Near-Optimal differentially private low-rank trace regression with guaranteed private initialization

Mengyue, Zha *

(March 26, 2024)

Abstract

We study differentially private (DP) estimation of a rank-r matrix $M \in \mathbb{R}^{d_1 \times d_2}$ under the trace regression model with Gaussian measurement matrices. Theoretically, the sensitivity of non-private spectral initialization is precisely characterized, and the differential-privacy-constrained minimax lower bound for estimating M under the Schatten-q norm is established. Methodologically, the paper introduces a computationally efficient algorithm for DP-initialization with a sample size of $n \geq \widetilde{O}(r^2(d_1 \vee d_2))$. Under certain regularity conditions, the DP-initialization falls within a local ball surrounding M. We also propose a differentially private algorithm for estimating M based on Riemannian optimization (DP-RGrad), which achieves a near-optimal convergence rate with the DP-initialization and sample size of $n \geq \widetilde{O}(r(d_1 + d_2))$. Finally, the paper discusses the non-trivial gap between the minimax lower bound and the upper bound of low-rank matrix estimation under the trace regression model. It is shown that the estimator given by DP-RGrad attains the optimal convergence rate in a weaker notion of differential privacy. Our powerful technique for analyzing the sensitivity of initialization requires no eigengap condition between r non-zero singular values.

1 Introduction

The trace regression model (Rohde and Tsybakov, 2011; Koltchinskii et al., 2011), as an extension of the standard regression model, has been widely applied in various fields such as ma-

*Department of Mathematics, Hong Kong University of Science and Technology, mzha@connect.ust.hk. Mengyue, Zha's research was supported by Hong Kong PhD Fellowship Scheme.

arXiv:2403.15999v1 [stat.ML] 24 Mar 2024

近似最优的差分隐私低秩迹回归，具有保证的隐私初始化

孟月，Zha *

(2024年3月26日)

摘要

我们研究了在具有高斯测量矩阵的迹回归模型下，对秩-r 矩阵 $M \in \mathbb{R}^{d_1 \times d_2}$ 的差分隐私（DP）估计。理论上，非隐私谱初始化的敏感性被精确地刻画，并在Schatten-q 范数下建立了估计 M 的差分隐私约束的最小最大下界。方法上，本文介绍了一种样本量为 $n \geq \widetilde{O}(r^2(d_1 \vee d_2))$的差分隐私初始化计算高效算法。在一定的正则条件下，差分隐私初始化落在围绕 M的局部球内。我们还提出了一种基于黎曼优化的差分隐私算法（DP-RGrad）来估计 M，该算法在差分隐私初始化和样本量 $n \geq \widetilde{O}(r(d_1 + d_2))$下实现了近似最优的收敛速度。最后，本文讨论了在迹回归模型下低秩矩阵估计的最小最大下界和上界之间的非平凡差距。结果表明，DP-RGrad给出的估计器在较弱的差分隐私概念下达到了最优的收敛速度。我们强大的技术用于分析初始化的敏感性，不需要 r 非零奇异值之间的特征值间隙条件。

1 引言

追踪回归模型（Rohde 和 Tsybakov， 2011； Koltchinskii 等人， 2011），作为标准回归模型的扩展，已被广泛应用于数学、统计学、计算机科学等各个领域。

*数学系，香港科技大学，mzha@connect.ust.hk。Zha 的研究由香港博士奖学金计划资助。

trix completion, compressed sensing, and multi-task learning (Negahban and Wainwright, 2011; Koltchinskii et al., 2011; Hamidi and Bayati, 2022). Previous studies have proposed both convex and non-convex approaches for optimal estimation procedures for the model. However, the increasing demand for privacy protection has added new complexities to this extensively studied problem. Differential privacy (DP) (Dwork et al., 2006), a framework for protecting individual privacy, has been widely adopted in industrial and governmental applications (Erlingsson et al., 2014; Apple Differential Privacy Team, 2017; Abowd et al., 2020). This paper aims to develop a near-optimal differentially private method for low-rank matrix estimation under the trace regression model.

Trace regression model Let $M \in \mathbb{R}^{d_1 \times d_2}$ be an unknown rank-r matrix and $X_i \in \mathbb{R}^{d_1 \times d_2}$ be the measurement matrix for $i = 1, \cdots, n$. Suppose the noisy observation y_i satisfies

$$y_i = \langle X_i, M \rangle + \xi_i \quad \text{for} \quad i = 1, \cdots, n, \tag{1}$$

where the model noise $\xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ and the inner product between X_i and M in Euclidean space is given by $\langle X_i, M \rangle := \mathrm{Tr}(X_i^\top M)$. The goal of the present paper is to estimate the unknown rank-r matrix $M \in \mathbb{R}^{d_1 \times d_2}$ under trace regression model defined by (1), subject to differential privacy, based on n independent observations $Z := \{(X_i, y_i)\}_{i=1}^n$.

Our approaches are built upon the Gaussian mechanism Dwork et al. (2006). The main difficulty in applying the Gaussian mechanism is sharply characterizing sensitivities of statistics whose privacy is under protection. Listed below are definitions of sensitivity, differential privacy (DP), and Gaussian mechanism. Interested readers may refer to Dwork et al. (2006, 2014); Vadhan (2017) for proofs and other details. Let $\|\cdot\|$ denotes the spectral norm and $\|\cdot\|_F$ denotes the Frobenius norm.

Sensitivity The sensitivity of a function f that maps a dataset Z into $\mathbb{R}^{d_1 \times d_2}$ is defined by $\Delta_f := \sup_{\text{neighbouring}(Z,Z')} \|f(Z) - f(Z')\|_{\mathrm{F}}$, where and the supremum is taken over all neighbouring datasets Z and Z' that differ by at most one observation.

Differential privacy Let $\varepsilon > 0$ and $0 < \delta < 1$, then we say the randomized algorithm A is (ε, δ)-differentially private if $\mathbb{P}\big(A(Z) \in \mathcal{Q}\big) \le e^\varepsilon \mathbb{P}\big(A(Z') \in \mathcal{Q}\big) + \delta$ for all neighbouring data sets Z, Z' and all subset $\mathcal{Q} \subset \mathbb{R}^{d_1 \times d_2}$.

trix 完成度, 压缩感知, 和多任务学习 (Negahban 和 Wainwright, 2011;Koltchinskii 等人, 2011; Hamidi 和 Bayati, 2022). 以往研究已经提出了凸和非凸方法用于模型的最佳估计程序。然而，对隐私保护的日益增长的需求给这个广泛研究的问题增加了新的复杂性。差分隐私 (DP) (Dwork 等人, 2006), 一个保护个人隐私的框架，已经在工业和政府应用中广泛采用 (Erlingsson 等人,2014; Apple 差分隐私团队, 2017; Abowd 等人, 2020)。本文旨在在迹回归模型下开发一种近优的差分隐私方法用于低秩矩阵估计。

迹回归模型 令 $M \in \mathbb{R}^{d_1 \times d_2}$ 为一个未知的秩-r 矩阵，$X_i \in \mathbb{R}^{d_1 \times d_2}$ 为 $i = 1 \cdots n$ 的测量矩阵, , 。假设噪声观测 y_i 满足

$$y_i = \langle X_i, M \rangle + \xi_i \quad \text{for} \quad i = 1, \cdots, n, \tag{1}$$

模型噪声在哪里 $\xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ 以及 X_i之间的内积 M在欧几里得空间中给出 $\langle X_i\ M \rangle := \mathrm{Tr}(X_i^\top M)$,。本文的目标是在 r 迹回归模型 $M \in \mathbb{R}^{d_1 \times d_2}$ 下估计未知秩- 1) 定义的 <style id='27'>矩阵 $Z := \{(X_i, y_i)\}_{i=1}^n$，基于 n <style id='29'>独立观测 $Z := \{(X_i, y_i)\}_{i=1}^n$。

我们的方法基于高斯机制 Dwork等人 (2006)。应用高斯机制的主要困难在于精确刻画受保护统计量的敏感性。下面列出了敏感性、差分隐私（DP）和高斯机制的定义。感兴趣的读者可以参考 Dwork等人 (2006， 2014)； Vadhan (2017) 了解证明和其他细节。设 $\|\cdot\|$ 表示谱范数， $\|\cdot\|_F$表示Frobenius范数。

灵敏度 函数 f 将数据集 Z 映射到 $\mathbb{R}^{d_1 \times d_2}$ 的灵敏度由 $\Delta_f := \sup_{\text{neighbouring}(Z,Z')} \|f(Z) - f(Z')\|_{\mathrm{F}}$ 定义，其中 和 是在所有相差至多一个观测值的相邻数据集 Z 和 Z' 上取的上确界。

差分隐私 设 $\varepsilon > 0$ 和 $0 < \delta < 1$，则称随机算法 A 是 (ε, δ)-差分隐私的，如果对于所有相邻数据集 Z Z' 和所有子集 $\mathcal{Q} \subset \mathbb{R}^{d_1 \times d_2}$，都有 $\mathbb{P}\big(A(Z) \in \mathcal{Q}\big) \le e^\varepsilon \mathbb{P}\big(A(Z') \in \mathcal{Q}\big) + \delta$。

Gaussian mechanism The randomized algorithm defined by $A(Z) = f(Z) + E$ is (ε, δ)-DP where $E \in \mathbb{R}^{d_1 \times d_2}$ has i.i.d. $N\left(0, 2\Delta_f^2 \varepsilon^{-2} \log(1.25/\delta)\right)$ entries.

RIP of Gaussian measurement matrices The sensitivity of any statistic involving $\{X_i\}_{i\in[n]}$ depends on the properties of the measurement matrices. Besides, it has been previously established since Candes and Tao (2005) that the restricted isometry property (RIP) on measurement matrices is crucial to the recovery of the unknown matrix M. Hence, assumptions on $\{X_i\}_{i\in[n]}$ are necessary and the present paper considers $\{X_i\}_{i\in[n]}$ with Gaussian design.

Assumption 1 (Gaussian design)**.** *The vectorization of measurement matrices $X_1, \ldots, X_n$ are independent Gaussian* $\mathrm{vec}(X_i) \sim \mathcal{N}(\mathbf{0}, \Lambda_i)$ *where Λ_i 's are known, symmetric and positive definite. There exist absolute constants $C_l, C_u > 0$ such that $C_l \le \lambda_{\min}(\Lambda_i) \le \lambda_{\max}(\Lambda_i) \le C_u$.*

The following Lemma 1 shows that under Assumption 1, the measurement matrices $\{X_i\}_{i=1}^n$ satisfy the restricted isometry property (RIP) with high probability, see the proof in E.1.

Lemma 1. *Under the Assumption 1, for any $B \in \mathbb{R}^{d_1 \times d_2}$ of rank r, there exist constants $c_1, c_2, c_3 > 0$ and $c_5 > c_4 > 0$ such that if $n \ge c_1 r(d_1 + d_2)$, with probability at least $1 - c_2 \exp(-c_3 r(d_1 + d_2))$, we have $c_4\sqrt{C_u C_l}\|B\|_{\mathrm{F}}^2 \le \frac{1}{n}\sum_{i=1}^n \langle X_i, B\rangle^2 \le c_5\sqrt{C_u C_l}\|B\|_{\mathrm{F}}^2$.*

Notations Suppose M is of rank-r and its singular value decomposition is of the form $M = U\Sigma V^\top \in \mathbb{R}^{d_1 \times d_2}$ where $U \in \mathbb{O}_{d_1,r}$, $V \in \mathbb{O}_{d_2,r}$ and $\Sigma = \mathrm{diag}\{\sigma_1 \cdots \sigma_r\}$ with $\sigma_1 \ge \cdots \ge \sigma_r$. Here, $\mathbb{O}_{d,r}$ denotes the set of $d \times r$ matrices satisfying $H^\top H = I_r$. Let $\kappa := \sigma_1/\sigma_r$ be the condition number and $\kappa_\xi := \sigma_\xi/\sigma_r$ be the signal-to-noise ratio. Let $\widetilde{O}$ stand for the typical big-O notation up to logarithmic factors and $\widetilde{O}_p(\cdot)$ stand for $\widetilde{O}$ holds with high probability.

1.1 Main results

The paper presents several key results related to differentially private low-rank matrix estimation. Firstly, we propose a private initialization $\widetilde{M}_0$ (as detailed in Algorithm 1). Secondly, we establish the privacy-constrained minimax lower bound under the general Shatten-q norm (as detailed in Theorem 2). Finally, we introduce a private estimator $\widetilde{M}_{l^*}$ (as detailed in Algorithm 2) that achieves the near-optimal convergence rate under the Frobenius norm. The sensitivity analysis of $\widetilde{M}_0$ heavily relies on a spectral representation formula for asymmetric matrices (See Lemma 2).

高斯机制 由 $A(Z) = f(Z) + E$ 定义的高斯机制是 (ε, δ)-DP，其中 $E \in \mathbb{R}^{d_1 \times d_2}$ 具有独立同分布 $N\left(0\ 2\Delta_f^2 \varepsilon^{-2} \log(1.25/\delta)\right)$，条目。

高斯测量矩阵的 RIP任何涉及 $\{X_i\}_{i\in[n]}$的统计量的敏感性取决于测量矩阵的性质。此外，由于 Candes 和 Tao (2005) 已经证明，测量矩阵上的限制等距性质 (RIP) 对未知矩阵的恢复至关重要 M。因此，关于 $\{X_i\}_{i\in[n]}$的假设是必要的，本文考虑 $\{X_i\}_{i\in[n]}$ 具有高斯设计。

假设 1（高斯设计）。测量矩阵 $X_1 \ \ldots \ X_n$, , 的向量化是独立的正态向量$\mathrm{c}(X_i) \sim \mathcal{N}(\mathbf{0}, \Lambda_i)$, 其中 Λ_i 是已知的、对称的和正定的。存在绝对常数 $C_l, C_u > 0$ 使得
$C_l \le \lambda_{\min}(\Lambda_i) \le \lambda_{\max}(\Lambda_i) \le C_u$.

以下引理 1表明在假设 1下，测量矩阵 $\{X_i\}_{i=1}^n$以高概率满足限制等距性质 (RIP)，参见 E.1中的证明。

引理 1. 在假设 *1*下，对于任何 $B \in \mathbb{R}^{d_1 \times d_2}$ 秩为 r，存在常数$c_1, c_2, c_3 > 0$和 $c_5 > c_4 > 0$，使得如果 $n \ge c_1 r(d_1 + d_2)$，以至少 $1 - c_2 \exp(-c_3 r(d_1 + d_2))$的概率，我们有
$c_4\sqrt{C_u C_l}\|B\|_{\mathrm{F}}^2 \sum_n^1 \quad {}_{i=1}^n \langle X_i\ B\rangle^2 \le c_5\sqrt{C_u C_l}\|B\|_{\mathrm{F}}^2, \quad .$

符号说明 假设 M 秩为-r ，其奇异值分解形式为 $M = U\Sigma V^\top \in \mathbb{R}^{d_1 \times d_2}$，其中 $U \in \mathbb{O}_{d_1,r}$, $V \in \mathbb{O}_{d_2,r}$和 $\Sigma = \mathrm{diag}\{\sigma_1 \cdots \sigma_r\}$，且 $\sigma_1 \ge \cdots \ge \sigma_r$。这里,$\mathbb{O}_{d\,r}$,表示满足 $H^\top H = I_r$的 $d \times r$ 矩阵的集合。令 $\kappa := \sigma_1/\sigma_r$ 为条件数， $\kappa_\xi := \sigma_\xi/\sigma_r$ 为信噪比。令 $\widetilde{O}$ 表示对数因子下的典型大O记号， $\widetilde{O}_p(\cdot)$ 表示 $\widetilde{O}$ 以高概率成立。

1.1 主要结果

本文提出了与差分隐私低秩矩阵估计相关的几个关键结果。首先，我们提出了一种隐私初始化 $\widetilde{M}_0$ (如算法1所述)。其次，我们在一般的Shatten$q$ 范数下建立了隐私约束的最小最大下界（如定理2所述）。最后，我们引入了一种隐私估计器 $\widetilde{M}_{l^*}$ (如算法2所述)，它在Frobenius范数下实现了近似最优的收敛速度。对 $\widetilde{M}_0$ 的敏感性分析严重依赖于非对称矩阵的谱表示公式（参见引理2）。

We prove in Corollary 1 that the private initialization $\widetilde{M}_0$ satisfies $\|\widetilde{M}_0 - M\|_F \leq \sqrt{2r}\|\widetilde{M}_0 - M\| \leq c_0\sigma_r$, with high probability (w.h.p.), for a small constant $0 < c_0 < 1$, provided that $n \geq \widetilde{O}\left((\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2)\right)$.

Theorem 2 establishes the DP-constrained minimax risk of estimating the rank-r matrix $M \in \mathbb{R}^{d_1 \times d_2}$ under model (1) and general Schatten-q norm. Specifically, the minimax risk under Frobenius norm is in the order of $\sigma_\xi\sqrt{\frac{r(d_1 \vee d_2)}{n}} + \sigma_\xi\frac{r(d_1 \vee d_2)}{n\varepsilon}$.

