Statistica Sinica Preprint No: SS-2026-0091	
Title	Differentially Private Truncation of Unbounded Data Via Public Second Moments
Manuscript ID	SS-2026-0091
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0091
Complete List of Authors	Zilong Cao, Xuan Bi and Hai Zhang
Corresponding Authors	Hai Zhang
E-mails	zhanghai@nwu.edu.cn
Notice: Accepted author version.	

Statistica Sinica

Differentially Private Truncation of Unbounded Data via Public Second Moments

Zilong Cao[1], Xuan Bi[2] and Hai Zhang[⋆,1]

[1]*School of Mathematics, Northwest University, China*

[2]*Department of Information and Decision Sciences, University of Minnesota, USA*

Abstract: Data privacy is important in the AI era, and differential privacy (DP) is one of the golden solutions. However, DP is typically applicable only if data have a bounded underlying distribution. We address this limitation by leveraging second-moment information from a small amount of public data. We propose Public-moment-guided Truncation (PMT), which transforms private data using the public second-moment matrix and applies a principled truncation whose radius depends only on non-private quantities: data dimension and sample size. This transformation yields a well-conditioned second-moment matrix, enabling its inversion with a significantly strengthened ability to resist the DP noise. Furthermore, we demonstrate the applicability of PMT by using penalized and generalized linear regressions. Specifically, we design new loss functions and algorithms, ensuring that solutions in the transformed space can be mapped back to

⋆Corresponding author (zhanghai@nwu.edu.cn).

This research was partially supported by the National Natural Science Foundation of China (No.12326615) and the Major Key Project of PCL under Grant PCL2024A06, and the Independent Research Project of the National Key Laboratory of Big Data and Decision.

the original domain. We have established improvements in the models' DP estimation through theoretical error bounds, robustness guarantees, and convergence results, attributing the gains to the conditioning effect of PMT. Experiments on synthetic and real datasets confirm that PMT substantially improves the accuracy and stability of DP estimators.

Key words and phrases: Differential privacy, Public data, Data truncation, Penalized regression, Generalized linear model

1. Introduction

Data privacy has become an increasingly critical challenge in today's data-driven world. Protecting large volumes of sensitive information is essential, especially when analytical results are derived from private data. Differential Privacy (DP; Dwork et al. (2006)) provides a rigorous mathematical framework that ensures the output of an algorithm is statistically insensitive to any single data point, thus offering strong privacy guarantees. To enhance the utility and composability of standard DP, Dong et al. (2022) introduced Gaussian Differential Privacy (GDP), which achieves the tightest known composition bounds for Gaussian mechanisms. GDP has since found broad applications across statistics (Awan and Vadhan, 2023; Avella-Medina et al., 2023; Zhao et al., 2025), machine learning (Bu et al., 2020b; Cao et al., 2023; Nasr et al., 2023), and other areas, demonstrating its

theoretical and practical effectiveness.

GDP relies on adding Gaussian noise to the output of a data-dependent algorithm, a process known as the Gaussian mechanism in differential privacy. However, when such mechanisms rely solely on private data, they often suffer from limited utility. More importantly, it requires data to be strictly bounded in order for GDP to be applicable (Bu et al., 2020a). One common practice for handling unbounded data is to apply truncation. However, this inevitably compromises data usefulness. On the one hand, using a small truncation radius can substantially distort the original data distribution. On the other hand, using a large truncation radius necessitates injecting a large noise to satisfy the same level of differential privacy guarantee, which can also degrade utility. In either case, truncation imposes unavoidable alterations to the data and compromise subsequent analysis. Public data, as a high-quality and privacy-free resource, has been increasingly leveraged to improve the performance and utility of DP algorithms in query release, synthetic data generation, learning, and prediction (Ji and Elkan, 2013; Nandi and Bassily, 2020; Bassily et al., 2020; Liu et al., 2021; Bi and Shen, 2023). More practically, many public datasets do not contain raw sensitive data, but rather privacy-preserving yet informative statistics, such as means or moments. For statistical estimation, Ferrando

et al. (2021) proposes a principled approach to combining public and private data. In differentially private Gaussian distribution estimation, Bie et al. (2022) shows that even a small amount of public data can substantially enhance performance. Public data has also been used to improve the utility of DP gradient-based learning algorithms, as demonstrated in works such as Kairouz et al. (2021), Amid et al. (2022), and Nasr et al. (2023). In contrast to these uses of public data as auxiliary samples or guidance for private estimation and optimization, our work exploits a public second-moment matrix as a geometric preconditioner for private data. This distinction allows to directly address two regression-specific bottlenecks under DP: sensitivity control for unbounded covariates and stability of noisy second-moment inversion.

In the context of differentially private regression, the literature mainly focuses on the linear regression, penalized linear regression, and generalized linear models. The prior studies on DP ordinary least squares (OLS) focus on linear regression under the assumption of bounded private data (Sheffet, 2017; Wang, 2018; Bernstein and Sheldon, 2019). DP penalized regression with the L_2 regularization also is studied, where injecting noise into the sufficient statistics can produce a DP-compliant solution (Wang, 2018; Bernstein and Sheldon, 2019). These methods typically treat the L_2

regularization as a tuning parameter to stabilize the inverse of the noisy second-moment matrix. Generalized linear models with DP guarantees is another direction in the literature, which can be solved via a convex loss, requiring iterative solving to the convex optimization problem. large body of work explore DP gradient-based optimization methods under bounded-data assumptions (Abadi et al., 2016; Ivkin et al., 2019; Wang and Zhang, 2019; Koloskova et al., 2023). More recently, DP Newton methods, which leverage second-order information, have gained attention for their faster convergence and reduced privacy budget consumption (Ganesh et al., 2023; Cao et al., 2023; Nasr et al., 2023). However, these approaches exhibit sensitivity to the ill-conditioning of the Hessian matrix, frequently leading to numerical instability during iterations, and struggle with appropriately adjusting the regularization value within the DP framework. Selecting an appropriate regularization is challenging, especially when it depends on private data. Small regularization is unable to resist the DP noises effectively, but large regularization leads to over-regularization and substantial estimation bias.

In this paper, we provide a practical solution for achieving differential privacy with unbounded data by leveraging a public second-moment matrix and demonstrating its effectiveness in regression settings. We propose Public-moment-guided Truncation (PMT), a transformation–truncation frame-

work that maps private data into an approximately isotropic space using the public second-moment matrix, where the l_2-norm of each transformed data point is bounded by $\sqrt{d(1+\log n)}$ with high probability, enabling principled truncation unrelated with private data. This transformation alleviates truncation distortion and produces a well-conditioned second-moment matrix, whose inverse is more stable and less sensitive to DP noise. Furthermore, we analyze the robustness of inverse second-moment estimation under differential privacy and prove that PMT significantly improves inverse accuracy by improving the condition number of the transformed matrix, reducing sensitivity to DP noise, weakening dependence on regularization, and lowering the private sample size requirement, with extensions to weighted second-moment matrices. To demonstrate its applicability, we develop two algorithms: DP-PMTRR for ridge regression, which combines PMT with sufficient statistics perturbation and a tailored loss function to obtain a closed-form and robust estimator, and DP-PMTLR for logistic regression, which integrates PMT into a DP Newton's method with a modified cross-entropy loss to improve convergence, numerical stability, and estimation accuracy without manual regularization tuning. More broadly, we establish the general applicability of PMT to penalized generalized linear models by constructing invariant loss functions in the transformed domain and show-

ing how solutions can be mapped back to the original parameter space. We provide rigorous theoretical guarantees, including formal DP estimator error bounds for DP ridge regression and stable convergence results for DP logistic regression, and show that PMT yields substantial utility improvements over private-data-only approaches by mitigating ill-conditioning, large inverse norms, and excessive regularization dependence.

2. Preliminaries and Motivation

In this section, we firstly remind readers of the background of differential privacy. Next, we illustrate the existing problems in DP and our motivation.

2.1 Background of Privacy

Definition 1 (Differential Privacy Dwork et al. (2014))**.** *A randomized algorithm* $\mathcal{M} : \mathcal{X}^n \to \mathcal{S}$ *satisfies* $(\epsilon, \delta)-$*differential privacy* ((ϵ, δ)*-DP), if for any neighboring datasets* $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ *(they differ in only one sample) and* $\forall S \subseteq \mathcal{S}, \epsilon > 0, \delta > 0$*, the following probability inequality hold*

$$\mathbb{P}[\mathcal{M}(\mathbf{X}) \in S] \leq \exp(\epsilon)\mathbb{P}[\mathcal{M}(\mathbf{X}') \in S] + \delta,$$

when $\delta = 0$ *means* ϵ*-DP.*

This definition strictly controls the distinguishability of outputs under neighboring datasets; hence, algorithms satisfying DP are insensitive to someone individual and protect the individual privacy. Dong et al. (2022) proposes another definition of DP as follows.

