
Distributed Differentially Private Data Analytics via Secure Sketching

Jakob Burkhardt¹ Hannah Keller¹ Claudio Orlandi¹ Chris Schwegelshohn¹

Abstract

We introduce the *linear-transformation model*, a distributed model of differentially private data analysis. Clients have access to a trusted platform capable of applying a public matrix to their inputs. Such computations can be securely distributed across multiple servers using simple and efficient secure multiparty computation techniques. The linear-transformation model serves as an intermediate model between the highly expressive *central model* and the minimal *local model*. In the central model, clients have access to a trusted platform capable of applying any function to their inputs. However, this expressiveness comes at a cost, as it is often prohibitively expensive to distribute such computations, leading to the central model typically being implemented by a single trusted server. In contrast, the local model assumes no trusted platform, which forces clients to add significant noise to their data. The linear-transformation model avoids the single point of failure for privacy present in the central model, while also mitigating the high noise required in the local model. We demonstrate that linear transformations are very useful for differential privacy, allowing for the computation of linear sketches of input data. These sketches largely preserve utility for tasks such as private low-rank approximation and private ridge regression, while introducing only minimal error, critically independent of the number of clients.

1. Introduction

Differential Privacy (DP) (Dwork et al., 2006b) has become the de-facto standard for ensuring the privacy of individuals whose data is used in data analytics. DP offers strong, provable guarantees, such as composability and resilience to

auxiliary information. In the classic *central model* of differential privacy, clients submit their data to a trusted central server, which processes the data and releases a (necessarily) noisy version of the result. Analysts can then use this result to perform queries on the data. Informally, differential privacy guarantees that the analyst cannot confidently determine whether any individual contributed their data, except with some small probability.

The central model requires trust that the server will not reveal any individual’s data and cannot be compromised by external actors. However, universally trusted servers may not exist in practice. In situations where privacy is critical but trusting a single server is not feasible, alternative approaches are available: One such approach is to use secure multiparty computation (MPC) techniques to distribute the central server’s computations across multiple servers. Under the assumption that a subset of these servers are honest, this setup guarantees the same level of privacy as the central model. However, MPC techniques often introduce significant computational overhead, especially when distributing complex or expressive computations.

Another option is to avoid relying on external servers altogether. In this case, each client must publish their data in a locally differentially private manner, by adding noise directly to their data. While this removes the need for a trusted server, it results in lower utility due to the increased noise. This approach is known as the *local model* of differential privacy (Kasiviswanathan et al., 2008), an example of which is randomized response (Warner, 1965).

Both the central and local models of differential privacy have strengths and drawbacks. The central model typically yields higher utility but requires strong trust in a single server or the costly use of MPC. The local model, by contrast, requires no trust in external parties but suffers from reduced utility. These tradeoffs have motivated significant research into finding a middle ground between these two models. One promising direction is to limit the expressiveness of the class of functions \mathcal{F} that can be executed in a trusted manner. While this restriction allows for more efficient distributed implementation, it is only useful if the DP mechanism built around such functions can still achieve good accuracy.

A notable example of such restricted expressiveness is the class of functions known as *shuffles*, where a central entity

¹Aarhus University, Denmark. Correspondence to: Hannah Keller <hkeller@cs.au.dk>.

(or a distributed system) randomly permutes the messages from clients before they reach the analysing server. This is known as the *shuffle model* (Bittau et al., 2017; Erlingsson et al., 2019; Cheu et al., 2019). The shuffle model has been used to implement many differentially-private mechanisms with reasonable utility (Erlingsson et al., 2019; Cheu et al., 2019). Surprisingly, the shuffle model is more expressive than initially expected: with a trusted shuffler, any function can be securely computed in two rounds under the assumption of an honest-majority (Beimel et al., 2020).

We therefore ask the following:

Question 1.1. *What is the least expressive class of functions \mathcal{F} that needs to be securely implemented in order to achieve computationally efficient distributed differential privacy with utility comparable to that of the central model?*

1.1. Our Contribution

In this work, we investigate the power of the *linear-transformation model* (LTM) for differentially-private mechanisms. In this model, the clients only have access to a trusted platform for performing arbitrary linear transformations of their collective inputs. Linear functions can be distributed extremely efficiently with secure computations. However, it is also known that the expressiveness of this class of functions is strictly limited.

We demonstrate the benefits of this linear transformation model (LTM) by showing that it can be used, in combinations with *linear sketches* to construct private summaries of a data set with noise comparable to that of central differentially private mechanisms. As in the local model, each client adds some amount of noise to their data before passing it onto the linear sketch. We then release the linear transform of the data. The benefit of having a trusted linear sketch is that the noise can be combined, thereby significantly reducing the amount of noise per client. All of our mechanisms require only a single round of communication from clients to server.

We demonstrate the effectiveness of this model by exemplarily applying it to problems in numerical linear algebra. A particularly powerful and expressive class of linear sketches with numerous linear algebra applications (Woodruff, 2014) are Johnson-Lindenstrauss (JL) transforms. In this paper, we focus on *low rank approximation* and *regularized linear regression*. Our contribution is to show how to design mechanisms for these problems, with the most widely used JL-transforms as the underlying secure sketch. For these problems, we achieve utility comparable to that of the central model with local privacy guarantees, assuming the JL-transform is computed securely.

For both of the following problems, we have the option of achieving ε differential privacy, via protocols based on

the Laplace mechanism, as well as (ε, δ) -differential privacy with slightly higher utility, via protocols based on the Gaussian mechanism.

Low Rank Approximation. We are given a data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$. Our goal is to compute an orthogonal projection $\mathbf{X} \in \mathbb{R}^{d \times r}$ minimizing the error $OPT = \min_{\mathbf{X}} \|\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. In the LTM, we give a (ε, δ) differentially private mechanism with multiplicative error $(1 + \alpha)$ and additive error $\tilde{O}(kd^{3/2}\alpha^{-3}\varepsilon^{-1}\log\delta^{-1})$. We also present a ε differentially private mechanism with multiplicative error $(1 + \alpha)$ and additive error $\tilde{O}(k^3d^{3/2}\alpha^{-6}\varepsilon^{-1})$. State of the art central model mechanisms (Dwork et al., 2014; Hardt & Roth, 2012) have no multiplicative error and additive error $O(k\sqrt{d}\cdot\varepsilon^{-1})$, which is optimal in terms of k and d . For the local model, we are only aware of the result by (Arora et al., 2018), which has a multiplicative error of $O(\log n \log^3 d)$ and additive error $O(k^2n \cdot \text{dpoly}(\varepsilon^{-1}, \log\delta^{-1}))$. There are several related questions we will survey in more detail in Appendix F.

Ridge Regression. We are given a data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, a target vector $\mathbf{b} \in \mathbb{R}^n$ and regularization parameter $\lambda > 0$. Our goal is to find $\mathbf{x} \in \mathbb{R}^d$ minimizing $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \cdot \|\mathbf{x}\|^2$, where $\|\cdot\|$ denotes the Euclidean norm of a vector. Each row of the data matrix, as well as the corresponding entry of \mathbf{b} resides with a client. Assume that $\lambda \geq \text{poly}(\varepsilon^{-1}, d, \log 1/\delta)$. In the LTM, there exists an (ε, δ) differentially private mechanism that can compute an $\hat{\mathbf{x}}$ with error $(1 + o(1)) \cdot (\|\mathbf{A}\mathbf{x}_{opt} - \mathbf{b}\|^2 + \lambda \cdot \|\mathbf{x}_{opt}\|^2) + \text{poly}(\varepsilon^{-1}, d, \log 1/\delta)$. Unlike the low rank approximation problem, we are not aware of other papers achieving comparable bounds in any model. Regression has been studied in the context of empirical risk minimization and some results also exist for comparing closeness of the computed predictor $\hat{\mathbf{x}}$ and that optimal \mathbf{x}_{opt} . While we view the low rank approximation mechanisms as our main result, we nevertheless believe this work can serve as a starting point for achieving bounds for ridge regression in terms of cost in any model including the central model.

Table 1 provides a comparison of the utility bounds for the problems in the local, central, shuffle and LTM models. To the best of our knowledge, there are no results on low rank approximation in the shuffle model.

As a warm-up to illustrate the potential inherent in the LTM, and to also highlight the comparison to the other models, we consider the basic problem of computing histograms and frequency of moments. More related work on problems related to low rank approximation and regression, as well as applications of linear sketching to differentially private mechanisms can be found in Appendix F.

Table 1. SOTA utility bounds for various models for frequency estimation and low rank approximation. Our bounds for low rank approximation have a multiplicative error $O(1 + \alpha_S)$, which we omitted for space.

	Local	Shuffle	LTM (this work)	Central
Frequency Est	$\tilde{O}(\sqrt{n})$ (Bassily et al., 2020) $\Omega(\sqrt{n})$ (Bassily & Smith, 2015)	$\tilde{\Theta}(\varepsilon^{-1})$ (Ghazi et al., 2021)	$\tilde{O}(\varepsilon^{-1})$	$\tilde{\Theta}(\varepsilon^{-1})$
Low-Rank App	$(\tilde{O}(\log n \log^3 d), O(k^2 nd))$ (Arora et al., 2018) $\Omega(\sqrt{n})$ (Bassily & Smith, 2015)	N.A.	$\tilde{O}(\frac{kd^{3/2}}{\alpha_S^3} \varepsilon^{-1})$	$\tilde{\Theta}(k\sqrt{d}\varepsilon^{-1})$ (Dwork et al., 2014)

2. Preliminaries

Notation. Column vectors are written in bold lowercase letters \mathbf{b} and matrices in bold uppercase letters \mathbf{A} . The transpose operator over vectors and matrices is \mathbf{b}^T and \mathbf{A}^T . For any vector \mathbf{b} , we denote by $\|\mathbf{b}\| = \sqrt{\sum_i b_i^2}$ its ℓ_2 -norm. For any matrix \mathbf{A} , we denote by $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j \mathbf{A}_{i,j}^2}$ its Frobenius norm. We denote the inner product between vectors $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}$. Two vectors \mathbf{a}, \mathbf{b} are orthogonal if their inner product is 0. An orthogonal matrix \mathbf{A} is a real square matrix where the columns have unit Euclidean norm and are pairwise orthogonal.

Subspace Preserving Sketches. We consider subspace-preserving sketches in this work, which are sparse Johnson-Lindenstrauss transforms. Of particular interest for us are algorithms that compute a subspace approximation of A without prior knowledge of A . Such algorithms are known as oblivious subspace embeddings. We will use a family of such embeddings known as *oblivious sparse norm-approximating projections* originally due to (Nelson & Nguyen, 2013), but since improved upon in several subsequent works.

Definition 2.1 (OSNAP (Nelson & Nguyen, 2013)). Let $\mathcal{D}_{m,n,s}^{\text{sketch}}$ be a distribution over random matrices $\{-1, 0, 1\}^{m \times n}$, where s entries of each column, chosen uniformly and independently for each column, are set to ± 1 with equal probability, and all other entries are set to 0. If $s = 1$, we write $\mathcal{D}_{m,n}^{\text{sketch}}$. We say that $S \sim \mathcal{D}_{m,n,s}^{\text{sketch}}$ is an (α, β, s, m) OSNAP for a $n \times d$ orthogonal matrix \mathbf{W} if with probability $1 - \beta$, for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and some precision parameter α

$$|\mathbf{a}^T \mathbf{W}^T \mathbf{S}^T \mathbf{S} \mathbf{W} \mathbf{b} - \mathbf{a}^T \mathbf{W}^T \mathbf{W} \mathbf{b}| \leq \alpha \cdot \|\mathbf{W} \mathbf{a}\| \|\mathbf{W} \mathbf{b}\|.$$

We give bounds on available choices of (α, β, s, m) for subspaces of rank k in Table 3 in the appendix.

Differential Privacy. Differential privacy (Dwork et al., 2006b) offers privacy guarantees to individuals contributing

their data to some randomized algorithm. We say that two datasets are neighboring if one can be obtained from the other by the replacement of a single individual with another individual.

Definition 2.2 (Differential Privacy in the central model (Dwork et al., 2006b)). Let $\varepsilon \geq 0$ and $\delta \in [0, 1]$. A randomized mechanism $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Y}$ is (ε, δ) -differentially private, if for all neighboring data sets $x, x' \in \mathcal{X}$ and all outputs $S \subseteq \mathcal{Y}$ it holds that

$$\Pr[\mathcal{M}(x) \in S] \leq e^\varepsilon \Pr[\mathcal{M}(x') \in S] + \delta,$$

where the probabilities are over the randomness of \mathcal{M} . If $\delta = 0$, then we simply say that the mechanism is ε -differentially private.

A probability distribution F is *infinitely divisible* if for any positive integer n , there is a random variable S_n which can be written as $S_n = \sum_{i \in [n]} X_i$ such that each of the X_i are random variables, independent and identically distributed, and S_n has the probability distribution F .

3. Privacy Guarantees in the LTM

We consider a model in which the trusted component can perform any public linear transformation of the inputs. The model includes three algorithms: (1) $R : \mathcal{X} \rightarrow \mathcal{Y}$ is a randomized encoder that takes a single user's data and outputs a randomized message, (2) $T : \mathcal{Y}' \rightarrow \mathcal{Y}^*$ is the idealized trusted component that performs a public linear transformation of its inputs, and (3) $A : \mathcal{Y}^* \rightarrow \mathcal{Z}$ is an analysis function that takes the results messages and estimates some function from these messages.

Note that standard definitions of differential privacy in the shuffle model only require $T(R(x_1), \dots, R(x_n))$ to be differentially private. This implicitly assumes that all clients are honest and do not collude with the adversary, in particular they are assumed not to leak the output of their randomizers publicly. This implies that existing definitions of differential privacy in the shuffle model could be satisfied even by (artificial) mechanisms in which a single client adds the whole

noise while the others do not randomize their messages at all. This is clearly a very weak privacy guarantee: in a setting in which a large number of clients participate in a differentially private data analysis, it is unrealistic to assume that the adversary does not control even a single client. Luckily, to the best of our knowledge, no proposed mechanisms in the literature suffer from these vulnerabilities, still this counterexample shows that the existing definition is not robust enough, and we therefore formalize a notion of differential privacy where we explicitly tolerate that a bounded number of clients might collude with the adversary. This refined definition could be applied to both the shuffle model and our model to ensure that the privacy guarantee holds even when a bounded number of clients might collude with the adversary. Similar observations were also made by (Talwar et al., 2023).

We assume that up to $t' < n$ clients and $t < k$ servers may collude in the semi-honest model, sharing information with each other while following the protocol.

Definition 3.1 (Trusted Computation Model for Differential Privacy). A tuple of algorithms $P = (R, T, A)$ is (ϵ, δ) -differentially private given corrupt clients C_{cor} if the output $\Pi_R(x_1, \dots, x_n) = T(R(x_1), \dots, R(x_n))$, as well as corrupted parties' randomizer output $R(x_i)$ for all $x_i \in C_{cor}$ satisfy (ϵ, δ) -differential privacy.

Multi-Central Model of Differential Privacy. (Steinke, 2020) introduce the multi-central model, which exactly defines this split-trust model that instantiates the trusted computation model above. They allow a semi-honest adversary that honestly follows the protocol to corrupt up to all but one servers and all but one client. We also operate in the same semi-honest setting, but define parameters t and t' for the number of tolerated server and client corruptions.

The view $\text{View}_{C_{cor}, S_{cor}}^{R, \Pi, A}(x)$ of the adversary consists of all the information available to the corrupted clients C_{cor} and servers S_{cor} plus the final output of the honest, uncorrupted servers. This view excludes only internal information of the trustworthy servers and clients. The resulting protocol is (ϵ, δ) -differentially private in the multi-central model if the adversary's view is (ϵ, δ) -indistinguishable from the output on a neighboring dataset.

Figure 1 further illustrates what is contained in the adversary's view. If one server and one client are corrupted, the adversary's view contains all incoming and outgoing messages from that server and that client marked in red in the figure.

Definition 3.2 (Instantiation of Trusted Computation Model for Differential Privacy with MPC). Let Π be a k -party MPC protocol that computes $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with perfect security, tolerating t corruptions. A tuple of algorithms $P = (R, \Pi, A)$ is (ϵ, δ) -differentially private if for all coalitions

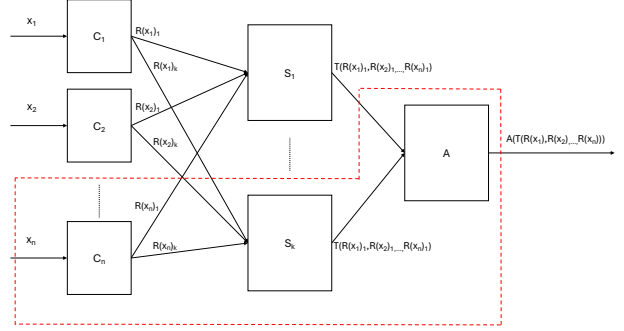


Figure 1. Adversary's View

Algorithm 1 Linear Transformation Model

- 1: **Input:** Individual input vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ representing values in \mathbb{R}^d
- 2: Each client $i \in [n]$ locally computes $R(\mathbf{x}_i)$ on their input \mathbf{x}_i .
- 3: Each client secret shares the resulting value using a linear secret sharing scheme, and sends one share to each server.
- 4: The servers jointly compute function T on the matrix, which results from concatenating all n vector secret shares.
- 5: The resulting matrix product is published and then taken as input to any analysis function A .
- 6: **Output:** Output $A(T(R(\mathbf{x}_1), \dots, R(\mathbf{x}_n)))$

S_{cor} of up to $t < k$ corrupt servers and for all possible S , all coalitions C_{cor} of $t' < n$ corrupt clients and all neighboring datasets x and x' :

$$\Pr[\text{View}_{S_{cor}, C_{cor}}^{R, \Pi, A}(x) \in S] \leq e^\epsilon \cdot \Pr[\text{View}_{S_{cor}, C_{cor}}^{R, \Pi, A}(x') \in S] + \delta$$

where the probability is over all the randomness in the algorithms (R, Π, A) .

Algorithm 1 describes how clients and servers jointly compute the linear transformation, allowing the transformed data to be published and used for analysis.

With this definition, we can give our first result showing that differential privacy is retained given a bounded number of corruptions. The proof can be found in Appendix D.