Finally, we show in Theorem 3 that with a sample size of $n \geq \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$ and any initialization satisfying (2), Algorithm 2 achieves geometric convergence rate. The private estimator $\widetilde{M}_{l^*}$ attains the near-optimal convergence rate

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq \widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + (\sigma_\xi + \sigma_r)\frac{r(d_1+d_2)}{n\varepsilon}\right).$$

1.2 Motivations and related works

The trace regression model has been extensively researched, resulting in well-known optimal procedures and theoretical properties. Both convex (Rohde and Tsybakov, 2011; Koltchinskii et al., 2011; Candes and Plan, 2011; Negahban and Wainwright, 2011) and non-convex methods (Burer and Montei 2003; Chen and Wainwright, 2015; Zheng and Lafferty, 2016; Wei et al., 2016) have achieved the optimal convergence rate of the order $\sigma_\xi\sqrt{\frac{r(d_1 \vee d_2)}{n}}$ without the constraint from differential privacy. However, the DP-constrained minimax rate of low-rank matrix estimation under the trace regression model is still unknown. (Near) Optimal DP-algorithms have been developed for statistical problems such as learning Gaussians Kamath et al. (2019); Kuditipudi et al. (2023); Brown et al. (2023) or heavy-tailed distributions Kamath et al. (2020), (sparse or generalized) linear regression Wang (2018); Cai et al. (2021, 2023), and PCA Blum et al. (2005); Dwork et al. (2014); Chaudhuri et al. (2012); Liu et al. (2022). Previous works on DP-regression Cai et al. (2021, 2023) assume that all measurements have bounded ℓ_2 norm. This assumption presents a significant limitation to studying the role of measurements play in the estimation error. Additionally, by treating measurements as a fixed vector or matrix, the statistical properties of measurements are disregarded. As a result, the opportunity for optimal statistical analysis subject to privacy concerns is inevitably lost. Recently, (McSherry and Mironov, 2009; Liu et al., 2015; Jain et al., 2018; Chien et al., 2021; Wang et al., 2023) propose gradient-descent-based algorithms for DP low-rank matrix completion. These algorithms have attained near-optimal sample complexity. However, the problem of sample-efficient, differentially private initialization remains under-explored. Additionally, it is unknown how to establish the minimax risk of low-

我们在推论 1 中证明了私有初始化 $\widetilde{M}_0$ 满足 $\|\widetilde{M}_0 - M\|_F \leq \sqrt{2r}\|\widetilde{M}_0 - M\| \leq c_0\sigma_r$, 以很高的概率 (w.h.p.)，对于一个小常数 $0 < c_0 < 1$，只要 $n \geq \widetilde{O}\ (\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2))$。

定理 2 建立了估计秩-r 矩阵 $M \in \mathbb{R}^{d_1 \times d_2}$ 在模型 (1) 和一般 Schatten-q 范数下的 DP-约束极小极大风险。具体来说，Frobenius 范数下的极小极大风险是 $\sigma_\xi\sqrt{\frac{r(d_1 \vee d_2)}{n}} + \sigma_\xi\frac{r(d_1 \vee d_2)}{n\varepsilon}$。

最后，我们在定理 3 中展示了样本大小为 $n \geq \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$ and any initialization satisfying (2), Algorithm 2 achieves geometric convergence rate. The private estimator $\widetilde{M}_{l^*}$ attains the near-optimal convergence rate

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq \widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + (\sigma_\xi + \sigma_r)\frac{r(d_1+d_2)}{n\varepsilon}\right).$$

1.2 Motivations andrelated works

迹回归模型已被广泛研究，形成了成熟的优化程序和理论性质。凸（Rohde和Tsybakov, 2011；Koltchinskii等人，2011；Candes和Plan，2011；Negahban和Wainwright, 2011）和非凸方法（Burer 2003；Chen和Wainwright，2015；Zheng和Lafferty，2016；Wei等人，2016）均在不满足差分隐私约束的情况下实现了 and Montei 最优收敛速度为 $\sigma_\xi\sqrt{\frac{r(d_1 \vee d_2)}{n}}$。然而，在迹回归模型下低秩矩阵估计的差分隐私约束下的最小最大率仍未知。（近似）最优差分隐私算法已被开发用于学习高斯分布 Kamath等人（2019）；Kuditipudi等人（2023）；Brown等人（2023）或重尾分布 Kamath等人（2020），（稀疏或广义）线性回归 Wang（2018）；Cai等人（2021，2023），以及PCA Blum等人（2005）；Dwork等人（2014）；Chaudhuri等人（2012）；Liu等人（2022）。先前关于差分回归 Cai等人（2021，2023）的工作假设所有测量值具有有界的 ℓ_2 范数。这一假设显著限制了研究测量值在估计误差中作用的可能性。此外，将测量值视为固定向量或矩阵，忽略了测量值的统计特性。结果，在隐私考虑下的最优统计分析机会不可避免地被丢失。最近，（McSherry和 Mironov，2009；Liu等人，2015；Jain等人，2018；Chien等人，2021；Wang等人，2023）提出了基于梯度下降的差分低秩矩阵补全算法。这些算法实现了近似最优的样本复杂度。然而，样本高效、差分隐私的初始化问题仍待深入探索。此外，如何建立低秩矩阵估计的最小最大风险尚不清楚。

rank matrix estimation with the constraints of differential privacy, especially when the matrix is asymmetric.

1.3 Organization

Section 2 proposes a DP-initialization algorithm and presents its privacy and utility guarantees. In Section 3, we establish a DP-constrained minimax lower bound (5) for estimating the rank-r matrix M under the trace regression model. Section 4 presents the DP-estimator based on non-convex optimization and derives the upper bound of the DP-estimator's error, as stated in (7). We discuss the score attack argument and the non-trivial gap between the upper bound of (7) and the DP-constrained minimax lower bound (5) in Section 5. Proofs are given in Appendix A to F.

2 DP-initialization

Section 2.1 presents an (ε, δ)-DP initialization $\widetilde{M}_0$, as stated in Algorithm 1. In Section 2.2, we introduce a spectral representation formula (See Lemma 2) that is crucial to sensitivity analysis on the initialization. With the help of the spectral representation formula, the privacy and utility guarantees of the DP-initialization $\widetilde{M}_0$ are given in Section 2.3.

2.1 Algorithm for DP-initialization

We begin with $\widehat{L} = n^{-1}\sum_{i=1}^{n} \operatorname{mat}\left(\Lambda_i^{-1}\operatorname{vec}(X_i)\right) y_i$, which is unbiased for M. Suppose that the leading-r left and right singular vectors of $\widehat{L}$ is given by the columns of $\widehat{U} \in \mathbb{O}_{d_1,r}$ and $\widehat{V} \in \mathbb{O}_{d_2,r}$, respectively. Then, $\widehat{M}_0 := \mathrm{SVD}_r(\widehat{L})$ is a non-private estimator for M. Let $\widehat{\Sigma} := \widehat{U}^\top \widehat{L} \widehat{V}$, then we have

$$\widehat{M}_0 = \mathrm{SVD}_r(\widehat{L}) = \widehat{U}\widehat{U}^\top \widehat{L}\widehat{V}\widehat{V}^\top = \widehat{U}\widehat{\Sigma}\widehat{V}^\top.$$

It is reasonable to think about privatizing $\widehat{U}$, $\widehat{V}$, and $\widehat{\Sigma}$, separately. We first privatize the empirical spectral projector $\widehat{U}\widehat{U}^\top$ and $\widehat{V}\widehat{V}^\top$ by Gaussian mechanism. Thanks to post-processing property Dwork et al. (2006), we obtain $\widetilde{U} \in \mathbb{O}^{d_1,r}$ and $\widetilde{V} \in \mathbb{O}^{d_2,r}$ whose columns are differentially private and orthogonal. Secondly, we privatize the $r \times r$ matrix $\widetilde{U}^\top \widehat{L} \widetilde{V}$ by Gaussian mechanism and obtain $\widetilde{\Sigma} \in \mathbb{R}^{r\times r}$ which is a private surrogate for $\widehat{\Sigma} = \widehat{U}^\top \widehat{L} \widehat{V}$. Finally, we take $\widetilde{M}_0 = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$ as the DP-initialization. We display the pseudo-code of the proposed

基于差分隐私约束的秩矩阵估计，尤其是在矩阵是非对称的情况下。

1.3 组织

第 2 节提出了一种DP初始化算法，并介绍了其隐私和效用保证。在 第 3节，我们建立了一个用于在迹回归模型下估计秩-r矩阵 M 的DP约束 minimax 下界（5）。第 4 节介绍了基于非凸优化的DP估计器，并推导出DP估计器误差的上界，如（第 7节所述）。我们在 第 5节讨论了评分攻击论点以及（第 7）的上界与DP约束 minimax 下界（第 5）之间的非平凡差距。证明在附录A 至F中给出。

2 DP初始化

第 2.1 节介绍了一种（ε, δ）-DP初始化 $\widetilde{M}_0$，如算法 1所述。在 第 2.2节，我们引入了一个谱表示公式（参见引理 2），该公式对于初始化的敏感性分析至关重要。借助谱表示公式，DP初始化 $\widetilde{M}_0$ 的隐私和效用保证在 第 2.3节给出。

2.1 DP初始化算法

我们从 $\widehat{L} = n^{-1}\sum_{i=1}^{n} \operatorname{mat}\left(\Lambda_i^{-1}\operatorname{vec}(X_i)\right) y_i$ 开始，它在 M 上是无偏的。假设 $\widehat{L}$ 的前导-r 左右奇异向量分别由 $\widehat{U} \in \mathbb{O}_{d_1\, r}$ 和 $\widehat{V} \in \mathbb{O}_{d_2\, r}$ 的列给出。那么，$\widehat{M}_0 := \mathrm{SVD}_r(\widehat{L})$ 是 M 的一个非私有估计量。令 $\widehat{\Sigma} := \widehat{U}^\top \widehat{L} \widehat{V}$，则我们有

$$\widehat{M}_0 = \mathrm{SVD}_r(\widehat{L}) = \widehat{U}\widehat{U}^\top \widehat{L}\widehat{V}\widehat{V}^\top = \widehat{U}\widehat{\Sigma}\widehat{V}^\top.$$

分别考虑私有化 $\widehat{U}$、$\widehat{V}$ 和 $\widehat{\Sigma}$ 是合理的。我们首先通过高斯机制私有化经验谱投影器 $\widehat{U}\widehat{U}^\top$ 和 $\widehat{V}\widehat{V}^\top$。得益于后处理性质 Dworket al. (2006)，我们得到 $\widetilde{U} \in \mathbb{O}^{d_1\, r}$ 和 $\widetilde{V} \in \mathbb{O}^{d_2\, r}$，它们的列是差分私有的且正交的。其次，我们通过高斯机制私有化 $r \times r$ 矩阵 $\widetilde{U}^\top \widehat{L} \widetilde{V}$ 并得到 $\widetilde{\Sigma} \in \mathbb{R}^{r\times r}$，它是一个 $\widehat{\Sigma} = \widehat{U}^\top \widehat{L} \widehat{V}$ 的私有替代。最后，我们取 $\widetilde{M}_0 = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$ 作为 DP初始化。我们展示了所提出算法的伪代码

DP-initialization in Algorithm 1. The privacy of $\widetilde{M}_0$ is guaranteed by the composition property Dwork et al. (2006).

Algorithm 1 Differentially private initialization for trace regression

Input: the data set $\{(X_i, y_i)\}_{i=1}^n$; the covariance matrices $\{\Lambda_i\}_{i=1}^n$; sensitivity $\Delta^{(1)}$, $\Delta^{(2)} > 0$; rank r; nuisance variance σ_ξ^2; privacy budget $\varepsilon > 0$ and $0 < \delta < 1$.

Output: (ε, δ)-differentially private initialization $\widetilde{M}_0$.

Compute the unbiased sample estimator $\widehat{L}$ and its top-r left and right singular vectors:

$$\widehat{L} \leftarrow n^{-1} \sum_{i=1}^{n} \operatorname{mat}\left(\Lambda_i^{-1} \operatorname{vec}(X_i)\right) y_i \quad \text{and} \quad \widehat{M}_0 = \widehat{U}\widehat{\Sigma}\widehat{V} \leftarrow \operatorname{SVD}_r(\widehat{L})$$

Compute $(\varepsilon/3, \delta/3)$-differentially private singular subspaces by adding artificial Gaussian noise:

$$\widetilde{U} \leftarrow \operatorname{SVD}_r\left(\widehat{U}\widehat{U}^\top + E_U\right) \text{with } (E_U)_{ij} = (E_U)_{ji} \overset{\text{i.i.d.}}{\sim} N\left(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le d_1$$

$$\widetilde{V} \leftarrow \operatorname{SVD}_r\left(\widehat{V}\widehat{V}^\top + E_V\right) \text{with } (E_V)_{ij} = (E_U)_{ji} \overset{\text{i.i.d.}}{\sim} N\left(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le d_2$$

Compute $(\varepsilon/3, \delta/3)$-differentially private estimates of singular values up to rotations:

$$\widetilde{\Sigma} \leftarrow \widetilde{U}^\top \widehat{L} \widetilde{V} + E_\Sigma \text{ with } (E_\Sigma)_{ij} = (E_\Sigma)_{ji} \overset{i.i.d.}{\sim} N\left(0, \frac{18\Delta^{(2)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le r$$

Compute (ε, δ)-differentially private initialization: $\widetilde{M}_0 \leftarrow \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$

Return: $\widetilde{M}_0$

To this end, we define the sensitivities of $\Delta^{(1)}$ and $\Delta^{(2)}$ appear in Algorithm 1. Let

$$\widehat{L}^{(i)} := n^{-1} \sum_{j \ne i}^{n} \operatorname{mat}\left(\Lambda_j^{-1} \operatorname{vec}(X_j)\right) Y_j + n^{-1} \operatorname{mat}\left(\Lambda_i^{-1} \operatorname{vec}(X_i')\right) y_i',$$

where (X_i', y_i') is an i.i.d. copy of (X_i, y_i). Then, the estimator $\widehat{L}^{(i)}$ differs with $\widehat{L}$ only by the i-th pair of observations. Suppose the top-r left and right singular vectors of $\widehat{L}^{(i)}$ are given by $U^{(i)}$ and $V^{(i)\top}$, respectively. The sensitivity of $\widehat{U}\widehat{U}^\top$ is defined by

$$\Delta_{\widehat{U}\widehat{U}^\top} = \sup_{\text{neighbouring}(Z,Z')} \left\| \widehat{U}(Z)\widehat{U}(Z)^\top - \widehat{U}(Z')\widehat{U}(Z')^\top \right\|_F = \max_{i \in [n]} \left\| \widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top} \right\|_F,$$

and the sensitivity $\Delta_{\widehat{V}\widehat{V}^\top}$ of $\widehat{V}\widehat{V}^\top$ is defined similarly. We refer to $\Delta^{(1)} \triangleq \Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$ as the sensitivity of singular subspaces and define the sensitivity

$$\Delta^{(2)} \triangleq \Delta_{\widetilde{U}^\top \widehat{L} \widetilde{V}} = \sup_{\text{neighbouring}(Z,Z')} \left\| \widetilde{U}\widehat{L}(Z)\widetilde{V}^\top - \widetilde{U}\widehat{L}(Z')\widetilde{V}^\top \right\|_F = \max_{i \in [n]} \left\| \widetilde{U}^\top \left(\widehat{L} - \widehat{L}^{(i)}\right) \widetilde{V} \right\|_F.$$

算法中的DP初始化1。隐私性 $\widetilde{M}_0$ 由组合属性保证Dwork等人(2006).

Algorithm 1 Differentially private initialization for trace regression

Input: the data set $\{(X_i, y_i)\}_{i=1}^n$; the covariance matrices $\{\Lambda_i\}_{i=1}^n$; sensitivity $\Delta^{(1)}$, $\Delta^{(2)} > 0$; rank r; nuisance variance σ_ξ^2; privacy budget $\varepsilon > 0$ and $0 < \delta < 1$.

Output: (ε, δ)-differentially private initialization $\widetilde{M}_0$.