Definition 2 (f-Differential Privacy Dong et al. (2022))**.** *Let X and X' be two neighboring datasets with domain $\mathcal{X}^n$ and a randomized function $\mathcal{M}: \mathcal{X}^n \to \mathbb{R}^p$. We say that $\mathcal{M}$ satisfies f-differential privacy (f-DP), if any α-level test on the hypotheses of $\mathcal{M}$ outputs, $H_0: \mathcal{M}$ based on X vs. $H_1: \mathcal{M}$ based on X', has power function $\beta(\alpha) \leqslant 1 - f(\alpha)$, where f is a convex, continuous, non-decreasing function satisfying $f(\alpha) \leqslant 1 - \alpha$ for all $\alpha \in [0, 1]$.*

Definition 3 (Guassian Differential Privacy Dong et al. (2022))**.** *If $\mathcal{M}$ satisfies f-DP and $f(\alpha) \geqslant \Phi(\Phi^{-1}(1-\alpha) - \mu)$ for all $\alpha \in [0, 1]$, where Φ is the cumulative distribution function of the standard normal distribution and μ is a constant, then $\mathcal{M}$ satisfies μ-Gaussian differential privacy (μ-GDP).*

There exists a relationship between (ϵ, δ)-DP and μ-GDP as the following lemma. These DPs hold two important properties: the composition and post-processing. (i) The *post-processing* property means that the output of a DP algorithm will not lose any privacy budget after any post-processing of not contacting the private data. (ii) The *composition* property means that

the privacy loss of two DP algorithms is the composition of their privacy losses. Because μ-GDP possesses the better composition property, we state it as the following theorem.

Lemma 1 (**Corollary 1** in Dong et al. (2022)). *A mechanism $\mathcal{M}$ satisfies μ-GDP if and only if it is $(\epsilon, \delta(\epsilon))$-DP for $\forall\ \epsilon \geqslant 0$, where* $\delta(\epsilon) = \Phi(-\frac{\epsilon}{\mu} + \frac{\mu}{2}) - e^{\epsilon}\Phi(-\frac{\epsilon}{\mu} - \frac{\mu}{2})$.

Theorem 1 (The T-fold *composition*, Dong et al. (2022)). *If $\mathcal{M}_i(\mathbf{X}, \mathcal{M}_{i-1}, ..., \mathcal{M}_1) : \mathcal{X}^n \times Y_1 \times \cdots \times Y_{i-1} \to Y_i$ is μ_i-GDP for $i = 1, .., T$, then $\mathcal{M}(\mathbf{X}) = \mathcal{M}_1 \circ \mathcal{M}_2 \circ ... \circ \mathcal{M}_T : \mathcal{X}^n \to Y_1 \times \cdots \times Y_T$ is $\mu = \sqrt{\sum_{i=1}^{T} \mu_i^2}$-GDP.*

And the simple way to achieve GDP is to add Gaussian noise to the output of the algorithm, called the Gaussian mechanism. The Gaussian mechanism requires the bounded algorithmic sensitivity. The sensitivity of an algorithm always depends on the domain of the private data. Particularly, when the data domain is unbounded, the sensitivity needs to be controlled by truncating the data. The related definition and theorem are as follows.

Definition 4 (Sensitivity). *The l_2-sensitivity of a function $h : \mathcal{X}^n \to \mathbb{R}^p$ is defined as*

$$\Delta_h = \max_{\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n} ||f(\mathbf{X}) - f(\mathbf{X}')||_2,$$

where $\mathbf{X}$ *and* $\mathbf{X}'$ *are neighboring datasets.*

Theorem 2 (Gaussian Mechanism in Dong et al. (2022))**.** *Let* $h : \mathcal{X}^n \to \mathbb{R}^p$ *be a function with* l_2*-sensitivity* Δ_h*. Let* $\mathbf{g} \in \mathbb{R}^p$ *be a standard normal random vector,* $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p\times p})$*. The Gaussian mechanism* $\mathcal{G}(\mathbf{X})$ *is defined as*

$$\mathcal{G}(\mathbf{X}) = h(\mathbf{X}) + \frac{\Delta_h}{\mu}\mathbf{g},$$

$\mathcal{G}(\mathbf{X})$ *is* μ*-GDP.*

2.2 Problems and Motivation

In this paper, we assume that both private and public data are independently and identically distributed from a sub-Gaussian distribution $subG(\Sigma)$ with second-moment matrix Σ, which covers many common bounded and unbounded distributions (Wainwright, 2019). However, the unboundedness leads to infinite sensitivity for fundamental statistics such as the sample mean and second-moment matrix, making direct application of differential privacy infeasible. A common remedy is data truncation,

$$T(\mathbf{x})_R = \begin{cases} \mathbf{x}, & \|\mathbf{x}\|_2 \le R, \\ \dfrac{\mathbf{x}}{\|\mathbf{x}\|_2} R, & \|\mathbf{x}\|_2 > R, \end{cases}$$

where R is the truncation radius. A larger radius preserves more information but increases sensitivity and DP noise, while a smaller radius reduces sensitivity at the cost of severe information loss. This trade-off raises a fundamental question:

How to truncate unbounded data with a principled radius?

The second-moment matrix and its inverse play a central role in many statistical procedures, including regression and Newton-type optimization methods. In the DP setting, the second-moment matrix is typically perturbed by Gaussian noise, and its inverse is computed based on the noisy version. When the original matrix is ill-conditioned, the perturbed inverse becomes unstable and highly sensitive to noise, often requiring heavy regularization that introduces substantial bias and degrades statistical utility. This leads to a second key question:

How to obtain a robust and accurate inverse second-moment matrix under differential privacy?

For isotropic sub-Gaussian data, the sample norm concentrates around $\sqrt{d}$, yielding a natural truncation radius of order $\sqrt{d(1+\log(1/\eta))}$ with high probability. However, for general non-isotropic data, this radius depends on the eigenvalues of Σ and may be excessively large, exacerbating

both truncation and inverse instability issues. This observation suggests that an appropriate transformation that makes the data approximately isotropic can simultaneously yield a principled truncation radius and improve the conditioning of the second-moment matrix. Consequently, the transformed data admit bounded norms and a more stable inverse, improving robustness and utility in subsequent DP procedures. Since private data cannot be used directly for this transformation, it is natural to consider whether the second-moment matrix estimated from public data can serve as an effective proxy. This motivates the development of a public-moment-guided transformation framework.

3. Public-moment-guided Truncation

In this section, we propose the public-moment-guided truncation (PMT) in Subsection 3.1. Subsection 3.2 will show the inverse second-moment matrix estimation of the transformed data is more robust and accurate. In particular, we show the detailed comparison between PMT and the private-data-only method in theoretical results.

3.1 Public-moment-guided Truncation Method

In this subsection, we propose **Algorithm 1**, which uses the public second-moment matrix to transform the private data to be an approximate isotropic form and then truncates the transformed data within a principled radius. **Theorem 3** guarantees the transformed data with an approximate isotropic form. **Corollary 1** shows the transformed data can be truncated with a principled radius.

Algorithm 1 Public-Moment-guided Truncation (PMT)

Input: Private dataset $\{\boldsymbol{\xi}_i\}_{i=1}^{n_\xi}$, the public second-moment matrix $\hat{\Sigma} = \frac{1}{n_\upsilon}\sum_{i=1}^{n_\upsilon} \boldsymbol{\upsilon}_i\boldsymbol{\upsilon}_i^T$, parameters d, n_ξ, n_υ and η.
Transform private data: $\tilde{\boldsymbol{\xi}}_i = \hat{\Sigma}^{-1/2}\boldsymbol{\xi}_i,\ i = 1, ..., n_\xi$.
Truncate data: for every transformed data $\tilde{\boldsymbol{\xi}}_i, i = 1, ..., n_\xi$,
while $i \in [n_\xi]$ **do**
  if $\|\tilde{\boldsymbol{\xi}}_i\|_2 \geq \sqrt{d(1+\log(\frac{2n_\xi}{\eta}))}$ **then**
    $\tilde{\boldsymbol{\xi}}_i \leftarrow \sqrt{d(1+\log(\frac{2n_\xi}{\eta}))} \cdot \frac{\tilde{\boldsymbol{\xi}}_i}{\|\tilde{\boldsymbol{\xi}}_i\|_2}$,
  else
    $\tilde{\boldsymbol{\xi}}_i$ is itself.
  end if
end while
Output: Dataset $\{\tilde{\boldsymbol{\xi}}_i\}_{i=1}^{n_\xi}$ and the public second-moment matrix $\hat{\Sigma}$.