Lemma 3.3. *If $P = (R, T, A)$ is (ϵ, δ) -differentially private with t' corrupt clients, and if Π is a perfectly secure k -party MPC protocol that computes T correctly while tolerating t corrupt servers, then $P = (R, \Pi, A)$ is (ϵ, δ) -differentially private.*

3.1. Frequency Moments

To illustrate the possibilities inherent to the LTM, we consider a simple application by way of estimating frequency moments. For simplicity, assume in this section that all clients are trustworthy, i.e. $t' = 0$, but the arguments can be straightforwardly extended to deal with arbitrary values of t' . Here, each client is given a number $[-\Delta, \Delta]$ and our goal is to estimate $F_k = \sum_{i=1}^n |x_i|^k$. In the local model, even if all entries are either 0 or 1, it is not possible to estimate F_1 without incurring an additive error of the order \sqrt{n} , that is the estimated value $\tilde{F}_1 = F_1 \pm O(\sqrt{n})$ for any privacy preserving local mechanism (Bassily & Smith, 2015). In contrast, simply computing F_k and adding an appropriate amount of noise yields an additive error that only depends on Δ, ε and δ .

We now consider the LTM. Let \mathbf{x}^k denote the vector of client entries with $\mathbf{x}_i^k = |x_i^k|$. We observe that $F_k = f_k(\mathbf{x}) = 1^T \mathbf{x}^k$. Suppose that every client i samples g_i from a distribution, and we denote $\tilde{\mathbf{x}}$ the vector of noisy client entries with $\tilde{\mathbf{x}}_i^k = |x_i^k| + g_i$. To guarantee differential privacy, we can use a differentially private additive noise mechanism $1^T(\tilde{\mathbf{x}}^k) = f_k(\mathbf{x}) + G$, where the noise distribution is infinitely divisible and $G = \sum_{i \in [n]} g_i$.

Many different infinitely divisible distributions can be used at this stage. One example is the geometric distribution, since the sum of negative binomial random variables follows a geometric distribution, which can be used in differentially private mechanisms (Ghazi et al., 2020). The Laplace distribution, which is commonly used for ε -DP mechanisms, is also infinitely divisible, since a zero-centered Laplace distribution can be formulated as a sum of differences between Gamma distributions. Other possible options are stable distributions, a subset of infinitely divisible distributions for which each g_i and G follow the same distribution, such as Gaussian or Cauchy. We provide parameter choices and an error bound examples for the Gaussian and Laplace distributions in Appendix B.

3.2. Privacy Preserving Mechanism for Oblivious Johnson Lindenstrauss Transforms

We instantiate the trusted computation model for differential privacy for specific choices for the functions in tuple (R, T, A) . We will show how to guarantee differential privacy for the tuples $(R_D^{\text{dense}}, T_S^{\text{dense}}, A)$ and (R_D, T_S, A) , which we will refer to as the dense sketching mechanism and the sparse sketching mechanism, respectively.

Randomizer function $R_D : \mathbb{R}^d \rightarrow \mathbb{R}^{ds}$. The randomizer takes as input a d -dimensional row vector \mathbf{x} , copies it s times and concatenates the result to form \mathbf{x}^s . Then noise is sampled independently from distribution D and added to each of the ds entries in the resulting vector.

$$R_D^{\text{dense}}(\mathbf{x}) = \mathbf{x}^s + \mathbf{g}, \mathbf{g} \sim D^{ds} \quad (1)$$

We define a second alternative to this randomizer function, which instead outputs 0^{ds} if $n < 8m \ln(dm/\delta) + t'$:

$$R_D(\mathbf{x}) = \begin{cases} 0^{ds} & n < 8m \ln(dm/\delta) + t' \\ R_D^{\text{dense}}(\mathbf{x}) & \text{else} \end{cases} \quad (2)$$

Transformation function $T_S : \mathbb{R}^{n \times ds} \rightarrow \mathbb{R}^{m \times d}$. This function takes as a parameter a matrix $\mathbf{S} \in \{-1, 0, 1\}^{m \times n}$ and takes as input a matrix $\mathbf{A} \in \mathbb{R}^{n \times ds}$, which corresponds to the vertical concatenation or stacking of n outputs from a randomizer function R . We parse \mathbf{A} as the horizontal concatenation of matrices $\mathbf{A}_i \in \mathbb{R}^{n \times d}$ for $i \in [s]$. Let s be the largest number of non-zero entries in a column of \mathbf{S} . Then \mathbf{S} can be decomposed such that $\mathbf{S} = \frac{1}{\sqrt{s}} \sum_i \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{m \times n}$ is a matrix with only one non-zero entry per column, and the assignment of non-zero entries to indices i is done uniformly at random. T_S outputs the sum of matrix products $\frac{1}{\sqrt{s}} \sum_{i \in [s]} \mathbf{S}_i \mathbf{A}_i$.

$$T_S(\mathbf{A}) = \frac{1}{\sqrt{s}} \sum_{i \in [s]} \mathbf{S}_i \mathbf{A}_i \quad (3)$$

We also define $T_S^{\text{dense}} : \mathbb{R}^{n \times dm} \rightarrow \mathbb{R}^{m \times d}$, where \mathbf{S} is a dense matrix rather than a sparse one, so we set $s = m$. Then \mathbf{S} can be decomposed such that $\mathbf{S} = \frac{1}{\sqrt{m}} \sum_i \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{m \times n}$ is a matrix with only one non-zero entry per column and exactly n/m non-zero entries per row.

Analysis function A . A takes a matrix in $\mathbb{R}^{m \times d}$, as well as any public values, and outputs some data analysis. Examples include low rank approximation and linear regression.

We now evaluate the noise distribution necessary to guarantee differential privacy in this model. We find that for sufficiently large n , we can apply the infinite divisibility of distributions to divide the noise into a number of pieces corresponding the number $(n - t')/2m$ of honest clients expected to contribute to any entry of a column in the resulting noisy matrix product. The concentration bound, formulated in Lemma D.1 and used to bound this number of honest clients, contributes to the necessary value of δ . Composition theorems can be applied on the columns of the resulting matrix product to yield the final differential privacy guarantee. We will later see that adding suitable noise to data results in only a small error in the resulting error for regularized linear regression and low rank approximation.

We are now ready to state our privacy guarantees in the LTM.

Theorem 3.4. Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $m \in [n]$, $t' < n$, $s \in [m]$. Let $\mathbf{S} \sim \mathcal{D}_{m,n,s}^{\text{sketch}}$ with s non-zero entries per column. Also let noise g_i be sampled from a symmetric distribution D such that $\sum_{i \in [(n-s-t')/(2m)]} g_i = G$. G follows a distribution D' that defines an additive noise mechanism that satisfies $(\varepsilon/(sd), \delta/(sd) - m \exp(-(n-s-t')/8m))$ -DP for function $f(\mathbf{A}) = (\mathbf{SA})_j + \mathbf{n}$ for $\mathbf{n} \sim D'^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of \mathbf{SA} . Then as long as input values are bounded above by η , the sparse sketching mechanism (R_D, T_S, A) , as defined by equations 2 and 3 and for any A , is (ε, δ) -differentially private in the trusted computation model for differential privacy with t' corrupt clients.

One additive mechanism to which this can easily be applied is the Gaussian mechanism. To do so, D is a normal distribution with variance:

$$\sigma^2 = \frac{4s^3\eta^2 \ln \left(\frac{1.25s}{\delta/d - m \exp(-\frac{(n-s-t')}{8m})} \right) md^2}{\varepsilon^2(n-s-t')}. \quad (4)$$

If we would like to obtain a result for mechanisms with $\delta = 0$, we can instead use R_D^{dense} and use a dense sketching matrix $\mathbf{S} \in \{-1, 1\}^{m \times n}$.

Theorem 3.5. Let $\varepsilon \geq 0$, $\delta \in [0, 1)$, $m \in [n]$, $t' < n$, $s \in [m]$. Let $\mathbf{S} \sim \{-1, 1\}^{m \times n}$ be a dense, uniformly sampled matrix. Also let R add noise g_i sampled from a symmetric distribution D such that $\sum_{i \in [n/m-t']} g_i = G$. G follows a distribution D' that defines an additive noise mechanism that satisfies $(\varepsilon/(md), \delta/(md))$ -DP for function $f(\mathbf{A}) = (\mathbf{SA})_j + \mathbf{n}$ for $\mathbf{n} \sim D'^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of \mathbf{SA} . Then as long as input values are bounded above by η , dense sketching mechanism $(R_D^{\text{dense}}, T_S^{\text{dense}}, A)$, as defined by equations 1 and 3 and for any A , is (ε, δ) -differentially private in the LTM for differential privacy with t' corrupt clients.

To instantiate this second theorem, D can be chosen as the distribution formulated as the difference of two samples of gamma distribution

$$\Gamma \left(\frac{1}{n/m - t'}, 2\eta m^2 d / \varepsilon \right). \quad (5)$$

The proofs can be found in Appendix D. Corollary 3.6 below follows directly from Theorem 3.4 and Lemma 3.3.

Corollary 3.6. Let (R_D, T_S, A) be the tuple of algorithms above. If T_S is computed correctly using an MPC protocol Π_S with perfect security tolerating t semi-honest corruptions, then the tuple of algorithms (R_D, Π_S, A) is (ε, δ) -differentially private with t' corrupt clients.

3.3. Cryptographic Assumptions and Relations to the Shuffle Model

The LTM can be securely distributed using simple cryptographic techniques for *secure multiparty computation* (MPC), such as *linear secret-sharing* (LSS). MPC ensures that no adversary corrupting all but one server learns more than the computation's output. (Dwork et al., 2006a) first explored combining differential privacy and MPC, focusing on distributed noise generation requiring server interaction. (Cheu & Yan, 2023) establish lower bounds for non-interactive multi-server mechanisms, while (Damgård et al., 2023) propose an interactive MPC protocol for selection in distributed trust models.

The linear transforms can be non-interactively computed locally by servers and are also known to all participating parties, that is, once initiated, the output of the protocol is fully deterministic. Linear transforms can be computed non-interactively by servers and are publicly known, making the protocol fully deterministic once initiated. This enables implementing the LTM with multiple central servers under the minimal assumption that at least one server remains honest. Using LSS has key advantages: clients send a single message per server proportional to the input size and sparsity s of the transformation matrix \mathbf{S} , servers communicate only with the data analyst, and the total workload scales only with the number of servers and matrix sparsity. Since LSS supports information-theoretic security, our system remains secure even against quantum adversaries. We describe the type of secret-sharing scheme used for secure linear transformations in Appendix H.

Secure aggregation and the shuffle model are two instantiations of intermediate trust models for differential privacy. Secure aggregation (Goryczka & Xiong, 2017; Bonawitz et al., 2017; Mugunthan et al., 2019; Apple & Google, 2021; Talwar et al., 2023) is a special case of the LTM, where the linear transformation applied is a sum, which is useful in federated learning. The shuffle model (Bittau et al., 2017; Erlingsson et al., 2019; Cheu et al., 2019) can be distributed among multiple servers using cryptographic protocols such as mixnets.

Implementing shuffles requires servers to communicate over a chain, increasing latency. In MPC-based shuffles, each server must sequentially shuffle and randomize the ciphertexts, leading to a round complexity proportional to the number of servers involved. In contrast, the LTM enables all servers to perform their computations in parallel, resulting in lower round complexity and latency.

Regarding the computational cost, servers in the LTM only perform simple linear operations on secret shares, such as matrix-vector multiplications, which are highly efficient. On the other hand, MPC-based shuffles typically rely on

mix-nets or similar constructions, where each server performs, for every data point, multiple expensive public-key operations (e.g., modular exponentiations with large moduli and exponents). This introduces an overhead that is orders of magnitude higher compared to simple matrix multiplications.

The total amount of data transmitted is comparable between the two models, but the size of individual messages differs substantially. In the LTM, secret shares can be as small as the plaintext itself; for example, when working with 64-bit types, each share is just 64 bits per server. In contrast, MPC-based shuffles transmit ciphertexts that are significantly larger due to the inherent expansion of public-key encryption. For example, even with elliptic curve ElGamal, each ciphertext would be at least 512 bits. As a result, the total communication volume in MPC shuffles is larger.

Finally, shuffles require computationally intensive public-key cryptography, which in turn requires computational assumptions. In contrast, LSS enables efficient, unconditional security and deterministic computation.¹

4. Numerical Linear Algebra in LTM

In this section we give the utility guarantees of our mechanism. We start with giving generic distortion bounds relating the spectrum of the noisy matrix to the spectrum of the original non-private matrix. Parameterizations of this mechanism depend on the underlying class of subspace embeddings. In this utility analysis, we limit ourselves to OSNAPs with m and s chosen according to (Cohen, 2016); other trade-offs are possible and can be found in Appendix C. Applications to specific problems such as regression and low rank approximation are given towards the end of the section.

Here we present the formal utility statements for ridge regression and low rank approximation. All proofs can be found in the Appendix E.

We will use the following spectral bounds for both linear regression and low rank approximation. We believe that there may be further applications and that the bounds themselves are therefore of independent interest.

Lemma 4.1. *Let $\mathbf{S} \in \frac{1}{\sqrt{s}} \cdot \{-1, 0, 1\}^{m \times n}$ be an (α, β, m, s) OSNAP with $\mathbf{S} = \frac{1}{\sqrt{s}} \sum_{i \in [s]} \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{n \times m}$ has exactly one non-zero entry per row. Let $\mathbf{G} = \sum_{i \in [s]} \mathbf{S}_i \mathbf{G}_i$ where every matrix $\mathbf{G}_i \in \mathbb{R}^{n \times d}$ has independent Gaussian entries $\mathcal{N}(0, \sigma^2)$. Further, let V be a set of d -dimensional vectors lying in a k -dimensional subspace. Then with probability at least $1 - \beta$ for some absolute*

¹The linear transformations we use are random but public and sampled before the computation.

constant η

$$\sup_{\mathbf{x} \in V} \frac{1}{s} \|\mathbf{G}\mathbf{x}\|_2^2 \leq \eta \cdot 2 \frac{n}{m} \sigma^2 \cdot \|\mathbf{x}\|^2 \cdot (\sqrt{(k + \log 1/\beta) \cdot m} + (k + \log 1/\beta)).$$

and

$$\frac{1}{s} \|\mathbf{G}\|_F^2 \leq \eta \cdot 2n\sigma^2 \cdot d \cdot \log 1/\beta$$

Related, though slightly weaker bounds can also be derived for the Gamma/Laplace mechanism. While it is possible to derive bounds for sparse Rademacher sketches, we focus on the dense constructions as these yield ε differentially private mechanisms (as opposed to (ε, δ) differentially private mechanisms).

Lemma 4.2. *Let $\mathbf{S} \in \frac{1}{\sqrt{s}} \cdot \{-1, 0, 1\}^{m \times n}$ be an (α, β, m, m) OSNAP with $\mathbf{S} = \frac{1}{\sqrt{m}} \sum_{i \in [s]} \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{m \times n}$ has exactly one non-zero entry per row and n/m non-zero entries per column. Let $\mathbf{L} = \frac{1}{\sqrt{m}} \sum_{i \in [m]} \mathbf{S}_i \mathbf{\Gamma}_i$ where every matrix $\mathbf{\Gamma}_i \in \mathbb{R}^{n \times d}$ has independent entries $X_i - Y_i$, where $X_i, Y_i \sim \Gamma(m/n, b)$. Then with probability at least $1 - \beta$ for some absolute constant η and*

$$\|\mathbf{L}\|_F^2 \leq \eta d m^2 b^2 \log^2 \frac{m}{\beta}$$

We first begin with our utility results for low rank approximation. Using the Gaussian mechanism, we obtain the following guarantee.

Theorem 4.3. *Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $t' < n/2$. Let $\mathbf{S} \sim \mathcal{D}_{m,n}^{\text{sketch}}$. Define A_k as an algorithm that computes $\arg\min_{\mathbf{X} \text{ rank } k, \mathbf{X} \in \mathbb{R}^{n \times d}} \|\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T\|_F^2$. An instantiation of the trusted computation model (R_D, T_S, A_k) specified in Equations 2 and 3, using noise distribution $D \sim \mathcal{N}(0, \sigma^2)$ according to Equation 4, computes an orthogonal projection \mathbf{X}' that with probability $1 - \beta$ achieves a multiplicative error $(1 + O(\alpha_S))$ and additive error $O\left(\frac{k d^{3/2}}{\alpha_S^3} \log^{3/2}(1/\beta) \varepsilon^{-1} \log^{1/2}(1/\delta)\right)$ for the low rank approximation problem.*

Using dense Rademacher sketches and the Gamma/Laplace mechanism, we instead obtain the following guarantee.

Theorem 4.4. *Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $t' < n/2$. Let $\mathbf{S} \sim \mathcal{D}_{m,n}^{\text{sketch}}$. Define A_k as an algorithm that computes $\arg\min_{\mathbf{X} \text{ rank } k, \mathbf{X} \in \mathbb{R}^{n \times d}} \|\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T\|_F^2$. An instantiation of the trusted computation model $(R_D^{\text{dense}}, T_S^{\text{dense}}, A_k)$ specified in Equations 1 and 3, using noise distribution D as the difference between two independent Γ random variables parameterized by Equation 5, computes an orthogonal projection \mathbf{X}' that with probability $1 - \beta$ with multiplicative error $(1 + O(\alpha_S))$ and additive error $O\left(\frac{k^3 d^{3/2}}{\alpha_S^6} \varepsilon^{-1} \log \frac{m}{\beta}\right)$.*

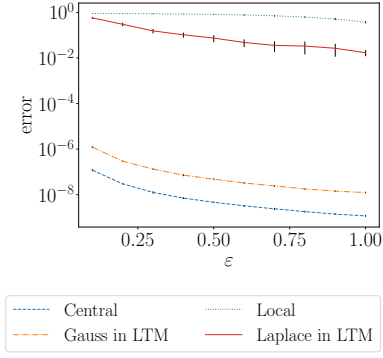


Figure 2. Error ψ as function of privacy parameter ε , with $n = 500000$, $d = 50$ and $k = 5$. The standard deviations are depicted by the vertical black lines and the y -axis is logarithmic in ψ .

For ridge regression, we achieve the following bound.

Theorem 4.5. *Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $t' < n$ and σ be chosen as described in Theorem D.3. Let $\mathbf{S} \sim \mathcal{D}_{m,n}^{\text{sketch}}$. Define A_λ that performs linear regression on the sketched noisy inputs, outputting $\arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2$, where the error is measured by $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2$. An instantiation of the LTM (R_D, T_S, A_λ) specified in Equations 2 and 3, using noise distribution $D \sim \mathcal{N}(0, \sigma^2)$ according to Equation 4, which tolerates t' corrupted clients, computes linear regression parameters for any $\alpha_S > 0$ with probability $1 - \beta$ with multiplicative error $1 + \tilde{O}(\alpha_S + p + p^2)$ and additive error $\tilde{O}(p + p^2)$, where $p = \frac{d^3 \alpha_S^{-5} \log^5 \frac{1}{\delta} \cdot \varepsilon^{-2} \log \frac{1}{\delta}}{\lambda}$.*

5. Experimental Evaluation

In this section we present an empirical evaluation of differential privacy mechanisms for the low-rank approximation problem in the central, the local and LTM models. The local and LTM mechanisms we consider are algorithmically identical up to their noise rate. An immediate question thus is, for which parametrizations of the noise rate we still achieve acceptable utility. In addition we implement a Gamma/Laplace based LTM mechanism, to evaluate the price of achieving pure ε -differential privacy. Additionally, we experimentally evaluated the minimal computational overhead introduced by the use of MPC for distributing the execution of the LTM. In Appendix G we provide additional experiments for the ridge regression problem as well as a more in depth description of the low-rank approximation experiments.