Compute the unbiased sample estimator $\widehat{L}$ and its top-r left and right singular vectors:

$$\widehat{L} \leftarrow n^{-1} \sum_{i=1}^{n} \operatorname{mat}\left(\Lambda_i^{-1} \operatorname{vec}(X_i)\right) y_i \quad \text{and} \quad \widehat{M}_0 = \widehat{U}\widehat{\Sigma}\widehat{V} \leftarrow \operatorname{SVD}_r(\widehat{L})$$

Compute $(\varepsilon/3, \delta/3)$-differentially private singular subspaces by adding artificial Gaussian noise:

$$\widetilde{U} \leftarrow \operatorname{SVD}_r\left(\widehat{U}\widehat{U}^\top + E_U\right) \text{with } (E_U)_{ij} = (E_U)_{ji} \overset{\text{i.i.d.}}{\sim} N\left(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le d_1$$

$$\widetilde{V} \leftarrow \operatorname{SVD}_r\left(\widehat{V}\widehat{V}^\top + E_V\right) \text{with } (E_V)_{ij} = (E_U)_{ji} \overset{\text{i.i.d.}}{\sim} N\left(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le d_2$$

Compute $(\varepsilon/3, \delta/3)$-differentially private estimates of singular values up to rotations:

$$\widetilde{\Sigma} \leftarrow \widetilde{U}^\top \widehat{L} \widetilde{V} + E_\Sigma \text{ with } (E_\Sigma)_{ij} = (E_\Sigma)_{ji} \overset{i.i.d.}{\sim} N\left(0, \frac{18\Delta^{(2)^2}}{\varepsilon^2}\log(3/\delta)\right), \forall 1 \le i \le j \le r$$

Compute (ε, δ)-differentially private initialization: $\widetilde{M}_0 \leftarrow \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$

Return: $\widetilde{M}_0$

为此，我们定义了 $\Delta^{(1)}$ 和 $\Delta^{(2)}$ 在算法1中出现的敏感性

$$\widehat{L}^{(i)} := n^{-1} \sum_{j \ne i}^{n} \operatorname{mat}\left(\Lambda_j^{-1} \operatorname{vec}(X_j)\right) Y_j + n^{-1} \operatorname{mat}\left(\Lambda_i^{-1} \operatorname{vec}(X_i')\right) y_i',$$

其中 (X_i', y_i') 是 (X_i, y_i) 的独立同分布副本。然后，估计器 $\widehat{L}^{(i)}$ 与 $\widehat{L}$ 只在 i-th 对观测值上不同。假设 $\widehat{L}^{(i)}$ 的顶部-r 左右奇异向量分别由 $U^{(i)}$ 和 $V^{(i)\top}$ 给出。$\widehat{U}\widehat{U}^\top$ 的敏感性定义为

$$\Delta_{\widehat{U}\widehat{U}^\top} = \sup_{\text{neighbouring}(Z,Z')} \left\| \widehat{U}(Z)\widehat{U}(Z)^\top - \widehat{U}(Z')\widehat{U}(Z')^\top \right\|_F = \max_{i \in [n]} \left\| \widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top} \right\|_F,$$

并且 $\widehat{V}\widehat{V}^\top$ 的敏感性 $\Delta_{\widehat{V}\widehat{V}^\top}$ 类似地定义。我们称 $\Delta^{(1)}$ $\Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$ 为奇异子空间的敏感性，并定义敏感性

$$\Delta^{(2)} \triangleq \Delta_{\widetilde{U}^\top \widehat{L} \widetilde{V}} = \sup_{\text{neighbouring}(Z,Z')} \left\| \widetilde{U}\widehat{L}(Z)\widetilde{V}^\top - \widetilde{U}\widehat{L}(Z')\widetilde{V}^\top \right\|_F = \max_{i \in [n]} \left\| \widetilde{U}^\top \left(\widehat{L} - \widehat{L}^{(i)}\right) \widetilde{V} \right\|_F.$$

As privatizing $\widehat{\Sigma} = \widehat{U}^\top \widehat{L} \widehat{V}$ by Gaussian mechanism, the scale of artificial noise avoids growing with an unnecessary $\sqrt{d_1} \vee \sqrt{d_2}$ but rather growing with a smaller quantity $\sqrt{r}$. This benefit motivates us to privatize $\widehat{U}$, $\widehat{V}$ and $\widehat{\Sigma}$, separately. However, it is technically challenging to characterize $\Delta^{(1)} = \Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$ due to the non-linear dependence of $\widehat{U}\widehat{U}^\top$ and $\widehat{V}\widehat{V}^\top$ on the dataset $Z = \{(X_i, y_i)\}_{i=1}^n$. To address this challenge, we introduce an explicit spectral representation formula (See Lemma 2) to obtain a sharp upper bound on the sensitivity of the singular subspaces.

2.2 Spetral representation formula

This section introduces a spectral representation formula for asymmetric matrices (See Lemma 2). To begin with, we quickly explain the standard *symmetric dilation* trick (See e.g., Section 2.1.17 in Tropp et al. (2015)) and define auxiliary operators used in Lemma 2.

Symmetric dilation and auxiliary operators For any $M \in \mathbb{R}^{d_1 \times d_2}$, the symmetric dialation M_* of M is a $(d_1 + d_2) \times (d_1 + d_2)$ matrix defined by $M_* := \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}$. It is easy to check that $M_* = M_*^\top$ and $\|M_*\| = \|M\|$. Further, if we assume that M is of rank r and has the form of SVD $M = U\Sigma V^\top \in \mathbb{R}^{d_1 \times d_2}$, then M_* is of rank-$2r$ and has eigendecomposition of the form

$$\frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \cdot \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix}^\top := U_{M^*} \Sigma_{M^*} U_{M^*}^\top.$$

The $2r$ eigenvectors of M_* is given by the columns of $U_{M^*} \in \mathbb{O}_{(d_1+d_2),2r}$. For integer $t \geq 1$, we define operators

$$Q^{-t} := U_{M^*} \Sigma_{M^*}^{-t} U_{M^*}^\top \quad \text{and} \quad Q^{-0} := Q^\perp \triangleq (U_{M^*})_\perp (U_{M^*})_\perp^\top = I_{d_1+d_2} - U_{M^*} U_{M^*}^\top.$$

Lemma 2 (Spectral representation formula)**.** *Let $M \in \mathbb{R}^{d_1 \times d_2}$ be any rank-r matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\widehat{L} = M + \Delta \in \mathbb{R}^{d_1 \times d_2}$ be a perturbation of M where $\Delta \in \mathbb{R}^{d_1 \times d_2}$ is the deviation matrix. Suppose the top-r left and right singular vectors of $\widehat{L}$ and M, are given by the columns of $\widehat{U}$, $\widehat{V}$ and U, V, respectively. Suppose that $2\|\Delta\| \leq \sigma_r$, then*

$$\begin{pmatrix} \widehat{U}\widehat{U}^\top - UU^\top & 0 \\ 0 & \widehat{V}\widehat{V}^\top - VV^\top \end{pmatrix} = \sum_{k \geq 1} \mathcal{S}_{M_*,k}(\Delta_*).$$

作为私有化 $\widehat{\Sigma} = \widehat{U}^\top \widehat{L} \widehat{V}$通过高斯机制，人工噪声的规模避免随着不必要地 $\sqrt{d_1} \vee \sqrt{d_2}$但而是随着更小的量 $\sqrt{r}$.这种优势激励我们进行私有化 $\widehat{U}$, $\widehat{V}$ 和 $\widehat{\Sigma}$，分别。然而，由于 $\Delta^{(1)} = \Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$对数据集 $\widehat{U}\widehat{U}^\top$和 $\widehat{V}\widehat{V}^\top$的非线性依赖，在技术上难以刻画 $Z = \{(X_i, y_i)\}_{i=1}^n$。为了应对这一挑战，我们引入了一个显式的谱表示公式（参见引理 2）来获得奇异子空间的敏感性的严格上界。

2.2 谱表示公式

本节介绍不对称矩阵的谱表示公式（参见引理2）。首先，我们简要解释标准的 对称扩展 技巧（例如，参见 Tropp 等人 (2015）中的第 2.1.17 节）并定义引理 2中使用的辅助算子。

对称膨胀和辅助算子 对于任何 $M \in \mathbb{R}^{d_1 \times d_2}$ $(d_1 + d_2) \times (d_1 + d_2)$ 矩阵，由 $M_* := \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}$， M 的对称膨胀 M_* 是一个定义。容易验证 $M_* = M_*^\top$ 和 $\|M_*\| = \|M\|$。此外，如果我们假设 M 的秩为 r 且具有 SVD 形式 $M = U\Sigma V^\top \in \mathbb{R}^{d_1 \times d_2}$，那么 M_* 的秩为 $2r$ 且具有如下特征分解

$$\frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \cdot \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix}^\top := U_{M^*} \Sigma_{M^*} U_{M^*}^\top.$$

The $2r$ 特征向量 of M_* is given by the columns of $U_{M^*} \in \mathbb{O}_{(d_1+d_2)\ 2r,\circ}$ 对于整数 $t \geq 1$，我们定义算子

$$Q^{-t} := U_{M^*} \Sigma_{M^*}^{-t} U_{M^*}^\top \quad \text{and} \quad Q^{-0} := Q^\perp \triangleq (U_{M^*})_\perp (U_{M^*})_\perp^\top = I_{d_1+d_2} - U_{M^*} U_{M^*}^\top.$$

引理 2（谱表示公式）。设 $M \in \mathbb{R}^{d_1 \times d_2}$ 是任何秩-r 矩阵，具有奇异值 $\sigma_1 \geq \cdots \geq \sigma_r > 0$，$\widehat{L} = M + \Delta \in \mathbb{R}^{d_1 \times d_2}$ 是 M 的扰动，其中 $\Delta \in \mathbb{R}^{d_1 \times d_2}$ 是偏差矩阵。假设 $\widehat{L}$ 和 M 的前 r 个左和右奇异向量分别由 $\widehat{U}$、 $\widehat{V}$ 和 U、 V 的列给出。假设 $2\|\Delta\| \leq \sigma_r$，则

$$\begin{pmatrix} \widehat{U}\widehat{U}^\top - UU^\top & 0 \\ 0 & \widehat{V}\widehat{V}^\top - VV^\top \end{pmatrix} = \sum_{k \geq 1} \mathcal{S}_{M_*,k}(\Delta_*).$$

Here, Δ_* is the symmetric dilation of $\Delta := \widehat{L} - M$ and the k-th order term $\mathcal{S}_{M_*,k}(\Delta_*)$ is a summation of $\binom{2k}{k}$ terms defined by $\mathcal{S}_{M_*,k}(\Delta_*) = \sum_{\mathbf{s}:s_1+\ldots+s_{k+1}=k}(-1)^{1+\tau(\mathbf{s})} \cdot Q^{-s_1}\Delta_* Q^{-s_2}\ldots\Delta_* Q^{-s_{k+1}}$, where $\mathbf{s} = (s_1,\ldots,s_{k+1})$ contains non-negative indices and $\tau(\mathbf{s}) = \sum_{j=1}^{k+1}\mathbb{I}(s_j > 0)$.

In Lemma 2, the spectral projectors $\widehat{U}\widehat{U}^\top$ and $\widehat{V}\widehat{V}^\top$ of the matrix $\widehat{L} = M + \Delta$, is explicitly represented in terms of the symmetric dilation of Δ, with the help of auxiliary operators Q^{-0} and Q^{-t} for integer $t \geq 1$. The proof of Lemma 2 is deferred to Appendix E.2. Note that Lemma 2 accommodates a diverging condition number and requires no eigengap condition between r non-zero singular values. In the proof of Theorem 1, we shall see that $\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}$ and $\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}$ are mainly contributed by the 1-st order approximation $\mathcal{S}_{M_*,1}(\Delta_*) - \mathcal{S}_{M_*,1}(\Delta_*^{(i)})$ where $\Delta_*^{(i)}$ is the symmetric dilation of $\Delta^{(i)} := \widehat{L}^{(i)} - M$.

2.3 Privacy and utility guarantees of the initialization

In this section, we study the privacy and utility guarantees of the initialization $\widetilde{M}_0$. Theorem 1 characterizes the sensitivities $\Delta^{(1)}$ and $\Delta^{(2)}$ needed to guarantee an (ε,δ)-DP $\widetilde{M}_0$, and present the upper bounds of $\|\widetilde{M}_0 - M\|$ and $\|\widetilde{M}_0 - M\|_F$. The proof of Theorem 1 is provided in Appendix A.

Theorem 1 (Privacy and utility guarantees of the initialization $\widetilde{M}_0$)**.** *Consider i.i.d. observations $Z = \{z_1,\cdots,z_n\}$ drawn from the trace regression model stated in (1) where $z_i := (X_i, y_i)$ for $i = 1,\cdots,n$. Let the true rank-r regression coefficients matrix be $M \in \mathbb{R}^{d_1\times d_2}$. Suppose that $\{X_i\}_{i\in[n]}$ satisfy the Assumption 1. Under the mild condition $n \geq \frac{\log^2 n}{(d_1\vee d_2)\log(d_1+d_2)}$, there exists absolute constants $C_1, C_2, C_3 > 0$ such that*

$$n \geq n_0 \triangleq C_1 C_l^{-1} r \left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2 (d_1 \vee d_2)\log(d_1+d_2);$$

if Algorithm 1 takes in sensitivities at least $\Delta^{(1)} = C_2\left(\frac{\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)}{\sigma_r}\right)\frac{\sqrt{r}}{n}\log n$ and $\Delta^{(2)} = C_2\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)\frac{\sqrt{r}}{n}\log n$, then Algorithm 1 is guaranteed to be (ε,δ)-DP. Moreover, the

这里，Δ_*是$\Delta := \widehat{L} - M$的对称膨胀，以及第 k阶项 $\mathcal{S}_{M_*\,k}(\Delta_*)$，是定义为 $\mathcal{S}_{M_*}$的 $\binom{2k}{k}$ 项的总和，$_k(\Delta_*)$ $= \sum_{\mathbf{s}:s_1+\ldots+s_{k+1}=k}(-1)^{1+\tau(\mathbf{s})} \cdot Q^{-s_1}\Delta_* Q^{-s_2}\ldots\Delta_* Q^{-s_{k+1}}$，其中 $\mathbf{s} = (s_1,\ldots,s_{k+1})$ 包含非负索引和 $\tau(\mathbf{s}) = \sum_{j=1}^{k+1}\mathbb{I}(s_j > 0)$。

在引理2中，矩阵 $\widehat{L} = M + \Delta$的谱投影器 $\widehat{U}\widehat{U}^\top$ 和 $\widehat{V}\widehat{V}^\top$，以对称膨胀 Δ的形式明确表示，并借助用于整数 $t \geq 1$的辅助算子 Q^{-0}和 Q^{-t}。引理2 的证明被推迟到附录E.2。请注意，引理2 适用于发散的条件数，并且不需要 r非零奇异值之间的特征值间隙。在定理1的证明中，我们将看到 $\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}$ 和$\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}$主要由1阶近似 $\mathcal{S}_{M_*\,1}(\Delta_*) - \mathcal{S}_{M_*\,1}(\Delta_*^{(i)})$，，贡献，其中 $\Delta_*^{(i)}$是 $\Delta^{(i)} := \widehat{L}^{(i)} - M$的对称膨胀。

2.3 初始化的隐私和效用保证

在本节中，我们研究初始化的隐私和效用保证 $\widetilde{M}_0$。定理 1刻画了保证一个（ε,δ）-DP $\widetilde{M}_0$所需的敏感度 $\Delta^{(1)}$ 和 $\Delta^{(2)}$，并给出了 $\|\widetilde{M}_0 - M\|$和 $\|\widetilde{M}_0 - M\|_F$的上界。定理 1的证明在附录 A中给出。

定理 1 (初始化的隐私和效用保证 $\widetilde{M}_0$)。 考虑从 (1) 中独立同分布抽取的观测值 $Z = \{z_1,\cdots,z_n\}$，其中 (1)，，。令真实的秩-r 回归系数矩阵为 $M \in \mathbb{R}^{d_1\times d_2}$。假设 $\{X_i\}_{i\in[n]}$满足假设 1。在温和的条件 $n \geq \frac{\log^2 n}{(d_1\vee d_2)\log(d_1+d_2)}$下，存在绝对常数 C_1 C_2 $C_3 > 0$，，使得

$$n \geq n_0 \triangleq C_1 C_l^{-1} r \left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2 (d_1 \vee d_2)\log(d_1+d_2);$$

如果 Algorithm 1至少输入灵敏度 $\Delta^{(1)} = C_2\left(\frac{\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)}{\sigma_r}\right)\frac{\sqrt{r}}{n}\log n$ and $\Delta^{(2)} = C_2\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)\frac{\sqrt{r}}{n}\log n$，那么 Algorithm 1保证是 (ε,δ)-DP。此外，

output $\widetilde{M}_0$ *of Algorithm 1 satisfies*

$$\begin{aligned}
&\|\widetilde{M}_0 - M\| \bigvee \left(\|\widetilde{M}_0 - M\|_F/\sqrt{2r}\right) \\
&\le \underbrace{C_3\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\frac{\sigma_1}{\sigma_r}\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}}_{e_1} \\
&\quad + \underbrace{C_3\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\left(\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon} + \frac{r}{n\varepsilon}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})}_{e_2},
\end{aligned}$$

with probability at least $1 - (d_1+d_2)^{-10} - n^{-9} - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$.

In Theorem 1, the sample size condition $n \ge n_0$ ensures that the spectral norm of perturbations is small enough, i.e., $\|\Delta\| + \max_{i\in[n]}\|\Delta^{(i)}\| \le \sigma_r/2$, to apply Lemma 2 and obtain a sharp characterization on $\Delta^{(1)} \triangleq \Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$. Theorem 1 also provides an upper bound on the sensitivity of *pseudo singular values*, which is of the order $\Delta^{(2)} \triangleq \Delta_{\widetilde{U}^\top\widehat{L}\widetilde{V}} \asymp \sigma_1\Delta^{(1)}$. Based on these results, Algorithm 1 outputs an (ε,δ)-DP initialization $\widetilde{M}_0$ under the sample size condition

$$n \ge \widetilde{O}\left((\kappa^2 r^2 + \kappa_\xi r)(d_1 \vee d_2)\right),$$

with an upper bound on the error $\|\widetilde{M}_0 - M\|$ consisting of two terms. The first term e_1 accounts for the statistical error of $\widehat{M}_0$ and is greater than the the optimal rate $\sigma_\xi\sqrt{\frac{d_1\vee d_2}{n}}$ (Koltchinskii, 2011). The second term e_2 can be further decomposed into the cost of privacy on the singular subspaces which is of the order $\widetilde{O}_p\left(\frac{\sigma_1}{\sigma_r}(\sigma_1\sqrt{r}+\sigma_\xi)\frac{\sqrt{(d_1\vee d_2)}}{n\varepsilon}\right)$, and the cost of privacy arises from privatizing the singular values by Gaussian mechanism which is of the order $\widetilde{O}_p\left((\sigma_1\sqrt{r}+\sigma_\xi)r/(n\varepsilon)\right)$.

Next, Corollary 1 gives the sample size required by a DP-initialization $\widetilde{M}_0$ that falls within a local ball of M. The proof of Corollary 1 is deferred to Appendix A.5.