Remark 1. *The algorithm only needs a second-moment matrix $\hat{\Sigma}$. Hence, we can weaken the public data requirement to a public second-moment estimation $\hat{\Sigma}$ which is easier to attain and more safe for privacy.*

Theorem 3 (Bound the second-moment matrix)**.** *Denote that a random*

vector $\boldsymbol{\xi} \in \mathbb{R}^{d\times 1} \sim subG(\Sigma)$, *where* $\Sigma = \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)$ *is the second-moment matrix. Suppose* $\boldsymbol{\Upsilon} \in \mathbb{R}^{n\times d}$ *is a data matrix whose elements* $\boldsymbol{\upsilon}_i's$ *are i.i.d.s sample drawn from* $subG(\Sigma)$ *and* $\hat{\Sigma} = \frac{1}{n}\boldsymbol{\Upsilon}^T\boldsymbol{\Upsilon}$ *is an estimation of the second-moment matrix. Then* $\tilde{\boldsymbol{\xi}} = \hat{\Sigma}^{-1/2}\boldsymbol{\xi} \sim subG(\hat{\Sigma}^{-1/2}\Sigma\hat{\Sigma}^{-1/2})$ *and, with at least probability* $1-2\eta$,

$$L\mathbf{I} \preceq \hat{\Sigma}^{-1/2}\Sigma\hat{\Sigma}^{-1/2} \preceq U\mathbf{I},$$

where $L = \frac{n}{(\sqrt{n}+O(\sqrt{d}+\sqrt{\log(\frac{1}{\eta})}))^2}$ *and* $U = \frac{n}{(\sqrt{n}-O(\sqrt{d}+\sqrt{\log(\frac{1}{\eta})}))^2}$.

Corollary 1 (Untility of truncation)**.** *Denote that random vectors* $\boldsymbol{\xi}_i \in \mathbb{R}^{d\times 1} \overset{i.i.d.}{\sim} subG(\Sigma)$, $i = 1, ..., n_\xi$, *where* $\Sigma = \mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)$ *is the second-moment matrix. Suppose* $\boldsymbol{\Upsilon} \in \mathbb{R}^{n_\upsilon\times d}$ *is a data matrix whose elements are i.i.d. samples drawn from* $subG(\Sigma)$, *i.i.d. and* $\hat{\Sigma} = \frac{1}{n_\upsilon}\boldsymbol{\Upsilon}^T\boldsymbol{\Upsilon}$ *is an estimation of the second-moment matrix. Let* $\tilde{\Sigma} = \hat{\Sigma}^{-1/2}\Sigma\hat{\Sigma}^{-1/2}$, *then* $\tilde{\boldsymbol{\xi}}_i = \hat{\Sigma}^{-1/2}\boldsymbol{\xi}_i \sim subG(\tilde{\Sigma})$ *and, with at least probability* $1-3\eta$,

$$\|\tilde{\boldsymbol{\xi}}_i\|_2^2 \le \mathrm{Tr}(\tilde{\Sigma}) + O(d\log(\frac{2n_\xi}{\eta})) \le O(d(1+\log(\frac{2n_\xi}{\eta}))), \quad \forall i \in [n_\xi].$$

3.2 Robust Private Second-moment Matrix

Next, we give the DP second-moment matrix of the transformed data and guarantee its μ-GDP.

Theorem 4 (Private Second-moment)**.** *Denote that private random* $\boldsymbol{\xi}_i \in \mathbb{R}^{d\times 1} \overset{i.i.d.}{\sim} subG(\Sigma)$, $i = 1, ..., n_\xi$, *where* $\Sigma = \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)$ *is the second-moment matrix. Suppose* $\boldsymbol{\Upsilon} \in \mathbb{R}^{n_\upsilon \times d}$ *is a public data matrix whose elements* $\boldsymbol{\upsilon}_i$*'s are i.i.d. samples drawn from* $subG(\Sigma)$*. If the transformed data* $\tilde{\boldsymbol{\xi}}_i$*'s are from* ***Algorithm 1*** *with* $1 > \eta > 0$ *, the DP transformation second-moment matrix*

$$\tilde{\Sigma}_{DP} = \frac{1}{n_\xi}\sum_{i=1}^{n_\xi} \tilde{\boldsymbol{\xi}}_i\tilde{\boldsymbol{\xi}}_i^T + \mathbf{G}$$

satisfies the μ*-GDP, where* $\mathbf{G} \sim SG_d(\sigma^2)$ *is a* $d-$*dimensional symmetric Gaussian random matrix with the parameter* $\sigma = \frac{2d(1+\log(\frac{2n_\xi}{\eta}))}{\mu\cdot n_\xi}$*. And with at least probability* $1 - O(\eta)$*, the* $\mathbf{G}$ *has bound as following:*

$$\|\mathbf{G}\|_2 \leq \frac{\sqrt{d^3\log(\frac{2d}{\eta})}(1+\log(\frac{2n_\xi}{\eta}))}{\mu\cdot n_\xi}. \tag{3.1}$$

There exist relationships between the transformed matrices and the original matrices. Moreover, we consider the regularized second-moment matrix, which is more general than the second-moment matrix. The following theorem shows the transformed data offer the more robust and accurate inverse second-moment matrix.

Theorem 5 (Robust DP Inverse)**.** *Denote that private random* $\boldsymbol{\xi}_i \in \mathbb{R}^{d\times 1} \overset{i.i.d.}{\sim} subG(\Sigma)$, $i = 1, ..., n_\xi$, *where* $\Sigma = \mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)$ *is the second-moment matrix*

and $\hat{\Sigma}_\xi = \frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T$. *Suppose* $\boldsymbol{\Upsilon} \in \mathbb{R}^{n_\upsilon \times d}$ *is a public data matrix whose elements* $\boldsymbol{\upsilon}_i$*'s are i.i.d. samples drawn from subG*(Σ) *and* $\hat{\Sigma}_\upsilon = \frac{1}{n_\upsilon}\boldsymbol{\Upsilon}^T\boldsymbol{\Upsilon}$ *is an estimation of the second-moment matrix. The transformed data* $\tilde{\boldsymbol{\xi}}_i s$ *and their data matrix* $\tilde{\boldsymbol{\Xi}} \in \mathbb{R}^{n_\xi \times d}$ *are from* ***Algorithm 1*** *with* $1 > \eta > 0$. *Considering the regularized second-moment matrix with the regularization parameter* λ

$$\tilde{\Sigma}_\xi + \lambda\hat{\Sigma}_\upsilon^{-1} = \frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\tilde{\boldsymbol{\xi}}_i\tilde{\boldsymbol{\xi}}_i^T + \lambda\hat{\Sigma}_\upsilon^{-1},$$

we have the following results:

1. Recovery. *With at least probability* $1 - O(\eta)$,

$$\hat{\Sigma}_\upsilon^{1/2}(\tilde{\Sigma}_\xi + \lambda\hat{\Sigma}_\upsilon^{-1})\hat{\Sigma}_\upsilon^{1/2} = \hat{\Sigma}_\xi + \lambda\mathbf{I}.$$

2. Inverse error. *When the public data satisfies* $\sqrt{n_\upsilon} \geqslant O(\sqrt{d} + \sqrt{\log(1/\eta)})$ *and the private data* n_ξ *makes* $\frac{\sqrt{d^3\log(\frac{2d}{\eta})}(1+\log(\frac{2n_\xi}{\eta}))}{(L(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^2+\lambda\|\hat{\Sigma}_\upsilon\|^{-1})\cdot\mu\cdot n_\xi} \leqslant \frac{1}{2}$, *with at least probability* $1 - O(\eta)$, *we have*

$$\|(\tilde{\Sigma}_\xi+\lambda\hat{\Sigma}_\upsilon^{-1}+\mathbf{G})^{-1}-(\tilde{\Sigma}_\xi+\lambda\hat{\Sigma}_\upsilon^{-1})^{-1}\|_2 \leqslant \frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\frac{(1+\log(\frac{2n_\xi}{\eta}))}{L^2(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^4+\lambda^2\|\hat{\Sigma}_\upsilon\|^{-2}}, \tag{3.2}$$

where $L = \frac{n_\upsilon}{(\sqrt{n_\upsilon}+O(\sqrt{d}+\sqrt{\log(\frac{1}{\eta})}))^2}$.

Moreover, with at least probability $1 - O(\eta)$, *we have the error bound*

about the original inverse regularized second-moment matrix

$$\|\hat{\Sigma}_v^{-1/2}(\tilde{\Sigma}_\xi+\lambda\hat{\Sigma}_v^{-1}+\mathbf{G})^{-1}\hat{\Sigma}_v^{-1/2}-(\hat{\Sigma}_\xi+\lambda\mathbf{I})^{-1}\|_2 \leqslant \frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\frac{\|\hat{\Sigma}_v^{-1}\|(1+\log(\frac{2n_\xi}{\eta}))}{L^2(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^4+\lambda^2\|\hat{\Sigma}_v\|^{-2}}. \tag{3.3}$$

It is worth noting that the terms involving public data, $\hat{\Sigma}_v$, are retained in the bound Eq.(3.3), but it typically has negligible impact on the overall utility of the inverse matrix. Then, we show that, based solely on private data, the utility of the inverse DP second-moment matrix is impacted by $\hat{\Sigma}_\xi$ or Σ. Without loss of generality, we assume that the l_2-norm of private data $\boldsymbol{\xi}_i \in \mathbb{R}^d$ is bounded by $\sqrt{\mathrm{Tr}(\Sigma)+d\log(\frac{n_\xi}{\eta})}$, *w.p.* $1-\eta>0$.