Setup. Our Gaussian mechanism in the LTM adds noise sampled from $\mathcal{N}(0, \sigma^2)$ to every entry in input data matrix \mathbf{A} . The variance σ^2 depends on privacy parameters ε and δ and is chosen proportional to n^{-p} for $p \in [0, 1]$. This enables us to interpolate between the local model ($p = 0$) and the LTM ($p = 1$). The Laplace mechanism in the LTM on

the other hand adds the sum of m samples from $\text{Lap}(0, 1/\varepsilon)$ to every entry in \mathbf{SA} . As a baseline for the central model, we implemented MOD-SULQ (Chaudhuri et al., 2012), an approach where $\mathbf{A}^T \mathbf{A}$ is perturbed by adding Gaussian noise. We measure the error in terms of the excess risk;

$$\psi = \frac{\|\mathbf{A} - \mathbf{A}\mathbf{X}'\mathbf{X}'^T\|_F^2 - \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F^2}{n},$$

where \mathbf{X}_{OPT} denotes the optimal solution and \mathbf{X}' denotes the solution after adding noise to the training data. We vary the privacy parameter ε and k based on the dataset at hand. For each dataset we measure ψ for the central model, the Laplace mechanism and the Gaussian mechanism with $p \in \{1, 0.9, 0.8, 0.7, 0.6, 0.5, 0\}$, by reporting the average ψ over 20 runs of the algorithms per dataset.

Datasets. We evaluate our mechanism on synthetic and real datasets, enabling us to vary the number of sampled points n and interpolating between datasets of different sizes.

The synthetic datasets have a large spectral gap from the k th to the $(k + 1)$ st singular value with random bases, which emphasizes the performance difference between the various private mechanisms. The parameters of the four real-world datasets from the UC Irvine Machine Learning Repository are given in Table 2. For more thorough descriptions on the datasets, see Appendix G.2.

Utility. Figure 3 shows our results for two privacy regimes ($\varepsilon \in \{0.1, 0.5\}$) with synthetic data and Table 2 shows the error for $\varepsilon = 0.05$ with real-world datasets. In all chosen parameter settings for synthetic data, we observe that as n grows, the error in the LTM asymptotically approaches the error in the central model, both on real and synthetic data sets (see Figure 3 and Table 2). Table 2 shows that on real-world datasets, our Gaussian approach performs significantly better than the Gaussian mechanism in the local model and is very close to the central MOD-SULQ (Chaudhuri et al., 2012) mechanism. Both with synthetic and real-world datasets, our Laplace approach performs better than the local model, though not as significantly as the Gaussian approach. However, this is the price of achieving pure ε -differential privacy with $\delta = 0$. Finally Figure 2 shows our results for a varying privacy budget ε . As expected, the error in the LTM stays between the one in the local and the central model, while it is decreasing in all models.

Running Time. The only overhead of the LTM over the local model is the execution of the linear transform in MPC, and comes from distributing $\mathbf{A} + \mathbf{G}_i \in \mathbb{R}^{n \times d}$ among the servers for $i \in [s]$ and then performing matrix multiplications $\mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ in MPC, where $\mathbf{S}_1, \dots, \mathbf{S}_s$ is a decomposition of some $\mathbf{S} \in \mathbb{R}^{m \times n}$ chosen according to (Nelson & Nguyen, 2013; Cohen, 2016) with sparsity s . We imple-

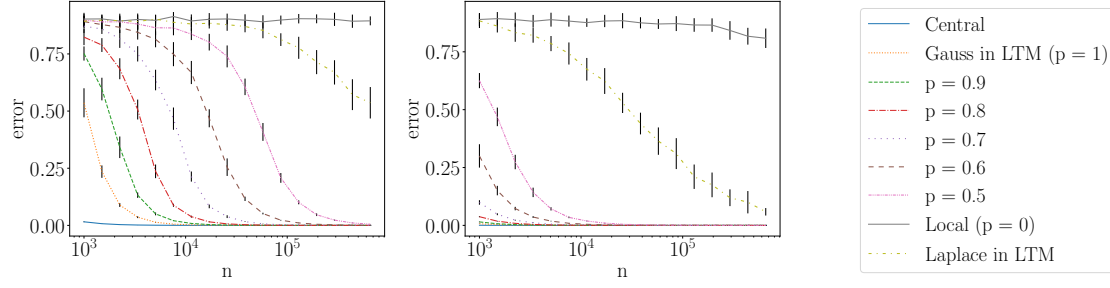


Figure 3. Plots depicting the asymptotic behavior of error ψ for $\varepsilon \in \{0.1, 0.5\}$ (left, right), with $d = 50$ and $k = 5$. The gray line depicts the error of the local mechanism, and the orange and lime lines depict our approach using Gaussian and Laplacian noise respectively. The other lines resemble different values of p when using Gaussian noise. The standard deviations are depicted by the vertical black lines and the x -axis is logarithmic in the number of clients n .

Table 2. Experimental evaluation of error ψ on real-world datasets. Here we are in the setting where $\varepsilon = 0.05$. For a more elaborate version of the table, including standard deviation, see Table 9 in the appendix.

Dataset	n	d	k	Local	LTM Laplace	LTM Gauss	Central
Power (Hebrail & Berard, 2012)	2049280	6	3	1.2×10^{-2}	4.0×10^{-3}	2.1×10^{-6}	1.8×10^{-6}
Elevation (Kaul, 2013)	434874	2	1	1.2×10^{-1}	1.5×10^{-2}	4.7×10^{-8}	4.5×10^{-8}
Ethylene (Fonollosa, 2015)	4178504	18	5	5.9×10^{-4}	5.8×10^{-4}	2.4×10^{-6}	7.9×10^{-6}
Songs (Bertin-Mahieux, 2011)	515345	89	15	5.6×10^{-7}	5.6×10^{-7}	5.6×10^{-7}	5.6×10^{-7}

mented our approach in MP-SPDZ (Keller, 2020) and provide our results in Appendix G.1. Notably, for $s = 1$ even with one million clients, the computation on each server lasts less than 2 seconds, which shows that securely distributing in the LTM using MPC techniques is a practically relevant approach.

Acknowledgments

Jakob Burkhardt and Chris Schwiegelshohn are partially supported by the Independent Research Fund Denmark (DFF) under a Sapere Aude Research Leader grant No 1051-00106B. This work was also supported by the European Research Council (ERC) under the European Unions’ Horizon 2020 research and innovation programme under grant agreement No 803096 (SPEC); the Danish Independent Research Council under Grant-ID DFF-2064-00016B (YOSO); DFF-3103-00077B (CryptoDigi) and the Danish Independent Research Council under Grant-ID 1051-001068 (ACBD).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

- Achlioptas, D. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003. doi: 10.1016/S0022-0000(03)00025-4. URL [https://doi.org/10.1016/S0022-0000\(03\)00025-4](https://doi.org/10.1016/S0022-0000(03)00025-4).
- Apple and Google. Exposure Notification Privacy-preserving Analytics (ENPA). White paper, 2021.
- Arora, R., Braverman, V., and Upadhyay, J. Differentially private robust low-rank approximation. In *Advances in Neural Information Processing Systems*, 2018.
- Arora, S., Hazan, E., and Kale, S. A fast random sampling algorithm for sparsifying matrices. In Díaz, J., Jansen, K., Rolim, J. D. P., and Zwick, U. (eds.), *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006, Barcelona, Spain, August 28-30 2006, Proceedings*, volume 4110 of *Lecture Notes in Computer Science*, pp. 272–279. Springer, 2006. doi: 10.1007/11830924_26. URL https://doi.org/10.1007/11830924_26.
- Balcan, M.-F., Du, S. S., Wang, Y., and Yu, A. W. An improved gap-dependency analysis of the noisy power method. In Feldman, V., Rakhlin, A., and

- Shamir, O. (eds.), *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 284–309, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR. URL <https://proceedings.mlr.press/v49/balcan16a.html>.
- Bassily, R. and Smith, A. D. Local, private, efficient protocols for succinct histograms. In Servedio, R. A. and Rubinfeld, R. (eds.), *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pp. 127–135. ACM, 2015. doi: 10.1145/2746539.2746632. URL <https://doi.org/10.1145/2746539.2746632>.
- Bassily, R., Smith, A. D., and Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, 2014.
- Bassily, R., Nissim, K., Stemmer, U., and Thakurta, A. Practical locally private heavy hitters. *J. Mach. Learn. Res.*, 21:16:1–16:42, 2020. URL <https://jmlr.org/papers/v21/18-786.html>.
- Becchetti, L., Bury, M., Cohen-Addad, V., Grandoni, F., and Schwiegelshohn, C. Oblivious dimension reduction for k -means: beyond subspaces and the johnson-lindenstrauss lemma. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2019.
- Beimel, A., Haitner, I., Nissim, K., and Stemmer, U. On the round complexity of the shuffle model. In *Theory of Cryptography: 18th International Conference, TCC 2020, Durham, NC, USA, November 16–19, 2020, Proceedings, Part II*, pp. 683–712, Berlin, Heidelberg, 2020. Springer-Verlag. ISBN 978-3-030-64377-5. doi: 10.1007/978-3-030-64378-2_24. URL https://doi.org/10.1007/978-3-030-64378-2_24.
- Bertin-Mahieux, T. YearPredictionMSD. UCI Machine Learning Repository, 2011. DOI: <https://doi.org/10.24432/C50K61>.
- Bittau, A., Erlingsson, Ú., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. Prochlo: Strong Privacy for Analytics in the Crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles, SOSP*, 2017.
- Blocki, J., Blum, A., Datta, A., and Sheffet, O. The johnson-lindenstrauss transform itself preserves differential privacy. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pp. 410–419. IEEE Computer Society, 2012. doi: 10.1109/FOCS.2012.67. URL <https://doi.org/10.1109/FOCS.2012.67>.
- Blum, A., Dwork, C., McSherry, F., and Nissim, K. Practical privacy: the sulq framework. In *Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '05*, pp. 128–138, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1595930620. doi: 10.1145/1065167.1065184. URL <https://doi.org/10.1145/1065167.1065184>.
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS*, 2017.
- Cai, T. T., Wang, Y., and Zhang, L. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5): 2825 – 2850, 2021. doi: 10.1214/21-AOS2058. URL <https://doi.org/10.1214/21-AOS2058>.
- Chaudhuri, K., Sarwate, A. D., and Sinha, K. Near-optimal differentially private principal components. In Bartlett, P. L., Pereira, F. C. N., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pp. 998–1006, 2012. URL <https://proceedings.neurips.cc/paper/2012/hash/f770b62bc8f42a0b66751fe636fc6eb0-Abstract.html>.
- Cheu, A. and Yan, C. Necessary Conditions in Multi-Server Differential Privacy. In *14th Innovations in Theoretical Computer Science Conference (ITCS)*, 2023.
- Cheu, A., Smith, A., Ullman, J., Zeber, D., and Zhilyaev, M. Distributed differential privacy via shuffling. In *Advances in Cryptology EUROCRYPT*. 2019.
- Clarkson, K. L. and Woodruff, D. P. Numerical linear algebra in the streaming model. In Mitzenmacher, M. (ed.), *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pp. 205–214. ACM, 2009. doi: 10.1145/1536414.1536445. URL <https://doi.org/10.1145/1536414.1536445>.
- Clarkson, K. L. and Woodruff, D. P. Low rank approximation and regression in input sparsity time. In Boneh, D., Roughgarden, T., and Feigenbaum, J. (eds.), *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pp. 81–90. ACM,

2013. doi: 10.1145/2488608.2488620. URL <https://doi.org/10.1145/2488608.2488620>.
- Cohen, M. B. Nearly tight oblivious subspace embeddings by trace inequalities. In Krauthgamer, R. (ed.), *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pp. 278–287. SIAM, 2016. doi: 10.1137/1.9781611974331.CH21. URL <https://doi.org/10.1137/1.9781611974331.ch21>.
- Cohen, M. B., Elder, S., Musco, C., Musco, C., and Persu, M. Dimensionality reduction for k-means clustering and low rank approximation. In Servedio, R. A. and Rubinfeld, R. (eds.), *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pp. 163–172. ACM, 2015. doi: 10.1145/2746539.2746569. URL <https://doi.org/10.1145/2746539.2746569>.
- Cramer, R., Damgård, I. B., and Nielsen, J. B. *Secure Multiparty Computation and Secret Sharing*. Cambridge University Press, 2015.
- Damgård, I., Keller, H., Nelson, B., Orlandi, C., and Pagh, R. Differentially private selection from secure distributed computing. In *The Web Conference, WWW*, 2023.
- Dimitrakakis, C., Nelson, B., Mitrokotsa, A., and Rubinfeld, B. I. P. Robust and private bayesian inference. In *Algorithmic Learning Theory*, pp. 291–305, Cham, 2014. Springer International Publishing. ISBN 978-3-319-11662-4.
- Dwork, C. and Roth, A. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*, 2014.
- Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT, 2006a*.
- Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, 2006b.
- Dwork, C., Talwar, K., Thakurta, A., and Zhang, L. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In Shmoys, D. B. (ed.), *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pp. 11–20. ACM, 2014. doi: 10.1145/2591796.2591883. URL <https://doi.org/10.1145/2591796.2591883>.
- Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., and Thakurta, A. Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity. In *Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2019.
- Fonollosa, J. Gas sensor array under dynamic gas mixtures. UCI Machine Learning Repository, 2015. DOI: <https://doi.org/10.24432/C5WP4C>.
- Foulds, J., Geumlek, J., Welling, M., and Chaudhuri, K. On the theory and practice of privacy-preserving bayesian data analysis. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, UAI’16*, pp. 192–201, Arlington, Virginia, USA, 2016. AUAI Press. ISBN 9780996643115.
- Ghazi, B., Kumar, R., Manurangsi, P., and Pagh, R. Private counting from anonymous messages: Near-optimal accuracy with vanishing communication overhead. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3505–3514. PMLR, 2020. URL <http://proceedings.mlr.press/v119/ghazi20a.html>.
- Ghazi, B., Golowich, N., Kumar, R., Pagh, R., and Velingker, A. On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In Canteaut, A. and Standaert, F. (eds.), *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part III*, volume 12698 of *Lecture Notes in Computer Science*, pp. 463–488. Springer, 2021. doi: 10.1007/978-3-030-77883-5_16. URL https://doi.org/10.1007/978-3-030-77883-5_16.
- Ghazi, B., Kamath, P., Kumar, R., Manurangsi, P., and Sealfon, A. On computing pairwise statistics with local differential privacy. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.
- Goryczka, S. and Xiong, L. A Comprehensive Comparison of Multiparty Secure Additions with Differential Privacy. *IEEE Transactions on Dependable and Secure Computing*, 2017.
- Hardt, M. and Price, E. The noisy power method: A meta algorithm with applications. In Ghahramani,

- Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/729c68884bd359ade15d5f163166738a-Paper.pdf.
- Hardt, M. and Roth, A. Beating randomized response on incoherent matrices. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pp. 1255–1268, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312455. doi: 10.1145/2213977.2214088. URL <https://doi.org/10.1145/2213977.2214088>.
- Hardt, M. and Roth, A. Beyond worst-case analysis in private singular vector computation. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pp. 331–340, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450320290. doi: 10.1145/2488608.2488650. URL <https://doi.org/10.1145/2488608.2488650>.
- Hebrail, G. and Berard, A. Individual household electric power consumption. UCI Machine Learning Repository, 2012. DOI: <https://doi.org/10.24432/C58K54>.
- Høgsgaard, M. M., Kamma, L., Larsen, K. G., Nelson, J., and Schwiiegelshohn, C. Sparse dimensionality reduction revisited. *CoRR*, abs/2302.06165, 2023. doi: 10.48550/ARXIV.2302.06165. URL <https://doi.org/10.48550/arXiv.2302.06165>.
- Kane, D. M. and Nelson, J. Sparsier johnson-lindenstrauss transforms. In Rabani, Y. (ed.), *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pp. 1195–1206. SIAM, 2012. doi: 10.1137/1.9781611973099.94. URL <https://doi.org/10.1137/1.9781611973099.94>.
- Kapralov, M. and Talwar, K. *On differentially private low rank approximation*, pp. 1395–1414. doi: 10.1137/1.9781611973105.101. URL <https://epubs.siam.org/doi/abs/10.1137/1.9781611973105.101>.
- Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? In *49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS, 2008.
- Kaul, M. 3D Road Network (North Jutland, Denmark). UCI Machine Learning Repository, 2013. DOI: <https://doi.org/10.24432/C5GP51>.
- Keller, M. MP-SPDZ: A versatile framework for multi-party computation. In Ligatti, J., Ou, X., Katz, J., and Vigna, G. (eds.), *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*, pp. 1575–1590. ACM, 2020. doi: 10.1145/3372297.3417872. URL <https://doi.org/10.1145/3372297.3417872>.
- Kenthapadi, K., Korolova, A., Mironov, I., and Mishra, N. Privacy via the johnson-lindenstrauss transform. *J. Priv. Confidentiality*, 5(1), 2013. doi: 10.29012/JPC.V5i1.625. URL <https://doi.org/10.29012/jpc.v5i1.625>.
- Kifer, D., Smith, A. D., and Thakurta, A. Private convex optimization for empirical risk minimization with applications to high-dimensional regression. In *The 25th Annual Conference on Learning Theory*, COLT, 2012.
- Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pp. 1302–1338, 2000.
- Liang, Y., Balcan, M., Kanchanapally, V., and Woodruff, D. P. Improved distributed principal component analysis. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 3113–3121, 2014. URL <https://proceedings.neurips.cc/paper/2014/hash/52947e0ade57a09e4a1386d08f17b656-Abstract.html>.
- Liu, X., Kong, W., and Oh, S. Differential privacy and robust statistics in high dimensions. In *Conference on Learning Theory*, 2022.
- Melis, L., Danezis, G., and De Cristofaro, E. Efficient private statistics with succinct sketches. 02 2016. doi: 10.14722/ndss.2016.23175.
- Meng, X. and Mahoney, M. W. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In Boneh, D., Roughgarden, T., and Feigenbaum, J. (eds.), *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pp. 91–100. ACM, 2013. doi: 10.1145/2488608.2488621. URL <https://doi.org/10.1145/2488608.2488621>.
- Milionis, J., Kalavasis, A., Fotakis, D., and Ioannidis, S. Differentially private regression with unbounded covariates. In Camps-Valls, G., Ruiz, F. J. R., and