Corollary 1. *Under the conditions stated in Theorem 1, as the sample size is sufficiently large*

$$\begin{aligned}
n \ge C_1 \max\Bigg\{ &\underbrace{C_l^{-1}\left(\frac{\sqrt{C_u r}\sigma_1 + \sigma_\xi}{\sigma_r}\right)^2 \kappa^2 r(d_1 \vee d_2)\log(d_1+d_2)}_{n_1}, \\
&\underbrace{\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\sigma_1 + \sigma_\xi}{\sigma_r}\right)\left(\kappa r\sqrt{d_1 \vee d_2} + r^{\frac{3}{2}}\right)\log n\frac{\log^{\frac{1}{2}}(\frac{3.75}{\delta})}{\varepsilon}}_{n_2}\Bigg\},
\end{aligned}$$

for some absolute constant $c_2 > 0$*, then we have, for some small constant* $0 < c_0 < 1$*,*

$$\|\widetilde{M}_0 - M\|_F \le \sqrt{2r}\|\widetilde{M}_0 - M\| \le c_0\sigma_r. \tag{2}$$

输出 $\widetilde{M}_0$ 算法 1 满足

$$\begin{aligned}
&\|\widetilde{M}_0 - M\| \bigvee \left(\|\widetilde{M}_0 - M\|_F/\sqrt{2r}\right) \\
&\le \underbrace{C_3\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\underbrace{\frac{\sigma_1}{\sigma_r}}_{e_1}\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}}_{} \\
&\quad + \underbrace{C_3\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\left(\underbrace{\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon}}_{e_2} + \frac{r}{n\varepsilon}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})}_{},
\end{aligned}$$

以至少 $1 - (d_1+d_2)^{-10} - n^{-9} - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$. 的概率

在定理 1中，样本量条件 $n \ge n_0$ 确保扰动谱范数足够小，即 $\|\Delta\| + \max_{i\in[n]}\|\Delta^{(i)}\| \le \sigma_r/2$， 以应用引理 2 并获得关于 $\Delta^{(1)}$ $\Delta_{\widehat{U}\widehat{U}^\top} \vee \Delta_{\widehat{V}\widehat{V}^\top}$的精确刻画。定理 1还提供了伪奇异值的灵敏度上界，该上界为 伪奇异值的阶 $\Delta^{(2)}$ $\Delta_{\widetilde{U}^\top\widehat{L}\widetilde{V}} \asymp \sigma_1\Delta^{(1)}$≜。基于这些结果，算法 1 在样本量条件下输出一个（ε,δ）-DP初始化 $\widetilde{M}_0$。

$$n \ge \widetilde{O}\left((\kappa^2 r^2 + \kappa_\xi r)(d_1 \vee d_2)\right),$$

存在误差上界 $\|\widetilde{M}_0 - M\|$由两项组成。第一项 e_1用于统计误差 $\widehat{M}_0$且大于最优率 $\sigma_\xi\sqrt{\frac{d_1\vee d_2}{n}}$ (Koltchinskii, 2011). 第二项 e_2 可以进一步分解为奇异子空间的隐私成本，其数量级为 $\widetilde{O}_p\left(\frac{\sigma_1}{\sigma_r}(\sigma_1\sqrt{r}+\sigma_\xi)\sqrt{\frac{(d_1\vee d_2)}{n\varepsilon}}\right)$，以及来自高斯机制对奇异值进行私有化的隐私成本，其数量级为 $\widetilde{O}_p\left((\sigma_1\sqrt{r}+\sigma_\xi)r/(n\varepsilon)\right)$。

接下来，引理 1 给出了DP初始化所需的样本量 $\widetilde{M}_0$ 该样本量位于半径为 M的局部球内。引理 1 的证明推迟到附录 A.5。

Corollary 1. *Under the conditions stated in Theorem 1, as the sample size is sufficiently large*

$$\begin{aligned}
n \ge C_1 \max\Bigg\{ &\underbrace{C_l^{-1}\left(\frac{\sqrt{C_u r}\sigma_1 + \sigma_\xi}{\sigma_r}\right)^2}_{n_1} \kappa^2 r(d_1 \vee d_2)\log(d_1+d_2), \\
&\underbrace{\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\sigma_1 + \sigma_\xi}{\sigma_r}\right)\left(\kappa r\underbrace{\sqrt{d_1 \vee d_2}}_{n_2} + r^{\frac{3}{2}}\right)\log n\frac{\log^{\frac{1}{2}}(\frac{3.75}{\delta})}{\varepsilon}}_{}\Bigg\},
\end{aligned}$$

对于某个绝对常数 $c_2 > 0$，然后对于某个小常数 $0 < c_0 < 1$，

$$\|\widetilde{M}_0 - M\|_F \le \sqrt{2r}\|\widetilde{M}_0 - M\| \le c_0\sigma_r. \tag{2}$$

In Corollary 1, the error due to $\|M\| \cdot \left(\left\|\widehat{V}\widehat{V}^\top - VV^\top\right\| + \left\|\widehat{U}\widehat{U}^\top - UU^\top\right\|\right)$ dominates the statistical error $\|\widehat{L} - M\|$ and the sample size n_1 is required to control these two terms; the sample size n_2 controls the error due to privatizing the singular subspaces and singular values. According to Corollary 1, as the sample size $n \geq \widetilde{O}\left((\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2)\right)$, the (ε, δ)-DP $\widetilde{M}_0$ is guaranteed to fall into a local ball surrounding M, as stated in (2). The condition (2) is a pre-requisite for Algorithm 2 to converge geometrically, as discussed in Theorem 3.

3 Minimax lower bounds

This section applies DP-Fano's lemma (See Lemma 3) to establish the DP-constrained minimax lower bound of estimating the matrix $M \in \mathbb{M}_r := \{M \in \mathbb{R}^{d_1 \times d_2} : \mathrm{rank}(M) = r\}$ under trace regression model

$$f_M(y_i|X_i) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(\frac{-\left(y_i - \langle X_i, M\rangle\right)^2}{2\sigma^2}\right); X_i \sim \mathcal{N}(\mathbf{0}, \Lambda_i). \tag{3}$$

Suppose we observe an i.i.d. sample $\{(X_i, y_i), (X_i', y_i')\}_{i\in[n]}$ of size $2n$ drawn from (3). Then, we have

$$\bar{y}_i := y_i + y_i' = \left\langle \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}, \begin{pmatrix} 0 & X_i \\ X_i'^\top & 0 \end{pmatrix} \right\rangle + \xi_i + \xi_i',$$

where the underlying matrix M_*. Let $f(X_i, X_i')$ be the joint distribution of X_i and X_i'; $f_{M_*}(\bar{y}_i \mid X_i, X_i')$ be the conditional distribution of $\bar{y}_i$ given X_i, X_i'; and denote the joint distribution of $\bar{y}_i$ and X_i, X_i' as $f_{M_*}(\bar{y}_i, X_i, X_i')$. It is clear that $f_{M_*}(\bar{y}_i \mid X_i, X_i')$ is given by the distribution of

$$\mathcal{N}(\left\langle \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}, \begin{pmatrix} 0 & X_i \\ X_i'^\top & 0 \end{pmatrix} \right\rangle, 2\sigma_\xi^2).$$

Let $\otimes$ represent the tensor product of marginal laws. For a given matrix $\Sigma = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}$ where $C\sigma \geq \sigma_1 \cdots \geq \sigma_r \geq c\sigma$ for some constants $\sigma > 0$ and $C > c > 0$, we consider the family of normal distribution under trace regression model:

$$\mathcal{P}_\Sigma := \left\{ \bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') : M_* = (U\Sigma V^\top)_*, U \in \mathbb{O}_{d_1,r}, V \in \mathbb{O}_{d_2,r} \right\}.$$

By definition, each distribution $P_{M_*} \in \mathcal{P}_\Sigma$ is indexed by $U_{M_*} = \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \in \mathbb{O}_{d_1+d_2,2r}$ whose columns are the r eigenvectors of M_*. Next, we employ DP-Fano's lemma to derive the

在推论 1中，由于 $\|M\| \cdot \left(\left\|\widehat{V}\widehat{V}^\top - VV^\top\right\| + \left\|\widehat{U}\widehat{U}^\top - UU^\top\right\|\right)$ 主导了统计误差 $\|\widehat{L} - M\|$ 和样本量 n_1 需要控制这两个项；样本量 n_2 控制了由于私有化奇异子空间和奇异值而产生的误差。根据推论 1，当样本量 $n \geq \widetilde{O}\left((\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2)\right)$，， (ε, δ)-DP $\widetilde{M}_0$保证会落入围绕 M的局部球内，如（2）所述。条件（2）是算法 2 几何收敛的先决条件，如定理 3所述。

3 Minimax lowerbounds

本节应用 DP-Fano 引理（参见引理 3）来建立估计矩阵 $M \in \mathbb{M}_r := \{M \in \mathbb{R}^{d_1 \times d_2} : \mathrm{rank}(M) = r\}$ 在迹回归模型下的 DP-约束 minimax 下界

$$f_M(y_i|X_i) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(\frac{-\left(y_i - \langle X_i, M\rangle\right)^2}{2\sigma^2}\right); X_i \sim \mathcal{N}(\mathbf{0}, \Lambda_i). \tag{3}$$

假设我们观察到来自 (3) 的 i.i.d. 样本 $\{(X_i, y_i), (X_i', y_i')\}_{i\in[n]}$ 大小为 $2n$。那么，我们有

$$\bar{y}_i := y_i + y_i' = \left\langle \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}, \begin{pmatrix} 0 & X_i \\ X_i'^\top & 0 \end{pmatrix} \right\rangle + \xi_i + \xi_i',$$

其中 M_* 是底层矩阵。令 $f(X_i, X_i')$ 是 X_i 和 X_i'; $f_{M_*}(\bar{y}_i \mid$ 的联合分布，X_i $X_i')$ 是给定 X_i 的 $\bar{y}_i$ 的条件分布， X_i'; 并记 $\bar{y}_i$ 和 X_i， X_i' 的联合分布为 $f_{M_*}(\bar{y}_i, X_i, X_i')$。显然，$f_{M_*}(\bar{y}_i \mid X_i, X_i')$ 由分布给出

$$\mathcal{N}(\left\langle \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}, \begin{pmatrix} 0 & X_i \\ X_i'^\top & 0 \end{pmatrix} \right\rangle, 2\sigma_\xi^2).$$

令 $\otimes$ 表示边缘定律的张量积。对于给定的矩阵 $\Sigma = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}$，其中 $C\sigma \geq \sigma_1 \cdots \geq \sigma_r \geq c\sigma$ 对于某些常数 $\sigma > 0$ 和 $C > c > 0$，我们考虑迹回归模型下的正态分布族：

$$\mathcal{P}_\Sigma := \left\{ \bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') : M_* = (U\Sigma V^\top)_*, U \in \mathbb{O}_{d_1,r}, V \in \mathbb{O}_{d_2,r} \right\}.$$

根据定义，每个分布 $P_{M_*} \in \mathcal{P}_\Sigma$ 都由 $U_{M_*} = \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \in \mathbb{O}_{d_1+d_2\ 2r}$, 索引，其列是 M_* 的 r 特征向量。接下来，我们采用 DP-Fano 引理来推导

minimax lower bound of estimating M_* by a sample drawn from $\mathcal{P}_\Sigma$. Let $\mathrm{KL}(\cdot\|\cdot)$ and $\mathrm{TV}(\cdot,\cdot)$ denote the Kullback-Leibler divergence and total variation distance between two distributions.

Lemma 3 (DP-Fano's lemma, Acharya et al. (2021)). *Let $\mathcal{P} := \{P : P = \mu^{(1)} \times \cdots \times \mu^{(n)}\}$ be a family of product measures indexed by a parameter from a pseudo-metric space (Θ, ρ). Denote $\theta(P) \in \Theta$ the parameter associated with the distribution P. Let $\mathcal{Q} = \{P_1, \cdots, P_N\} \subset \mathcal{P}$ contain N probability measures and there exist constants $\rho_0, l_0, t_0 > 0$ such that for all $i \neq i' \in [N]$, $\rho\left(\theta(P_i), \theta(P_{i'})\right) \geqslant \rho_0, \quad \mathrm{KL}\left(P_i \| P_{i'}\right) \leq l_0, \quad \sum_{k\in[n]} \mathrm{TV}\left(\mu_i^{(k)}, \mu_{i'}^{(k)}\right) \leq t_0$, where $P_i = \mu_i^{(1)} \times \cdots \times \mu_i^{(n)}$ and $P_{i'} = \mu_{i'}^{(1)} \times \cdots \times \mu_{i'}^{(n)}$. Suppose $\delta \lesssim e^{-n}$, then*

$$\inf_{A\in\mathcal{A}_{\varepsilon,\delta}(\mathcal{P})} \sup_{P\in\mathcal{P}} \mathbb{E}_A\, \rho(A, \theta(P)) \geqslant \max\left\{\frac{\rho_0}{2}\left(1 - \frac{l_0 + \log 2}{\log N}\right), \frac{\rho_0}{4}\left(1 \bigwedge \frac{N-1}{\exp\left(4\varepsilon t_0\right)}\right)\right\}, \tag{4}$$

where the infimum is taken over all the (ε, δ)-DP randomized algorithm defined by $\mathcal{A}_{\varepsilon,\delta}(\mathcal{P}) := \{A : Z \mapsto \Theta$ and A is (ε, δ)-differentially private for all $Z \sim P \in \mathcal{P}\}$.

To apply Fano's lemma, we need to construct a large subset with well-separated elements for $\mathbb{O}_{d_1+d_2,2r}$. By Lemma 6, there exists a subset $\mathcal{S}_q^{(d_1+d_2)} \subset \mathbb{O}_{d_1+d_2,2r}$ with cardinality $\left|\mathcal{S}_q^{(d_1+d_2)}\right| \geq 2^{2r(d_1+d_2-2r)}$ such that for any $H \neq H' \in \mathcal{S}_q^{(d_1+d_2)}$,

$$\|HH^\top - H'H'^\top\|_q \gtrsim \tau\varepsilon_0(2r)^{1/q} \quad \text{and} \quad \|HH^\top - H'H'^\top\|_{\mathrm{F}} \lesssim 2\sqrt{r}\varepsilon_0,$$

for some small constants $\tau, \varepsilon_0 > 0$, where $\|\cdot\|_q$ denotes the Schatten-q norm. We then consider the family of distributions

$$\mathcal{P}_\sigma = \left\{\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') : M_* = \sigma HH^\top, H \in \mathcal{S}_q^{(d_1+d_2)}\right\} \subset \mathcal{P}_\Sigma,$$

whose cardinality $N := |\mathcal{P}_\sigma| \geq 2^{2r(d_1+d_2-2r)}$. Let $M_* = \sigma HH^\top$ and $M_*' = \sigma H'H'^\top$. As shown in Appendix B, for any $H \neq H' \in \mathcal{S}_q^{(d_1+d_2)}$, we have

$$\mathrm{KL}\left(\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') \| \bigotimes_{i=1}^n f_{M_*'}(\bar{y}_i, X_i, X_i')\right) \lesssim \frac{n}{\sigma_\xi^2} C_u \sigma^2 r \varepsilon_0^2,$$

and $\sum_{k\in[n]} \mathrm{TV}\left(f_{M_*}(\bar{y}_i, X_i, X_i'), f_{M_*'}(\bar{y}_i, X_i, X_i')\right) \lesssim n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0$. To this end, we obtain Theorem 2 by applying Lemma 3 with the bounded KL divergence and TV distance, together with the facts that $\mathcal{P}_\sigma \subset \mathcal{P}_\Sigma$. The proof of Theorem 2 is deferred to Appendix B.

Theorem 2. *Consider a sample of size n drawn from the distribution $P_{M^*} \in \mathcal{P}_\Sigma$, then for any $\delta \lesssim e^{-n}$ and any $q \in [1, \infty]$, there exists a constant $c > 0$*

$$\inf_{\widetilde{M}} \sup_{P\in\mathcal{P}_\Sigma} \mathbb{E}\left\|\widetilde{M} - M\right\|_q \geq c\frac{\sigma_\xi}{\sqrt{C_u}}\left(r^{1/q}\sqrt{\frac{d_1 \vee d_2}{n}} + r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1 \vee d_2}{n\varepsilon}\right) \bigwedge r^{1/q}\sigma,$$

M_* 的minimax下界，通过从 $\mathcal{P}_\Sigma$中抽取的样本进行估计。令$\mathrm{KL}(\cdot\|\cdot)$和$\mathrm{TV}(\cdot,\cdot)$表示两个分布之间的Kullback-Leibler散度及总变差距离。

引理3（DP-Fano引理，Acharya等人（2021））。令 $\mathcal{P} := \{P : P = \mu^{(1)} \times \cdots \times \mu^{(n)}\}$是由伪度量空间$(\Theta, \rho)$中参数索引的产品测度族。记$\theta(P) \in \Theta$为与分布 P相关的参数。令 $\mathcal{Q} = \{P_1, \cdots, P_N\} \subset \mathcal{P}$包含$N$概率测度，并且存在常数 $\rho_0, l_0, t_0 > 0$，使得对于所有 $i = i' \in [N]$,

$\rho\ \theta\ P_i\ \ \theta\ P_{i'} \quad \geqslant \rho_0 \qquad P_i \| P_{i'}\ \ \leq l_0 \qquad \left(\mu_i^{(k)}, \mu_{i'}^{(k)}\right) \leq t_0 \qquad P_i = \not{\mu}_i^{(1)} \times \cdots \times$，

满足(( )) ( )) , KL ( ) , $\sum_{k\in[n]} \mathrm{TV}$，其中$\mu_i^{(n)}$和 $P_{i'} = \mu_{i'}^{(1)} \times \cdots \times \mu_{i'}^{(n)}$。假设 $\delta \lesssim e^{-n}$，则

$$\inf_{A\in\mathcal{A}_{\varepsilon,\delta}(\mathcal{P})} \sup_{P\in\mathcal{P}} \mathbb{E}_A\, \rho(A, \theta(P)) \geqslant \max\left\{\frac{\rho_0}{2}\left(1 - \frac{l_0 + \log 2}{\log N}\right), \frac{\rho_0}{4}\left(1 \bigwedge \frac{N-1}{\exp\left(4\varepsilon t_0\right)}\right)\right\}, \tag{4}$$

在所有定义的 (ε, δ)-DP 随机算法上取下确界，其中 $\{A : Z \to \Theta$ 和 A 对所有 $Z \sim P \in \mathcal{P}\}$ 是 (ε, δ)-差分隐私的。.$\{A : Z \to \Theta$ 和 A 是 (ε, δ)-differentially private for all $Z \sim P \in \mathcal{P}\}$.

