Optimality of Matrix Mechanism on ℓ_p^p-metric

[1]Jingcheng Liu, [2]Jalaj Upadhyay, [1]Zongrui Zou

[1]Nanjing University, [2]Rutgers

Matrix Mechanism and Different Error Metric

Many fundamental analyses can be cast as a set of linear queries: given an input $x \in \mathbb{R}^n$, a set of m linear queries can be represented as the rows of a matrix $A \in \mathbb{R}^{m\times n}$. The answer to the set of queries is simply the matrix-vector product Ax. Here, $x, x' \in \mathbb{R}^n$ are neighboring if $\|x - x'\|_1 \le 1$. When these queries are answered using a privacy-preserving algorithm, $\mathcal{M}$, the performance of the algorithm is usually measured in terms of its *absolute error* or *mean squared error*.

absolute error

$$\mathsf{err}_{\ell_\infty}(\mathcal{M}, A, n) := \max_{x\in\mathbb{R}^n} \mathbb{E}\left[\|\mathcal{M}(x) - Ax\|_\infty\right]$$

mean squared error

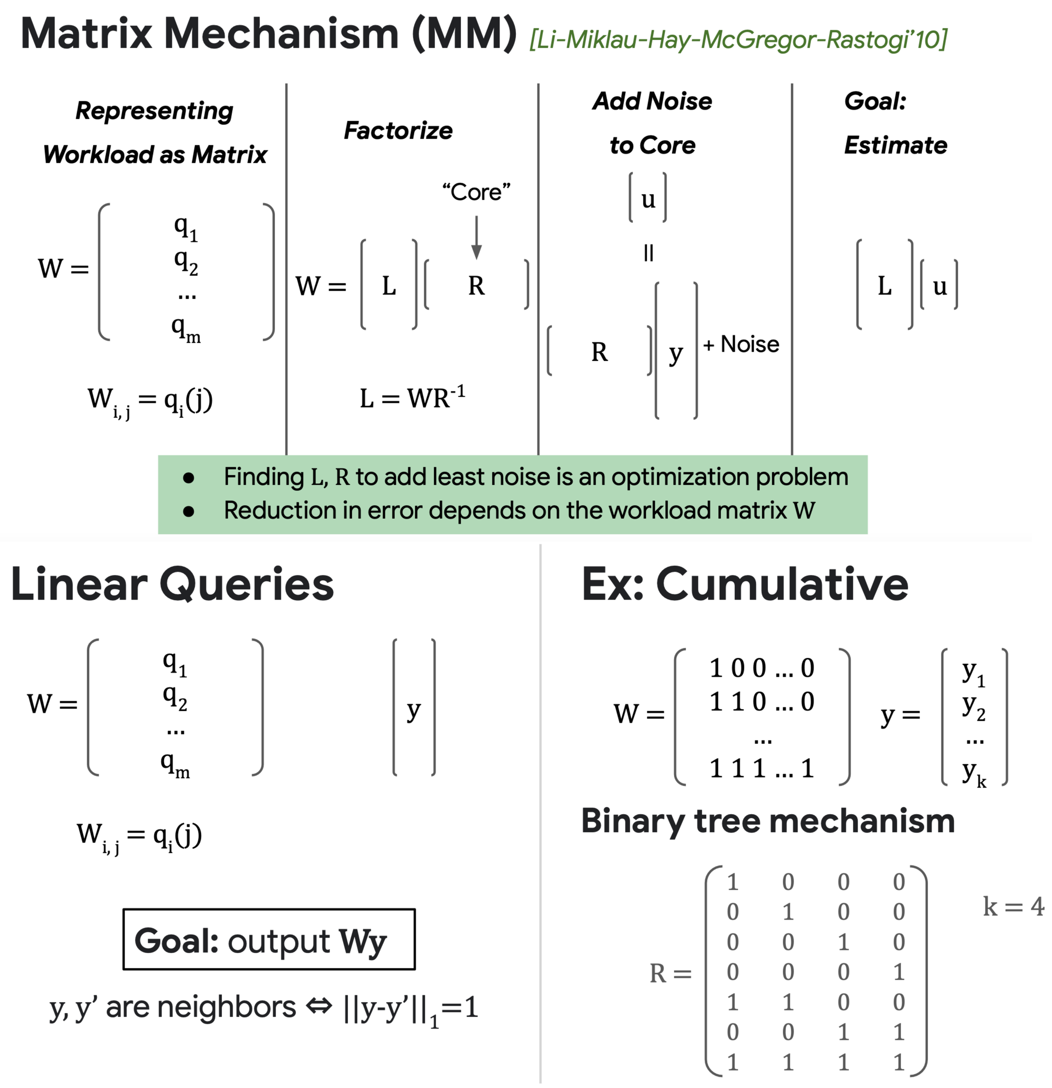
$$\mathsf{err}_{MSE}(\mathcal{M}, A, n) := \max_{x\in\mathbb{R}^n} \mathbb{E}\left[\tfrac{1}{n}\|\mathcal{M}(x) - Ax\|_2^2\right]$$