- Valera, I. (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3242–3273. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/milionis22a.html>.
- Minami, K., Arai, H., Sato, I., and Nakagawa, H. Differential privacy without sensitivity. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/a7aeed74714116f3b292a982238f83d2-Paper.pdf.
- Mugunthan, V., Polychroniadou, A., Byrd, D., and Balch, T. H. Smpai: Secure multi-party computation for federated learning. In *Proceedings of the NeurIPS 2019 Workshop on Robust AI in Financial Services*, 2019.
- Nelson, J. and Nguyễn, H. L. Osnap: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 2013.
- Nikolov, A. Private query release via the johnson-lindenstrauss transform. In Bansal, N. and Nagarajan, V. (eds.), *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pp. 4982–5002. SIAM, 2023. doi: 10.1137/1.9781611977554.CH182. URL <https://doi.org/10.1137/1.9781611977554.ch182>.
- Pagh, R. and Thorup, M. Improved utility analysis of private countsketch. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 25631–25643. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/a47f5cdf1469751597d78e803fc590f-Paper-Conference.pdf.
- Pisier, G. *The volume of convex bodies and Banach space geometry*. Cambridge Tracts in Mathematics. 94, 1999.
- Sarlós, T. Improved approximation algorithms for large matrices via random projections. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pp. 143–152. IEEE Computer Society, 2006. doi: 10.1109/FOCS.2006.37. URL <https://doi.org/10.1109/FOCS.2006.37>.
- Sheffett, O. Old techniques in differentially private linear regression. In Garivier, A. and Kale, S. (eds.), *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pp. 789–827. PMLR, 22–24 Mar 2019. URL <https://proceedings.mlr.press/v98/sheffett19a.html>.
- Stausholm, N. M. Improved differentially private euclidean distance approximation. In Libkin, L., Pichler, R., and Guagliardo, P. (eds.), *PODS’21: Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Virtual Event, China, June 20-25, 2021*, pp. 42–56. ACM, 2021a. doi: 10.1145/3452021.3458328. URL <https://doi.org/10.1145/3452021.3458328>.
- Stausholm, N. M. Improved differentially private euclidean distance approximation. In Libkin, L., Pichler, R., and Guagliardo, P. (eds.), *PODS’21: Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Virtual Event, China, June 20-25, 2021*, pp. 42–56. ACM, 2021b. doi: 10.1145/3452021.3458328. URL <https://doi.org/10.1145/3452021.3458328>.
- Steinke, T. Multi-central differential privacy, 2020. URL <https://arxiv.org/abs/2009.05401>.
- Talwar, K., Wang, S., McMillan, A., Jina, V., Feldman, V., Basile, B., Cahill, A., Chan, Y. S., Chatzidakis, M., Chen, J., Chick, O., Chitnis, M., Ganta, S., Goren, Y., Granqvist, F., Guo, K., Jacobs, F., Javidbakht, O., Liu, A., Low, R., Mascenik, D., Myers, S., Park, D., Park, W., Parsa, G., Pauly, T., Priebe, C., Rishi, R., Rothblum, G., Scaria, M., Song, L., Song, C., Tarbe, K., Vogt, S., Winstrom, L., and Zhou, S. Samplable anonymous aggregation for private federated data analysis, 2023. URL <https://arxiv.org/abs/2307.15017>.
- Upadhyay, J. The price of privacy for low-rank factorization. In *Advances in Neural Information Processing Systems*, 2018.
- Varshney, P., Thakurta, A., and Jain, P. (nearly) optimal private linear regression for sub-gaussian data via adaptive clipping. In Loh, P.-L. and Raginsky, M. (eds.), *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 1126–1166. PMLR, 02–05 Jul 2022. URL <https://proceedings.mlr.press/v178/varshney22a.html>.
- Vu, D. and Slavkovic, A. B. Differential privacy for clinical trial data: Preliminary evaluations. In *ICDM Workshops 2009, IEEE International Conference on Data Mining*, 2009.

- Wang, D., Gaboardi, M., and Xu, J. Empirical risk minimization in non-interactive local differential privacy revisited. In *Advances in Neural Information Processing Systems*, 2018a.
- Wang, D., Smith, A. D., and Xu, J. Noninteractive locally private learning of linear models via polynomial approximations. In *International Conference on Algorithmic Learning Theory*, 2018b. URL <https://api.semanticscholar.org/CorpusID:67856419>.
- Wang, D., Hu, L., Zhang, H., Gaboardi, M., and Xu, J. Generalized linear models in non-interactive local differential privacy with public data. *Journal of Machine Learning Research*, 24(132):1–57, 2023. URL <http://jmlr.org/papers/v24/21-0523.html>.
- Wang, Y. Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI*, 2018.
- Wang, Y., Fienberg, S. E., and Smola, A. J. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *Proceedings of the 32nd International Conference on Machine Learning, ICML*, 2015.
- Warner, S. L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 1965.
- Woodruff, D. P. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10(1-2):1–157, 2014. doi: 10.1561/04000000060. URL <https://doi.org/10.1561/04000000060>.
- Zhao, F., Qiao, D., Redberg, R., Agrawal, D., El Abbadi, A., and Wang, Y.-X. Differentially private linear sketches: efficient implementations and applications. In *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22*, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.
- Zheng, K., Mou, W., and Wang, L. Collect at once, use effectively: making non-interactive locally private learning possible. In *Proceedings of the 34th International Conference on Machine Learning, ICML*, 2017.

A. Differential Privacy Basics

We first define the Gaussian and Laplace mechanisms, which we use later.

Lemma A.1 (The Gaussian Mechanism (Dwork & Roth, 2014)). *Let $f : \mathcal{X} \rightarrow \mathbb{R}^k$ be a function and let $\varepsilon \geq 0$ and $\delta \in [0, 1]$. The Gaussian mechanism adds to each of the k components of the output, noise sampled from $N(0, \sigma^2)$ with*

$$\sigma^2 \geq \frac{2(\Delta_2 f)^2 \ln(1.25/\delta)}{\varepsilon^2},$$

where $\Delta_2 f = \max_{x \sim x'} \|f(x) - f(x')\|_2$ denotes the ℓ_2 sensitivity of function f . The Gaussian mechanism is (ε, δ) differentially private.

Lemma A.2 (The Laplace Mechanism (Dwork et al., 2006b)). *Let $f : \mathcal{X} \rightarrow \mathbb{R}^k$ be a function and let $\varepsilon \geq 0$. The Laplace mechanism adds to each of the k components of the output, noise sampled from $\text{Lap}(\Delta_1 f / \varepsilon)$, where $\Delta_1 f = \max_{x \sim x'} |f(x) - f(x')|$ denotes the ℓ_1 sensitivity of function f . The Laplace mechanism is ε differentially private.*

We will use sequential composition of differentially private mechanisms.

Lemma A.3 (Sequential Composition (Dwork et al., 2006b)). *Let $\mathcal{M}_i : \mathcal{X} \rightarrow \mathcal{Y}_i$ be an $(\varepsilon_i, \delta_i)$ -differentially private mechanism for $i \in [k]$. Then mechanism $\mathcal{M} : \mathcal{X} \rightarrow \prod_{i=1}^k \mathcal{Y}_i$ defined as $\mathcal{M}(x) = (\mathcal{M}_1(x), \dots, \mathcal{M}_k(x))$, is $(\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)$ -differentially private.*

B. Frequency Estimation

Since Gaussians are popular in differentially private mechanisms, we will describe the parameter choices for this distribution. Let g_i be Gaussian distributed with zero mean and variance $\sigma^2 \geq \frac{2\Delta^k \ln(1.25/\delta)}{n \cdot \varepsilon^2}$. Then we have $\tilde{F}_k = 1^T(\mathbf{x}^k + g) = F_k + \sum_{i=1}^n g_i$. We now observe that $\sum_{i=1}^n g_i$ is Gaussian distributed with zero mean and variance $n \cdot \sigma^2 \geq \frac{2\Delta^k \ln(1.25/\delta)}{\varepsilon^2}$. Thus, assuming a trusted computation of $1^T g$, \tilde{F}_k is a differentially private estimate of F_k with error $\frac{2\Delta^k \ln(1.25/\delta)}{\varepsilon^2}$.

We also describe the analogous parameter choices for the Laplace mechanism. Let g_i be sample from the distribution defined as the difference between two Gamma distributions $\Gamma(\frac{1}{n}, \frac{2\Delta^k}{\varepsilon})$. Again, we have $\tilde{F}_k = 1^T(\mathbf{x}^k + g) = F_k + \sum_{i=1}^n g_i$. We now observe that $\sum_{i=1}^n g_i$ is Laplace distributed, centered at 0 and with scale $\frac{2\Delta^k}{\varepsilon}$. Thus, assuming a trusted computation of $1^T g$, \tilde{F}_k is a differentially private estimate of F_k with error $\frac{2\Delta^k}{\varepsilon}$.

C. Sparse Oblivious Subspace Embeddings

Oblivious subspace embeddings (OSEs) (Sarlós, 2006), allow for faster approximation algorithms for problems in linear algebra. Achlioptas (Achlioptas, 2003) provided the first Johnson-Lindenstrauss transform with some amount of sparsity. The first embedding with asymptotically smaller number of non-zeros than dense Johnson-Lindenstrauss transforms was probably due to (Kane & Nelson, 2012), who also applied them in the context of subspace embeddings. In a remarkable result, the super-sparse version with only 1 non-zero entries per column, was first analyzed in (Clarkson & Woodruff, 2013) and improved independently in (Meng & Mahoney, 2013; Nelson & Nguyễn, 2013), by increasing the target dimension m . In (Nelson & Nguyễn, 2013), the sparse Johnson-Lindenstrauss transform (Kane & Nelson, 2012) was studied with a wider range of parameters. Subsequent works (Cohen, 2016; Høgsgaard et al., 2023) further analyze and improve the relationship between m and s , emphasizing different parameters. Table 3 provides the exact interplay between m and s in those works. (Liang et al., 2014) proposes an approach to boost the success probability β of an OSE, which gives an alternative to increasing the target dimension.

When dealing with rank k approximation, we will condition on

$$(1 - \alpha_S) \|\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T\|_F^2 \leq \|\mathbf{S}(\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T)\|_F^2 \leq (1 + \alpha_S) \|\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}^T\|_F^2$$

for all rank k orthogonal matrices \mathbf{X} . When dealing with regression, we will condition on

$$(1 - \alpha_S) \|\mathbf{A}\mathbf{x}' - \mathbf{b}\|^2 \leq \|\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})\|^2 \leq (1 + \alpha_S) \|\mathbf{A}\mathbf{x}' - \mathbf{b}\|^2$$

for all $\mathbf{x} \in \mathbb{R}^d$. For the parameters given here, this is true with the probability given in table 3 (assuming $k = d$ in the case of regression). For the case of regression in particular, it is sometimes beneficial to select α_S as large as possible. A sufficiently largest value of α_S such that \mathbf{S} still provide a subspace embedding guarantee is $1/3$. Throughout this paper, we will sometimes bound α_S by 1.

We further will give our utility proofs using the sparsity/target dimension bounds from (Cohen, 2016). Other tradeoffs are possible, but they have worse bounds for most ranges of parameters. The results in (Meng & Mahoney, 2013; Nelson & Nguyễn, 2013; Høgsgaard et al., 2023) give different trade-offs depending on which set of parameters are considered the most dominant. Notably, the $s = 1$ sketches of (Meng & Mahoney, 2013; Nelson & Nguyễn, 2013) result in an additive error of $\tilde{O}\left(\frac{k^4 d^3}{\alpha_S^4 \beta^2} \varepsilon^{-2} \log \frac{1}{\delta}\right)$ for low rank approximation and a term

$\tilde{O}\left(\frac{d^7 \varepsilon^{-2} \log \frac{1}{\delta}}{\alpha_S^5 \beta^2 \lambda} + \frac{d^{14} \varepsilon^{-4} \log^2 \frac{1}{\delta}}{\alpha_S^6 \beta^4 \lambda^2}\right)$ in both the multiplicative and additive error. The utility guarantees using the bounds from (Høgsgaard et al., 2023) change only by logarithmic factors compared to those in Theorems 4.3 and 4.5.

D. Privacy Proofs

Proof Sketch of Lemma 3.3. This proof can be seen as analogous to the security proof of Theorem 4 in (Talwar et al., 2023). Recall that each server learns nothing more from a perfectly secure multi-party computation protocol other than what is implied by their own input and the output of the function being evaluated. Then there is a simulator $\text{Sim}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(x)$ whose output is (ε, δ) indistinguishable from the adversary's view $\text{View}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(x)$. Note that the secret shares that are part of the view of the corrupted servers in the protocol Π are independent of the input values of the corrupted clients, and all can be easily simulated by random sampling.

A bit more formally, let x be an input dataset, where x_i is the input value contributed by client i and \tilde{x}_i is the vector of input shares for server i that serves as input to Π . There is a mapping from $x \in \mathcal{X}^n$ to $\tilde{x} \in \tilde{\mathcal{X}}^k$. Denote $X = \{x_i\}_{i \in C_{\text{cor}}}$ and $X' = \{x'_i\}_{i \in C_{\text{cor}}}$. Then for any two datasets x, x' that differ in one entry x_i with $i \notin C_{\text{cor}}$:

$$\begin{aligned} & \Pr[\text{View}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(x) \in S] \\ & \leq \Pr[\text{Sim}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(f(R(x_1), \dots, R(x_n)), X) \in S] \\ & \leq e^\varepsilon \Pr[\text{Sim}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(f(R(x'_1), \dots, R(x'_n)), X') \in S] + \delta \\ & = e^\varepsilon \Pr[\text{View}_{S_{\text{cor}}, C_{\text{cor}}}^{R, \Pi, A}(x') \in S] + \delta \end{aligned}$$

□

Lemma D.1. Let $\gamma \in (0, 1)$, and let $\mathbf{S} \sim \mathcal{D}_{m, n}^{\text{sketch}}$. Then the probability that \mathbf{S} has at least $(1 - \gamma) \frac{n}{m}$ and at most $(1 + \gamma) \frac{n}{m}$ non-zero entries in every row is bounded by:

$$\Pr\left(\exists i \in [m] : \left|X_i - \frac{n}{m}\right| > \gamma \cdot \frac{n}{m}\right) < 2m \exp\left(-\frac{\gamma^2 n}{2m}\right),$$

where X_i denotes the number of non-zero entries in row i .

Proof. We only give a proof of the lower bound, as a proof of the upper bound is completely analogous. Let X_{ij} denote the indicator random variable that is 1 if S_{ij} is non-zero and 0 otherwise. Note that each entry of \mathbf{S} can be thought of as an independent Bernoulli random variable X_{ij} , and the number of non-zero entries in a row is the sum of n of these random variables. Fixing a row $k \in [m]$, we can use a Chernoff bound for sums of Bernoulli random variables to

Table 3. Trade-off between the target dimension m and sparsity s of sparse OSEs that are generated such that they have s non-zeros entries per column. Here α denotes the accuracy of the OSE and β denotes its fail probability. Variable k is the parameter to k -rank approximation, which can be replaced by dimension d for linear regression. The paper by (Achlioptas, 2003) was the first to propose the construction, while its application to preserving subspaces was given by (Clarkson & Woodruff, 2009).

Paper	Target Dimension m	Sparsity s	With Probability
(Achlioptas, 2003)	$O(\frac{k+\log 1/\beta}{\alpha^2})$	m	$1 - \beta$
(Nelson & Nguyen, 2013)	$O(\frac{k^2}{\alpha^2\beta})$	1	$1 - \beta$
(Cohen, 2016)	$O(\frac{k \log(k/\beta)}{\alpha^2})$	$O(\frac{\log(k/\beta)}{\alpha})$	$1 - \beta$
(Høgsgaard et al., 2023)	$O(\frac{k}{\alpha^2})$	$O(\frac{1}{\alpha} \cdot (\frac{k}{\log(1/\alpha)} + k^{2/3} \log^{1/3}(k)))$	$1 - 2^{-k^{2/3}}$

get a concentration bound around the expected number of non-zero entries $X_k = \sum_{j=1}^n X_{kj}$ in row k . In order to do this, we first need the expected value of X_k ;

$$\mathbb{E}[X_k] = \sum_{j=1}^n \Pr(X_{kj} = 1) = \frac{n}{m}.$$

Thus a Chernoff bound gives us that

$$\Pr\left(X_k < (1 - \gamma)\frac{n}{m}\right) < \exp\left(-\frac{\gamma^2 n}{2m}\right).$$

Applying a union bound over all rows of \mathbf{S} then gives us exactly what we were to prove;

$$\begin{aligned} & \Pr\left(\exists i \in [m] : X_i < (1 - \gamma)\frac{n}{m}\right) \\ & \leq \sum_{k=1}^m \Pr\left(X_k < (1 - \gamma)\frac{n}{m}\right) \\ & < m \exp\left(-\frac{\gamma^2 n}{2m}\right). \end{aligned}$$

□

To prove Theorem 3.4, we first prove a simpler version of the theorem below.

Lemma D.2. *Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $m \in [n]$, $t' < n$. Let $\mathbf{S} \sim \mathcal{D}_{m,n}^{\text{sketch}}$ with one non-zero entry per column. Also let $g_i \sim D$ be samples from a distribution such that $\sum_{i \in [(n-t')/(2m)]} g_i = G$. G follows a distribution D' that defines an additive noise mechanism that satisfies $(\varepsilon/d, \delta/d - m \exp(-(n-t')/8m))$ -DP for function $f(\mathbf{A}) = (\mathbf{S}\mathbf{A})_j + \mathbf{n}$ with $\mathbf{n} \sim D^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of $\mathbf{S}\mathbf{A}$. Then as long as input values are bounded above by η , tuple of algorithms (R_D, T_S, A) , as defined by equations 2 and 3 and for any A , is (ε, δ) -differentially private in the trusted computation model for differential privacy with t' corrupt clients.*

Proof of Lemma D.2. Let ε, δ, R_D and T_S be given as described in the theorem. If $\delta/d \leq m \exp(-(n-t')/8m)$,

then the R_D outputs 0^d , which is entirely independent of the input and is therefore (ε, δ) differentially private.

Otherwise, $\mathcal{M}(\mathbf{A}) = T_S(R_D(\mathbf{a}_1), \dots, R_D(\mathbf{a}_n)) = \mathbf{S}(\mathbf{A} + \mathbf{G})$ where \mathbf{a}_i denotes row i of $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{G} \leftarrow D^{n \times d}$, and we now prove that this is (ε, δ) differentially private.

Note that $\mathbf{S}_i(\mathbf{A} + \mathbf{G})$ is a d -dimensional vector, when \mathbf{S}_i refers to the i -th row of \mathbf{S} . If $(\mathbf{S}_i(\mathbf{A} + \mathbf{G}))_j$ is $(\varepsilon/d, \delta/d)$ -DP, then sequential composition (Lemma A.3) gives that \mathcal{M} is (ε, δ) differentially private.