要应用 Fano 定理，我们需要为 $\mathbb{O}_{d_1+d_2,2r}$ 构造一个元素间隔良好的大子集。根据引理 6，存在一个基数 $\left|\mathcal{S}_q^{(d_1+d_2)}\right| \geq 2^{2r(d_1+d_2-2r)}$ 的子集 $\mathcal{S}_q^{(d_1+d_2)} \subset \mathbb{O}_{d_1+d_2\ 2r}$,使得对于任何 $H = H' \in \mathcal{S}_q^{(d_1+d_2)}$,

$$\|HH^\top - H'H'^\top\|_q \gtrsim \tau\varepsilon_0(2r)^{1/q} \quad \text{and} \quad \|HH^\top - H'H'^\top\|_{\mathrm{F}} \lesssim 2\sqrt{r}\varepsilon_0,$$

对于某些小的常数 $\tau\ \varepsilon_0 > 0$，其中 $\|\cdot\|_q$ 表示 Schatten-q 范数。然后我们考虑分布族

$$\mathcal{P}_\sigma = \left\{\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') : M_* = \sigma HH^\top, H \in \mathcal{S}_q^{(d_1+d_2)}\right\} \subset \mathcal{P}_\Sigma,$$

其基数 $N := |\mathcal{P}_\sigma| \geq 2^{2r(d_1+d_2-2r)}$。令 $M_* = \sigma HH^\top$ 和 $M_*' = \sigma H'H'^\top$。如附录 B所示，对于任何 $H = H' \in \mathcal{S}_q^{(d_1+d_2)}$, 我们有

$$\mathrm{KL}\left(\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i, X_i, X_i') \| \bigotimes_{i=1}^n f_{M_*'}(\bar{y}_i, X_i, X_i')\right) \lesssim \frac{n}{\sigma_\xi^2} C_u \sigma^2 r \varepsilon_0^2,$$

和 $\sum^{k\in[n]} \mathrm{TV}\left(f_{M_*}(\bar{y}_i, X_i, X_i')\ f_{M_*'}(\bar{y}_i, X_i, X_i')\right) \lesssim n^{\frac{\sqrt{C_u}\sigma}{\sigma_\xi}}\ \sqrt{r}\varepsilon_0$，. 为此，我们通过应用引理 2 3，并结合有界KL散度与TV距离的事实，以及 $\mathcal{P}_\sigma \subset \mathcal{P}_\Sigma$。定理 2 的证明推迟到附录 B。

定理 2。考虑从分布 $P_{M^*} \in \mathcal{P}_\Sigma$ 中抽取的样本 n，那么对于任何$\delta \lesssim e^{-n}$ 和任何 $q \in [1, \infty]$, 存在一个常数 $c > 0$

$$\inf_{\widetilde{M}} \sup_{P\in\mathcal{P}_\Sigma} \mathbb{E}\left\|\widetilde{M} - M\right\|_q \geq c\ \sqrt{C_u}^{\sigma_\xi}\left(r^{1/q}\sqrt{\begin{matrix} d_1 \vee d_2 \\ n \end{matrix}} + r^{\frac{1}{2}+\frac{1}{q}}\begin{matrix} d_1 \vee d_2 \\ n\varepsilon \end{matrix}\right) \bigwedge r^{1/q}\sigma,$$

where the infimum is taken over all possible (ε, δ)-DP algorithms. It suffices to choose $q = 1, 2, \infty$ to obtain the bounds in the nuclear norm, Frobenius norm, and spectral norm, respectively. For example, when $q = 2$, there exists a constant $c > 0$

$$\inf_{\widetilde{M}} \sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}\left\|\widetilde{M} - M\right\|_F \geq c\Big(\underbrace{\frac{\sigma_\xi}{\sqrt{C_u}} \sqrt{\frac{r(d_1 \vee d_2)}{n}}}_{l_1} + \underbrace{\frac{\sigma_\xi}{\sqrt{C_u}} \frac{r(d_1 \vee d_2)}{n\varepsilon}}_{l2} \Big) \bigwedge r^{1/2}\sigma. \tag{5}$$

In Theorem 2, the lower bound (5) consists of two terms, the statistical error l_1 and the cost of privacy l_2. The next section proposes a DP-estimator that attains the minimax lower bound (5), up to an additional factor σ_r and some logarithmic factors. As a supplement to DP-Fano's Lemma which works for $\delta \lesssim e^{-n}$, we also try the score attack argument, which is valid for a wider range of $\delta \lesssim n^{1+\gamma}$ where $\gamma > 0$ is a constant. Theorem 5 presents the DP-constrained lower bound established by the score attack argument. The content and proof of Theorem 5 are deferred to Appendix D. We also point out that it is trivial to derive the minimax lower bound of the case $d_1 = d_2 = d$ based on DP-Fano's Lemma since there is no need to apply the trick of symmetrization.

4 Upper bounds with differential privacy

In this section, we present Algorithm 2, DP-RGrad, and show that DP-RGrad attains the near-optimal convergence rate for differentially privately estimating low-rank matrices under the trace regression model. Our approach is based on privatizing the Riemannian gradient descent (RGrad) by the Gaussian mechanism. Interested readers may refer to Vandereycken (2013); Edelman et al. (1998); Adler et al. (2002); Absil et al. (2008) for the basics of RGrad. Let the estimate we obtain after l iterations be the rank-r matrix $M_l \in \mathbb{R}^{d_1 \times d_2}$ whose SVD has the form $M_l = U_l \Sigma_l V_l^\top$. It is well-known in Absil et al. (2008); Vandereycken (2013) that the tangent space of $\mathbb{M}_r$ at M_l is given by $\mathbb{T}_l := \left\{Z \in \mathbb{R}^{d_1 \times d_2} : Z = U_l R^\top + L V_l^\top, R \in \mathbb{R}^{d_2 \times r}, L \in \mathbb{R}^{d_1 \times r}\right\}$. The projection of the gradient G_l onto $\mathbb{T}_l$ is $\mathcal{P}_{\mathbb{T}_l}(G_l) = U_l U_l^\top G_l + G_l V_l V_l^\top - U_l U_l^\top G_l V_l V_l^\top$, which is of rank at most $2r$. Let the noisy gradient descent on the tangent space be $M_l - \eta_l \mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l} N_l$ where $N_l \in \mathbb{R}^{d_1 \times d_2}$ is the Gaussian noise matrix. Then, we retract it back to $\mathbb{M}_r$ and obtain

$$M_{l+1} = \mathrm{SVD}_r\left(M_l - \eta_l \mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l} N_l\right). \tag{6}$$

We update the estimate as defined in (6) for $1 = 0, \cdots, l^* - 1$ where l^* is the total number of iterations. Thanks to the composition property and Gaussian mechanism, we only need to ensure

在所有可能的 (ε, δ)-DP 算法上取下确界。只需选择 $q = 1, 2, \infty$ 即可分别获得核范数、Frobenius 范数和谱范数的界限。例如，当 $q = 2$ 时，存在一个常数 $c > 0$

$$\inf_{\widetilde{M}} \sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}\left\|\widetilde{M} - M\right\|_F \geq c\Big(\underbrace{\frac{\sigma_\xi}{\sqrt{C_u}} \sqrt{\frac{r(d_1 \vee d_2)}{n}}}_{l_1} + \underbrace{\frac{\sigma_\xi}{\sqrt{C_u}} \frac{r(d_1 \vee d_2)}{n\varepsilon}}_{l2} \Big) \bigwedge r^{1/2}\sigma. \tag{5}$$

在定理 2 中，下界 (5) 包含两项，即统计误差 l_1 和隐私成本 l_2。下一节提出了一种 DP-估计器，它达到了 minimax 下界 (5)，最多相差一个附加因子 σ_r 和一些对数因子。作为适用于 $\delta \lesssim e^{-n}$ 的 DP-Fano 引理的补充，我们还尝试了评分攻击论证，该论证适用于更广泛的 $\delta \lesssim n^{1+\gamma}$，其中 $\gamma > 0$ 是一个常数。定理 5 提出了由评分攻击论证建立的 DP-约束下界。定理 5 的内容和证明被推迟到附录 D。我们还指出，基于 DP-Fano 引理推导该情况的最小最大下界是显而易见的，因为不需要应用对称化的技巧。

4 上界与差分隐私

在本节中，我们介绍算法2，DP-RGrad，并证明DP-RGrad在迹回归模型下对低秩矩阵进行差分隐私估计时达到近最优收敛速度。我们的方法基于高斯机制对黎曼梯度下降（RGrad）进行隐私化。感兴趣的读者可参考Vandereycken (2013)；Edelman et al. (1998)；Adler et al. (2002)；Absil et al. (2008)了解RGrad的基础知识。设经过 l 次迭代后得到的估计为秩-r 矩阵 $M_l \in \mathbb{R}^{d_1 \times d_2}$，其SVD形式为$M_l = U_l \Sigma_l V_l^\top$。在Absil et al. (2008)；Vandereycken (2013)中众所周知，$\mathbb{M}_r$ 在 M_l 处的切空间由 $\mathbb{T}_l := \left\{Z \in \mathbb{R}^{d_1 \times d_2} : Z = U_l R^\top + L V_l^\top, R \in \mathbb{R}^{d_2 \times r}, L \in \mathbb{R}^{d_1 \times r}\right\}$给出。梯度 G_l在 $\mathbb{T}_l$上的投影为 $\mathcal{P}_{\mathbb{T}_l}(G_l) = U_l U_l^\top G_l + G_l V_l V_l^\top - U_l U_l^\top G_l V_l V_l^\top$，其秩至多为 $2r$。设切空间上的带噪声梯度下降为 $M_l - \eta_l \mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l} N_l$，其中 $N_l \in \mathbb{R}^{d_1 \times d_2}$ 是高斯噪声矩阵。然后，我们将它回缩到 $\mathbb{M}_r$ 并得到

$$M_{l+1} = \mathrm{SVD}_r\left(M_l - \eta_l \mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l} N_l\right). \tag{6}$$

我们根据（6）中定义的方式更新估计值，$1 = 0$、$\cdots$、$l^* - 1$ 其中 l^* 是迭代的总次数。由于组合属性和高斯机制，我们只需要确保

that each iteration is $(\varepsilon/l^*, \delta/l^*)$-DP. For trace regression model defined in (1), empirical mean squared loss is defined as $\mathcal{L}_n(M_l; Z) := \frac{1}{2n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i)^2$ and the empirical Euclidean gradient is $G_l := \nabla\mathcal{L}_n(M_l; Z) = \frac{1}{n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i) X_i$. The sensitivity of the l-th iteration is $\Delta_l := \max_{\text{neighbouring}(Z,Z')} \|M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z)) - [M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z'))]\|_F$.

Algorithm 2 DP-RGrad for trace regression

Input: the loss function $\mathcal{L}$; the data set $\{(X_i, y_i)\}_{i=1}^n$; sensitivities $\{\Delta_l\}_{l\in[l^*]}$; DP-initialization $\widetilde{M}_0$; rank r; nuisance variance σ_ξ^2; privacy budget $\varepsilon > 0, \delta \in (0,1)$.

Output: (ε, δ)-differentially private estimate M_{l^*} for trace regression.

Initialization: $M_0 \longleftarrow \widetilde{M}_0$.

for $l + 1 \in [l^*]$ **do**

$$M_{l+1} \longleftarrow \text{SVD}_r\left(M_l - \eta_l\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\right),$$

where G_l is the empirical Euclidean gradient

$$G_l := \nabla\mathcal{L}_n(M_l; Z) = \frac{1}{n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i) X_i,$$

and $N_l \in \mathbb{R}^{d_1\times d_2}$ has entries i.i.d. to

$$\mathcal{N}\left(0, \frac{2\Delta_l^2 l^{*2}}{\varepsilon^2}\log\left(\frac{1.25l^*}{\delta}\right)\right).$$

Return: $\widetilde{M}_{l^*} \longleftarrow M_{l^*}$

Theorem 3 establishes the error bound of the estimator $\widetilde{M}_{l^*}$ given by Algorithm 2. The proof of Theorem 3 is deferred to Appendix C.

Theorem 3. *Consider i.i.d. observations $Z = \{z_1, \cdots, z_n\}$ drawn from the trace regression model stated in (1) where the true low-rank regression coefficients matrix being $M \in \mathbb{M}_r$. Here, $z_i := (X_i, y_i)$ for $i = 1, \cdots, n$ and we assume that $\{X_i\}_{i\in[n]}$ satisfy the Assumption 1. Under the Assumption 1 and the condition that $(d_1 + d_2) > \log n$, suppose that Algorithm 2 takes in an (ε, δ)-DP initialization such that for some small constant $0 < c_0 < 1$, $\|\widetilde{M}_0 - M\|_F \le \sqrt{2r}\|\widetilde{M}_0 - M\| \le c_0\sigma_r$, and the sensitivities Δ_l take the value*

$$\Delta_l = C_3\frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2)\log n}, \quad \text{for all} \quad l = 1, \cdots, l^*,$$

for some absolute constant $C_3 > 0$, then we have, Algorithm 2 is $(2\varepsilon, 2\delta)$-differentially private.

那每个迭代是 $(\varepsilon/l^*, \delta/l^*)$-DP。对于在 (1), 中定义的跟踪回归模型，经验均方损失定义为 $\mathcal{L}_n(M_l; Z) := \frac{1}{2n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i)^2$ 并且经验欧几里得梯度是 $G_l := \nabla\mathcal{L}_n(M_l; Z) = \frac{1}{n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i) X_i$. l-th 迭代的敏感性是 $\Delta_l := \max_{\text{neighbouring}(Z,Z')} \|M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z)) - [M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z'))]\|_F$.

Algorithm 2 DP-RGrad for trace regression

Input: the loss function $\mathcal{L}$; the data set $\{(X_i, y_i)\}_{i=1}^n$; sensitivities $\{\Delta_l\}_{l\in[l^*]}$; DP-initialization $\widetilde{M}_0$; rank r; nuisance variance σ_ξ^2; privacy budget $\varepsilon > 0, \delta \in (0,1)$.

Output: (ε, δ)-differentially private estimate M_{l^*} for trace regression.

Initialization: $M_0 \longleftarrow \widetilde{M}_0$.

for $l + 1 \in [l^*]$ **do**

$$M_{l+1} \longleftarrow \text{SVD}_r\left(M_l - \eta_l\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\right),$$

where G_l is the empirical Euclidean gradient

$$G_l := \nabla\mathcal{L}_n(M_l; Z) = \frac{1}{n}\sum_{i=1}^n (\langle X_i, M_l\rangle - y_i) X_i,$$

and $N_l \in \mathbb{R}^{d_1\times d_2}$ has entries i.i.d. to

$$\mathcal{N}\left(0, \frac{2\Delta_l^2 l^{*2}}{\varepsilon^2}\log\left(\frac{1.25l^*}{\delta}\right)\right).$$

Return: $\widetilde{M}_{l^*} \longleftarrow M_{l^*}$

定理 3 建立了估计器的误差界限 $\widetilde{M}_{l^*}$ 由算法 2给出。定理 3 的证明被推迟到附录 C。

定理 3。考虑 i.i.d. 观测 $Z = \{z_1, \cdots, z_n\}$ 从在 (1) 中陈述的跟踪回归模型中抽取，其中真实的低秩回归系数矩阵是 $M \in \mathbb{M}_r$。这里,$z_i := (X_i, y_i)$ 对于 $i = 1, \cdots, n$ 并且我们假设 $\{X_i\}_{i\in[n]}$ 满足假设 1。在假设 1 和 $(d_1 + d_2) > \log n$ 的条件下，假设算法 2 接收一个 (ε δ $0 < c_0 < 1$ $\|\widetilde{M}_0 - M\|_F \le$,)-DP 初始化，使得对于某个小的常数 ，$2r\|\widetilde{M}_0 - M\| \le c_0\sigma_r$，并且敏感性 Δ_l 取值

$$\Delta_l = C_3\frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2)\log n}, \quad \text{for all} \quad l = 1, \cdots, l^*,$$

对于某个绝对常数 $C_3 > 0$，那么我们有，算法 2 是 $(2\varepsilon, 2\delta)$-差分隐私。

Moreover, as the sample size

$$n \geq c_4 \max\left\{ \underbrace{c_1 r(d_1+d_2)}_{n_3}, \quad \underbrace{\eta^2 \kappa_\xi^2 C_u r(d_1+d_2)\log(d_1+d_2)}_{n_4}, \right. \\ \left. \underbrace{\eta\sqrt{C_u}\left(\kappa_\xi+\sqrt{C_u}\right) r(d_1+d_2)\log^{3/2}(n)\frac{\log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}{\varepsilon}}_{n_5} \right\},$$

for some small constant $0 < c_4 < 1$*, number of iteration* $l^* = O(\log n)$*, and the step size* $0 < \eta < 1$*, we have the output* $\widetilde{M}_{l^*}$ *of Algorithm 2 satisfies*

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq \underbrace{C_4\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2)}_{u_1} \\ + \underbrace{C_4\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{3/2} n \log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}_{u_2}.$$

with probability at least

$$1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right) - (d_1+d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r} \\ - \left((d_1+d_2)^{-10} + \exp(-(d_1+d_2)) + n^{-9} + \exp\left(-10C_u(d_1+d_2)\right) n^{-9}\right)\log n.$$

According to Theorem 3, the upper bound of $\left\|\widetilde{M}_{l^*} - M\right\|_F$ can be decomposed into the the statistical error u_1 and the cost of privacy u_2. The term u_1 matches the the optimal rate $l_1 \sim \sigma_\xi\sqrt{\frac{r(d_1\vee d_2)}{n}}$, only up to logarithmic factors. However, the term u_2 differs from the theoretical lower bound of the cost of privacy $l_2 \sim \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}$, by a non-trivial factor σ_r, apart from logarithmic factors. In conclusion, Theorem 3 shows that as the sample size $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$, the estimator $\widetilde{M}_{l^*}$ given by Algorithm 2 attains the near-optimal convergence rate

$$\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + (\sigma_\xi+\sigma_r)\frac{r(d_1+d_2)}{n\varepsilon}\right). \tag{7}$$

The sample size requirement of Theorem 3 has the following explanations. The sample size n_3 is required to guarantee that the RIP condition stated in Lemma 1 occurs with high probability. The sample size n_4 is necessary to control the statistical error contributed by $\sum_{i\in[n]}\xi_i X_i$ in each iteration where ξ_i is the model noise. The sample size n_5 arises from controlling the cost of privacy due to $\mathcal{P}_{\mathbb{T}_l}N_l$ in each iteration.

