to that of Lemma 5.1 in Bassily et al. (2014), except that we change the construction by adding $\mathbf{0}$ as our dummy points. For completeness we include it here.

Lemma 4.2 (Lemma 5.1 in Bassily et al. (2014))**.** *Let $n, p \geq 2$ and $\epsilon > 0$. There is a number $n^* = \Omega(\min(n, \frac{p}{\epsilon}))$ such that for any ϵ-differentially private algorithm $\mathcal{A}$, there is a dataset $\mathcal{D} = \{d_1, ..., d_n\} \subset \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ with $||\sum_{i=1}^n d_i||_2 = n^*$ such that, with probability at least $1/2$ (taken over the algorithm random coins), we have*

$$||\mathcal{A}(\mathcal{D}) - q(\mathcal{D})||_2 = \Omega(\min(1, \frac{p}{n\epsilon})) \tag{8}$$

where $q(\mathcal{D}) = \frac{1}{n} \sum_{i=1}^n d_i$

Proof. By using a standard packing argument we can construct $K = 2^{\frac{p}{2}}$ points $d^{(1)}, ..., d^{(K)}$ in $\{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ such that for every distinct pair $d^{(i)}, d^{(j)}$ of these points, we have

$$||d^{(i)} - d^{(j)}||_2 \geq \frac{1}{8} \tag{9}$$

It is easy to show the existence of such set of points using the probabilistic method (for example, the Gilbert-Varshamov construction of a linear random binary code).

Fix $\epsilon > 0$ and define $n^\star = \frac{p}{20\epsilon}$. Let's first consider the case where $n \leq n^\star$. We construct K datasets $\mathcal{D}^{(1)}, ..., \mathcal{D}^{(K)}$ where for each $i \in [K]$, $\mathcal{D}^{(i)}$ contains n copies of $d^{(i)}$. Note that $q(\mathcal{D}^{(i)}) = d^{(i)}$, we have that for all $i \neq j$,

$$||q(\mathcal{D}^{(i)}) - q(\mathcal{D}^{(j)})||_2 \geq \frac{1}{8} \tag{10}$$

Let $\mathcal{A}$ be any ϵ-differentially private algorithm. Suppose that for every $\mathcal{D}^{(i)}, i \in [K]$, with probability at least $1/2$, $||\mathcal{A}(\mathcal{D}^{(i)}) - q(\mathcal{D}^{(i)})||_2 < \frac{1}{16}$,i.e.,$Pr[\mathcal{A}(\mathcal{D}^{(i)}) \in B(\mathcal{D}^{(i)})] \geq \frac{1}{2}$ where for any dataset $\mathcal{D}$, $B(\mathcal{D})$ is defined as

$$B(\mathcal{D}) = \{x \in R^p : ||x - q(\mathcal{D})||_2 < \frac{1}{16}\} \tag{11}$$

Note that for all $i \neq j$, $\mathcal{D}^{(i)}$ and $\mathcal{D}^{(j)}$ differs in all their n entries. Since $\mathcal{A}$ is ϵ-differentially private, for all $i \in [K]$, we have $Pr[A(\mathcal{D}^{(1)}) \in B(\mathcal{D}^{(i)})] \geq \frac{1}{2}e^{-\epsilon n}$. Since all $B(\mathcal{D}^{(i)})$ are mutually disjoint, then

$$\frac{K}{2}e^{-\epsilon n} \leq \sum_{i=1}^{K} Pr[\mathcal{A}(\mathcal{D}^{(1)}) \in B(\mathcal{D}^{(i)})] \leq 1 \tag{12}$$

which implies that $n > n^\star$ for sufficiently large p, contradicting the fact that $n \leq n^\star$. Hence, there must exist a dataset $\mathcal{D}^{(i)}$ on which A makes an ℓ_2-error on estimating $q(\mathcal{D})$ which is at least $1/16$ with probability at least $1/2$. Note also that the ℓ_2 norm of the sum of the entries of such $\mathcal{D}^{(i)}$ is n.

Next, we consider the case where $n > n^\star$. As before, we construct $K = 2^{\frac{p}{2}}$ datasets $\tilde{\mathcal{D}}^{(1)}, \cdots, \tilde{\mathcal{D}}^{(K)}$ of size n where for every $i \in [K]$, the first $n^\star$ elements of each dataset $\tilde{\mathcal{D}}^{(i)}$ are the same as dataset $\mathcal{D}^{(i)}$ from before whereas the remaining $n - n^\star$ elements are $\mathbf{0}$.