Now consider $d = 1$ and let mechanism $\mathcal{M}_1 : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be defined as $\mathcal{M}_1(\mathbf{a}) = \mathbf{S}(\mathbf{a} + \mathbf{g})$ for $\mathbf{a} \in \mathbb{R}^n$, with $\mathbf{g} \leftarrow D^n$.

\mathbf{S} is sampled such that it has only one non-zero entry per column; thus, the columns of \mathbf{S} are orthogonal. This means that the entries of $(\mathbf{S}(\mathbf{a} + \mathbf{g}))_j$ and $(\mathbf{S}(\mathbf{a} + \mathbf{g}))_{j'}$, for $j \neq j'$, have disjoint support. Therefore, the privacy guarantee can be analyzed independently for each entry of $\mathbf{S}(\mathbf{a} + \mathbf{g})$.

Denote by E the event that the support of $\mathbf{S}(\mathbf{a} + \mathbf{g})_j$ has at least $\frac{n-t'}{2m}$ uncorrupted clients, whose set of indices we denote N . We can formulate $\mathbf{S}(\mathbf{a} + \mathbf{g})_j = (\mathbf{S}\mathbf{a})_j + (\mathbf{S}\mathbf{g})_j = (\mathbf{S}\mathbf{a})_j + \sum_{k \in [n]} \mathbf{S}_{j,k} \mathbf{g}_k = (\mathbf{S}\mathbf{a})_j + \sum_{k \in N} \mathbf{S}_{j,k} \mathbf{g}_k + \sum_{k \in [n] \setminus N} \mathbf{S}_{j,k} \mathbf{g}_k$. Also note that $\sum_{k \in N} \mathbf{S}_{j,k} \mathbf{g}_k$ contains at least $\frac{n-t'}{2m}$ terms in event E , the sum of which suffice for guaranteeing DP.

Due to the fact that D is symmetric, $-1 \cdot \mathbf{g}_k$ follows the same distribution as \mathbf{g}_k , and $\sum_{k \in N} \mathbf{S}_{j,k} \mathbf{g}_k$ follows the same distribution as $\sum_{k \in N} \mathbf{g}_k$. Due to the infinite divisibility of D' , the sum of the first $\frac{n-t'}{2m}$ terms of $\sum_{k \in N} \mathbf{g}_k$ are samples from D' .

Then in event E , \mathcal{M}_1 is $(\varepsilon, \delta - m \exp(-(n-t')/8m))$ differentially private using the additive noise mechanism that adds noise from D' . Setting $\gamma = 1/2$, Lemma D.1 gives that E occurs with probability at least $1 - m \exp(-\frac{n-t'}{8m})$. Since event E does not occur with probability at most $m \exp(-(n-t')/8m)$, \mathcal{M}_1 is $(\varepsilon, \delta - m \exp(-(n-t')/8m))$ differentially private except with probability $m \exp(-(n-t')/8m)$. Then \mathcal{M}_1 is (ε, δ) -differentially private.

□

□

Proof of Theorem 3.4. Let ε, δ, R_D and T_S be given as described in the theorem. If $\delta/d \leq m \exp(-(n-s-t')/8m)$, then the R_D outputs 0^d , which is entirely independent of the input and is therefore (ε, δ) differentially private.

Otherwise, $\mathcal{M}(\mathbf{A}) = T_S(R_D(\mathbf{a}_1), \dots, R_D(\mathbf{a}_n)) = \frac{1}{\sqrt{s}} \sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ where \mathbf{a}_i denotes row i of $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{G}_i \leftarrow D^{n \times d}$. Here, \mathbf{S} has been decomposed such that $\mathbf{S} = \frac{1}{\sqrt{s}} \sum_{i=1}^s \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{m \times n}$ is a matrix with only one non-zero entry per column.

We argue $(\varepsilon/s, \delta/s)$ -DP based on Lemma D.2. Note that the non-zero entry per column in each \mathbf{S}_i is not sampled uniformly at random from all possible rows; rather, for \mathbf{S}_i is sampled uniformly at random from the remaining $n-i$ rows (removing the one that was already chosen to be non-zero, sampling without replacement). Therefore, we replace n with $n-s$ when applying Lemma D.2.

By sequential composition, releasing $\mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ for all $i \in [s]$ satisfies (ε, δ) -DP, and thus by post-processing, the scaled sum of these is also (ε, δ) -DP. □

Corollary D.3. Let $\varepsilon \geq 0, \delta \in (0, 1), t' < n, m \in [n], s \in [n]$ and D be the Gaussian distribution $\mathcal{N}(0, \sigma^2)$ with:

$$\sigma^2 = \frac{4s^3\eta^2 \ln(1.25s/((\delta/d - m \exp(-(n-s-t')/8m))))md^2}{\varepsilon^2(n-s-t')}$$

Let $\mathbf{S} \sim \mathcal{D}_{m,n,s}^{\text{sketch}}$. Then as long as input values are bounded above by η , tuple of algorithms (R_D, T_S, A) , as defined by equations 2 and 3 and for any A , is (ε, δ) -differentially private in the trusted computation model for differential privacy with t' corrupt clients.

Proof of Corollary D.3. We apply Theorem 3.4, setting D to a Gaussian distribution with the variance specified in the corollary. First note that zero-centered Gaussians are symmetric around the origin. Due to infinite divisibility of Gaussians, $\sum_{i \in [(n-s-t')/(2m)]} g_i = G$ when $g_i \sim D$ is Gaussian distributed with variance $\sum_{i \in [(n-s-t')/(2m)]} \sigma^2 = (n-s-t')\sigma^2/(2m)$.

Notice that the ℓ_2 -sensitivity of algorithm T_S is $\Delta_2 T_S = \max_{\mathbf{A}, \mathbf{A}'} \|\mathbf{S}\mathbf{A} - \mathbf{S}\mathbf{A}'\|_2 = 2\eta\sqrt{s}$, where $\mathbf{A}, \mathbf{A}' \in \mathbb{R}^{n \times d}$ differ in a single row, and every entry of \mathbf{A}, \mathbf{A}' is bounded above by η .

Based on the Gaussian mechanism (Lemma A.1), the additive noise mechanism adding Gaussian noise satisfies $(\varepsilon/(sd), \delta/(sd) - m \exp(-(n-s-t')/8m))$ -DP for function $f(\mathbf{A}) = (\mathbf{S}\mathbf{A})_j + \mathbf{n}$ with $\mathbf{n} \sim \text{Lap}(0, 2\eta m^2 d/\varepsilon)^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of $\mathbf{S}\mathbf{A}$.

Lemma D.4. Let $\varepsilon \geq 0, \delta \in [0, 1), m \in [n], t' < n$. Let $\mathbf{S} \in \{-1, 0, 1\}^{m \times n}$ with one non-zero entry per column and n/m non-zero entries per row. Also let noise g_i be sampled from a symmetric distribution D such that $\sum_{i \in [(n)/m-t']} g_i = G$. G follows a distribution D' that defines an additive noise mechanism that satisfies $(\varepsilon/d, \delta/d)$ -DP for function $f(\mathbf{A}) = (\mathbf{S}\mathbf{A})_j + \mathbf{n}$ with $\mathbf{n} \sim D'^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of $\mathbf{S}\mathbf{A}$. Then as long as input values are bounded above by η , tuple of algorithms $(R_D^{\text{dense}}, T_S, A)$, as defined by equations 1 and 3 and for any A , is (ε, δ) -differentially private in the trusted computation model for differential privacy with t' .

Proof of Lemma D.4. Let $\varepsilon, \delta, R_D^{\text{dense}}$ and T_S be given as described in the theorem.

We show that $\mathcal{M}(\mathbf{A}) = T_S(R(\mathbf{a}_1), \dots, R(\mathbf{a}_n)) = \mathbf{S}(\mathbf{A} + \mathbf{G})$ where \mathbf{a}_i denotes row i of $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{G} \leftarrow D^{n \times d}$, is (ε, δ) differentially private.

Note that $\mathbf{S}_i(\mathbf{A} + \mathbf{G})$ is a d -dimensional vector, when \mathbf{S}_i refers to the i -th row of \mathbf{S} . If $(\mathbf{S}_i(\mathbf{A} + \mathbf{G}))_j$ is $(\varepsilon/d, \delta/d)$ -DP, then sequential composition (Lemma A.3) gives that \mathcal{M} is (ε, δ) differentially private.

Consider $d = 1$ and let mechanism $\mathcal{M}_1 : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be defined as $\mathcal{M}_1(\mathbf{a}) = \mathbf{S}(\mathbf{a} + \mathbf{g})$ for $\mathbf{a} \in \mathbb{R}^n$, with $\mathbf{g} \leftarrow D^n$.

\mathbf{S} is sampled such that it has only one non-zero entry per column; thus, the columns of \mathbf{S} are orthogonal. This means that the entries of $(\mathbf{S}(\mathbf{a} + \mathbf{g}))_j$ and $(\mathbf{S}(\mathbf{a} + \mathbf{g}))_{j'}$ for $j \neq j'$, have disjoint support. Therefore, the privacy guarantee can be analyzed independently for each entry of $\mathbf{S}(\mathbf{a} + \mathbf{g})$.

Note that the support of $\mathbf{S}(\mathbf{a} + \mathbf{g})_j$ has at least $\frac{n}{m} - t'$ uncorrupted clients, whose set of indices we denote N . We can formulate $\mathbf{S}(\mathbf{a} + \mathbf{g})_j = (\mathbf{S}\mathbf{a})_j + (\mathbf{S}\mathbf{g})_j = (\mathbf{S}\mathbf{a})_j + \sum_{k \in [n]} \mathbf{S}_{j,k} g_k = (\mathbf{S}\mathbf{a})_j + \sum_{k \in N} \mathbf{S}_{j,k} g_k + \sum_{k \in [n] \setminus N} \mathbf{S}_{j,k} g_k$.

Due to the fact that D is symmetric, $-1 \cdot g_k$ follows the same distribution as g_k , and $\sum_{k \in N} \mathbf{S}_{j,k} g_k$ follows the same distribution as $\sum_{k \in N} g_k$. Due to the infinite divisibility of D' , the sum of the first $\frac{n}{m} - t'$ terms of $\sum_{k \in N} g_k$ are samples from D' .

Then \mathcal{M}_1 is (ε, δ) -differentially private. □

Proof of Theorem 3.5. Let $\varepsilon, \delta, R_D^{\text{dense}}$ and T_S^{dense} be given as described in the theorem.

We know $\mathcal{M}(\mathbf{A}) = T_S(R_D(\mathbf{a}_1), \dots, R_D(\mathbf{a}_n)) = \frac{1}{\sqrt{m}} \sum_{i \in [m]} \mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ where $\mathbf{G}_i \leftarrow D^{n \times d}$. Recall from Equation 3 that \mathbf{S} has been decomposed such that

$\mathbf{S} = \frac{1}{\sqrt{m}} \sum_i^m \mathbf{S}_i$, where $\mathbf{S}_i \in \{-1, 0, 1\}^{m \times n}$ is a matrix with only one non-zero entry per column and n/m non-zero entries per row.

We argue $(\varepsilon/s, \delta/s)$ -DP based on Lemma D.4.

By sequential composition, releasing $\mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ for all $i \in [s]$ satisfies (ε, δ) -DP, and thus by post-processing, the scaled sum of these is also (ε, δ) -DP. \square

Corollary D.5. *Let $\varepsilon \geq 0$, $t' < n$, $m \in [n]$, $s \in [n]$, and let D be the distribution formulated as the difference of two samples of gamma distribution $\Gamma(\frac{1}{n/m-t'}, b)$ with parameter*

$$b = 2\eta m^2 d / \varepsilon$$

Let $\mathbf{S} \sim \{-1, 1\}^{m \times n}$ sampled uniformly. Then as long as input values are bounded above by η , $(R_D^{\text{dense}}, T_S, A)$ is ε -differentially private in the trusted computation model for differential privacy with t' corrupt clients C' .

Proof of Corollary D.5. We apply Theorem 3.5, setting D to the difference between two samples of the gamma distribution $\Gamma(\frac{1}{n/m-t'}, 2\eta m^2 d / \varepsilon)$. Note that the resulting distribution is centered at 0 and symmetric. Due to infinite divisibility of Laplace distributions, $\sum_{i \in [n/m-t']} g_i = G$ when $g_i \sim D$ is Laplace distributed with mean zero and scale parameter $b = 2\eta m^2 d / \varepsilon$.

Notice that the ℓ_1 -sensitivity of algorithm T_S is $\Delta_1 T_S = \max_{\mathbf{A}, \mathbf{A}'} \|\mathbf{S}\mathbf{A}_j - \mathbf{S}\mathbf{A}'_j\|_1 = 2\eta m$, where $\mathbf{A}, \mathbf{A}' \in \mathbb{R}^{n \times d}$ differ in a single row, and every entry of \mathbf{A}, \mathbf{A}' is bounded above by η .

Based on the Laplace mechanism (Lemma A.2), the additive noise mechanism adding Laplace noise satisfies $\varepsilon/(dm)$ -DP for function $f(\mathbf{A}) = (\mathbf{S}\mathbf{A})_j + \mathbf{n}$ with $\mathbf{n} \sim \text{Lap}(0, 2\eta m^2 d / \varepsilon)^m$ on input $\mathbf{A} \in \mathbb{R}^{n \times d}$, where the subscript refers to one column of $\mathbf{S}\mathbf{A}$. \square

E. Utility Proofs

Lemma E.1. *Let $\mathbf{G} \in \mathbb{R}^{n \times d}$ such that all entries in \mathbf{G} are independently sampled from a normal distribution with mean 0 and variance at most σ^2 . Further, let V be a set of d -dimensional vectors lying in a k -dimensional subspace. Then with probability at least $1 - \beta$ for some absolute constant η*

$$\sup_{\mathbf{x} \in V} \|\mathbf{G}\mathbf{x}\|_2^2 \leq n \cdot \sigma^2 \cdot \|\mathbf{x}\|^2 + \eta \cdot \sigma^2 \cdot \|\mathbf{x}\|^2 \cdot (\sqrt{(k + \log 1/\beta) \cdot n} + (k + \log 1/\beta)).$$

Moreover, with probability at least $1 - \beta$

$$\|\mathbf{G}\|_F^2 \leq \eta \cdot \sigma^2 \cdot n \cdot d \cdot \log 1/\beta.$$

Proof. Denote by \mathbf{G}_i the i th row of \mathbf{G} . We start by observing that $\mathbf{G}_i^T \mathbf{x}$ is Gaussian distributed with mean 0 and variance $\sigma^2 \cdot \|\mathbf{x}\|^2$. Thus, we focus our attention on controlling $\frac{\|\mathbf{G}\mathbf{x}\|_2^2}{\sigma^2 \cdot \|\mathbf{x}\|^2}$, that is, we assume that \mathbf{x} is a unit vector and that the entries of \mathbf{G} are standard normal Gaussian random variables. The overall claim then follows by rescaling. Consider an ε -net N_ε of V , that is, for every $\mathbf{x} \in V$ with unit norm there exists a vector $\mathbf{x}' \in N_\varepsilon$ with $\|\mathbf{x} - \mathbf{x}'\| \leq \varepsilon$. Such nets exist with $|N_\varepsilon| \leq \exp(\eta \cdot k \log 1/\gamma)$ for some absolute constant η , see (Pisier, 1999). Suppose $\gamma = 1/4$. Using concentration bounds for sums of Gaussians (see for example Lemma 1 of (Laurent & Massart, 2000)), we then have for any $\|\mathbf{x}'\|$ with probability $1 - \beta$

$$\begin{aligned} & \mathbb{P}[\exists \mathbf{x}' \in N_\gamma : \|\mathbf{G}\mathbf{x}'\|^2 - \mathbb{E}[\|\mathbf{G}\mathbf{x}'\|^2] \\ & \geq 2(\sqrt{(\log N_\gamma + \log 1/\beta) \cdot n} + (\log N_\gamma + \log 1/\beta))] \\ & \leq |N_\gamma| \cdot \exp(-(\eta \cdot k + \log 1/\beta)) \leq \beta. \end{aligned}$$

for some absolute constant η .

We now extend this argument to all vectors, using an argument from (Arora et al., 2006) (Lemma 4 of that reference). Let \mathbf{U} be an orthogonal basis of V . Then our goal is to control $\|\mathbf{G}\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{U}^T \mathbf{G}^T \mathbf{G} \mathbf{U} \mathbf{x}$. Define the matrix $\mathbf{B} := \mathbf{U}^T \mathbf{G}^T \mathbf{G} \mathbf{U} - n \cdot \mathbf{I}$, where \mathbf{I} is the identity matrix. Note that $\mathbb{E}[\mathbf{U}^T \mathbf{G}^T \mathbf{G} \mathbf{U}] = n \cdot \mathbf{I}$ and that $\|\mathbf{B}\mathbf{x}\|$ is the deviation of $\|\mathbf{G}\mathbf{x}\|$ around its expectation. Let $\|\mathbf{B}\|_{op} := \sup_{\mathbf{x} \in V} \|\mathbf{B}\mathbf{x}\|$.

$$\begin{aligned} \|\mathbf{B}\|_{op} &= \langle \mathbf{B}\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{B}\mathbf{x}', \mathbf{x}' \rangle + \langle \mathbf{B}(\mathbf{x}' + \mathbf{x}), \mathbf{x}' - \mathbf{x} \rangle \\ &\leq 2(\sqrt{(\log N_\gamma + \log 1/\beta) \cdot n} + (\log N_\gamma + \log 1/\beta)) \\ &\quad + \|\mathbf{B}\|_{op} \cdot \|\mathbf{x}' + \mathbf{x}\| \|\mathbf{x}' - \mathbf{x}\| \\ &\leq 2(\sqrt{(\log N_\gamma + \log 1/\beta) \cdot n} + (\log N_\gamma + \log 1/\beta)) \\ &\quad + 2\gamma \cdot \|\mathbf{B}\|_{op} \\ &\leq 2(\sqrt{(\log N_\gamma + \log 1/\beta) \cdot n} + (\log N_\gamma + \log 1/\beta)) \\ &\quad + 1/2 \cdot \|\mathbf{B}\|_{op} \end{aligned}$$

Rearranging implies that $\|\mathbf{B}\|_{op} \leq 4(\sqrt{(\log N_\gamma + \log 1/\beta) \cdot n} + (\log N_\gamma + \log 1/\beta))$. The first claim now follows by rescaling η .