In this paper, we initiate the study of ℓ_p^p-error metric that seamlessly interpolate between $p = 2$ (squared error) to $p = \infty$ (absolute error):

ℓ_p^p error

$$\mathsf{err}_{\ell_p^p}(\mathcal{M}, A) := \max_{x\in\mathbb{R}^n} \left(\mathbb{E}\left[\|\mathcal{M}(x) - Ax\|_p^p\right]\right)^{1/p}$$

One popular mechanism for privately answering linear queries under different error metrics is the *matrix mechanism*, also known as the *factorization mechanism*:



Optimality of Matrix Mechanism

In this paper, we show that the optimal matrix mechanism is also optimal among all differentially private mechanisms with respect to the ℓ_p^p metric, up to logarithmic factors:

Matrix Mechanism is optimal on ℓ_p^p metric

Fix $A \in \mathbb{R}^{m\times n}$ be a matrix representing m linear queries, and let $\mathcal{M} : \mathbb{R}^n \to \mathbb{R}^m$ be any (ϵ, δ)-DP algorithm. Then, there exists a factorization of $A = LR$ such that $\mathcal{M}_{\text{matrix}}(x) = L(Rx + z)$ with $z \sim \mathcal{N}(\mathbf{0}, \|R\|_{1\to 2}^2 \mathbb{I}_k)$ preserves (ϵ, δ)-DP and that

$$\mathsf{err}_{\ell_p^p}(\mathcal{M}_{\text{matrix}}, A) \lesssim \mathsf{err}_{\ell_p^p}(\mathcal{M}, A) \cdot \mathsf{polylog}(1/\delta, m).$$

Here, $\mathbb{I}_k \in \mathbb{R}^{k\times k}$ is the identity matrix.

An (ϵ, δ)-DP lower bound on General Linear Queries

Our main claim is a lower bound on general (ϵ, δ)-differentially private mechanisms for answering linear queries in high privacy regimes in terms of certain *factorization norms* in [NT24] defined below

$\gamma_{(p)}$ norm

$$\gamma_{(p)}(A) := \min_{LR=A} \left\{ \sqrt{\mathsf{tr}_{p/2}(LL^\top)}\|R\|_{1\to 2} \right\}, \text{ where}$$

$$\mathsf{tr}_p(U) := \begin{cases} \left(\sum_{i=1}^d U_{ii}^p\right)^{1/p} & p < \infty \\ \max_{i\in[d]} |U_{ii}| & p = \infty \end{cases}$$

is the p-*trace*.

Equipped with this definition, we state our lower bound:

Lower bound for (ϵ, δ)-DP

Fix any $n, m \in \mathbb{N}$, $\epsilon \in (0, \frac{1}{2})$, $0 \le \delta \le 1$ and $p \in [2, \infty)$. For any query matrix $A \in \mathbb{R}^{m\times n}$, if a mechanism $\mathcal{M} : \mathbb{R}^n \to \mathbb{R}^m$ preserves (ϵ, δ)-differential privacy, then there exists a universal constant C',

$$\mathsf{err}_{\ell_p^p}(\mathcal{M}, A) \ge \frac{(1-\tilde{\delta})\gamma_{(p)}(A)}{C'\epsilon}, \quad \text{where} \quad \tilde{\delta} = O_\varepsilon(\delta).$$

Exact Lower Bound on Private Prefix Sum

The meaning of $\gamma_{(p)}(A)$ is not immediately apparent. Thus, as one of its applications, we study explicit lower bound (with respect to n instead of $\gamma_{(p)}(A)$) for some special types of queries that are widely used in the community of privacy.

Tight Bounds in Prefix Sum

For any $n \in \mathbb{N}$ and any $p \in [2, \infty)$, the matrix mechanism, $\mathcal{M}_{\text{fact}}$, achieves the following error guarantee while preserving (ϵ, δ)-differential privacy:

$$\mathsf{err}_{\ell_p^p}(\mathcal{M}_{\text{fact}}, A_{\text{prefix}}, n) = \tilde{O}\left(\frac{n^{1/p}}{\epsilon}\right),$$

and there is no (ϵ, δ)-differentially private mechanism $\mathcal{M}$ that achieves

$$\mathsf{err}_{\ell_p^p}(\mathcal{M}, A_{\text{prefix}}, n) = o\left(\frac{(1-\delta)n^{1/p}\log(n)}{e^{3\epsilon} - 1}\right).$$

Define

$$\kappa(A) := \min_{\theta^\top\theta=1} \left(\max_{x\in B_1^n} \theta^\top Ax - \min_{x\in B_1^n} \theta^\top Ax\right)$$