此外，由于样本量

$$n \geq c_4 \max\left\{ \underbrace{c_1 r(\underbrace{d_1+d_2}_{n_3})}_{}, \quad \underbrace{\eta^2 \kappa_\xi^2 C_u r(\underbrace{d_1+d_2}_{n_4})\log(d_1+d_2)}_{}, \right. \\ \left. \underbrace{\eta\sqrt{C_u}\left(\kappa_\xi+\sqrt{C_u}\right) r(\underbrace{d_1+d_2}_{n_5})\log^{3/2}(n)\frac{\log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}{\varepsilon}}_{} \right\},$$

对于某个小的常数 $0 < c_4 < 1$，迭代次数 $l^* = O(\log n)$，以及步长$0 < \eta < 1$，我们有输出 $\widetilde{M}_{l^*}$ 算法 2 满足

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq \underbrace{C_4\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}}_{u_1}\log^{1/2}(d_1+d_2) \\ + \underbrace{C_4\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}}_{u_2}\log^{3/2} n \log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right).$$

至少以概率

$$1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right) - (d_1+d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r} \\ - \left((d_1+d_2)^{-10} + \exp(-(d_1+d_2)) + n^{-9} + \exp\left(-10C_u(d_1+d_2)\right) n^{-9}\right)\log n.$$

根据定理 3，的上限可以分解为统计误差 $\left\|\widetilde{M}_{l^*} - M\right\|_F$ 和隐私成本 u_1。项 u_1 与最优率 $l_1 \sim \sigma_\xi\sqrt{\frac{r(d_1\vee d_2)}{n}}$ 匹配，仅相差对数因子。然而，项 u_2 与隐私成本的理论下限 $l_2 \sim \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}$ 相差一个非平凡的因子σ_r，除了对数因子。总之，定理 3 表明随着样本量 $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$，估计量 $\widetilde{M}_{l^*}$ 由算法 2 给出达到近似最优的收敛率

$$\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + (\sigma_\xi+\sigma_r)\frac{r(d_1+d_2)}{n\varepsilon}\right). \tag{7}$$

定理的样本量要求有如下解释。样本量 3 需要保证引理 1 中所述的RIP条件以高概率发生。样本量 n_3需要控制由 $\sum_{i\in[n]}\xi_i X_i$ 在每个迭代中贡献的统计误差，其中 ξ_i 是模型噪声。样本量 n_4 源于控制每个迭代中由于 $\mathcal{P}_{\mathbb{T}_l}N_l$ 造成的隐私成本。

5 Discussion

In this section, we discuss the non-trivial gap σ_r between $u_2 \sim (\sigma_\xi + \sigma_r)\frac{r(d_1+d_2)}{n\varepsilon}$ and $l_2 \sim \sigma_\xi \frac{r(d_1 \vee d_2)}{n\varepsilon}$. Note that l_2 is free of σ_r while u_2 contains the factor σ_r arising from sensitivities

$$\Delta_l \asymp \frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2)\log n} \quad \text{for} \quad l = 1, \cdots, l^*.$$

The quantity $\sigma_r\sqrt{C_u}$ of Δ_l arises from $\|\langle M_l - M, X_i\rangle\|_{\psi_2} \le \sqrt{C_u}\,\|M_l - M\|_F$, as elaborated in (S.12). Here, $\|\cdot\|_{\psi_2}$ denotes the sub-Gaussian norm. Therefore, we cannot get rid of the factor σ_r once the measurement matrices $\{X_i\}_{i\in[n]}$ are subject to differential privacy. In many real applications, however, the measurement matrices $\{X_i\}_{i\in[n]}$ are fixed with deterministic designs. People publish $\{X_i\}_{i\in[n]}$ to the public with little concern on the privacy of $\{X_i\}_{i\in[n]}$. Although the exposure of $\{X_i\}_{i\in[n]}$ alone will not reveal any information on M, the privacy of M suffers from leakage when the public has access to the joint observations $\{(X_i, y_i)\}_{i\in[n]}$. We, therefore, introduce the following notion of privacy for neighboring datasets sharing the same measurement matrix.

Definition 1 (weak (ε,δ)-differential privacy)**.** *The algorithm A that maps Z into $\mathbb{R}^{d_1\times d_2}$ is weak (ε,δ)-differentially private over the dataset Z if $\mathbb{P}\big(A(Z)\in\mathcal{Q}\big) \le e^\varepsilon \mathbb{P}\big(A(Z')\in\mathcal{Q}\big)+\delta$, for all neighbouring data set Z, Z' sharing the same measurement X and all subset $\mathcal{Q}\subset\mathbb{R}^{d_1\times d_2}$.*

In Theorem 7, Appendix F, we show that as $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$, the estimator $\widetilde{M}_{l^*}$ given by Algorithm 2 attains the optimal convergence rate $\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}\right)$ in the sense of weak differential privacy. The analogs of Theorem 1, Corollary 1 and Theorem 3 in the sense of weak differential privacy can be found as Theorem 6, Corollary 2 and Theorem 7 in Appendix F. It is interesting to explore in future work whether the score attack argument or DP-Fano's Lemma can be generalized to include the non-trivial factor σ_r.

Acknowledgement

The author would like to express sincere gratitude to Hong Kong PhD Fellowship Scheme and Hong Kong RGC GRF Grant 16301622 for providing financial support for this research. The author also wishes to acknowledge the invaluable guidance provided by Prof. Dong, Xia throughout the research process. Additionally, the author would like to extend heartfelt thanks to Mr. Zetao, Fei for his constructive criticism during the paper revision. Their contributions have been instrumental in the successful completion of this research.

5 讨论

在本节中，我们讨论 σ_r 和 $u_2 \sim (\sigma_\xi + \sigma_r)^{\frac{r(d_1+d_2)}{n\varepsilon}}$ 之间的非平凡差距 $l_2 \sim \sigma_\xi^{\frac{r(d_1\vee d_2)}{n\varepsilon}}$。请注意，$l_2$ 没有 σ_r，而 u_2 包含来自敏感性的因子 σ_r

$$\Delta_l \asymp \frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2)\log n} \quad \text{for} \quad l = 1, \cdots, l^*.$$

数量 σ_r C_u的 $\sqrt{\Delta_l}$源于 $\|\langle M_l - M\ X_i\rangle\|_{\psi_2} \le$ $C_u\,\|M_l - M\|_F$，，如 (S.12) 中所述。这里，$\|\cdot\|_{\psi_2}$表示次高斯范数。因此，一旦测量矩阵 $\{X_i\}_{i\in[n]}$ 受到差分隐私的保护，我们就无法消除因子σ_r。然而，在许多实际应用中，测量矩阵 $\{X_i\}_{i\in[n]}$ 具有确定性设计并固定。人们将 $\{X_i\}_{i\in[n]}$ 公开发布，而很少关注 $\{X_i\}_{i\in[n]}$ 的隐私。尽管单独暴露 $\{X_i\}_{i\in[n]}$ 不会泄露任何关于 M 的信息，但当公众可以访问联合观测 $\{(X_i, y_i)\}_{i\in[n]}$ 时， M 的隐私会受到泄露。因此，我们为共享相同测量矩阵的相邻数据集引入以下隐私概念。

定义 1 (弱 (ε,δ)-差分隐私)。将 Z 映射到 $\mathbb{R}^{d_1\times d_2}$ 的算法 A 在数据集 Z *if* $\mathbb{P}\big(A(Z)\in\mathcal{Q}\big) \le e^\varepsilon \mathbb{P}\big(A(Z')\in\mathcal{Q}\big)+\delta$ 上是弱 (ε,δ)-差分隐私的，对于所有共享相同测量 X 的相邻数据集 Z, Z' 和所有子集 $\mathcal{Q}\subset\mathbb{R}^{d_1\times d_2}$。

在定理 7，附录 F，我们证明，当 $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$， 估计器 $\widetilde{M}_{l^*}$ 由算法 2 给出时，在弱差分隐私的意义下达到最优收敛速度 $\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}\right)$。定理 1，推论 1 和定理 3在弱差分隐私的意义下的类似物可以在附录 6，推论 2 和定理7 中找到 F。在未来的工作中探索分数攻击论证或 DP-Fano 引理是否可以推广以包含非平凡因子 σ_r。

致谢

作者谨向香港博士奖学金计划及香港研究资助局一般研究资助金16301622表达诚挚的感谢，感谢其为本研究提供资金支持。作者亦希望感谢董晓教授在整个研究过程中提供的宝贵指导。此外，作者衷心感谢费泽涛先生在论文修改期间提出的建设性意见。他们的贡献对于本研究的成功完成至关重要。

A Proof of Theorem 1

The proof of Theorem 1 consists of four parts. In Part A.1, we list several existing results that are useful in the proofs later. In Part A.2, Lemma 2 works as the main technique to derive the sensitivity $\Delta^{(1)}$. Part A.3 derives the sensitivity $\Delta^{(2)}$. Part A.4 establishes the upper bounds of $\|\widetilde{M}_0 - M\|$ and $\|\widetilde{M}_0 - M\|_F$ based on the $\Delta^{(1)}$ and $\Delta^{(2)}$.

A.1 Part 1

The following Theorem 4 proposed by Proposition 2, Koltchinskii (2011) will be frequently used to establish the upper bound of the spectral norm of a summation of independent random matrices.

Theorem 4 (Bernstein's inequality, Koltchinskii (2011)). *Let $B_1, \cdots, B_n$ be independent $d_1 \times d_2$ matrices such that for some $\alpha \geq 1$ and all $i \in [n]$*

$$\mathbb{E}B_i = 0, \quad \|\Lambda_{\max}(B_i)\|_{\Psi_\alpha} =: K < +\infty.$$

Let

$$S^2 := \max\left\{\Lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}B_i B_i^\top\right)/n, \Lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}B_i^\top B_i\right)/n\right\}.$$

Then, for some constant $C > 0$ and for all $t > 0$,

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n B_i\right\| \geq CS\sqrt{\frac{t+\log(d_1+d_2)}{n}} + CK\log^{1/\alpha}\left(\frac{K}{S}\right)\frac{t+\log(d_1+d_2)}{n}\right) \leq \exp(-t).$$

Theorem 4 applies to bound the spectral norm of $\Delta := \widehat{L} - M$. The existing result for the case of heavy-tailed noise can be found in Theorem 6, Shen et al. (2023). Adapting the existing result to the case of Gaussian noise, we have that for some absolute constant $C_0 > 0$,

$$\|\Delta\| = \|\widehat{L} - M\| \leq C_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}, \tag{S.1}$$

with probability at least $1 - (d_1+d_2)^{-10}$. The following Lemma originated from Lemma 18, Shen et al. (2023), is useful to analyze the matrix permutation due to singular value decomposition.

Lemma 4 (Matrix Permutation, Shen et al. (2023)). *Let $M \in \mathbb{R}^{d_1 \times d_2}$ be a rank-r matrix with singular value decomposition of the form $M = U\Sigma V^\top$ where $\Sigma = \operatorname{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$ with*

定理的证明1

定理的证明1 分为四个部分。在部分 A.1，我们列出了几个在后续证明中有用的现有结果。在部分 A.2，引理 2 作为主要技术来推导灵敏度 $\Delta^{(1)}$。部分 A.3推导灵敏度 $\Delta^{(2)}$。部分 A.4 基于 $\|\widetilde{M}_0 - M\|$和 $\|\widetilde{M}_0 - M\|_F$建立了 $\Delta^{(1)}$和 $\Delta^{(2)}$的上界。

A.1 部分 1

以下定理 4 由命题 2 提出， Koltchinskii (2011) 将被频繁用于建立独立随机矩阵和的谱范数的上界。

定理 4 (伯恩斯坦不等式， Koltchinskii (2011))。 设 $B_1, \cdots, B_n$ 为独立的 $d_1 \times d_2$矩阵，使得对于某些 $\alpha \geq 1$ 和所有 $i \in [n]$

$$\mathbb{E}B_i = 0, \quad \|\Lambda_{\max}(B_i)\|_{\Psi_\alpha} =: K < +\infty.$$

Let

$$S^2 := \max\left\{\Lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}B_i B_i^\top\right)/n, \Lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}B_i^\top B_i\right)/n\right\}.$$

然后，对于某个常数 $C > 0$ 以及对于所有 $t > 0$,

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n B_i\right\| \geq CS\sqrt{\frac{t+\log(d_1+d_2)}{n}} + CK\log^{1/\alpha}\left(\frac{K}{S}\right)\frac{t+\log(d_1+d_2)}{n}\right) \leq \exp(-t).$$

定理 4 适用于界定 $\Delta := \widehat{L} - M$ 的谱范数。对于重尾噪声的情况，现有结果可以在定理 6 中找到， Shen 等人 (2023)。将现有结果应用于高斯噪声的情况，我们得到对于某个绝对常数 $C_0 > 0$,

$$\|\Delta\| = \|\widehat{L} - M\| \leq C_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}, \tag{S.1}$$

至少以概率 $1 - (d_1+d_2)^{-10}$。以下引理源自引理 18，Shen 等人 (2023)，对于由于奇异值分解而进行的矩阵置换分析很有用。

引理 4 (矩阵置换， Shen 等人 (2023))。 设 $M \in \mathbb{R}^{d_1 \times d_2}$ 是一个秩-r 矩阵，其奇异值分解形式为 $M = U\Sigma V^\top$ 其中 $\Sigma = \operatorname{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$,

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. *For any* $\widehat{M} \in \mathbb{R}^{d\times d}$ *satisfying* $\|\widehat{M} - M\|_{\mathrm{F}} < \sigma_r/8$, *then*

$$\left\|\mathrm{SVD}_r(\widehat{M}) - M\right\| \leq \|\widehat{M} - M\| + 40\frac{\|\widehat{M} - M\|^2}{\sigma_r},$$
$$\left\|\mathrm{SVD}_r(\widehat{M}) - M\right\|_{\mathrm{F}} \leq \|\widehat{M} - M\|_{\mathrm{F}} + 40\frac{\|\widehat{M} - M\|\|\widehat{M} - M\|_{\mathrm{F}}}{\sigma_r},$$

and

$$\left\|\widehat{U}\widehat{U}^\top - UU^\top\right\| \leq \frac{8}{\sigma_r}\|\widehat{M} - M\|, \quad \left\|\widehat{V}\widehat{V}^\top - VV^\top\right\| \leq \frac{8}{\sigma_r}\|\widehat{M} - M\|,$$

where the leading r *left singular vectors of* $\widehat{M}$ *are given by the columns of* $\widehat{U} \in \mathbb{R}^{d_1\times r}$ *and the leading* r *right singular vectors of* $\widehat{M}$ *are given by the columns of* $\widehat{V} \in \mathbb{R}^{d_2\times r}$.

A.2 Part 2

The second part aims to derive the sensivitity

$$\Delta^{(1)} := \max_{i\in[n]}\left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F \vee \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right).$$

Before moving on, we present Lemma 5, which provides conclusions on Δ and $\Delta^{(i)}$, frequently used in the proof later. The proof of Lemma 5 can be found in Appendix E.4.

Lemma 5. *Under model (1), Assumption 1, and the condition* $n \geq \frac{\log^2 n}{(d_1\vee d_2)\log(d_1+d_2)}$, *there exists some absolute constant* $C_0, C_1 > 0$ *such that the event*

$$\mathcal{E}_* := \left\{\max_{i\in[n]}\left\|\Delta - \Delta^{(i)}\right\| \leq C_0\cdot n^{-1}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n\right\}$$
$$\bigcap\left\{\|\Delta\| + \max_{i\in[n]}\|\Delta^{(i)}\| \leq C_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1\vee d_2)\log(d_1+d_2)}{n}}\right\},$$

holds with probability at least $1-(d_1+d_2)^{-10} - n^{-9}$. *Conditioned on the event* $\mathcal{E}_*$, *as the sample size satisfies*

$$n \geq C_1 C_l^{-1} r\left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2 (d_1\vee d_2)\log(d_1+d_2), \tag{S.2}$$

we have

$$\|\Delta_*\| \vee \max_{i\in[n]}\|\Delta_*^{(i)}\| = \|\Delta\| \vee \max_{i\in[n]}\|\Delta^{(i)}\| \leq \frac{\sigma_r}{8\sqrt{2r}} < \frac{\sigma_r}{5+\delta} < \frac{\sigma_r}{2}, \tag{S.3}$$

and

$$\|\Delta_*\|_F \vee \max_{i\in[n]}\|\Delta_*^{(i)}\|_F = \max_{i\in[n]}\|\Delta\|_F \vee \|\Delta^{(i)}\|_F \leq \frac{\sigma_r}{8},$$

for some constant $\delta > 0$, *where* $\Delta_*^{(i)}$ *is the symmetric dilation of* $\Delta^{(i)} := L^{(i)} - M$.