For the second claim, we observe that $\|\mathbf{G}\|_F^2$ has a Gaussian distribution with mean 0 and variance $n \cdot d \cdot \sigma^2$. The same concentration inequality we applied above also implies that the probability that $\|\mathbf{G}\|_F$ exceeds $2\sqrt{\sigma^2 \cdot n \cdot d \cdot \log 1/\beta}$ is at most $1 - \beta$. \square

Proof of Lemma 4.1. We first argue that \mathbf{G} is Gaussian distributed. Each G_i has independent Gaussian entries and multiplying a Gaussian with a random Rademacher does not change the distribution. Therefore, the entries of $S_i G_i$ are likewise Gaussian distributed, with mean 0 and variance at most $(1+\gamma) \frac{n}{m} \sigma^2 \leq 2 \frac{n}{m} \sigma^2$ due to Lemma D.1. Concluding, the variance of the entries of $\mathbf{G} = \sum_{i \in [s]} S_i G_i$ is therefore at most $2s \frac{n}{m} \sigma^2$. Applying Lemma E.1, we then have with probability $1 - \beta$

$$\sup_{\mathbf{x} \in V} \|\mathbf{G}\mathbf{x}\|_2^2 \leq 2 \frac{n}{m} \sigma^2 \cdot \|\mathbf{x}\|^2 + \eta \cdot 2s \frac{n}{m} \sigma^2 \cdot \|\mathbf{x}\|^2 \cdot (\sqrt{(k + \log 1/\beta) \cdot m} + (k + \log 1/\beta)).$$

and

$$\|\mathbf{G}\|_F^2 \leq \eta \cdot 2sn\sigma^2 \cdot d \cdot \log 1/\beta$$

as desired \square

Proof of Lemma 4.2. We have $\|\mathbf{L}\|_F \leq \sqrt{m} \cdot \|\max_{i \in [m]} \mathbf{S}_i \mathbf{\Gamma}_i\|_F$, thus it is sufficient to give a high probability bound on $\|\mathbf{S}_i \mathbf{\Gamma}_i\|_F$.

The entries of $\mathbf{S}_i \mathbf{\Gamma}_i$ are Laplace distributed entries with mean 0 and scale parameter b by infinite divisibility of a Laplace random variable into Γ distributed random variables. Thus the expected value of $\|\mathbf{S}_i \mathbf{\Gamma}_i\|_F^2$ is $md \cdot 2b^2$.

To prove high concentration, let us consider the moments of the Laplace distribution. For a Laplace random variable L with mean 0 and scale b , we have $\mathbb{E}[L^{2\ell}] = (2\ell)!b^{2\ell}$. Then due to Hoelder's inequality

$$\begin{aligned} \mathbb{E}[\|\mathbf{S}_i \mathbf{\Gamma}_i\|_F^{2\ell}] &= \mathbb{E}\left[\left(\sum_{j=1}^{md} L_j^2\right)^\ell\right] \\ &\leq \mathbb{E}\left[\left(\sum_{j=1}^{md} L_j^{2\ell}\right)\right] \cdot \left(\sum_{j=1}^{md} 1^{\frac{\ell}{\ell-1}}\right)^{\ell-1} \\ &= (2\ell)!b^{2\ell} \cdot (md)^\ell \end{aligned}$$

Applying Markov's inequality, we then have for any $\gamma > 1$

$$\begin{aligned} &\mathbb{P}[\|\mathbf{S}_i \mathbf{\Gamma}_i\|_F^{2\ell} > \gamma \cdot md \cdot 2b^{2\ell}] \\ &= \mathbb{P}\left[\|\mathbf{S}_i \mathbf{\Gamma}_i\|_F^{2\ell} > (\gamma \cdot md \cdot 2b^{2\ell})^\ell\right] \\ &\leq \frac{(2\ell)!b^{2\ell}(md)^\ell}{(\gamma \cdot md \cdot 2b^{2\ell})^\ell} = \frac{(2\ell)!}{(2\gamma)^\ell} \\ &\leq \frac{(2\ell)^{2\ell}}{(2\gamma e^2)^\ell} \sqrt{6\pi\ell} \end{aligned}$$

For $\gamma > 4\ell^2$ and $\ell = \log \frac{m}{\beta}$, this term is at most β/m . We can therefore conclude that with probability $1 - \beta$, $\max_{i \in [m]} \mathbf{S}_i \mathbf{\Gamma}_i \leq 4 \log \frac{m}{\beta}$ and thus $\|\mathbf{L}\|_F^2 \leq \eta m^2 db^2 \log^2 \frac{m}{\beta}$, for some absolute constant η . \square

Proof of Theorem 4.3. We will use the following inequality. The applications to vector and matrix norms are straightforward corollaries.

Lemma E.2 (Generalized Triangle Inequality (Becchetti et al., 2019)). *For any two real numbers a, b and any $\alpha > 0$*

$$|a^2 - b^2| \leq \alpha \cdot a^2 + \left(1 + \frac{1}{\alpha}\right) \cdot (a - b)^2.$$

Finally, we require the following consequence of subspace embedding properties.

Lemma E.3 (Paraphrased from Definition 1, Lemma 11, and Theorem 12 (Cohen et al., 2015)). *Let \mathbf{S} be drawn from an OSNAP distribution. Then for any matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and any rank k orthogonal matrix $\mathbf{X} \in \mathbb{R}^{d \times k}$, we have*

$$\|\mathbf{SA} - \mathbf{SAXX}^T\|_F = (1 \pm \alpha) \cdot \|\mathbf{A} - \mathbf{AXX}^T\|_F.$$

Without noise, the low rank approximation outputs $\mathbf{X}_{\text{OPT}} = \arg\min_{\mathbf{X} \text{ rank } k, \mathbf{X} \in \mathbb{R}^{n \times d}} \|\mathbf{A} - \mathbf{AXX}^T\|_F^2$, and after sketching with Gaussian noise of variance σ^2 outputs $\mathbf{X}' = \arg\min_{\mathbf{X} \text{ rank } k, \mathbf{X} \in \mathbb{R}^{n \times d}} \|(\mathbf{A} + \mathbf{G}) - (\mathbf{A} + \mathbf{G})\mathbf{X}\mathbf{X}^T\|_F^2$. Let \mathbf{X}' be the matrix returned by the mechanism. We have

$$\begin{aligned} &\|\mathbf{A} - \mathbf{AX}'\mathbf{X}'^T\|_F \\ &\leq (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i \mathbf{A} - \mathbf{S}_i \mathbf{AX}'\mathbf{X}'^T \right\|_F \\ &\leq (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i (\mathbf{A} + \mathbf{G}_i) - \mathbf{S}_i (\mathbf{A} + \mathbf{G}_i) \mathbf{X}'\mathbf{X}'^T \right\|_F \\ &\quad + (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i \mathbf{G}_i - \mathbf{S}_i \mathbf{G}_i \mathbf{X}'\mathbf{X}'^T \right\|_F \end{aligned}$$

where the first inequality follows from the subspace embedding property. By optimality of \mathbf{X}' for the low rank approximation problem on $\sum_{i \in [s]} \mathbf{S}_i (\mathbf{A} + \mathbf{G}_i)$, we then

have

$$\begin{aligned}
 & \|\mathbf{A} - \mathbf{A}\mathbf{X}'\mathbf{X}'^T\|_F \\
 & \leq (1 + \alpha_S) \sqrt{\frac{1}{s}} \\
 & \quad \cdot \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} + \mathbf{G}_i) - \mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T \right\|_F \\
 & \quad + (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i - \mathbf{S}_i\mathbf{G}_i\mathbf{X}'\mathbf{X}'^T \right\|_F \\
 & \leq (1 + \alpha_S)^2 \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F \\
 & \quad + (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i - \mathbf{S}_i\mathbf{G}_i\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T \right\|_F \\
 & \quad + (1 + \alpha_S) \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i - \mathbf{S}_i\mathbf{G}_i\mathbf{X}'\mathbf{X}'^T \right\|_F
 \end{aligned}$$

where the inequality follows from Lemma E.3. Now notice that $\mathbb{I} - \mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T$ and $\mathbb{I} - \mathbf{X}'\mathbf{X}'^T$ are orthogonal projections, multiplying by which cannot increase the Frobenius norm of a matrix. Therefore, for $\mathbf{X} = \mathbf{X}'$ or $\mathbf{X} = \mathbf{X}_{\text{OPT}}$:

$$\begin{aligned}
 & \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i - \mathbf{S}_i\mathbf{G}_i\mathbf{X}\mathbf{X}^T \right\|_F \\
 & = \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i(\mathbb{I} - \mathbf{X}\mathbf{X}^T) \right\|_F \leq \sqrt{\frac{1}{s}} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i \right\|_F
 \end{aligned}$$

Then:

$$\begin{aligned}
 & \|\mathbf{A} - \mathbf{A}\mathbf{X}'\mathbf{X}'^T\|_F \\
 & \leq (1 + \alpha_S)^2 \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F \\
 & \quad + (1 + \alpha_S) 2 \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i \right\|_F \tag{6}
 \end{aligned}$$

Using Lemma 4.1, we have $\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i\mathbf{G}_i \right\|_F^2 \leq 2n \cdot \sigma^2 \cdot d \log(1/\beta)$. To simplify the error, observe that for the sketching matrix by (Cohen, 2016) (see Table 3), we can set $s = \frac{\log k}{\alpha_S}$ and $m = \frac{k \log k}{\alpha_S^2} \log(1/\beta)$. Using the bound on the variance from Theorem D.3 and plugging this in above, we can simplify the additive and multiplicative errors.

$$\begin{aligned}
 & \|\mathbf{A} - \mathbf{A}\mathbf{X}'\mathbf{X}'^T\|_F^2 \\
 & \leq (1 + \alpha_S)^2 \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F^2 \\
 & \quad + \eta(1 + \alpha_S) \sqrt{\sigma^2 \cdot n \cdot d \cdot \log \frac{1}{\beta}} \\
 & \leq (1 + 3\alpha_S) \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F^2 \\
 & \quad + \tilde{O} \left(\frac{kd^{3/2}}{\alpha_S^3} \log^{3/2}(1/\beta) \varepsilon^{-1} \log^{1/2}(1/\delta) \right)
 \end{aligned}$$

The stated bound follows by rescaling α_S by a factor of 3. \square

Proof of Theorem 4.4. The proof follows the same line of reasoning as Theorem 4.3. The main difference is that upon reaching Equation 6, we must bound $\frac{1}{m} \left\| \sum_{i \in [m]} \mathbf{S}_i\mathbf{G}_i \right\|_F^2$ via Lemma 4.2 rather than using Lemma 4.1. Assuming the semi-trusted corruption model with at most t corruptions, each entry of \mathbf{G}_i consists of $X_j - Y_j$ where $X_j, Y_j \sim \Gamma(1/(n-t), b)$. Consequently, the entries of $\sum_{i \in [m]} \mathbf{S}_i\mathbf{G}_i$ are a sum of $m \cdot \lceil \frac{n}{n-t} \rceil$ matrices where each entry is either 0 or a Laplacian distributed random variable with mean 0 and scale b . Using Lemma 4.2, we therefore have with probability $1 - \beta$.

$$\begin{aligned}
 & \frac{1}{m} \left\| \sum_{i \in [m]} \mathbf{S}_i\mathbf{G}_i \right\|_F^2 \\
 & \in O(1) \cdot m^2 \cdot \frac{n}{n-t} db^2 \log^2 \frac{m \cdot \frac{n}{n-t}}{\beta} \\
 & \in O(1) \cdot m^2 \cdot db^2 \log^2 \frac{m}{\beta},
 \end{aligned}$$

where we used $\frac{n}{n-t} \leq 2$ by assumption. Since \mathbf{S} is a dense Rademacher matrix, we have $m = \frac{k + \log \beta^{-1}}{\alpha_S^2}$ (Achlioptas, 2003; Woodruff, 2014) and $s = m$, and $b = 2\eta m^2 d/\varepsilon$ following Equation 5.

Thus we obtain, with probability $1 - \beta$, a $(1 + \alpha_S)$ multiplicative approximation using the same calculation as in Theorem 4.3 and a suitable rescaling, and an additive $O(\frac{k^3 d^{3/2}}{\alpha_S^6} \varepsilon^{-1} \log^2 \frac{k}{\alpha_S \beta})$ approximation. \square

Proof of Theorem 4.5. Let ε , δ , R_D and T_S be given as in the theorem. We consider the output $A_{\text{reg}}(T_S(R_D(\mathbf{a}_1), \dots, R_D(\mathbf{a}_n)))$.

If $n \leq 8m \ln(dm/\delta) + t'$, then the R_D outputs 0^d , then the optimal $\mathbf{x}' = 0^d$, leading to error $\|\mathbf{A}\mathbf{x}' - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}'\|_2^2 = \|\mathbf{b}\|_2^2 \leq \eta' n \leq \eta' m \ln(dm/\delta) + t'$.

Otherwise, $\mathcal{M}(\mathbf{A}) = T_S(R_D(\mathbf{a}_1), \dots, R_D(\mathbf{a}_n)) = \frac{1}{\sqrt{s}} \sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ where \mathbf{a}_i denotes row i of $\mathbf{A} \in$

$\mathbb{R}^{n \times d}$ and $\mathbf{G} \leftarrow \mathcal{N}(0, \sigma^2)^{n \times d}$, is (ε, δ) differentially private.

To streamline the presentation, we give the analysis without corrupted clients. Adding corrupted clients merely changes the analysis along the same lines as Theorem 4.3.

We first control the terms $\|\sum_{i \in [s]} \mathbf{S}_i(\mathbf{G}_i \mathbf{x}' - \mathbf{g}_i)\|^2$ and $\|\sum_{i \in [s]} \mathbf{S}_i(\mathbf{G}_i \mathbf{x}_{\text{OPT}} - \mathbf{g}_i)\|^2$. Using Lemma 4.1 with an added coordinate of -1 to both \mathbf{x}' and \mathbf{x}_{OPT} , we get

$$\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{G}_i \mathbf{x} - \mathbf{g}_i) \right\|^2 \leq \eta \cdot \sigma^2 n d \log \frac{1}{\beta} (\|\mathbf{x}\|^2 + 1) \quad (7)$$

for either $\mathbf{x} = \mathbf{x}_{\text{OPT}}$ or $\mathbf{x} = \mathbf{x}'$ and for a sufficiently large constant η . Then we have using Lemma E.3

$$\begin{aligned} & \|\mathbf{A} \mathbf{x}' - \mathbf{b}\|^2 + \lambda \|\mathbf{x}'\|^2 \\ & \leq (1 + \alpha_S) \|\mathbf{S}(\mathbf{A} \mathbf{x}' - \mathbf{b})\|^2 + \lambda \|\mathbf{x}'\|^2 \\ & = (1 + \alpha_S) \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} \mathbf{x}' - \mathbf{b}) \right\|^2 + \lambda \|\mathbf{x}'\|^2 \\ & = \lambda \|\mathbf{x}'\|^2 + (1 + \alpha_S) \frac{1}{s} \\ & \quad \cdot \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i - \mathbf{G}_i) \mathbf{x}' - \mathbf{b} + \mathbf{g}_i - \mathbf{g}_i) \right\|^2 \\ & \leq (1 + \alpha_S) \frac{1}{s} \\ & \quad \left((1 + \alpha_S) \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}' - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 \right. \\ & \quad \left. + \left(1 + \frac{1}{\alpha_S}\right) \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{G}_i \mathbf{x}' - \mathbf{g}_i) \right\|^2 \right) + \lambda \|\mathbf{x}'\|^2 \\ & \leq (1 + \alpha_S)^2 \cdot \left(\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}' - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 \right. \\ & \quad \left. + \lambda \cdot \|\mathbf{x}'\|^2 \right) + 2 \cdot \left(1 + \frac{1}{\alpha_S}\right) \eta \sigma^2 n d \log \frac{1}{\beta} (\|\mathbf{x}'\|^2 + 1) \end{aligned} \quad (8)$$

where the second to last inequality follows by applying Lemma E.2 and the final inequality follows from Equation 7. By optimality of \mathbf{x}' for the instance $\frac{1}{s} \|\sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} \mathbf{x}' + \mathbf{G}_i - (\mathbf{b} + \mathbf{g}_i))\|^2 + \lambda \|\mathbf{x}'\|^2$, we then have

$$\begin{aligned} & \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}' - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 + \lambda \cdot \|\mathbf{x}'\|^2 \\ & \leq \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2 \end{aligned}$$

which likewise implies

$$\begin{aligned} \|\mathbf{x}'\| & \leq \frac{1}{\lambda} \cdot \left(\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 \right. \\ & \quad \left. + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2 \right). \end{aligned}$$

Insertion both bounds back into Equation 9, we obtain

$$\begin{aligned} & \|\mathbf{A} \mathbf{x}' - \mathbf{b}\|^2 + \lambda \|\mathbf{x}'\|^2 \\ & \leq (1 + \alpha_S)^2 \\ & \quad \left(\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 + \lambda \|\mathbf{x}_{\text{OPT}}\|^2 \right) \\ & \quad + 2 \left(1 + \frac{1}{\alpha_S}\right) \eta \sigma^2 n d \log \frac{1}{\beta} + 2 \left(1 + \frac{1}{\alpha_S}\right) \frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} \\ & \quad \cdot \left(\frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 \right. \\ & \quad \left. + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2 \right) \end{aligned} \quad (10)$$

We now turn our attention to $\|\sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i))\|^2$. Using Lemma E.2 and Equation 7, we have

$$\begin{aligned} & \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i((\mathbf{A} + \mathbf{G}_i) \mathbf{x}_{\text{OPT}} - (\mathbf{b} + \mathbf{g}_i)) \right\|^2 + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2 \\ & \leq (1 + \alpha_S) \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b}) \right\|^2 \\ & \quad + \left(1 + \frac{1}{\alpha_S}\right) \frac{1}{s} \left\| \sum_{i \in [s]} \mathbf{S}_i(\mathbf{G}_i \mathbf{x}_{\text{OPT}} - \mathbf{g}_i) \right\|^2 + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2 \\ & \leq (1 + \alpha_S)^2 \|\mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b}\|^2 \\ & \quad + \left(1 + \frac{1}{\alpha_S}\right) \eta \sigma^2 n d \log \frac{1}{\beta} (\|\mathbf{x}_{\text{OPT}}\|^2 + 1) + \lambda \|\mathbf{x}_{\text{OPT}}\|^2 \\ & \leq (1 + \alpha_S)^2 (\|\mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}_{\text{OPT}}\|^2) \\ & \quad + \left(1 + \frac{1}{\alpha_S}\right) \frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} (\lambda \|\mathbf{x}_{\text{OPT}}\|^2 + \lambda) \\ & \leq \left((1 + \alpha_S)^2 + \left(1 + \frac{1}{\alpha_S}\right) \frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} \right) \\ & \quad (\|\mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b}\|^2 + \lambda \cdot \|\mathbf{x}_{\text{OPT}}\|^2) \\ & \quad + \left(1 + \frac{1}{\alpha_S}\right) \eta \sigma^2 n d \log \frac{1}{\beta} \end{aligned}$$