Note that any two distinct datasets $\tilde{\mathcal{D}}^{(i)}, \tilde{\mathcal{D}}^{(j)}$ in this collection differ in exactly $n^\star$ entries. Let $\mathcal{A}$ be any ϵ-differentially private algorithm for answering q. Suppose that for every $i \in [K]$, with probability at least $1/2$, we have that

$$||\mathcal{A}(\tilde{\mathcal{D}}^{(i)}) - q(\tilde{\mathcal{D}}^{(i)})||_2 < \frac{n^\star}{16n} \tag{13}$$

Note that for all $i \in [K]$, we have that $q(\tilde{\mathcal{D}}^{(i)}) = \frac{n^*}{n} q(\mathcal{D}^{(i)})$. Now, we define an algorithm $\tilde{\mathcal{A}}$ for answering q on datasets $\mathcal{D}$ of size $n^\star$ as follows. First, $\tilde{\mathcal{A}}$ appends $\mathbf{0}$ as above to get a dataset $\tilde{\mathcal{D}}$ of size n. Then, it runs $\mathcal{A}$ on $\tilde{\mathcal{D}}$ and outputs $\frac{n^* \mathcal{A}(\tilde{\mathcal{D}})}{n}$. Hence, by the post-processing propertry of differential privacy, $\tilde{\mathcal{A}}$ is ϵ-differentially private since $\mathcal{A}$ is ϵ-differentially private. Thus for every $i \in [K]$, with probability at least $1/2$, we have that $||\tilde{\mathcal{A}}(\mathcal{D}^{(i)}) - q(\mathcal{D}^{(i)})||_2 < \frac{1}{16}$. However, this contradicts our result in the first part of the proof. Therefore, there must exist a dataset $\tilde{\mathcal{D}}^{(i)}$ in the above collection such that, with probability at least $1/2$,

$$||\mathcal{A}(\tilde{\mathcal{D}}^{(i)}) - q(\tilde{\mathcal{D}}^{(i)})||_2 \geq \frac{n^\star}{16n} \geq \frac{p}{320\epsilon n} \tag{14}$$

Note that the ℓ_2 norm of the sum of entries of such $\tilde{D}^{(i)}$ is always $n^\star$. □

Lemma 4.2 basicly says that it's impossible to estimate the average of $d_1, ..., d_n$ with accuracy $o(\min(1, \frac{p}{n\epsilon}))$ with an ϵ-DP algorithm. We are ready to prove our main theorem of this section.

Theorem 4.3 (Lower bound for ϵ-differentially private algorithms)**.** *Let $n, p \geq 2$ and $\epsilon > 0$. For every ϵ-differentially private algorithm with output $\theta^{priv} \in \mathbb{R}^p$, there is a dataset $\mathcal{D} = \{d_1, ..., d_n\} \subset \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ such that, with probability at least $1/2$ (over the algorithm random coins), we must have that*

$$L(\theta^{priv}; \mathcal{D}) - \min_{\theta} L(\theta; \mathcal{D}) = \Omega(\min(1, \frac{p}{n\epsilon})) \tag{15}$$

Proof. Let $\mathcal{A}$ be an ϵ-differentially private algorithm for minimizing L and let θ^{priv} denote its output, define $r := \theta^{priv} - \theta^*$. First, observe that for any $\theta \in \mathbb{R}^p$ and dataset $\mathcal{D}$ as constructed in

Lemma 4.2 (recall that $\mathcal{D}$ consists of n^* copies of a vector $d \in \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p$ and $n - n^*$ copies of $\mathbf{0}$).