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. 对于任何满足 $\|\widehat{M} - M\|_{\mathrm{F}} < \sigma_r/8$ 的 $\widehat{M} \in \mathbb{R}^{d\times d}$，则

$$\left\|\mathrm{SVD}_r(\widehat{M}) - M\right\| \leq \|\widehat{M} - M\| + 40\frac{\|\widehat{M} - M\|^2}{\sigma_r},$$
$$\left\|\mathrm{SVD}_r(\widehat{M}) - M\right\|_{\mathrm{F}} \leq \|\widehat{M} - M\|_{\mathrm{F}} + 40\frac{\|\widehat{M} - M\|\|\widehat{M} - M\|_{\mathrm{F}}}{\sigma_r},$$

and

$$\left\|\widehat{U}\widehat{U}^\top - UU^\top\right\| \leq \frac{8}{\sigma_r}\|\widehat{M} - M\|, \quad \left\|\widehat{V}\widehat{V}^\top - VV^\top\right\| \leq \frac{8}{\sigma_r}\|\widehat{M} - M\|,$$

其中 $\widehat{M}$ 的前 r 个左奇异向量由 $\widehat{U} \in \mathbb{R}^{d_1\times r}$ 的列给出，且 $\widehat{M}$ 的前 r 个右奇异向量由 $\widehat{V} \in \mathbb{R}^{d_2\times r}$ 的列给出。

A.2 部分 2

第二部分旨在推导灵敏度

$$\Delta^{(1)} := \max_{i\in[n]}\left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F \vee \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right).$$

在继续之前，我们介绍引理 5，它提供了关于 Δ 和 $\Delta^{(i)}$ 的结论，这些结论在后面的证明中经常使用。引理 5 的证明可以在附录 E.4 中找到。

引理 5. 在模型 (1) 下，假设 1，以及条件 $n \geq \frac{\log^2 n}{(d_1\vee d_2)\log(d_1+d_2)}$，存在某个绝对常数 C_0 $C_1 > 0$,, 使得事件

$$\mathcal{E}_* := \left\{\max_{i\in[n]}\left\|\Delta - \Delta^{(i)}\right\| \leq C_0\cdot n^{-1}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n\right\}$$
$$\bigcap\left\{\|\Delta\| + \max_{i\in[n]}\|\Delta^{(i)}\| \leq C_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1\vee d_2)\log(d_1+d_2)}{n}}\right\},$$

以至少 $1-(d_1+d_2)^{-10} - n^{-9}$ 的概率保持。在事件 $\mathcal{E}_*$ 的条件下，随着样本量满足

$$n \geq C_1 C_l^{-1} r\left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2 (d_1\vee d_2)\log(d_1+d_2), \tag{S.2}$$

我们有

$$\|\Delta_*\| \vee \max_{i\in[n]}\|\Delta_*^{(i)}\| = \|\Delta\| \vee \max_{i\in[n]}\|\Delta^{(i)}\| \leq \frac{\sigma_r}{8\sqrt{2r}} < \frac{\sigma_r}{5+\delta} < \frac{\sigma_r}{2}, \tag{S.3}$$

and

$$\|\Delta_*\|_F \vee \max_{i\in[n]}\|\Delta_*^{(i)}\|_F = \max_{i\in[n]}\|\Delta\|_F \vee \|\Delta^{(i)}\|_F \leq \frac{\sigma_r}{8},$$

对于某个常数 $\delta > 0$，其中 $\Delta_*^{(i)}$ 是 $\Delta^{(i)} := L^{(i)} - M$ 的对称膨胀。

The following analysis is proceeded under the sample size condition (S.2) and is mainly conditioned on the event $\mathcal{E}_*$ which happens with probability at least $1-(d_1+d_2)^{-10}-n^{-9}$.

Step 1: expansion. Conditioned on $\mathcal{E}_*$, we are able to apply Lemma 2 to Δ_* and $\Delta_*^{(i)}$ and get

$$\begin{pmatrix} \widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top} & 0 \\ 0 & \widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top} \end{pmatrix} = \sum_{k\geq 1}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 1}\mathcal{S}_{M_*,k}(\Delta_*^{(i)}).$$

Our goal is to bound $\Delta^{(1)}$ which satisfies

$$\begin{aligned}\Delta^{(1)} &\leq \max_{i\in[n]}\left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F + \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right) \\ &\leq \max_{i\in[n]}\left(\left\|\mathcal{S}_{M_*,1}(\Delta_*) - \mathcal{S}_{M_*,1}(\Delta_*^{(i)})\right\|_F + \left\|\sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*^{(i)})\right\|_F\right).\end{aligned}$$

Step 2: bounding the first order term. By the definition of $\mathcal{S}_{M_*,1}(\Delta_*)$ and $\mathcal{S}_{M_*,1}(\Delta_*^{(i)})$,

$$\begin{aligned}&\max_{i\in[n]}\|\mathcal{S}_{M_*,1}(\Delta_*) - \mathcal{S}_{M_*,1}(\Delta_*^{(i)})\| = \max_{i\in[n]}\|Q^{-1}(\Delta-\Delta^{(i)})^\top Q_\perp + Q_\perp(\Delta-\Delta^{(i)})^\top Q^{-1}\| \\ &\leq \frac{2\sqrt{r}}{\sigma_r}\max_{i\in[n]}\|\Delta-\Delta^{(i)}\| \leq C_4\frac{\sqrt{r}}{n\sigma_r}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)\log n,\end{aligned} \tag{S.4}$$

conditioned on the event $\mathcal{E}_*$, for some absolute constant $C_4>0$.

Step 3: bounding the higher order terms. Let I_k be the index set for terms in $\mathcal{S}_{M_*,k}$

$$I_k = \left\{\mathbf{s}:\mathbf{s}=(s_1,\ldots,s_{k+1}), \sum_{m=1}^{k+1}s_m = k, s_m\geq 0 \quad \forall m\in[k+1]\right\},$$

with the cardinality $|I_k| = \binom{2k}{k}$. We define

$$\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right) := Q^{-s_1}\Delta_*^{(i)}Q^{-s_2}\cdots Q^{-s_l}\left(\Delta_*-\Delta_*^{(i)}\right)Q^{s_{l+1}}\cdots Q^{-s_k}\Delta_* Q^{s_{k+1}},$$

for $k\geq 2, \mathbf{s}=(s_1,\cdots,s_{k+1})\in I_k$ and $l\in[k]$. Since $|I_k| = \binom{2k}{k}$, the higher order terms $\max_{i\in[n]}\left\|\sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*,k}\left(\Delta_*^{(i)}\right)\right\|_{\mathrm{F}}$

$$= \max_{i\in[n]}\left\|\sum_{k\geq 2}\sum_{\mathbf{s}\in I_k}\sum_{l\in[k]}\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \leq \max_{i\in[n]}\sum_{k\geq 2}\binom{2k}{k}\sum_{l\in[k]}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right)\right\|_{\mathrm{F}}. \tag{S.5}$$

It is sufficient to find a upper bound of $\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right)\right\|_{\mathrm{F}}$. Denote

$$D_{\max} := C_1\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1\vee d_2)\log(d_1+d_2)}{n}},$$

在样本量条件下（S.2）进行以下分析，主要基于事件 $\mathcal{E}_*$ 发生的概率至少为 $1-(d_1+d_2)^{-10}-n^{-9}$。

步骤1：展开。 在 $\mathcal{E}_*$条件下，我们能够应用引理 2 到 Δ_* 和 $\Delta_*^{(i)}$并得到

$$\begin{pmatrix} \widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top} & 0 \\ 0 & \widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top} \end{pmatrix} = \sum_{k\geq 1}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 1}\mathcal{S}_{M_*,k}(\Delta_*^{(i)}).$$

我们的目标是界定满足 $\Delta^{(1)}$ 的

$$\begin{aligned}\Delta^{(1)} &\leq \max_{i\in[n]}\left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F + \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right) \\ &\leq \max_{i\in[n]}\left(\left\|\mathcal{S}_{M_*,1}(\Delta_*) - \mathcal{S}_{M_*,1}(\Delta_*^{(i)})\right\|_F + \left\|\sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*^{(i)})\right\|_F\right).\end{aligned}$$

步骤2：界定一阶项。根据 $\mathcal{S}_{M_*\,1}(\Delta_*)$ 的定义，和 $\mathcal{S}_{M_*\,1}(\Delta_*^{(i)})$，

$$\begin{aligned}&\max_{i\in[n]}\|\mathcal{S}_{M_*,1}(\Delta_*) - \mathcal{S}_{M_*,1}(\Delta_*^{(i)})\| = \max_{i\in[n]}\|Q^{-1}(\Delta-\Delta^{(i)})^\top Q_\perp + Q_\perp(\Delta-\Delta^{(i)})^\top Q^{-1}\| \\ &\leq \frac{2\sqrt{r}}{\sigma_r}\max_{i\in[n]}\|\Delta-\Delta^{(i)}\| \leq C_4\frac{\sqrt{r}}{n\sigma_r}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1+\sigma_\xi\right)\log n,\end{aligned} \tag{S.4}$$

在事件 $\mathcal{E}_*$ 的条件下，对于某个绝对常数 $C_4>0$。

步骤3：界定高阶项。设 I_k 是 $\mathcal{S}_{M_*,\,k}$ 中项的指标集。

$$I_k = \left\{\mathbf{s}:\mathbf{s}=(s_1,\ldots,s_{k+1}), \sum_{m=1}^{k+1}s_m = k, s_m\geq 0 \quad \forall m\in[k+1]\right\},$$

具有基数 $|I_k| = \binom{2k}{k}$。我们定义

$$\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right) := Q^{-s_1}\Delta_*^{(i)}Q^{-s_2}\cdots Q^{-s_l}\left(\Delta_*-\Delta_*^{(i)}\right)Q^{s_{l+1}}\cdots Q^{-s_k}\Delta_* Q^{s_{k+1}},$$

对于 $k\geq 2\ \mathbf{s}=(s_1,\cdots,s_{k+1})\in I_k$，和 $l\in[k]$。由于 $|I_k| = \binom{2k}{k}$，高阶项 $\max_{i\in[n]}\left\|\sum_{k\geq 2}\mathcal{S}_{M_*\,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*\,k}\left(\Delta_*^{(i)}\right)\right\|_{\mathrm{F}}$,

$$= \max_{i\in[n]}\left\|\sum_{k\geq 2}\sum_{\mathbf{s}\in I_k}\sum_{l\in[k]}\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \leq \max_{i\in[n]}\sum_{k\geq 2}\binom{2k}{k}\sum_{l\in[k]}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_*-\Delta_*^{(i)}\right)\right\|_{\mathrm{F}}. \tag{S.5}$$

找到 $\left\|\mathcal{T}_{M_*\,k\,\mathbf{s}\,l}\ \ \Delta_*-\Delta_*^{(i)}\right\|_{\mathrm{F}}$, , , 的一个上界就足够了。记

$$D_{\max} := C_1\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1\vee d_2)\log(d_1+d_2)}{n}},$$

which appeared in the event $\mathcal{E}_*$ as an upper bound of $\|\Delta\| + \max_{i\in[n]} \left\|\Delta^{(i)}\right\|$.

Conditioned on $\mathcal{E}_*$, for all $i \in [n]$, $k \geq 2, \mathbf{s} \in I_k$ and $l \in [k]$,

$$\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta - \Delta^{(i)}\right)\right\|_{\mathrm{F}} \leq \sqrt{2r}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta - \Delta^{(i)}\right)\right\| \leq \frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\left(\frac{D_{\max}}{\sigma_r}\right)^{k-1},$$

where the first inequality is because $\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_* - \Delta_*^{(i)}\right)$ is of rank at most $2r$. Let a be a function defined by $a(k) = \binom{2k}{k}k$, then $a(2) = 12$ and $\frac{a(k+1)}{a(k)} \leq 5$ for all integer $k \geq 2$,

$$\begin{aligned}
&\max_{i\in[n]} \sum_{k\geq 2}\binom{2k}{k}\sum_{l\in[k]}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_* - \Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \\
&\leq \max_{i\in[n]}\binom{4}{2}\sum_{k\geq 0} 5^k\left(\frac{\|D_{\max}\|}{\sigma_r}\right)^k\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\left(\frac{D_{\max}}{\sigma_r}\right) \\
&\leq \max_{i\in[n]} a(2)\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\sum_{k\geq 0}\left(\frac{5}{5+\delta}\right)^k\left(\frac{D_{\max}}{\sigma_r}\right) \\
&\leq \max_{i\in[n]} a(2)\left(\frac{5+\delta}{\delta}\right)\left(\frac{D_{\max}}{\sigma_r}\right)\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|,
\end{aligned} \tag{S.6}$$

where the last step is due to (S.3), which is guaranteed by the sample size condition (S.2) together with the event $\mathcal{E}_*$. Combining (S.5) and (S.6), since $\left\|\Delta_* - \Delta_*^{(i)}\right\| = \left\|\Delta - \Delta^{(i)}\right\|$, conditioned on the event $\mathcal{E}_*$, we have

$$\max_{i\in[n]}\left\|\sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*,k}\left(\Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \leq C_4\left(\frac{12}{\delta}\right)\frac{\sqrt{2r}}{n}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n,$$

for some absolute constant $C_3 > 0$. In conclusion, conditioned on $\mathcal{E}_*$, as the sample size $n \geq C_1 C_l^{-1} r\left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2(d_1 \vee d_2)\log(d_1 + d_2)$, for some absolute constant $C_1 > 0$, we have $\Delta^{(1)} \leq C_4\frac{\sqrt{r}}{n}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n$, for some absolute constant C_4.

Let $E_U \in \mathbb{R}^{d_1\times d_1}$ be a symmetric matrix where the entries $(E_U)_{ij}$ i.i.d. to $\mathcal{N}(0, \frac{18\Delta^{(1)2}}{\varepsilon^2}\log(\frac{3.75}{\delta}))$ for $1 \leq i \leq j \leq d_1$. Then, conditioned on the event $\mathcal{E}_*$ and (S.2), for some absolute constant $\widetilde{C}_4 > 0$, $\|E_U\| \leq \widetilde{C}_4\frac{\sqrt{rd_1}}{n\varepsilon}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta})$, with probability at least $1 - e^{-d_1}$. Moreover, by Gaussian mechanism, $\widehat{U}\widehat{U}^\top + E_U$ is $(\varepsilon/3, \delta/3)$-DP and thus $\widetilde{U}$ is also $(\varepsilon/3, \delta/3)$-DP thanks to the post-processing property of differential privacy. Furthermore, by Davis-Kahan's

在事件 $\mathcal{E}_*$ 中出现，作为 $\|\Delta\| + \max_{i\in[n]}\left\|\Delta^{(i)}\right\|$ 的上界。

在 $\mathcal{E}_*$ 的条件下，对于所有 $i \in [n]$, $k \geq 2$, $\mathrm{s} \in I_k$ 和 $l \in [k]$,

$$\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta - \Delta^{(i)}\right)\right\|_{\mathrm{F}} \leq \sqrt{2r}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta - \Delta^{(i)}\right)\right\| \leq \frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\left(\frac{D_{\max}}{\sigma_r}\right)^{k-1},$$

其中第一个不等式是因为 $\mathcal{T}_{M_*\,k\,\mathbf{s}\,l}\left(\Delta_* - \Delta_*^{(i)}\right)$, , , 的秩至多为 $2r$。令 a 是一个由 $a(k) = \binom{2k}{k}k$ 定义的函数，则 $a(2) = 12$ 并且 $\frac{a(k+1)}{a(k)} \leq 5$ 对于所有整数 $k \geq 2$,

$$\begin{aligned}
&\max_{i\in[n]} \sum_{k\geq 2}\binom{2k}{k}\sum_{l\in[k]}\left\|\mathcal{T}_{M_*,k,\mathbf{s},l}\left(\Delta_* - \Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \\
&\leq \max_{i\in[n]}\binom{4}{2}\sum_{k\geq 0} 5^k\left(\frac{\|D_{\max}\|}{\sigma_r}\right)^k\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\left(\frac{D_{\max}}{\sigma_r}\right) \\
&\leq \max_{i\in[n]} a(2)\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|\sum_{k\geq 0}\left(\frac{5}{5+\delta}\right)^k\left(\frac{D_{\max}}{\sigma_r}\right) \\
&\leq \max_{i\in[n]} a(2)\left(\frac{5+\delta}{\delta}\right)\left(\frac{D_{\max}}{\sigma_r}\right)\frac{\sqrt{2r}}{\sigma_r}\left\|\Delta_* - \Delta_*^{(i)}\right\|,
\end{aligned} \tag{S.6}$$

其中最后一步是由于 (S.3)，这由样本量条件 (S.2) 与事件 $\mathcal{E}_*$。结合 (S.5) 和 (S.6)，由于 $\left\|\Delta_* - \Delta_*^{(i)}\right\| = \left\|\Delta - \Delta^{(i)}\right\|$，在事件 $\mathcal{E}_*$ 的条件下，我们有

$$\max_{i\in[n]}\left\|\sum_{k\geq 2}\mathcal{S}_{M_*,k}(\Delta_*) - \sum_{k\geq 2}\mathcal{S}_{M_*,k}\left(\Delta_*^{(i)}\right)\right\|_{\mathrm{F}} \leq C_4\left(\frac{12}{\delta}\right)\frac{\sqrt{2r}}{n}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n,$$

对于某个绝对常数 $C_3 > 0$。总之，在 $\mathcal{E}_*$ 的条件下，随着样本大小 $n \geq C_1 C_l^{-1} r\left(\frac{\sigma_\xi + \sqrt{C_u r}\sigma_1}{\sigma_r}\right)^2(d_1 \vee d_2)\log(d_1 + d_2)$，对于某个绝对常数 $C_1 > 0$，我们有 $\Delta^{(1)} \leq C_4\frac{\sqrt{r}}{n}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n$，对于某个绝对常数 C_4。

设 $E_U \in \mathbb{R}^{d_1\times d_1}$ 是一个对称矩阵，其中元素 $(E_U)_{ij}$ 独立同分布于 $\mathcal{N}(0, \frac{18\Delta^{(1)2}}{\varepsilon^2}\log(\frac{3.75}{\delta}))$，对于 $1 \leq i \leq j \leq d_1$。然后，在事件 $\mathcal{E}_*$ 和 (S.2) 的条件下，对于某个绝对常数 $\widetilde{C}_4 > 0$ $\|E_U\| \leq \widetilde{C}_4\frac{\sqrt{rd_1}}{n\varepsilon}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)$, $\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta})$，概率至少为 $1 - e^{-d_1}$。此外，通过高斯机制，$\widehat{U}\widehat{U}^\top + E_U$ 是 $(\varepsilon/3, \delta/3)$-DP，并且因此 $\widetilde{U}$ 也是 $(\varepsilon/3, \delta/3)$-DP，这得益于差分隐私的后处理特性。此外，通过 Davis-Kahan 的

Theorem, for some absolute constant $\tilde{c}_0 > 0$

$$\begin{aligned}\left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\| &\le 1 \wedge \left(\left\|\widehat{U}\widehat{U}^\top - UU^\top\right\| + \|E_U\|\right) \overset{(a)}{\le} 1 \wedge \left(\left(\frac{8}{\sigma_r}\|\widehat{L} - M\| \wedge 1\right) + \|E_U\|\right) \\ &\le 1 \wedge \tilde{c}_0\left(\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}\right. \\ &\quad \left. + \sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\frac{\sqrt{r d_1}}{n\varepsilon}\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)\right),\end{aligned} \tag{S.7}$$

where we apply Lemma 4 in step (a).