Inserting this into Equation 11 and collecting all the terms,

we obtain

$$\begin{aligned}
 & \| \mathbf{A} \mathbf{x}' - \mathbf{b} \|^2 + \lambda \| \mathbf{x}' \|^2 \\
 & \leq \left(\frac{32}{\alpha_S} \left(\frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} + \frac{2}{\alpha_S} \left(\frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} \right)^2 \right) \right. \\
 & \quad \left. + (1 + \alpha_S)^4 \right) (\| \mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b} \|^2 + \lambda \| \mathbf{x}_{\text{OPT}} \|^2) \\
 & \quad + \frac{32}{\alpha_S} \left(\frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} + \frac{2}{\alpha_S} \left(\frac{\eta \sigma^2 n d \log \frac{1}{\beta}}{\lambda} \right)^2 \right)
 \end{aligned} \tag{12}$$

By our choice of σ^2 from Theorem D.3 and using the sketching matrix of (Cohen, 2016) (see Table 3), we have $\sigma^2 \cdot n \in O(d^2 \log^4 d \cdot \alpha_S^{-4} \cdot \log^4 \frac{1}{\beta} \varepsilon^{-2} \cdot \log \frac{1}{\delta})$. Thus, we have

$$\begin{aligned}
 & \| \mathbf{A} \mathbf{x}' - \mathbf{b} \|^2 + \lambda \| \mathbf{x}' \|^2 \\
 & \leq (\| \mathbf{A} \mathbf{x}_{\text{OPT}} - \mathbf{b} \|^2 + \lambda \| \mathbf{x}_{\text{OPT}} \|^2) \cdot \left(1 + 15 \alpha_S \right. \\
 & \quad \left. + \tilde{O} \left(\frac{d^3 \alpha_S^{-5} \log^5 \frac{1}{\beta} \cdot \varepsilon^{-2} \log \frac{1}{\delta}}{\lambda} \right. \right. \\
 & \quad \left. \left. + \frac{d^6 \alpha_S^{-10} \log^{10} \frac{1}{\beta} \cdot \varepsilon^{-4} \log^2 \frac{1}{\delta}}{\lambda^2} \right) \right) \\
 & \quad + \tilde{O} \left(\frac{d^3 \alpha_S^{-5} \log^5 \frac{1}{\beta} \cdot \varepsilon^{-2} \log \frac{1}{\delta}}{\lambda} \right. \\
 & \quad \left. + \frac{d^6 \alpha_S^{-10} \log^{10} \frac{1}{\beta} \cdot \varepsilon^{-4} \log^2 \frac{1}{\delta}}{\lambda^2} \right)
 \end{aligned}$$

□

Note that we can get a constant multiplicative error as long as the regularization factor λ depends on a sufficiently large polynomial in d , while being independent of n . Generally such a relationship still has good generalization properties when training a regression model. If λ is sufficiently large, one should choose α_S to be as small as possible, minimizing the tradeoff between α_S and $\frac{\alpha_S^{-5}}{\lambda}$. If λ is not sufficiently large, one should choose $\alpha_S = 1/3$ such that \mathbf{S} is an oblivious subspace embedding, but the target dimension does not have a prohibitively large dependency on the sketch distortion.

F. Related Work

There is substantial work on the shuffle model, which also aims at facilitating differential privacy in a distributed setting. We discuss the relationship between the LTM and the

shuffle model in more detail in Section 3. For the problems we studied here, there is an abundance of prior work discussed as follows.

Low Rank Approximation: There are various ways in which one could formulate the low rank approximation problem. The setting which is most important to us is the seminal paper (Dwork et al., 2014), who achieved a worst case additive error of the order $k\sqrt{d}$ for outputting an orthogonal projection matrix V_k in the row space of \mathbf{A} . This bound is also optimal. In the local model where each client holds a row of the data matrix \mathbf{A} , and we wish to output an orthogonal projection in the column space of \mathbf{A} , (Upadhyay, 2018) gave an algorithm with an additive error of the order \sqrt{n} , which matches the lower bound by (Bassily & Smith, 2015) up to lower order terms. This is different from computing a projections in the row space of \mathbf{A} considered by (Dwork et al., 2014) and indeed our own results. The only known local algorithm to do likewise in the local model is due to (Arora et al., 2018), with a substantially larger multiplicative and additive error compared to our results. To the best of our knowledge, no single round shuffle protocol improving over the local bounds is known.

Low rank approximation and its sister problem PCA, have seen substantial attention in settings more loosely related to our work. Much of the work on private low rank approximation (Blum et al., 2005; Chaudhuri et al., 2012; Dwork et al., 2014; Hardt & Roth, 2012; 2013; Hardt & Price, 2014; Balcan et al., 2016; Kapralov & Talwar) considers the data to be fixed, often using Gaussian noise. In addition, a strong spectral gap assumption, typically of the form that the first singular value is substantially larger than the second, is crucial to the analysis. We note that all algorithms operating with this assumption do not yield worst case bounds.

When there is no spectral gap, we compare to the central model exponential mechanism approach from (Chaudhuri et al., 2012), implemented by (Kapralov & Talwar). (Liu et al., 2022) also provide a solution to this problem; however, it is unfortunately computationally intractable.

Ridge Regression The previous work most related to ours is due to (Sheffet, 2019). The author gave a mechanism that preserved the entire spectrum of a data matrix in a private manner and also showed that the returned regression vector $\hat{\mathbf{x}}$ is close to the optimal \mathbf{x} assuming that the matrix is well conditioned.

Most previous work (Kifer et al., 2012; Cai et al., 2021; Bassily et al., 2014; Wang et al., 2015; 2018a;b; 2023; Wang, 2018; Varshney et al., 2022; Milionis et al., 2022; Liu et al., 2022; Zheng et al., 2017) study private linear regression in the context of risk minimization, where the algorithm is given i.i.d. samples from some unknown distribution and

aims at computing a solution with good out of sample performance. Even without privacy constraints risk minimization yields an additive error that depends on n . Another line of work studies various loss functions, including regression in a Bayesian setting (Minami et al., 2016; Foulds et al., 2016; Dimitrakakis et al., 2014). We are not aware of any prior work on linear regression in the local DP model.

DP via Johnson-Lindenstrauss Transforms. The Johnson-Lindenstrauss lemma has been previously used in differential privacy. (Blocki et al., 2012) studies the DP guarantee a JL transform itself gives in the central model, in the context of cut queries and directional variance queries. (Kenthapadi et al., 2013) and (Stausholm, 2021a) use JL transforms together with Gaussian noise in the context of DP Euclidean distance approximation, to decrease the error. This was subsequently improved in (Stausholm, 2021b), using Laplace noise. The difference to our work is that the transform is used to reduce d instead of n , by having every client apply it locally before adding noise. In (Nikolov, 2023) the authors make use of JL transforms to achieve private query release, by applying it before releasing a bundle of queries. (Ghazi et al., 2023) studies pairwise statistics in the local model of differential privacy and uses JL transforms to reduce client-sided dimensionality.

Differentially Private Non-Oblivious Sketches. Non-Oblivious sketches, such as CountSketch, have also been used in combination with either distributed or central DP. (Zhao et al., 2022; Pagh & Thorup, 2022) provide an analysis of differentially private CountSketch in the central model. (Melis et al., 2016) consider differentially private CountSketch and Count-Min Sketch in a distributed setting, but they reduce the dimensionality of each input, which can be done by users locally, rather than reducing the dependence on the number of users, as we do.

G. Experimental Evaluation

Find here more in depth descriptions of the experiments we carried out. Section G.1 provides the running time experiments and in Section G.2 we discuss our utility investigation both for ridge regression and low-rank approximation.

G.1. Running Time of LTM

As mentioned in the main body, the only overhead of the LTM over the local model is the execution of the linear transform in MPC, and comes from distributing $\mathbf{A} + \mathbf{G}_i \in \mathbb{R}^{n \times d}$ among the servers for $i \in [s]$ and then performing matrix multiplications $\mathbf{S}_i(\mathbf{A} + \mathbf{G}_i)$ in MPC, where $\mathbf{S}_1, \dots, \mathbf{S}_s$ is a decomposition of some $\mathbf{S} \in \mathbb{R}^{m \times n}$ chosen according to (Nelson & Nguyen, 2013; Cohen, 2016) with sparsity s . We implemented our approach in MP-SPDZ (Keller, 2020), a

popular and easy to use MPC framework, for multiple combinations of n and s , using S servers from Amazon Web Services t3.large instances. We use additive secret sharing over the ring of integers modulo 2^{64} , which results in no communication between servers, and semi-honest security against an adversary that corrupts all but one servers.

We fix the dimensionality $d = 10$ and $m = 50$, and vary the number of clients $n \in \{100000, 250000, 500000, 750000, 1000000\}$, the sparsity $s \in \{1, 10, 20, 30, 40, 50\}$ of the sketch (which dictates how many linear transformations we need to apply) and the number of servers $S \in \{2, 3\}$. The data matrix \mathbf{A} is generated at random such that all entries are smaller than 2^{32} and the decomposition $\mathbf{S}_1, \dots, \mathbf{S}_s$ of some sketch \mathbf{S} is generated according to our mechanism. Every parameter setting is then evaluated by running the protocol 10 times and averaging over the running times. Table 4 shows the runtime for each server and the total communication load for one particular combination of parameters, which includes three servers: Even with one million clients, the computation on each servers lasts less than 2 seconds. We conclude that the use of MPC in the LTM does not hinder the computational efficiency of our proposed mechanisms, showing that they can easily be used in practice. Table 5 reports communication and time required when running the protocol with two servers, and Table 6 reports time and communication requirements when varying s .

Table 4. Computation cost T_{MPC} and communication cost C_{MPC} of the LTM using MPC for varying number of clients n . Here $m = 100$, $d = 10$, $s = 1$ and we are working with $S = 3$ servers.

n	T_{MPC} (sec)	C_{MPC} (MB)
100000	0.172 ± 0.019	24.024
250000	0.457 ± 0.058	60.024
500000	0.861 ± 0.102	120.024
750000	1.219 ± 0.128	180.024
1000000	1.641 ± 0.165	240.024

Table 5. Computation cost T_{MPC} and communication cost C_{MPC} of the LTM using MPC for varying number of clients n . Here $m = 100$, $d = 10$, $s = 1$ and we are working with $S = 2$ servers.

n	T_{MPC} (sec)	C_{MPC} (MB)
100000	0.177 ± 0.028	16.016
250000	0.453 ± 0.036	40.016
500000	0.824 ± 0.101	80.016
750000	1.240 ± 0.157	120.016
1000000	1.683 ± 0.130	160.016

G.2. Interpolation between Local and Central DP

In contrast to the central model of differential privacy, the performance of differentially private outputs in the local

Table 6. Computation cost T_{MPC} and communication cost C_{MPC} of the LTM using MPC for varying sparsity s . Here $m = 100$, $d = 10$, $n = 500000$ and we are working with $S = 3$ servers.

s	T_{MPC} (sec)	C_{MPC} (MB)
1	0.851 ± 0.099	80.016
10	6.863 ± 0.388	800.016
20	13.002 ± 0.445	1600.016
30	20.356 ± 1.171	2400.016
40	28.023 ± 0.660	3200.016
50	33.708 ± 1.384	4000.016

model decreases with increasing number of clients n . For regression and matrix factorization problems, our theoretical results for the LTM and the Gaussian mechanism suggest that this performance decrease can be avoided without sacrificing privacy under reasonable assumptions. It is natural to ask for which parametrizations of the Gaussian mechanism the performance remains acceptable. If more noise still yields good performance, we can tolerate more corrupted clients. Therefore, we evaluate empirically how the error of ridge regression and low-rank approximation develops as n grows in practice and against synthetic benchmarks.

The experiments described here thus aim to demonstrate that, as the number of clients increases, the asymptotic error of differential privacy mechanisms in the LTM is between the error in the central and local models.

G.2.1. RIDGE REGRESSION

We here investigate the error for the ridge regression problem due to its wide spectrum of possible error bounds from Theorem 4.5, depending on problem parameters, as well as due to its singular dependency on a multiplicative error. Therefore, we use approximation factor as an error measure:

$$\phi = \frac{\|\mathbf{Ax}' - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}'\|_2^2}{\|\mathbf{Ax}_{\text{OPT}} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}_{\text{OPT}}\|_2^2}$$

where \mathbf{x}_{OPT} denotes the optimal solution and \mathbf{x}' denotes the solution after adding noise to the training data.

We evaluate the error for synthetic datasets, as well as for four real-world datasets. It is not possible to interpolate between two different datasets in order to gain insights into asymptotic behaviour. For synthetic datasets, varying the number of sampled points n allows us to interpolate between datasets of different sizes. Real-world datasets are useful to gauge the relative performance of our mechanisms, and to benchmark our synthetic datasets against the behavior of the mechanisms in practice.

Setup. As a baseline for the central model, we implemented the so-called Sufficient Statistics Perturbation (SSP) algorithm (Vu & Slavkovic, 2009). Operating between SSP

and a locally private algorithm, our LTM model approach adds Gaussian noise sampled from $\mathcal{N}(0, \sigma^2)$ to every entry of \mathbf{A} and \mathbf{b} . Our choice of σ^2 is proportionate to varying powers of n^{-p} , ranging from $p = 1$ (the LTM model) to $p = 0$ (the local model).

We generate synthetic data by first sampling every entry in \mathbf{A} from $\mathcal{N}(0, 1)$, then sampling \mathbf{x}_{OPT} such that all entries are sampled from $\mathcal{N}(0, \mu^2)$, and finally setting $\mathbf{b} = \mathbf{Ax}_{\text{OPT}}$. We vary parameters $d \in \{3, 10, 50\}$, $\varepsilon \in \{0.01, 0.03, 0.05, 0.1\}$, $\lambda \in \{1, 10, 100\}$, and $\mu^2 \in \{1, n, n^2\}$. For all combinations of those parameters, we generate 17 synthetic datasets of sizes $\{1000^{i \cdot 1.5} | i = 0, \dots, 16\}$. For each of those we then measure ϕ for the SSP mechanism and the Gaussian mechanism with $p \in \{1, 0.9, 0.8, 0.7, 0.6, 0.5, 0\}$, by running the algorithms 30 times per dataset and reporting the average ϕ .

Real-World Datasets In addition to synthetic datasets, we also evaluated our mechanism for ridge regression on 4 datasets marked as suitable for regression on the UC Irvine Machine Learning Repository. The following table provides their number of entries n and dimensionality d . A more thorough description of the datasets is found below the table.

Dataset	n	d
Power (Hebrail & Berard, 2012)	2049280	6
Elevation (Kaul, 2013)	434874	2
Ethylene (Fonollosa, 2015)	4178504	18
Songs (Bertin-Mahieux, 2011)	515345	89

- The first dataset consists of electric power consumption measurements in one household (Hebrail & Berard, 2012) and the feature we try to predict is `sub_metering_3`. We ignore the date and time features and the data points that had missing values. This leaves us with 6 features (plus the one we are predicting) and 2049280 data points.
- The Elevation dataset (Kaul, 2013) consists of 434874 open street map elevation measurements from North Jutland, Denmark. We predict the elevation from the longitude and latitude features.
- The Ethylene dataset (Fonollosa, 2015) contains recordings of sensors exposed to a mixture of gas. We trained on the part where the sensors were exposed to a mixture of Ethylene and CO in air. The feature we are predicting is the last one `TGS2620`, which leaves us with $d = 18$ and $n = 4178504$.
- The Songs dataset (Bertin-Mahieux, 2011), consists of 89 audio features that are meant to predict the release year of a song.

Software and Hardware All in the main body described mechanisms, are implemented in Python 3.6.9 making use of numpy. Experiments were executed on an Ubuntu 18.04 LTS machine, with an Intel Core i7-10510U CPU clocked at 1.8GHz and 16GB of RAM.

Results and Interpretation. Figure 4 provides our experimental results based on synthetic data, and Table 7 shows the error resulting from real-world datasets. As expected, our approach performs asymptotically better than the Gaussian mechanism in the local model, but worse than SSP in the central model. We also found that as n increases, the error of our approach asymptotically approaches the error in the central model. For $p = 0.9, 0.8, 0.7, 0.6$, the error eventually decreases for a sufficiently large n for the privacy regimes we considered ($\epsilon \in \{0.01, 0.03, 0.05, 0.1\}$).

For $p = 0.5$ though we saw a significant jump towards the local model in terms of asymptotic behaviour. In all the settings we considered, $p = 0.5$ showed an asymptotic increase of the error. We tested on synthetic datasets of up to $n = 40$ million, and the error also increases in this regime.

Table 7 shows the results we got by applying the local, our mechanism and SSP on real-world data. They include standard deviations, though for many of the results those are so small that they appear as 0 in the table. We found, that our mechanism performs significantly better than the Gaussian mechanism for local privacy. In most it even performs so good, that there is basically no difference to the central model.

Table 7. Experimental evaluation of error ϕ on real-world datasets, including standard deviations. Here we are in the setting where $\epsilon = 0.03$ and $\lambda = 10$.

Dataset	Local	Our	Central
Power	$2.364 \pm .007$	$1.055 \pm .000$	$1.001 \pm .001$
Elevation	$1.939 \pm .008$	$1.000 \pm .000$	$1.000 \pm .000$
Ethylene	$3.125 \pm .002$	$1.000 \pm .000$	$1.000 \pm .000$
Songs	$20.64 \pm .016$	$1.000 \pm .000$	$1.000 \pm .000$

Varying λ On Figure 5, we provide plots for the same setting of parameters as on Figure 4 ($d = 10$ and $\mu^2 = n$). Though, we also vary $\lambda \in \{1, 10, 100\}$ here. The experiments show that varying λ does not change the asymptotic behavior of any of the mechanisms we investigated, in the sense that a variance proportional to $n^{-0.6}$ or lower eventually had decreasing error, while a variance proportional to $n^{0.5}$ produced increasing errors. The increase of λ does however produce a significantly lower error in all settings (note the y -axes on the figure). Our theoretical results likewise predict this behaviour, as while the error bounds of

Theorem 4.5 improve with increasing λ , they do not affect the dependency on n . Thus, we view the experiments as a confirmation that the theoretical bounds, while potentially improvable, express the correct asymptotic relationship between parameters and approximation bounds.

G.2.2. LOW-RANK APPROXIMATION

In addition to the error in ridge regression, we investigate the error when performing low-rank approximation. Compared to ridge regression, the additive error plays a more significant role (see Theorem 4.3), so we measure the error in terms of the excess risk as opposed to the approximation factor:

$$\psi = \frac{\|\mathbf{A} - \mathbf{A}\mathbf{X}'\mathbf{X}'^T\|_F^2 - \|\mathbf{A} - \mathbf{A}\mathbf{X}_{\text{OPT}}\mathbf{X}_{\text{OPT}}^T\|_F^2}{n},$$

where \mathbf{X}_{OPT} denotes the optimal solution and \mathbf{X}' denotes the solution after adding noise to the training data. Again, we evaluate the error for synthetic datasets as well as four real-world datasets. We utilize the same software, hardware and real-world datasets as for ridge regression (see Table 8). For a thorough description of the datasets, consult the above section on ridge regression. We ignore the feature we were predicting when performing ridge regression i.e. matrix \mathbf{A} is the same in both settings.