Theorem 6. *Denote that private random* $\boldsymbol{\xi}_i \in \mathbb{R}^{d\times 1}$, $\|\boldsymbol{\xi}_i\| \leqslant \sqrt{\mathrm{Tr}(\Sigma)+d\log(\frac{n_\xi}{\eta})}$, $i=1,...,n_\xi$ *with the second-moment* $\Sigma=\mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)$ *and* $\hat{\Sigma}_\xi=\frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T$. *Considering the regularized second-moment matrix with the regularization parameter* λ

$$\hat{\Sigma}_\xi+\lambda\mathbf{I},$$

and its μ*-GDP form*

$$\hat{\Sigma}_\xi+\lambda\mathbf{I}+\mathbf{G},$$

where $\mathbf{G}\sim SG_d(\sigma^2)$, $\sigma=\frac{2(\mathrm{Tr}(\Sigma)+d\log(\frac{n_\xi}{\eta}))}{\mu\cdot n_\xi}$. *If* n_ξ *makes* $\frac{\sqrt{d^3\log(\frac{2d}{\eta})}(d^{-1}\mathrm{Tr}(\Sigma)+\log(\frac{n_\xi}{\eta}))}{\mu\cdot n_\xi\left(\lambda_{min}(\Sigma)\left(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}})\right)^2+\lambda\right)} \leqslant \frac{1}{2}$, *we have the error bound of the inverse regularized second-moment matrix,*

with at least probability $1 - O(\eta)$,

$$\|(\hat{\Sigma}_\xi+\lambda\mathbf{I}+\mathbf{G})^{-1}-(\hat{\Sigma}_\xi+\lambda\mathbf{I})^{-1}\| \leqslant \frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\frac{(d^{-1}\mathrm{Tr}(\Sigma)+\log(\frac{n_\xi}{\eta}))}{\lambda_{min}^2(\Sigma)\big(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}})\big)^4+\lambda^2},$$

where $\lambda_{\min}(\Sigma)$ *represents the smallest eigenvalue of* Σ.

We compare Theorem 5 and Theorem 6 to highlight the advantage of incorporating public moments. More details are provided in the Supplementary Material. For enough private and public sample size n_ξ, n_υ and a small regularization parameter λ, the simplified error bound in Theorem 6 is

$$\frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\|\Sigma^{-1}\|\cdot(\bar{\kappa}(\Sigma)+\|\Sigma^{-1}\|\log(\frac{n_\xi}{\eta})),$$

where $\bar{\kappa}(\Sigma) = d^{-1}\sum_{i=1}^d \kappa_i(\Sigma) \geqslant 1$, $\kappa_i(\Sigma) = \frac{\lambda_i(\Sigma)}{\lambda_{\min}(\Sigma)} \geqslant 1$ and $\lambda_i(\cdot)$ is the i-th eigenvalue. But the simplified error bound of the public-moment-guided approach in Theorem 5 is

$$\frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\log(\frac{n_\xi}{\eta})\cdot\|\Sigma^{-1}\|,$$

where the public data n_υ makes L tend to 1 and $\|\hat{\Sigma}_\upsilon^{-1}\|$ tend to $\|\Sigma^{-1}\|$.

The non-public approach depends critically on both the spectral norm $\|\Sigma^{-1}\|$ and the average condition number $\bar{\kappa}(\Sigma)$, indicating substantial degra-

dation when the underlying second-moment matrix is ill-conditioned. In contrast, the public-moment-guided approach reduces the impact of Σ and substantially improves both robustness and accuracy.

4. Application to Penalized Regression

In this section, we apply the proposed method to accommodate penalized regression. We consider the linear regression model:

$$\mathbf{y}_\xi = \boldsymbol{\Xi}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\Xi}$ is a $n \times d$ data matrix and $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1^T \ ... \ \boldsymbol{\xi}_n^T]^T$, $\boldsymbol{\xi}_i \sim subG(\Sigma) \in \mathbb{R}^{d\times 1}$ is the i-th sample. $\mathbf{y}_\xi$ is the response variable vector. The noise vector $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_{n\times n})$. For ridge regression, the general loss function is defined as

$$\mathcal{L}(\boldsymbol{\beta}; \boldsymbol{\Xi}, \mathbf{y}_\xi) = \frac{1}{2n}||\mathbf{y}_\xi - \boldsymbol{\Xi}\boldsymbol{\beta}||_2^2 + \frac{\lambda}{2}||\boldsymbol{\beta}||_2^2,$$

where λ is the regularization parameter. The analytical solution of the ridge regression is given by

$$\hat{\boldsymbol{\beta}} = (\frac{\boldsymbol{\Xi}^T\boldsymbol{\Xi}}{n} + \lambda\mathbf{I}_{d\times d})^{-1}\frac{\boldsymbol{\Xi}^T\mathbf{y}_\xi}{n}.$$

Next, we give the DP ridge regression based on PMT (DP-PMTRR, **Algorithm 2**). In theory, **Theorem 7** guarantees that the *recover* operation in the *line* 6 of **Algorithm 2** is right. **Theorem 8** provides the privacy guarantee and the estimator error bound. The content of private-data-only method, DP-RR, is provided in the Supplementary Material.

Algorithm 2 Differentially Private PMT Ridge Regression (DP-PMTRR)

1: **Input:** Private dataset $\{(\boldsymbol{\xi}_i, y_{\xi_i})^T \in \mathbb{R}^{d+1}\}_{i=1}^{n_\xi}$, public dataset $\{(\boldsymbol{\upsilon}_i, y_{\upsilon_i})^T \in \mathbb{R}^{d+1}\}_{i=1}^{n_\upsilon}$. Parameters μ, λ, d, n_ξ, n_υ and η.

2: **Transform covariates:**
$$(\{\tilde{\boldsymbol{\xi}}_i\}_{i=1}^{n_\xi}, \hat{\Sigma}_\upsilon) = PMT(\{\boldsymbol{\xi}_i\}_{i=1}^{n_\xi}, \{\boldsymbol{\Upsilon}_i\}_{i=1}^{n_\upsilon}, d, n_\xi, n_\upsilon, \eta).$$

3: **Transform responses:**
$$(\{\tilde{y}_{\boldsymbol{\xi}_i}\}_{i=1}^{n_\xi}, \hat{\sigma}_\upsilon^2) = PMT(\{y_{\boldsymbol{\xi}_i}\}_{i=1}^{n_\xi}, \{y_{\upsilon_i}\}_{i=1}^{n_\upsilon}, 1, n_\xi, n_\upsilon, \eta).$$

4: **Private parameter:**
$$\sigma_1 = \frac{2d(1+\log(\frac{2n_\xi}{\eta}))}{\mu \cdot n_\xi},\ \sigma_2 = \frac{2\sqrt{d}(1+\log(\frac{2n_\xi}{\eta}))}{\mu \cdot n_\xi}.$$

5: **Gaussian mechanism and estimator:**
$$\tilde{\boldsymbol{\beta}}^{DP} = \left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\boldsymbol{\Xi}}}{n_\xi} + \lambda\hat{\Sigma}_\upsilon^{-1} + \mathbf{G}\right)^{-1}\left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\mathbf{y}}_\xi}{n_\xi} + \mathbf{g}\right),$$
where $\mathbf{G} \sim SG_d(\sigma_1^2)$ and $\mathbf{g} \sim \mathcal{N}(0, \sigma_2^2\mathbf{I})$.

6: **Recover:** $\bar{\boldsymbol{\beta}}^{DP} \leftarrow \hat{\sigma}_\upsilon \cdot \hat{\Sigma}_\upsilon^{-1/2} \cdot \tilde{\boldsymbol{\beta}}^{DP}$.

7: **Output:** DP estimator $\bar{\boldsymbol{\beta}}^{DP}$.

Remark 2. *Note that* $\tilde{y}_{\boldsymbol{\xi}_i} = \hat{\sigma}_\upsilon^{-1} y_{\boldsymbol{\xi}_i}$ *and* $\hat{\sigma}_\upsilon = \sqrt{\frac{1}{n_\upsilon}\sum_{i=1}^{n_\upsilon} y_{\upsilon_i}^2}$ *is the root of second-moment estimation of the public response* $\mathbf{y}_\upsilon$.