$$L(\theta^*; \mathcal{D}) = \frac{n - n^*}{n} \max\{0, ||\theta^*||_2 - 1\} + \frac{n^*}{n} \min_{||y||_2 \leq 1} (-\langle y, d\rangle + ||\theta^* - y||_2) = -\frac{n^*}{n} \tag{16}$$

when $\theta^* = d$, and also

$$\begin{aligned} L(\theta^{priv}; \mathcal{D}) &= \frac{n - n^*}{n} \max\{0, ||\theta^{priv}||_2 - 1\} + \frac{n^*}{n} \min_{||y||_2 \leq 1} (-\langle y, d\rangle + ||\theta^{priv} - y||_2) \\ &\geq \frac{n^*}{n} \min_{||y||_2 \leq 1} (-\langle y, d\rangle + ||\theta^{priv} - y||_2) \\ &= \frac{n^*}{n} \min_{||y||_2 \leq 1} (-\langle y, d\rangle + ||r + d - y||_2) \\ &\text{(because } \theta^* = d\text{)} \\ &\geq \frac{n^* \min\{1, ||r||_2^2\}}{8n} - \frac{n^*}{n} \end{aligned}$$

the last inequality follows by discussing the norm of $y - d$. If $||y - d||_2 \leq ||r||_2/2$, then

$$||r + d - y||_2 \geq ||r||_2/2 \geq \min\{1, ||r||_2^2\}/2 \tag{17}$$

combining with the fact that $|\langle y, d\rangle| \leq 1$ proves the last inequality.

If $||y - d||_2 \geq ||r||_2/2$, then we have $\min_{||y||_2 \leq 1} -\langle y, d\rangle \geq -1 + \frac{||r||_2^2}{8}$. To prove this, we assume $d = e_1$ without loss of generality and $y - d = (x_1, ..., x_p)$ where $\sum_{i=1}^p x_i^2 \geq ||r||_2^2/4$. Since $\|y\|_2 = ||y - d + d||_2 \leq 1$, we must have

$$1 + \sum_{i=1}^p x_i^2 + 2x_1 \leq 1 \tag{18}$$

Thus $-\langle y, d\rangle = -1 - \langle y - d, d\rangle = -1 - x_1 \geq -1 + ||r||_2^2/8$ as desired, which finishes the discussion on the second case.

From the above result we have that

$$L(\theta^{priv}; \mathcal{D}) - L(\theta^*; \mathcal{D}) \geq \frac{n^* \min\{1, ||r||_2^2\}}{8n} \tag{19}$$

To proceed, suppose for the sake of a contradiction, that for every dataset $\mathcal{D} = \{d_1, ..., d_n\} \subset \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ with $||\sum_{i=1}^n d_i||_2 = n^*$, with probability more than $1/2$, we have that $||\theta^{priv} - \theta^*||_2 = ||r||_2 \neq \Omega(1)$. Let $\tilde{\mathcal{A}}$ be an ϵ-differentially private algorithm that first runs $\mathcal{A}$ on the data and then outputs $\frac{n^*}{n}\theta^{priv}$. Recall that $q(\mathcal{D}) = \frac{n^*}{n}\theta^*$, this implies that for every dataset $\mathcal{D} = \{d_1, ..., d_n\} \subset \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ with $||\sum_{i=1}^n d_i||_2 = n^*$, with probability more than $1/2$, $||\tilde{\mathcal{A}}(\mathcal{D}) - q(\mathcal{D})||_2 \neq \Omega(\min(1, \frac{p}{n\epsilon}))$ which contradicts Lemma 4.2. Thus, there must exists a dataset $\mathcal{D} = \{d_1, ..., d_n\} \subset \{\frac{1}{\sqrt{p}}, -\frac{1}{\sqrt{p}}\}^p \cup \{\mathbf{0}\}$ with $||\sum_{i=1}^n d_i||_2 = n^*$, such that with pr obability more than $1/2$, we have $\|r\|_2 = ||\theta^{priv} - \theta^*||_2 = \Omega(1)$, and as a result

$$L(\theta^{priv}; \mathcal{D}) - L(\theta^*; \mathcal{D}) = \Omega(\min(1, \frac{p}{n\epsilon})) \tag{20}$$

□

5 CONCLUSION

In this paper, we study differentially private convex ERM in the unconstrained case and give the first tight lower bounds for approximate-DP ERM for general loss functions. Our results also directly imply a same lower bound for the constrained case, improving the classic lower bound in Bassily et al. (2014) by $\log(1/\delta)$. We also give an $\Omega(\frac{p}{n\epsilon})$ lower bound for unconstrained pure-DP ERM which recovers the result in the constrained case. Our techniques enrich the quite limited tools in constructing lower bounds in the private setting and we hope they can find future use, especially for those problems which are not (easily) reducible from one-way marginals. Designing better algorithms for general (un)constrained DP-ERM based on our insights would also be an interesting and meaningful direction, which we leave as future work.