Setup. As for ridge regression, our mechanism in the LTM adds Gaussian noise sampled from $\mathcal{N}(0, \sigma^2)$ to every entry of \mathbf{A} , dependent on privacy parameters ϵ and δ . By choosing σ^2 proportional to n^{-p} for $p \in [0, 1]$, we interpolate between the local model ($p = 0$) and the central model. We also implement a Gamma/Laplace based mechanism in the LTM, which adds the sum of m samples from $\text{Lap}(0, 1/\epsilon)$ to every entry in \mathbf{SA} , in order to investigate the $\delta = 0$ regime in the LTM. As a baseline for the central model, we implemented MOD-SULQ, an approach presented by (Chaudhuri et al., 2012), where $\mathbf{A}^T \mathbf{A}$ is perturbed by adding Gaussian noise.

We generate synthetic data by first producing a matrix \mathbf{A}' where every entry is sampled from $\mathcal{N}(0, 1)$ and then changing its singular values such that there are exactly k big ones and the rest are small. More specifically, we set $\mathbf{A} = \mathbf{U}'\mathbf{\Sigma}'\mathbf{V}'$, where $\mathbf{U}'\mathbf{\Sigma}'\mathbf{V}'$ is the SVD of \mathbf{A} and $\mathbf{\Sigma}$ is a diagonal matrix with the first k values set to $\sqrt{n/k}$ and the rest set to $1/n$. We vary parameters $\epsilon \in \{0.05, 0.1, 0.5\}$ and $k \in \{5, 10\}$ for $d = 50$. For all combinations of parameters, we generate 17 synthetic datasets of sizes $\{1000^{i \cdot 1.5} | i = 0, \dots, 16\}$. For each of those we then measure ψ for the central mechanism, the Laplace mechanism and the Gaussian mechanism with $p \in \{1, 0.9, 0.8, 0.7, 0.6, 0.5, 0\}$, by running the algorithms 20 times per dataset and reporting the average ψ .

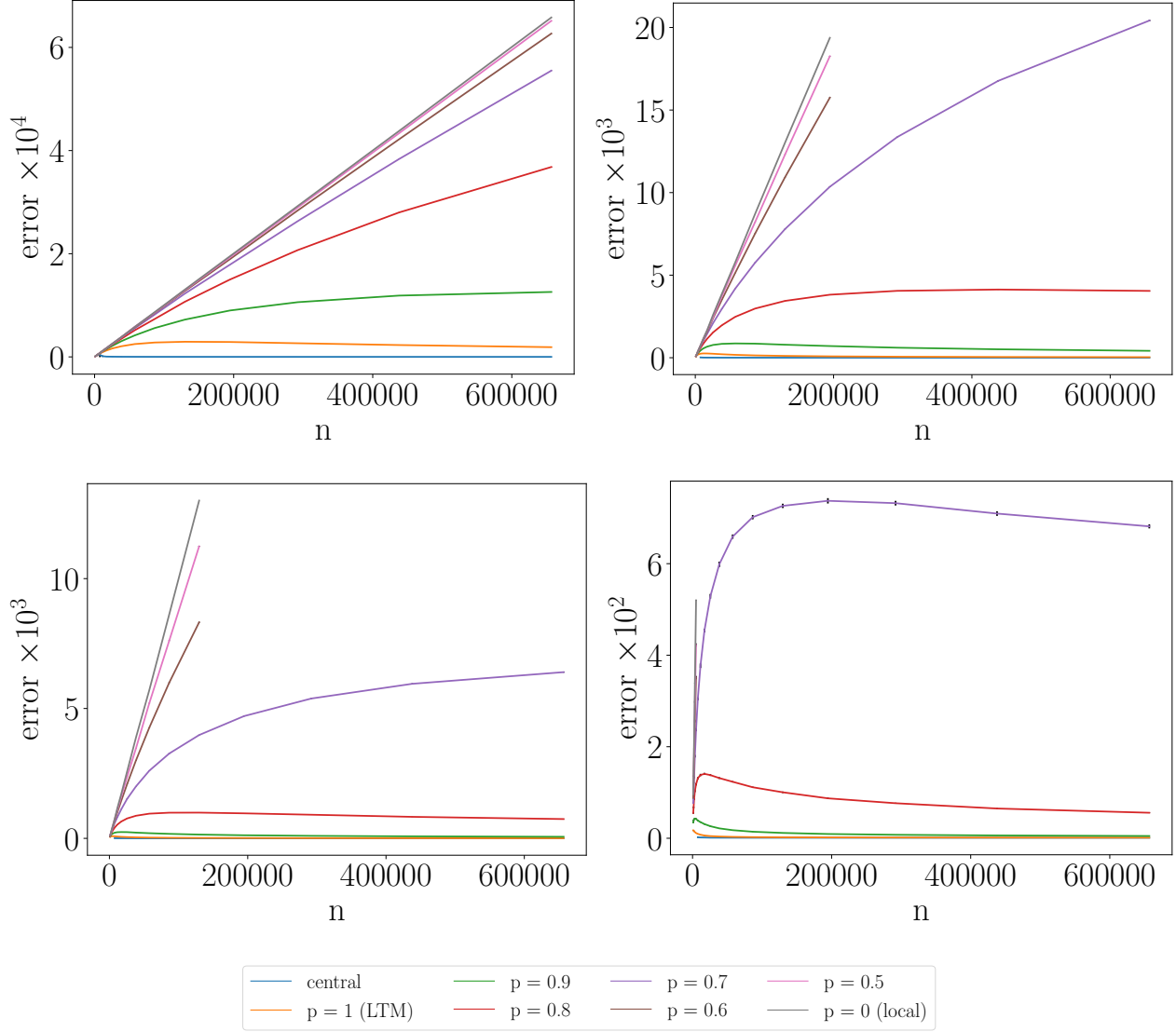


Figure 4. Plots depicting the asymptotic behavior of error ϕ for $\varepsilon \in \{0.01, 0.03, 0.05, 0.1\}$ (top left, top right, bottom left, bottom right), with $d = 10$, $\lambda = 10$ and $\mu^2 = n$. The grey line depicts the error of the local mechanism, the blue line does it in the central model and the orange one depicts our approach. The other lines resemble different values of p . In most cases the standard deviations are so small, that it is not possible to see those.

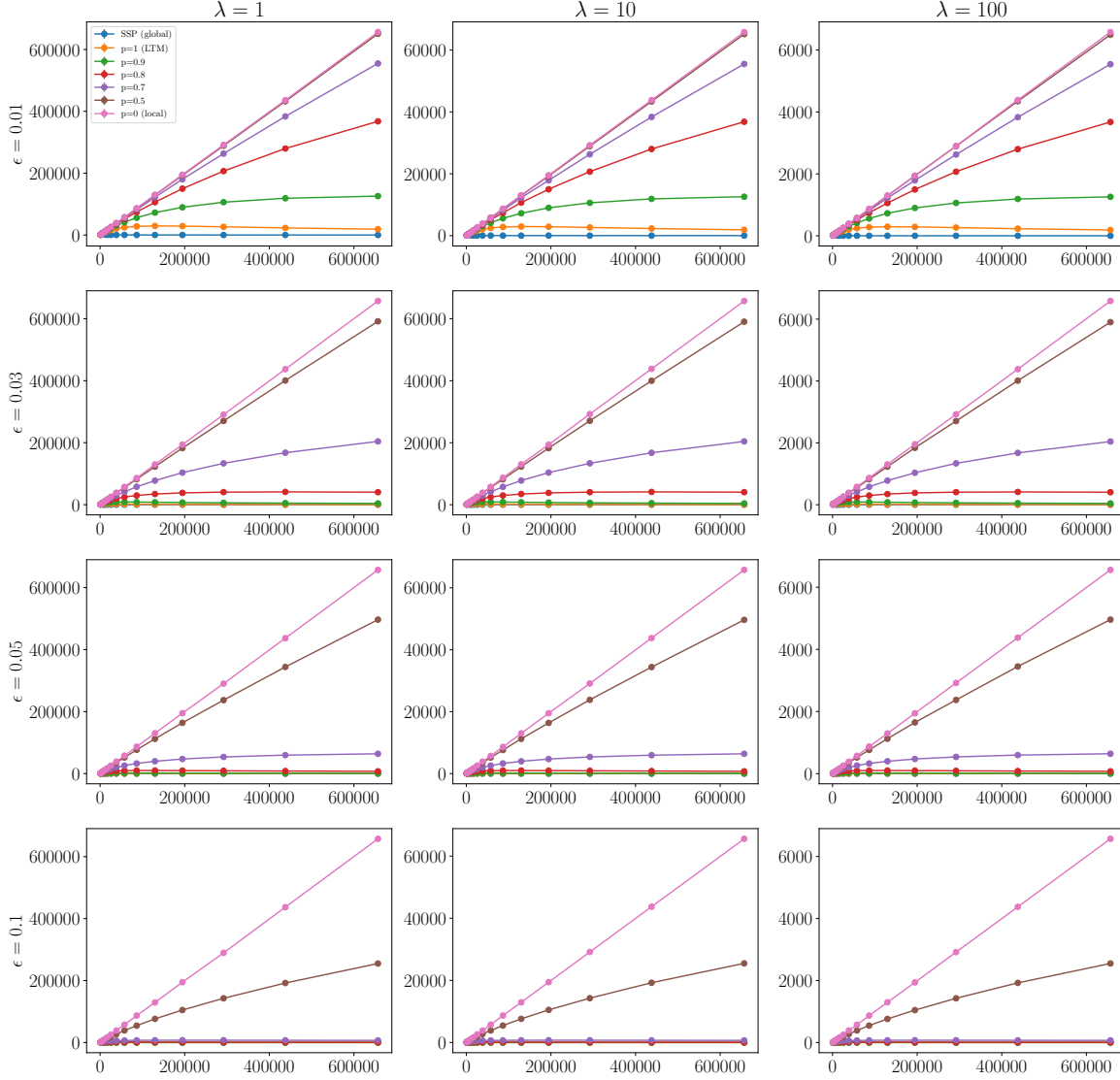


Figure 5. Plots depicting the error ϕ as n increases, for $\varepsilon \in \{0.01, 0.03, 0.05, 0.1\}$ (top to bottom) and $\lambda \in \{1, 10, 100\}$ (left to right). The middle row depicts the same choice of parameters as Figure 4.

Results and Interpretation. Figure 6 shows our experimental results for some of the parameters mentioned above, based on synthetic data and Table 9 provides the error resulting from real-world datasets summarized in Table 8. As expected and in par with ridge regression, our approach performs better than the Gaussian mechanism in the local model but worse than MOD-SULQ in the central model. We observed that as n grows, the error of our approach asymptotically approaches the error in the central model. In the privacy regimes we considered ($\varepsilon \in \{0.05, 0.1, 0.5\}$), the error of intermediate mechanisms ($p \in [0, 1]$), decreases with increasing n as well.

Table 9 shows the experimental results of running our mechanisms, the Gaussian mechanism in the local model and MOD-SULQ in the central model on real-world data. The k 's we used for the individual datasets are provided in Table 8. Similar to ridge regression, we found that our Gaussian mechanism in the LTM performs significantly better than the Gaussian mechanism in the local model. Both with synthetic and real-world datasets, our Laplace approach also performs better than the local model, though not as significantly as the Gaussian approach. However, this is the price of achieving pure ε -differential privacy with $\delta = 0$.

Table 8. Parameters of real-world datasets.

Dataset	n	d	k
Power (Hebrail & Berard, 2012)	2049280	6	3
Elevation (Kaul, 2013)	434874	2	1
Ethylene (Fonollosa, 2015)	4178504	18	5
Songs (Bertin-Mahieux, 2011)	515345	89	15

H. Distributed Computation of Linear Sketches

The linear transformation we consider is motivated by efficient secure distributed cryptographic tools from secure multi-party computation; in particular, we can use linear secret-sharing. We require this secret sharing scheme to be information-theoretically secure. We use an additive secret sharing scheme, such that clients C_1, \dots, C_n generate k shares of their noisy inputs x^1, \dots, x^n , where k is the number of servers that compute the linear transformation. We use $[x^i]$ to denote a secret sharing of some client C_i 's noisy input x^i for some $i \in [n]$. For some field \mathbb{F} of size p and some prime number p , $[x^i]$ consists of shares $x_1^i, \dots, x_k^i \in \mathbb{F}$ such that $\sum_{j=1}^k x_j^i = x^i$. To split a secret into k shares, a client can sample $k - 1$ random field elements x_1^i, \dots, x_{k-1}^i and compute the last secret share as $x^i - \sum_{j=1}^{k-1} x_j^i$. Therefore, it also seems intuitive that an adversary that sees all but one of the shares knows nothing about the input. For every $j \in [k]$, server S_j receives $\{x_j^i\}_{i \in [n]}$ from the respective parties. We define these shares

such that the security of our distributed protocol does not rely on any computational assumption.

Since the linear transformation is public, servers can apply the transformation locally to their secret shares without any need to communicate with each other. Each server then reveals their shares of the resulting linear transformation, such that the linear sketch can be revealed in the clear. For more details on MPC based on secret sharing, see e.g., (Cramer et al., 2015).

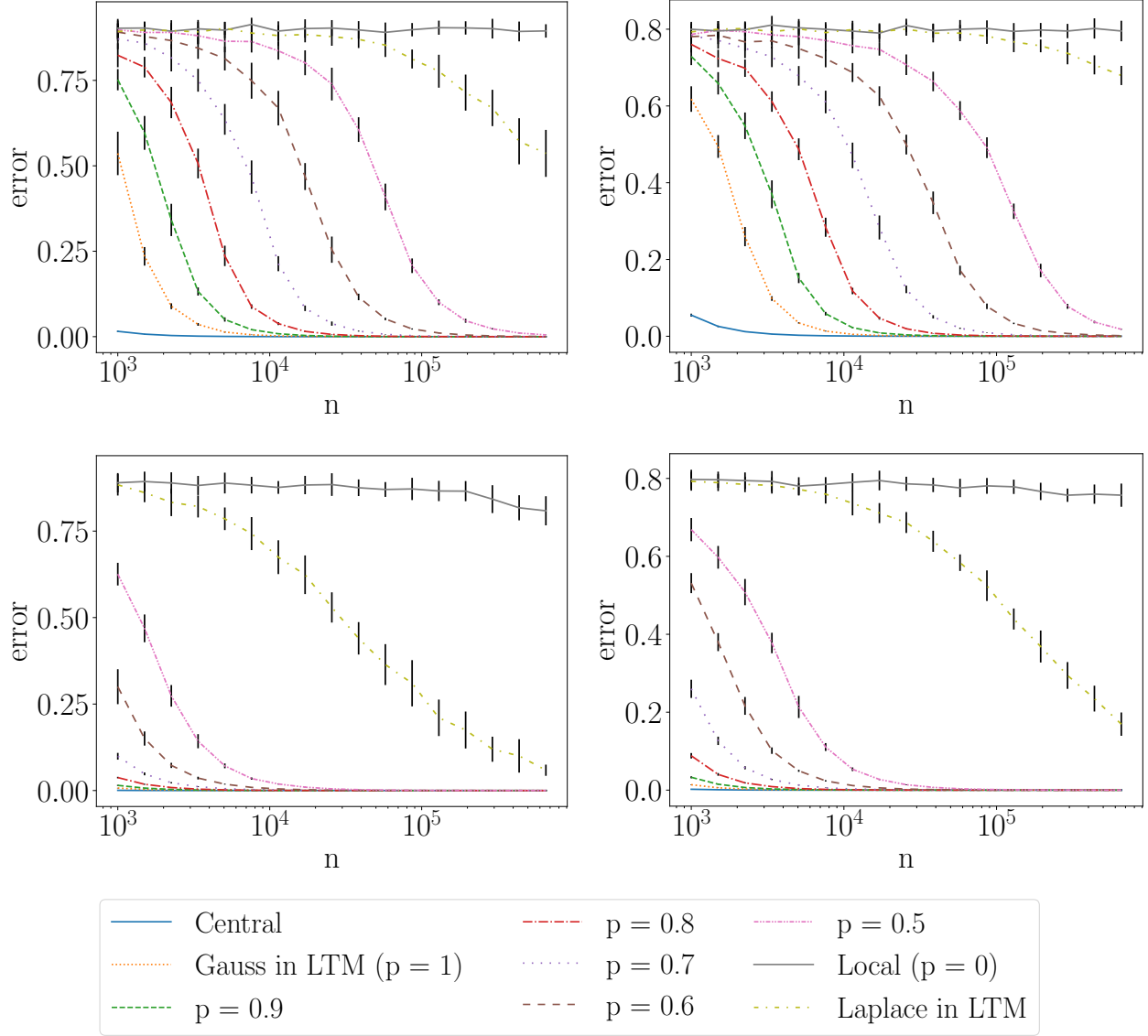


Figure 6. Plots depicting the asymptotic behavior of error ψ for $\epsilon \in \{0.1, 0.5\}$ (top, bottom) and $k \in \{5, 10\}$ (left, right), with $d = 50$. The gray line depicts the error of the local mechanism, and the orange and lime lines depict our approach using Gaussian and Laplacian noise respectively. The other lines resemble different values of p when using Gaussian noise. The standard deviations are depicted by the vertical black lines and the x -axis is logarithmic in the number of clients n .

Table 9. Experimental evaluation of error ψ on real-world datasets, including standard deviations. Here we are in the setting where $\varepsilon = 0.05$.

Dataset	Local	LTM Laplace	LTM Gauss	Central
Power	$(1.206 \pm 0.651) \times 10^{-2}$	$(4.036 \pm 2.448) \times 10^{-3}$	$(2.108 \pm 2.400) \times 10^{-6}$	$(1.763 \pm 0.621) \times 10^{-6}$
Elevation	$(1.193 \pm 0.896) \times 10^{-1}$	$(1.520 \pm 4.484) \times 10^{-2}$	$(4.718 \pm 4.201) \times 10^{-8}$	$(4.465 \pm 4.836) \times 10^{-8}$
Ethylene	$(5.877 \pm 1.256) \times 10^{-4}$	$(5.821 \pm 1.074) \times 10^{-4}$	$(2.420 \pm 0.486) \times 10^{-6}$	$(7.901 \pm 1.909) \times 10^{-6}$
Songs	$(5.617 \pm 0.283) \times 10^{-7}$	$(5.629 \pm 0.311) \times 10^{-7}$	$(5.556 \pm 0.347) \times 10^{-7}$	$(5.613 \pm 0.324) \times 10^{-7}$