Theorem 7 (Equivalent ridge estimator)**.** *The setting is same as* ***Algorithm 2****. Let the original ridge regression be* $\hat{\boldsymbol{\beta}} = \left(\frac{\boldsymbol{\Xi}^T\boldsymbol{\Xi}}{n_\xi} + \lambda\mathbf{I}\right)^{-1}\left(\frac{\boldsymbol{\Xi}^T\mathbf{y}_\xi}{n_\xi}\right)$ *and the PMT ridge regression is* $\tilde{\boldsymbol{\beta}} = \left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\boldsymbol{\Xi}}}{n_\xi} + \lambda\hat{\Sigma}_\upsilon^{-1}\right)^{-1}\left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\mathbf{y}}_\xi}{n_\xi}\right)$*, then with at least*

probability $1-O(\eta)$*, we have*

$$\hat{\boldsymbol{\beta}} = \hat{\sigma}_{\upsilon}\hat{\Sigma}_{\upsilon}^{-1/2}\tilde{\boldsymbol{\beta}}. \tag{4.4}$$

Remark 3. *The PMT ridge regression estimator* $\tilde{\boldsymbol{\beta}} = \left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\boldsymbol{\Xi}}}{n_\xi}+\lambda\hat{\Sigma}_{\upsilon}^{-1}\right)^{-1}\left(\frac{\tilde{\boldsymbol{\Xi}}^T\tilde{\mathbf{y}}_\xi}{n_\xi}\right)$ *is the minimizer of the loss function* $\mathcal{L}(\hat{\Sigma}_{\upsilon}^{-1/2}\boldsymbol{\beta};\boldsymbol{\Xi},\tilde{\mathbf{y}}_\xi) = \frac{1}{2n}||\tilde{\mathbf{y}}_\xi - \tilde{\boldsymbol{\Xi}}\boldsymbol{\beta}||_2^2 + \frac{\lambda}{2}||\hat{\Sigma}_{\upsilon}^{-1/2}\boldsymbol{\beta}||_2^2$*. Moreover, it holds regularization parameter* λ *invariant.*

Theorem 8 (DP-PMTRR)**.** ***Algorithms 2*** *satisfies* $\sqrt{2}\mu$*-GDP. When the number of public data makes* $\sqrt{n_{\upsilon}} \geqslant O(\sqrt{d}+\sqrt{\log(1/\eta)})$ *and the number of the private data* n_ξ *makes* $\frac{\sqrt{d^3\log(\frac{2d}{\eta})}(1+\log(\frac{2n_\xi}{\eta}))}{(L(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^2+\lambda\|\hat{\Sigma}_{\upsilon}\|^{-1})\cdot\mu\cdot n_\xi} \leqslant \frac{1}{2}$*, the DP estimator* $\tilde{\boldsymbol{\beta}}^{DP}$ *satisfies, with at least probability* $1-O(\eta)$*,*

$$\|\tilde{\boldsymbol{\beta}}^{DP}-\tilde{\boldsymbol{\beta}}\| \leqslant O\Big(\frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\frac{\hat{\sigma}_{\upsilon}^{-1}\|\hat{\Sigma}_{\upsilon}^{-1/2}\|\|\Sigma\|\|\boldsymbol{\beta}\|(1+\log(\frac{2n_\xi}{\eta}))}{L^2(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^4+\lambda^2\|\hat{\Sigma}_{\upsilon}\|^{-2}}\Big),$$

where $L = \frac{n_{\upsilon}}{(\sqrt{n_{\upsilon}}+O(\sqrt{d}+\sqrt{\log(\frac{1}{\eta})}))^2}$*. Moreover, based on* ***Theorem 7****, the output* $\bar{\boldsymbol{\beta}}^{DP}$ *satisfies, with at least probability* $1-O(\eta)$*,*

$$\|\bar{\boldsymbol{\beta}}^{DP}-\hat{\boldsymbol{\beta}}\| \leqslant O\Big(\frac{\sqrt{d^3\log(\frac{2d}{\eta})}}{\mu n_\xi}\cdot\frac{\|\hat{\Sigma}_{\upsilon}^{-1}\|\|\Sigma\|\|\boldsymbol{\beta}\|(1+\log(\frac{2n_\xi}{\eta}))}{L^2(1-O(\sqrt{\frac{d}{n_\xi}}+\sqrt{\frac{\log(1/\eta)}{n_\xi}}))^4+\lambda^2\|\hat{\Sigma}_{\upsilon}\|^{-2}}\Big).$$

Remark 4. *The improvements achieved by DP-PMTRR are analogous to*

those in ***Theorem 5****, where the influence of the average condition number* $\bar{\kappa}(\Sigma)$ *and the matrix norm* $\|\Sigma^{-1}\|$ *is effectively removed. Moreover, DP-PMTRR significantly reduces the sensitivity of* $\hat{\Sigma}_{\xi y}$ *from* $d\left(\bar{\kappa}(\Sigma)+\log(n_\xi)\right)\|\boldsymbol{\beta}\|$ *to* $\sqrt{d}\log(n_\xi)$*. This eliminates the dependence on the unknown* $\boldsymbol{\beta}$ *and* Σ*.*

5. Application to Generalized Linear Models

In this section, we study the application to a classical generalized linear model, logistic regression. Then, we extend to generalized linear models provided in the Supplementary Material.

5.1 Application to Logistic Regression

We consider DP logistic regression based on the PMT and the transformed data. Logistic regression is a classical generalized linear model that needs to be solved by iterative optimization. We consider logistic regression model:

$$\begin{aligned} y_i &\sim Bernoulli(p_i) \\ p_i &= \frac{1}{1+e^{-\boldsymbol{\xi}_i^T\boldsymbol{\beta}}}, \quad i=1,2,\cdots,n_\xi. \end{aligned} \tag{5.5}$$

Denote $\boldsymbol{\Xi}$ is a $n\times d$ data matrix and $\boldsymbol{\Xi}=[\boldsymbol{\xi}_1^T \ \ldots \ \boldsymbol{\xi}_n^T]^T$, $\boldsymbol{\xi}_i \sim subG(\Sigma)\in \mathbb{R}^{d\times 1}$ is the i-th sample. $\mathbf{y}_\xi \in \{0,1\}^{n_\xi}$ is the respond variable. For logistic

regression, the loss function with the regularization is defined as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta};\boldsymbol{\Xi}) &= -\frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\Big[y_i\log(p_i)+(1-y_i)\log(1-p_i)\Big]+\frac{\lambda}{2}||\boldsymbol{\beta}||_2^2 \\ &= -\frac{1}{n_\xi}\sum_{i=1}^{n_\xi} l_i(\boldsymbol{\xi}_i^T\boldsymbol{\beta})+\frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2.\end{aligned} \tag{5.6}$$

where $l_i(\boldsymbol{\xi}_i^T\boldsymbol{\beta}) = y_i\boldsymbol{\xi}_i^T\boldsymbol{\beta} - \log(1+\exp(\boldsymbol{\xi}_i^T\boldsymbol{\beta}))$ and λ is the regularization parameter.

When the data are transformed, the loss function is

$$\boldsymbol{\beta}' = \arg\min_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta};\tilde{\boldsymbol{\Xi}}) = \arg\min_{\boldsymbol{\beta}}\Big\{-\frac{1}{n_\xi}\sum_{i=1}^{n_\xi} l_i(\tilde{\boldsymbol{\xi}}_i^T\boldsymbol{\beta})+\frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2\Big\}.$$

That results in the different estimation, $\boldsymbol{\beta}' \neq \arg\min_{\boldsymbol{\beta}}\mathcal{L}(\boldsymbol{\beta};\boldsymbol{\Xi})$. We give the following theorem and redefine a new loss function so as to get the same estimation as the original one.

Theorem 9 (Equivalent estimation)**.** *Define the following loss function*

$$\tilde{\mathcal{L}}(\boldsymbol{\beta};\tilde{\boldsymbol{\Xi}}) = -\frac{1}{n_\xi}\sum_{i=1}^{n_\xi} l_i(\tilde{\boldsymbol{\xi}}_i^T\boldsymbol{\beta})+\frac{\lambda}{2}\|\hat{\Sigma}_v^{-1/2}\boldsymbol{\beta}\|_2^2, \tag{5.7}$$

where $\hat{\Sigma}_v$ is the second-moment estimation of the public data, and its min-

imizer $\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \tilde{\mathcal{L}}(\boldsymbol{\beta}; \tilde{\boldsymbol{\Xi}})$. *Then we have the equivalent optimization*

$$\hat{\Sigma}_{\upsilon}^{-1/2}\tilde{\boldsymbol{\beta}} \;=\; \hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}; \boldsymbol{\Xi}).$$

Especially, for Newton's method, we also have the equation $\hat{\Sigma}_{\upsilon}^{-1/2}\tilde{\boldsymbol{\beta}}^{(t)} = \hat{\boldsymbol{\beta}}^{(t)}$ *at each* t *iteration.*

Algorithm 3 Differentially Private PMT Logistic Regression (DP-PMTLR)

1: **Input:** Private dataset $\{(\boldsymbol{\xi}_i, y_i) \in \mathbb{R}^d \times \{0,1\}\}_{i=1}^{n_\xi}$, public dataset $\{\boldsymbol{\upsilon}_i \in \mathbb{R}^d\}_{i=1}^{n_\upsilon}$. Parameters μ, λ, d, n_ξ, n_υ and η.
2: **Transform covariates:**
$(\{\tilde{\boldsymbol{\xi}}_i\}_{i=1}^{n_\xi}, \hat{\Sigma}_\upsilon) = PMT(\{\boldsymbol{\xi}_i\}_{i=1}^{n_\xi}, \{\boldsymbol{\upsilon}_i\}_{i=1}^{n_\upsilon}, d, n_\xi, n_\upsilon, \eta)$.
3: **Private parameter:**

$$\sigma_1 = \frac{\sqrt{T}d(1+\log(\frac{2n_\xi}{\eta}))}{2\mu n_\xi},\ \sigma_2 = \frac{2\sqrt{Td(1+\log(\frac{2n_\xi}{\eta}))}}{\mu n_\xi}.$$