REFERENCES

Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: Optimal rates in ℓ_1 geometry. *arXiv preprint arXiv:2103.01516*, 2021.

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

Raef Bassily, Vitaly Feldman, Kunal Talwar, and Abhradeep Guha Thakurta. Private stochastic convex optimization with optimal rates. In *Advances in Neural Information Processing Systems*, pages 11282–11291, 2019.

Raef Bassily, Vitaly Feldman, Cristóbal Guzmán, and Kunal Talwar. Stability of stochastic gradient descent on nonsmooth convex losses. *arXiv preprint arXiv:2006.06914*, 2020.

Raef Bassily, Cristóbal Guzmán, and Anupama Nandi. Non-euclidean differentially private stochastic convex optimization. *arXiv preprint arXiv:2103.01278*, 2021.

Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5):1897–1905, 1998.

Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. *SIAM Journal on Computing*, 47(5):1888–1938, 2018.

Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *NIPS*, volume 8, pages 289–296. Citeseer, 2008.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(3), 2011.

S Cobzas and C Mustata. Norm-preserving extension of convex lipschitz functions. *J. Approx. Theory*, 24(3):236–244, 1978.

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

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

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*, pages 439–449, 2020.

Kazuto Fukuchi, Quang Khai Tran, and Jun Sakuma. Differentially private empirical risk minimization with input perturbation. In *International Conference on Discovery Science*, pages 82–90. Springer, 2017.

Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 705–714, 2010.

Roger Iyengar, Joseph P Near, Dawn Song, Om Thakkar, Abhradeep Thakurta, and Lun Wang. Towards practical differentially private convex optimization. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 299–316. IEEE, 2019.

Prateek Jain and Abhradeep Guha Thakurta. (near) dimension independent risk bounds for differentially private learning. In *International Conference on Machine Learning*, pages 476–484. PMLR, 2014.

Peter Kairouz, Mónica Ribero, Keith Rush, and Abhradeep Thakurta. Dimension independence in unconstrained private erm via adaptive preconditioning. *arXiv preprint arXiv:2008.06570*, 2020.

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.

Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

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.

Janardhan Kulkarni, Yin Tat Lee, and Daogao Liu. Private non-smooth empirical risk minimization and stochastic convex optimization in subquadratic steps. *arXiv preprint arXiv:2103.15352*, 2021.

Benjamin IP Rubinstein, Peter L Bartlett, Ling Huang, and Nina Taft. Learning in a large function space: Privacy-preserving mechanisms for svm learning. *arXiv preprint arXiv:0911.5708*, 2009.

Shuang Song, Kamalika Chaudhuri, and Anand D Sarwate. Stochastic gradient descent with differentially private updates. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 245–248. IEEE, 2013.

Shuang Song, Thomas Steinke, Om Thakkar, and Abhradeep Thakurta. Evading the curse of dimensionality in unconstrained private glms. In *International Conference on Artificial Intelligence and Statistics*, pages 2638–2646. PMLR, 2021.

Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *arXiv preprint arXiv:1501.06095*, 2015.

Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Nearly-optimal private lasso. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*, pages 3025–3033, 2015.

Gábor Tardos. Optimal probabilistic fingerprint codes. *Journal of the ACM (JACM)*, 55(2):1–24, 2008.

Neal R Wagner. Fingerprinting. In *1983 IEEE Symposium on Security and Privacy*, pages 18–18. IEEE, 1983.

Di Wang, Minwei Ye, and Jinhui Xu. Differentially private empirical risk minimization revisited: Faster and more general. In *Advances in Neural Information Processing Systems*, pages 2722–2731, 2017.

Puyu Wang, Yunwen Lei, Yiming Ying, and Hai Zhang. Differentially private sgd with non-smooth loss. *arXiv preprint arXiv:2101.08925*, 2021.

Xi Wu, Fengan Li, Arun Kumar, Kamalika Chaudhuri, Somesh Jha, and Jeffrey Naughton. Bolt-on differential privacy for scalable stochastic gradient descent-based analytics. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pages 1307–1322, 2017.

Jiaqi Zhang, Kai Zheng, Wenlong Mou, and Liwei Wang. Efficient private erm for smooth objectives. *arXiv preprint arXiv:1703.09947*, 2017.