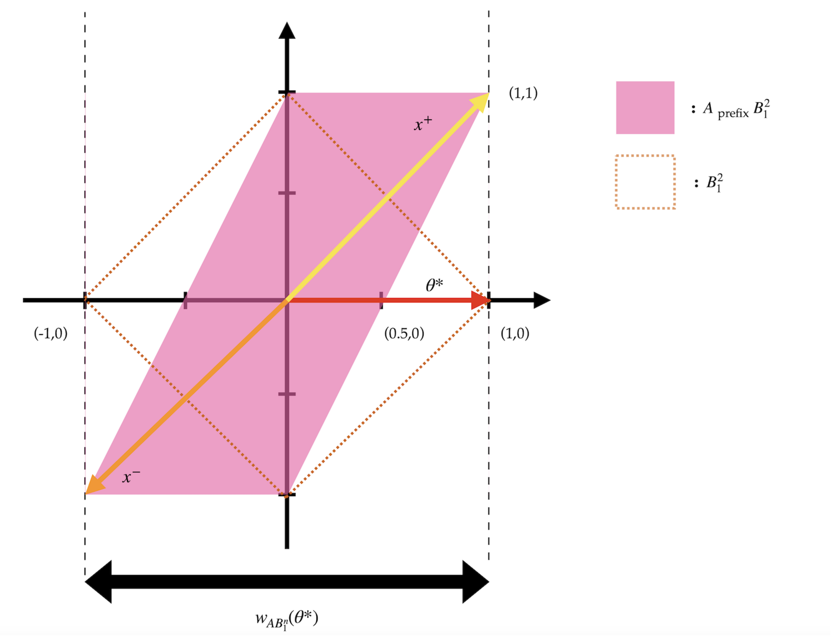
be the width of the most narrow direction of the sensitivity polytope AB_1^m, we also give the following geometric characterization of A_{prefix}, which could be of independent interest:

Geometric characterization of prefix sum

Let A_{prefix} be a lower-triangular matrix with non-zero entry equal to one, then $\kappa(A_{\text{prefix}}) = 2$.

Exact Lower Bound on Private Prefix Sum cont.

The following diagram gives an intuition of the above lemma:



Exact Lower Bound on Private Parity Queries

We also characterize the lower bound on privately answering parity queries. The theorem recovers the lower bound in Section 8 of Henzinger et al. for $p = 2$ and Section 3.6 of Edmonds et al. when $p \to \infty$.

Tight Bounds in Parity Queries

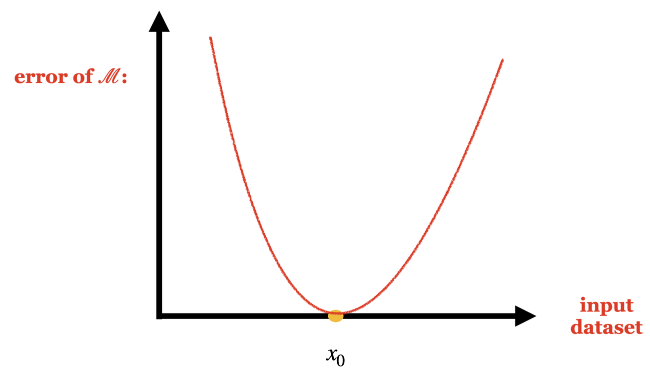
Let $\mathcal{Q}_{d,w}^P$ be the collection of parity queries. For any (ϵ, δ)-differentially private mechanism $\mathcal{M}$ for answering queries in $\mathcal{Q}_{d,w}^P$, the worst case ℓ_p^p error

$$\mathsf{err}_{\ell_p^p}\left(\mathcal{M}, \mathcal{Q}_{d,w}^P, \binom{d}{w}\right) = \Omega\left(\frac{(1-\delta)}{e^{3\epsilon} - 1}\binom{d}{w}^{1/2+1/p}\right).$$

Further, this lower bound can be achieved by trivial Gaussian mechanism.

Instance Optimality v.s. Worst Case Optimality

Se note that we only give a worst case lower bound over all $x \in \mathbb{R}^n$ by the definition of ℓ_p^p error metric. To understand why we cannot get a instance-optimal lower bound, consider a trivial mechanism M_{x_0} such that for any $x \in \mathbb{R}^n$, it always outputs Ax_0 where $x_0 \in \mathbb{R}^n$ is any given dataset. Clearly M_{x_0} is not an oblivious additive noise mechanism, and it preserves perfect differential privacy, and perfect accuracy on the input x_0.



In Nikolov et al., the authors study unbiased mechanism, and show that the Gaussian mechanism is indeed instance-optimal over all such unbiased mechanisms, by giving an asymmetric lower bound saying that if an unbiased mechanism performs well in an input x_0, then it must perform worse in some other inputs x' where x' neighboring x_0. It is still open if such an asymmetric lower bound exists for general linear queries over all general mechanisms.

Contact: zou.zongrui@smail.nju.edu.cn