4: $\boldsymbol{\beta}^{(0)} = \mathbf{0}$
5: **for** $t = 0, ..., T$ **do**
6: **Gaussian mechanism and Newton update:**

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \Big(\nabla^2\tilde{\mathcal{L}}(\boldsymbol{\beta}^{(t)}; \tilde{\boldsymbol{\Xi}}) + \mathbf{G}\Big)^{-1}\Big(\nabla\tilde{\mathcal{L}}(\boldsymbol{\beta}^{(t)}; \tilde{\boldsymbol{\Xi}}) + \mathbf{g}\Big),$$

where $\mathbf{G} \sim SG_d(\sigma_1^2)$ and $\mathbf{g} \sim \mathcal{N}(0, \sigma_2^2\mathbf{I})$.
7: **end for**
8: **Recover:** $\bar{\boldsymbol{\beta}}^{DP} \leftarrow \hat{\Sigma}_\upsilon^{-1/2} \cdot \boldsymbol{\beta}^{(T)}$.
9: **Output:** DP estimator $\bar{\boldsymbol{\beta}}^{DP}$.

Next, we propose differentially private logistic regression based on PMT, **Algorithm 3**. In the logistic regression, the responses $y_i \in \{0, 1\}$ are bounded naturally and aren't transformed. In theory, **Theorem 10** guarantees the differential privacy and **Theorem 11** provides its convergence.

The content of private-data-only method, DP-LR, is provided in the Supplementary Material.

Theorem 10 (Privacy). ***Algorithm 3*** *satisfies* $\sqrt{2}\mu$*-GDP.*

Theorem 11 (DP-PMTLR). *Suppose* $\boldsymbol{\Xi}_i \in subG(\Sigma)$*, the minimizer* $\tilde{\boldsymbol{\beta}}$*,* $\boldsymbol{\beta}^{(0)} \in \mathcal{B}_r(\tilde{\boldsymbol{\beta}})$ *and* $\|\nabla\tilde{\mathcal{L}}(\boldsymbol{\beta}^{(0)};\tilde{\boldsymbol{\Xi}})\| \leqslant \min\{\gamma_L r, \frac{\gamma_L^2}{C_{\tilde{\xi}}}\}$*, where* $\gamma_L = \tau_0 L(1 - O(\sqrt{\frac{d}{n_\xi}} + \sqrt{\frac{\log(1/\eta)}{n_\xi}}))^2 + \lambda\|\hat{\Sigma}_\upsilon\|^{-1}$*. Let* $\sqrt{n_\upsilon} \geqslant O(\sqrt{d} + \sqrt{\log(1/\eta)})$*,* n_ξ *makes* $\frac{\sqrt{Td^3\log(\frac{2Td}{\eta})}(1+\log(\frac{2n_\xi}{\eta}))}{\mu\cdot n_\xi}$ *small enough and* $T = O(\log\log(n_\xi))$*. The* T^{th} *DP Newton's method satisfies* $\|\boldsymbol{\beta}^{(T)} - \tilde{\boldsymbol{\beta}}\| \leqslant O\Big(\frac{\sqrt{Td^3\log(\frac{2Td}{\eta})}(1+\log(\frac{2n_\xi}{\eta}))}{\mu\cdot n_\xi\cdot\gamma_L^2}\Big)$*, w.p.* $1-\eta$*.*

Remark 5. *Due to the Newton's iteration invariant in* ***Theorem 9****, we also conclude the convergence of the DP estimator* $\bar{\boldsymbol{\beta}}^{DP}$*.*

6. Experiments

In this section, we set up separate subsections for DP ridge regression and logistic regression evaluated in simulations and real-world datasets. For real data, DP-RR/DP-LR uses full-sample moments as oracle calibration and is reported only as a favorable utility benchmark, not as an end-to-end DP implementation. The Supplement shows more details and experiments.

6.1 DP Ridge Regression

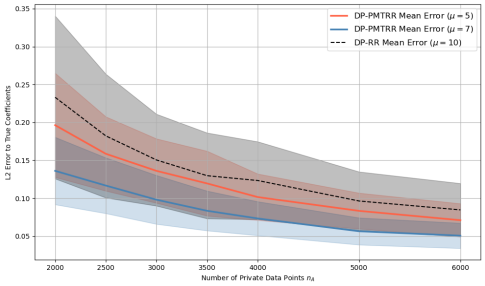
6.1.1 Simulations

In the simulation, we generate data from a linear model using a feature matrix $\mathbf{X}$ with the dimension $d = 10$ and its distribution is $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the noise $\omega \sim \mathcal{N}(0, (0.05)^2)$ and the parameter $\boldsymbol{\beta} \sim \mathcal{N}(0, \mathbf{I})$. Namely,

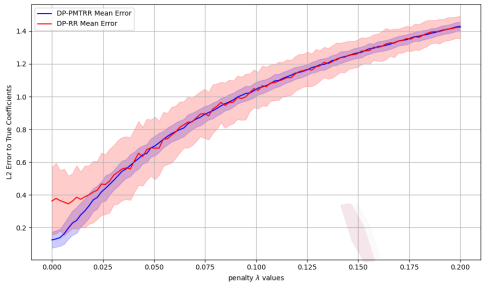
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\omega}.$$

According the linear model setting, we generate the private data $(\boldsymbol{\Xi}, \mathbf{y}_\xi) \in \mathbb{R}^{n_\xi \times (d+1)}$ and the public data $(\boldsymbol{\Upsilon}, \mathbf{y}_\upsilon) \in \mathbb{R}^{n_\upsilon \times (d+1)}$. Typically, we denote the second moment of feature data as $\Sigma = \Psi + \mu\mu^T$. We present the averaged l_2-norm errors between the true model parameters and the DP estimations, with each experiment being conducted 300 times.

Firstly, we evaluate the proposed method against the change of different privacy parameters μ and the numbers of the private data n_ξ, as shown in Figure 1(a). In the simulation, we fix the public data $n_\upsilon = 20$, the regularization parameter $\lambda = 0.01$, and the probability parameter $\eta = 0.05$. Secondly, we explore the impact of the regularization parameter, seeing Figure 1(b). We fix the private data $n_\xi = 1e4$, the public data $n_\upsilon = 100$, the privacy parameter $\mu = 3$, and the probability parameter $\eta = 0.05$.



(a) Different private data sizes and privacy parameters



(b) Different regularization parameters

Figure 1: Simulations about ridge regression.

Figure 1(a) illustrates the averaged errors (depicted as lines) and the standard deviations of errors (shown as shaded areas). Overall, the results indicate that DP-PMTRR performs better accuracy and robustness than DP-RR, even with a smaller privacy-budget, and well across different privacy parameters and private data sizes. Notably, the method exhibits clear decreasing trends in averaged errors and standard deviations as the size of private data and the privacy parameters increase. These verify our theoretical results. Figure 1(b) presents the averaged errors (shown as lines) and the standard deviations of errors (shown as shaded areas). Overall, increasing the regularization parameter λ tends to reduce the standard deviation of the errors, indicating improved robustness, but at the cost of increased averaged errors. In particular, these reveal a trade-off in DP-RR: there exists an optimal λ that minimizes the averaged error, but this choice does

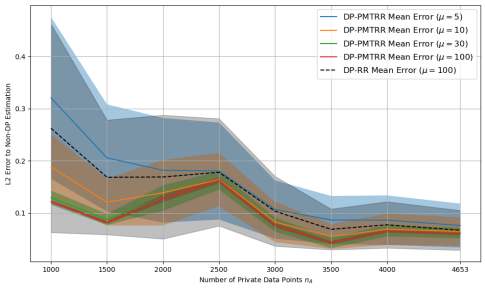
not offer significant robustness. Conversely, choosing a larger λ improves robustness but results in higher averaged error. By contrast, both the accuracy and robustness of our approach are less sensitive to the choice of λ.

6.1.2 Real-world Datasets

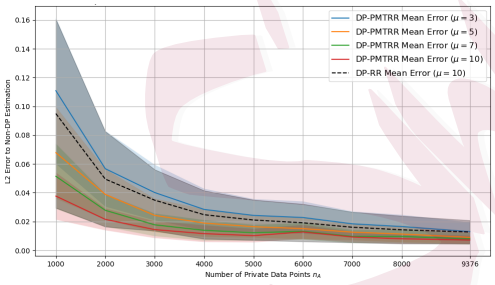
We use two real-world datasets from UCI, White-wine Quality Cortez et al. (2009) and Combined Cycle Power Plant Tfekci and Kaya (2014). The goal of the White-wine Quality dataset is to model wine quality based on physicochemical tests, which includes 4898 samples with 11 continuous physicochemical features and one integer target (quality, scored between 0 and 10) and is viewed as a linear regression task. The Combined Cycle Power Plant dataset contains 9568 samples collected from a Combined Cycle Power Plant over 6 years (2006-2011). The 4 features consist of hourly average ambient variables to predict the net hourly electrical energy output of the plant. We separate them into two parts: private dataset and public dataset. We present the averaged l_2-norm errors between the DP and the non-DP estimations, with each experiment being conducted 300 times.

We first evaluate the proposed method under varying privacy parameters μ and private sample sizes n_ξ, as shown in Figure 2(a)(b). For the white-wine and power-plant datasets, we set $(n_\upsilon, \lambda, \eta) = (245, 0, 1e-3)$

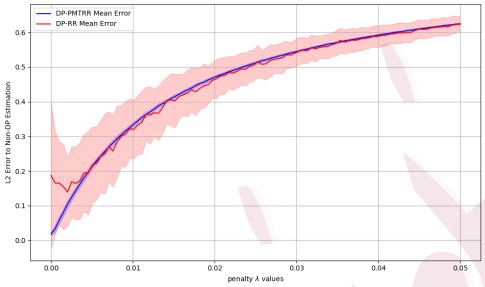
and $(192, 0, 1e-3)$, respectively. We then study the effect of regularization in Figure 2(c)(d). For this experiment, the fixed parameters are $(n_\xi, n_\upsilon, \mu, \eta) = (4649, 245, 20, 1e-3)$ for white wine and $(9376, 192, 3, 0.05)$ for power plant.



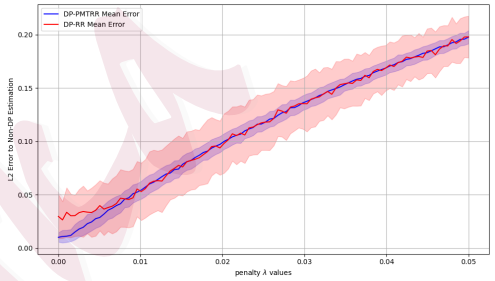
(a) White-wine Quality



(b) Combined Cycle Power Plant



(c) White-wine Quality



(d) Combined Cycle Power Plant

Figure 2: Real-world experiments with different private data sizes and privacy parameters in the top. And the bottom shows the experiments about different regularization values.

Figure 2(a)(b) illustrates the averaged errors (depicted as lines) and the standard deviations of errors (shown as shaded areas). The results in real-world datasets indicate that DP-PMTRR performs as well as the previous simulation. Figure 2(c)(d) exists with the same effect in the simulation

about the regularization. The real-world experiments verify our method in practice.

6.2 DP Logistic Regression

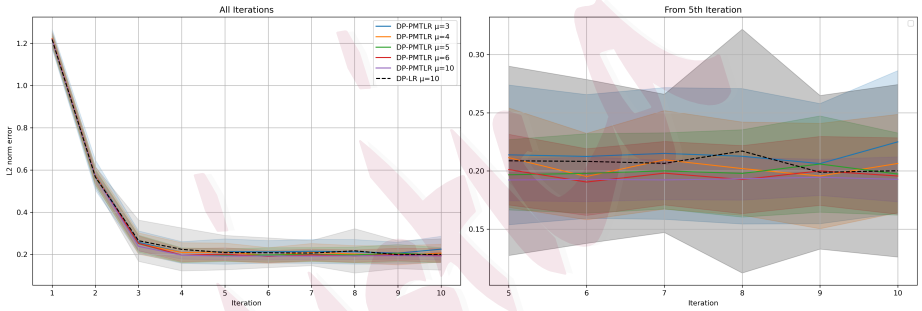
6.2.1 Simulations

We design a simulation of the logistic regression model, where the feature data $\mathbf{X}$ with the dimension $d = 8$ and the true model parameters are generated the same as the simulation in Subsection 6.1. And the responses are generated according to these settings. All experiments are repeated 100 times, and we show the averaged l_2-norm errors between the true model parameters and the DP estimations at every iteration.

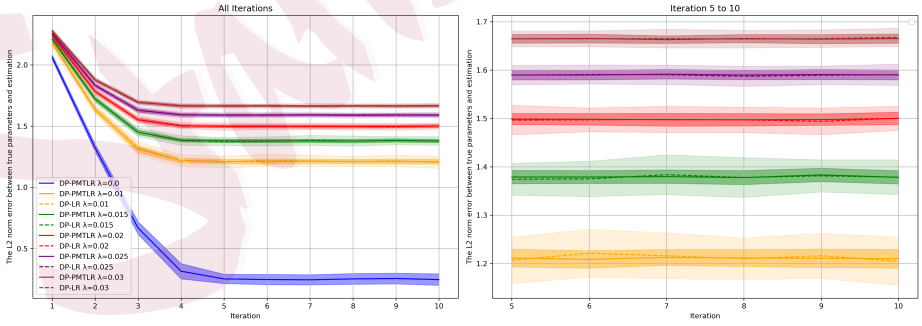
Firstly, we compare the DP-LR and DP-PMTLR (our method) with different privacy parameters μ, as shown in Figure 3(a). In the simulation, we fix the private data $n_{\upsilon} = 1e4$, the public data $n_{\upsilon} = 100$, the regularization parameter $\lambda = 1e{-}3$, and the probability parameter $\eta = 0.05$. Secondly, we explore the impact of the regularization parameter, seeing Figure 3(b). In the simulation, we fix the private data $n_{\xi} = 8000$, the public data $n_{\upsilon} = 100$, the privacy parameter $\mu = 10$, and the probability parameter $\eta = 0.05$.

Figure 3(a) displays the averaged errors (as lines) and the standard deviations of errors (as shaded areas). We typically show the final iterations

in Figure 3(a) to allow for detailed comparison. Overall, increasing the privacy parameter—corresponding to weaker privacy protection and lower noise—leads to more accurate DP estimations and more stable iterations. Notably, Figure 3(a) shows that even when using smaller privacy parameters (i.e., stronger privacy protection and higher noise), DP-PMTLR achieves lower errors and greater robustness than DP-LR. In contrast, DP-LR with the largest privacy parameter still exhibits higher error and less robustness. These results demonstrate that our method outperforms standard DP logistic regression in both utility and robustness.



(a) Different privacy parameters.



(b) Different regularization parameters.

Figure 3: Simulation comparisons of DP-PMTLR and DP-LR.

Figure 3(b) illustrates the averaged errors (shown as lines) and the standard deviations of errors (depicted as shaded areas), where the narrow and dark shaded area is the standard deviation of DP-PMTLR. The result verifies DP-PMTLR is more robust in iterations. Increasing the regularization parameter λ reduces the standard deviations, indicating improved robustness, but also results in larger averaged errors. Notably, DP-LR fails to converge when $\lambda = 0$ (original regularization parameter), while DP-PMTLR remains robust and converges successfully under the same condition. These results demonstrate that our approach effectively resolves the trade-off about the regularization and that its utility and robustness are only weakly dependent on the choice of λ.

6.2.2 Real-world Datasets

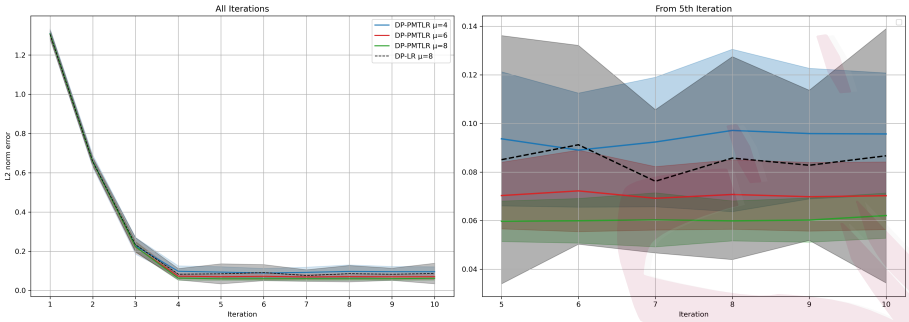
About the real-world datasets, we use two datasets from UCI, Bank Marketing (Moro et al. (2014)) and Banknote Authentication (Lohweg, Volker (2012)). The Bank Marketing dataset is related to direct marketing campaigns (phone calls) of a Portuguese banking institution, including 45211 samples and 16 features. The classification goal is to predict if the client will subscribe to a term deposit. We set aside 10% for the public dataset. All experiments are repeated 100 times, and we show the averaged l_2-norm

errors between the DP and the non-DP estimations at every iteration. The experiment about the dataset Banknote Authentication is provided in the Supplementary Material.

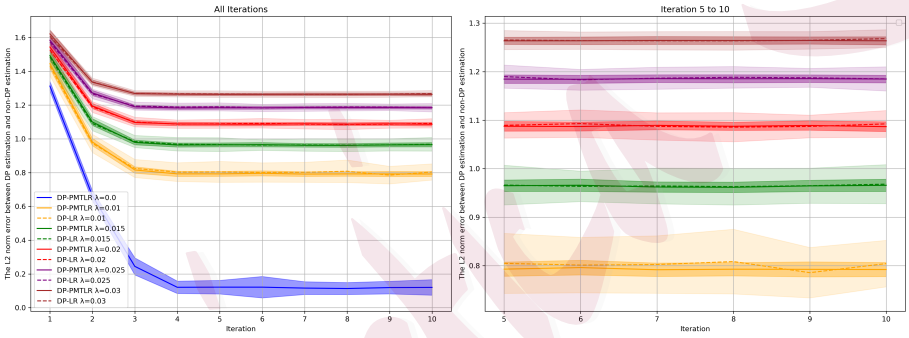
Firstly, we compare the DP-LR and DP-PMTLR (our method) with different privacy parameters μ, as shown in Figure 4(a). In the real datasets, we use all private data and public data. In the Bank Marketing dataset, we fix the regularization parameter $\lambda = 0$ and the probability parameter $\eta = 1e-3$. Secondly, we explore the impact of the regularization parameter in the real-world datasets, seeing Figure 4(b). We fix the private data, the public data, and the probability parameter $\eta = 1e - 3$. The privacy parameters is $\mu = 3$.

Figure 4(a) displays the averaged errors (as lines) and the standard deviations of errors (as shaded areas). We typically show the final iterations in Figure 4(a) to allow for detailed comparison. Overall, this result is similar to Figure 3(a), where our method also outperforms standard DP logistic regression in both utility and robustness in practice. Figure 4(b) illustrates the averaged errors (shown as lines) and the standard deviations of errors (depicted as shaded areas), where the narrow and dark shaded area is the standard deviation of DP-PMTLR. All results verify DP-PMTLR is more robust in iterations. In the real-world experiments, increasing the

regularization parameter λ also reduces the standard deviations, indicating improved robustness, but also results in larger averaged errors.



(a) Different privacy parameters.



(b) Different regularization parameters.

Figure 4: Real-world experiments of DP-PMTLR and DP-LR.

Notably, in Figure 4(b), DP-LR fails to converge when $\lambda = 0$ (original regularization parameter), while DP-PMTLR remains robust and converges successfully under the same condition. This illustrates that DP-LR must abandon the prior regularization parameter for stability; however, DP-PMTLR can make use of it and achieves a tuning-free effect. These results

indicate that our method adequately addresses the trade-off concerning regularization, and its utility and resilience exhibit minimal dependence on the selection of λ.

7. Conclusion

In this paper, we address the challenge of unbounded private data under differential privacy by leveraging the second-moment matrix estimated from public data. We propose a public-moment-guided truncation (PMT) method that transforms private data into an approximately isotropic space and applies principled truncation with a radius determined solely by non-private quantities. This transformation produces a well-conditioned second-moment matrix and significantly improves the robustness and accuracy of its DP inverse by removing dependence on the average condition number and reducing sensitivity to both the inverse norm and regularization.

Based on this framework, we develop loss functions and algorithms for ridge regression and logistic regression under differential privacy. The resulting estimators achieve improved robustness, tighter error bounds, and more stable optimization, particularly by mitigating the need for heavy regularization. These improvements are supported by rigorous theoretical analysis and extensive experiments, demonstrating that even a small

amount of public data can substantially enhance the utility of DP regression methods.

Our results highlight the value of incorporating public information to improve differentially private estimation. Future work includes extending this framework to other DP algorithms and exploring the use of additional forms of public information, such as other public statistics or pretrained models, to further enhance DP performance.

8. Supplementary Material

The supplementary material provides a comparison between PMT-based and private-data-only inverse second-moment estimation, introduces DP-RR and DP-LR baselines, and includes the key notation, complete proofs, additional theoretical discussions, and supplementary experiments.

References

Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318.

Amid, E., A. Ganesh, R. Mathews, S. Ramaswamy, S. Song, T. Steinke, V. M. Suriyakumar, O. Thakkar, and A. Thakurta (2022). Public data-assisted mirror descent for private model

training. In *International Conference on Machine Learning*, pp. 517–535. PMLR.

Avella-Medina, M., C. Bradshaw, and P.-L. Loh (2023). Differentially private inference via noisy optimization. *The Annals of Statistics 51*(5), 2067–2092.

Awan, J. and S. Vadhan (2023). Canonical noise distributions and private hypothesis tests. *The Annals of Statistics 51*(2), 547–572.

Bassily, R., S. Moran, and A. Nandi (2020). Learning from mixtures of private and public populations. *Advances in neural information processing systems 33*, 2947–2957.

Bernstein, G. and D. R. Sheldon (2019). Differentially private bayesian linear regression. *Advances in Neural Information Processing Systems 32*.

Bi, X. and X. Shen (2023). Distribution-invariant differential privacy. *Journal of econometrics 235*(2), 444–453.

Bie, A., G. Kamath, and V. Singhal (2022). Private estimation with public data. *Advances in neural information processing systems 35*, 18653–18666.

Bu, Z., J. Dong, Q. Long, and W. Su (2020a, 07). Deep learning with gaussian differential privacy. *Harvard Data Science Review 2020*.

Bu, Z., J. Dong, Q. Long, and W. J. Su (2020b). Deep learning with gaussian differential privacy. *Harvard data science review 2020*(23), 10–1162.

Cao, Z., X. Guo, and H. Zhang (2023). Privacy-preserving distributed learning via newton algorithm. *Mathematics 11*(18), 3807.

REFERENCES

Cortez, P., A. Cerdeira, T. Almeida, F.and Matos, and J. Reis (2009). Wine Quality. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C56S3T.

Dong, J., A. Roth, and W. J. Su (2022). Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 84*(1), 3–37.

Dwork, C., F. McSherry, K. Nissim, and A. D. Smith (2006). Calibrating noise to sensitivity in private data analysis. *J. Priv. Confidentiality 7*, 17–51.

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

Ferrando, C., J. Gillenwater, and A. Kulesza (2021). Combining public and private data. *arXiv preprint arXiv:2111.00115*.

Ganesh, A., M. Haghifam, T. Steinke, and A. Guha Thakurta (2023). Faster differentially private convex optimization via second-order methods. *Advances in Neural Information Processing Systems 36*, 79426–79438.

Ivkin, N., D. Rothchild, E. Ullah, I. Stoica, R. Arora, et al. (2019). Communication-efficient distributed sgd with sketching. *Advances in Neural Information Processing Systems 32*.

Ji, Z. and C. Elkan (2013). Differential privacy based on importance weighting. *Machine learning 93*, 163–183.

Kairouz, P., M. R. Diaz, K. Rush, and A. Thakurta (2021). (nearly) dimension independent private erm with adagrad rates via publicly estimated subspaces. In *Conference on Learning*

Theory, pp. 2717–2746. PMLR.

Koloskova, A., H. Hendrikx, and S. U. Stich (2023). Revisiting gradient clipping: Stochastic bias and tight convergence guarantees. In *International Conference on Machine Learning*, pp. 17343–17363. PMLR.

Liu, T., G. Vietri, T. Steinke, J. Ullman, and S. Wu (2021). Leveraging public data for practical private query release. In *International Conference on Machine Learning*, pp. 6968–6977. PMLR.

Lohweg, Volker (2012). Banknote authentication. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C55P57.

Moro, S., P. Rita, and P. Cortez (2014). Bank Marketing. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5K306.

Nandi, A. and R. Bassily (2020). Privately answering classification queries in the agnostic pac model. In *Algorithmic Learning Theory*, pp. 687–703. PMLR.

Nasr, M., J. Hayes, T. Steinke, B. Balle, F. Tramèr, M. Jagielski, N. Carlini, and A. Terzis (2023). Tight auditing of differentially private machine learning. In *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 1631–1648.

Nasr, M., S. Mahloujifar, X. Tang, P. Mittal, and A. Houmansadr (2023). Effectively using public data in privacy preserving machine learning. In *International Conference on Machine Learning*, pp. 25718–25732. PMLR.

REFERENCES

Sheffet, O. (2017). Differentially private ordinary least squares. In *International Conference on Machine Learning*, pp. 3105–3114. PMLR.

Tfekci, P. and H. Kaya (2014). Combined Cycle Power Plant. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5002N.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge university press.

Wang, P. and H. Zhang (2019). Distributed logistic regression with differential privacy. *Sci. Sin. Inform. doi 10.*

Wang, Y.-X. (2018). Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain. *arXiv preprint arXiv:1803.02596*.

Zhao, W., X. Zhu, and L. Zhu (2025). Minimax rates of convergence for sliced inverse regression with differential privacy. *Computational Statistics & Data Analysis 201*, 108041.

Zilong Cao, School of Mathematics, Northwest University, Xi'an, China

E-mail: nwu_czl@stumail.nwu.edu.cn

Xuan Bi, Department of Information and Decision Sciences, University of Minnesota, USA

E-mail: xbi@umn.edu

Hai Zhang, School of Mathematics, Northwest University, Xi'an, China

E-mail:zhanghai@nwu.edu.cn