Let $E_V \in \mathbb{R}^{d_2 \times d_2}$ be a symmetric matrix with $(E_V)_{ij}$ i.i.d. to $\mathcal{N}(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(\frac{3.75}{\delta}))$ for $1 \le i \le j \le d_2$, then for some absolute constant $\widetilde{C}_4 > 0$, conditioned on the event $\mathcal{E}_*$ and (S.2), we have $\|E_V\| \le \widetilde{C}_4 \frac{\sqrt{r d_2}}{n\varepsilon}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})$, with probability at least $1 - e^{-d_2}$. Moreover, by Gaussian mechanism, $\widehat{V}\widehat{V}^\top + E_U$ is $(\varepsilon/3, \delta/3)$-DP and $\widetilde{V}$ is also $(\varepsilon/3, \delta/3)$-DP thanks to the post-processing property of differential privacy. Furthermore, by Davis-Kahan's Theorem, for some absolute constant $\tilde{c}_0 > 0$

$$\begin{aligned}\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| &\le 1 \wedge \tilde{c}_0\left(\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}\right. \\ &\quad \left. + \sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\frac{\sqrt{r d_2}}{n\varepsilon}\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)\right).\end{aligned} \tag{S.8}$$

A.3 Part 3

Given the $(\varepsilon/3, \delta/3)$-DP singular vectors $\widetilde{U} \in \mathbb{O}^{d_1 \times r}$ and $\widetilde{V} \in \mathbb{O}^{d_2 \times r}$ obtained in Part A.2, we derive the sensitivity $\Delta^{(2)} := \max_{i \in [n]} \left\|\widetilde{U}^\top\left(\widehat{L} - \widehat{L}^{(i)}\right)\widetilde{V}\right\|_F$. Conditioned on the event $\mathcal{E}_*$,

$$\begin{aligned}\Delta^{(2)} &:= \max_{i\in[n]}\left\|\widetilde{U}^\top\left(\widehat{L} - \widehat{L}^{(i)}\right)\widetilde{V}\right\|_F \le \max_{i\in[n]}\sqrt{r}\left\|\widehat{L} - \widehat{L}^{(i)}\right\| \\ &= \max_{i\in[n]}\sqrt{r}\left\|\widehat{L} - M + M - \widehat{L}^{(i)}\right\| = \max_{i\in[n]}\sqrt{r}\left\|\Delta - \Delta^{(i)}\right\| \\ &\le C_3 \cdot n^{-1}\sqrt{C_l^{-1}}\sqrt{r}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n.\end{aligned}$$

Let E_Σ be a $r \times r$ matrix with entries i.i.d. to $\mathcal{N}(0, 18\Delta^{(2)^2}\log(\frac{3.75}{\delta})/\varepsilon^2)$, then

$$\|E_\Sigma\| \le \widetilde{C}_4 \cdot \frac{r}{n\varepsilon}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right), \tag{S.9}$$

for some absolute constant $\widetilde{C}_4 > 0$ with probability at least 10^{-20r}. Moreover, by Gaussian mechanism, $\widetilde{\Sigma} = \widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma$ is $(\varepsilon/3, \delta/3)$-differentially private. Thanks to the composition property of differential privacy, the output of Algorithm 1 $\widetilde{M}_0 = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$, is (ε, δ)-differentially private.

定理，对于某个绝对常数 $\tilde{c}_0 > 0$

$$\begin{aligned}\left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\| &\le 1 \wedge \left(\left\|\widehat{U}\widehat{U}^\top - UU^\top\right\| + \|E_U\|\right) \overset{(a)}{\le} 1 \wedge \left(\left(\frac{8}{\sigma_r}\|\widehat{L} - M\| \wedge 1\right) + \|E_U\|\right) \\ &\le 1 \wedge \tilde{c}_0\left(\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}\right. \\ &\quad \left. + \sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\frac{\sqrt{r d_1}}{n\varepsilon}\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)\right),\end{aligned} \tag{S.7}$$

我们在哪里应用引理 4在步骤（a）。

设 $E_V \in \mathbb{R}^{d_2 \times d_2}$ 是一个对称矩阵，其中 $(E_V)_{ij}$ 独立同分布于 $\mathcal{N}(0, \frac{18\Delta^{(1)^2}}{\varepsilon^2}\log(\frac{3.75}{\delta}))$ 对于 $1 \le i \le j \le d_2$，那么对于某个绝对常数 $\widetilde{C}_4 > 0$，在事件 $\mathcal{E}_*$ 和 (S.2) 的条件下，我们有 $\|E_V\| \le \widetilde{C}_4 \frac{\sqrt{r d_2}}{n\varepsilon}\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})$，概率至少为 $1 - e^{-d_2}$。此外，通过高斯机制，$\widehat{V}\widehat{V}^\top + E_U$ 是 $(\varepsilon/3, \delta/3)$-DP，并且 $\widetilde{V}$ 也通过差分隐私的后处理属性是 $(\varepsilon/3, \delta/3)$-DP。此外，根据 Davis-Kahan 定理，对于某个绝对常数 $\tilde{c}_0 > 0$

$$\begin{aligned}\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| &\le 1 \wedge \tilde{c}_0\left(\sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}}\right. \\ &\quad \left. + \sqrt{C_l^{-1}}\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)\frac{\sqrt{r d_2}}{n\varepsilon}\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)\right).\end{aligned} \tag{S.8}$$

A.3 第3部分

给定 $(\varepsilon/3, \delta/3)$-DP奇异向量 $\widetilde{U} \in \mathbb{O}^{d_1 \times r}$和 $\widetilde{V} \in \mathbb{O}^{d_2 \times r}$在A.2部分获得的A.2，我们推导出灵敏度 $\Delta^{(2)} := \max_{i \in [n]} \left\|\widetilde{U}^\top \widehat{L} - \left(\widehat{L}^{(i)}\right)\widetilde{V}\right\|_F$.在事件 $\mathcal{E}_*$条件下，

$$\begin{aligned}\Delta^{(2)} &:= \max_{i\in[n]}\left\|\widetilde{U}^\top\left(\widehat{L} - \widehat{L}^{(i)}\right)\widetilde{V}\right\|_F \le \max_{i\in[n]}\sqrt{r}\left\|\widehat{L} - \widehat{L}^{(i)}\right\| \\ &= \max_{i\in[n]}\sqrt{r}\left\|\widehat{L} - M + M - \widehat{L}^{(i)}\right\| = \max_{i\in[n]}\sqrt{r}\left\|\Delta - \Delta^{(i)}\right\| \\ &\le C_3 \cdot n^{-1}\sqrt{C_l^{-1}}\sqrt{r}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n.\end{aligned}$$

设 E_Σ是一个元素独立同分布于 $\mathcal{N}(0, 18\Delta^{(2)^2}\log(\frac{3.75}{\delta})/\varepsilon^2)$的 $r \times r$矩阵，则

$$\|E_\Sigma\| \le \widetilde{C}_4 \cdot \frac{r}{n\varepsilon}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n \log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right), \tag{S.9}$$

对于某个绝对常数 $\widetilde{C}_4 > 0$ 以至少 10^{-20r}的概率。此外，通过高斯机制，$\widetilde{\Sigma} = \widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma$是 $(\varepsilon/3, \delta/3)$-差分隐私的。由于差分隐私的组合性质，算法 1 $\widetilde{M}_0 = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$的输出， 是 (ε, δ)-差分隐私的。

A.4 Part 4

In this part, we derive the upper bound of $\|\widetilde{M}_0 - M\|$. Note that

$$\|\widetilde{M}_0 - M\| = \left\|\widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top - U\Sigma V^\top\right\| = \left\|\widetilde{U}(\widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma)\widetilde{V}^\top - UU^\top MVV^\top\right\|,$$

is a $d_1 \times d_2$ matrix of rank at most $2r$. Since

$$\begin{aligned}
&\left\|\widetilde{U}(\widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma)\widetilde{V}^\top - UU^\top MVV^\top\right\| \le \left\|\widetilde{U}\widetilde{U}^\top\widehat{L}\widetilde{V}\widetilde{V}^\top - UU^\top MVV^\top\right\| + \left\|\widetilde{U}E_\Sigma\widetilde{V}^\top\right\| \\
&\le \left\|\left(\widetilde{U}\widetilde{U}^\top - UU^\top\right)\widehat{L}\widetilde{V}\widetilde{V}^\top\right\| + \left\|UU^\top\left(\widehat{L} - M\right)\widetilde{V}\widetilde{V}^\top\right\| + \left\|UU^\top M\left(\widetilde{V}\widetilde{V}^\top - VV^\top\right)\right\| + \left\|\widetilde{U}E_\Sigma\widetilde{V}^\top\right\| \\
&\le \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\left\|\widehat{L}\right\| + \left\|\widehat{L} - M\right\| + \|M\|\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \|E_\Sigma\| \\
&\le \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\left\|\widehat{L} - M\right\| + \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\|M\| + \left\|\widehat{L} - M\right\| + \|M\|\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \|E_\Sigma\| \\
&\le 2\left\|\widehat{L} - M\right\| + \|M\|\left(\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\right) + \|E_\Sigma\|,
\end{aligned} \tag{S.10}$$

it is sufficient to plug in the upper bound (S.1) of $\left\|\widehat{L} - M\right\|$, $\|M\| = \sigma_1$, as well as the upper bounds (S.7) of $\|\widetilde{U}\widetilde{U}^\top - UU^\top\|$, (S.8) of $\|\widetilde{V}\widetilde{V}^\top - VV^\top\|$ and (S.9) of $\|E_\Sigma\|$. In conclusion, conditioned on the event $\mathcal{E}_*$, as the sample size

$$n \ge C_1 C_l^{-1} r\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)^2 (d_1 \vee d_2)\log(d_1 + d_2),$$

for some absolute constant $C_1 > 0$, we have

$$\begin{aligned}
\|\widetilde{M}_0 - M\| \le\ & C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\frac{\sigma_1}{\sigma_r}\sqrt{\frac{(d_1 \vee d_2)\log(d_1 + d_2)}{n}} \\
&+ C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\,\sigma_1 + \sigma_\xi\right)\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon}\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta}) \\
&+ C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\frac{r}{n\varepsilon}\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta}),
\end{aligned}$$

for some absolute constant $C_5 > 0$ with probability at least $1 - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$.

A.5 Proof of Corollary 1

The proof of Corollary 1 is obtained by setting the upper bound of $\|\widetilde{M}_0 - M\|_F \le \sqrt{2r}\|\widetilde{M}_0 - M\|$ given by Theorem 1 smaller than the order of σ_r.

B Proof of Theorem 2

We first present some preliminary results on the KL-divergence and total variation distance between Gaussian distributions. Let $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$ be two p-dimensional multivariate

A.4 第4部分

在本部分中，我们推导出 $\|\widetilde{M}_0 - M\|$ 的上界。请注意，

$$\|\widetilde{M}_0 - M\| = \left\|\widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top - U\Sigma V^\top\right\| = \left\|\widetilde{U}(\widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma)\widetilde{V}^\top - UU^\top MVV^\top\right\|,$$

$d_1 \times d_2$ 是一个秩至多为 $2r$ 的矩阵。由于，

$$\begin{aligned}
&\left\|\widetilde{U}(\widetilde{U}^\top\widehat{L}\widetilde{V} + E_\Sigma)\widetilde{V}^\top - UU^\top MVV^\top\right\| \le \left\|\widetilde{U}\widetilde{U}^\top\widehat{L}\widetilde{V}\widetilde{V}^\top - UU^\top MVV^\top\right\| + \left\|\widetilde{U}E_\Sigma\widetilde{V}^\top\right\| \\
&\le \left\|\left(\widetilde{U}\widetilde{U}^\top - UU^\top\right)\widehat{L}\widetilde{V}\widetilde{V}^\top\right\| + \left\|UU^\top\left(\widehat{L} - M\right)\widetilde{V}\widetilde{V}^\top\right\| + \left\|UU^\top M\left(\widetilde{V}\widetilde{V}^\top - VV^\top\right)\right\| + \left\|\widetilde{U}E_\Sigma\widetilde{V}^\top\right\| \\
&\le \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\left\|\widehat{L}\right\| + \left\|\widehat{L} - M\right\| + \|M\|\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \|E_\Sigma\| \\
&\le \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\left\|\widehat{L} - M\right\| + \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\|M\| + \left\|\widehat{L} - M\right\| + \|M\|\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \|E_\Sigma\| \\
&\le 2\left\|\widehat{L} - M\right\| + \|M\|\left(\left\|\widetilde{V}\widetilde{V}^\top - VV^\top\right\| + \left\|\widetilde{U}\widetilde{U}^\top - UU^\top\right\|\right) + \|E_\Sigma\|,
\end{aligned} \tag{S.10}$$

它足以将上界（S.1）代入 $\left\|\widehat{L} - M\right\|$ $\|M\| = \sigma_1$，以及上界（S.7）的 $\|\widetilde{U}\widetilde{U}^\top - UU^\top\|$，(S.8) 的 $\|\widetilde{V}\widetilde{V}^\top - VV^\top\|$和（S.9）的 $\|E_\Sigma\|$。总之，在事件 $\mathcal{E}_*$ 的条件下，随着样本量

$$n \ge C_1 C_l^{-1} r\left(\frac{\sqrt{C_u r}\,\sigma_1 + \sigma_\xi}{\sigma_r}\right)^2 (d_1 \vee d_2)\log(d_1 + d_2),$$

对于某个绝对常数 $C_1 > 0$，我们有，

$$\begin{aligned}
\|\widetilde{M}_0 - M\| \le\ & C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\frac{\sigma_1}{\sigma_r}\sqrt{\frac{(d_1 \vee d_2)\log(d_1 + d_2)}{n}} \\
&+ C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\,\sigma_1 + \sigma_\xi\right)\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon}\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta}) \\
&+ C_5\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\frac{r}{n\varepsilon}\log n\log^{\frac{1}{2}}(\frac{3.75}{\delta}),
\end{aligned}$$

对于某个绝对常数 $C_5 > 0$，以至少 $1 - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$ 的概率，

A.5 推论证明 1

推论证明 1是通过将 $\|\widetilde{M}_0 - M\|_F \le \sqrt{2r}\|\widetilde{M}_0 - M\|$由定理 1 的上界设置为 σ_r的阶数更小。

B 定理证明 2

我们首先介绍高斯分布之间KL散度与总变差距离的一些预备结果。设 $\mathcal{N}(\mu_1, \Sigma_1)$ 和 $\mathcal{N}(\mu_2, \Sigma_2)$ 是两个 p-维多元

Gaussians, then

$$\begin{aligned}&\mathrm{KL}\left(\mathcal{N}\left(\mu_1,\Sigma_1\right)\|\mathcal{N}\left(\mu_2,\Sigma_2\right)\right)\\&=\frac{1}{2}\left(\mathrm{Tr}\left(\Sigma_2^{-1}\Sigma_1-I_p\right)+\left(\mu_2-\mu_1\right)^\top\Sigma_2^{-1}\left(\mu_2-\mu_1\right)+\log\left(\frac{\det\Sigma_2}{\det\Sigma_1}\right)\right).\end{aligned}$$

Let $M_*=\sigma HH^\top$ and $M_*'=\sigma H'H'^\top$ where $H\neq H'\in\mathcal{S}_q^{(d_1+d_2)}$, then

$$\begin{aligned}&\mathrm{KL}\left(f_{M_*}(\bar{y}_i,X_i,X_i')\|f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\\&=\mathbb{E}_{f_{M_*}(\bar{y}_i,X_i,X_i')}\left[\log\frac{f_{M_*}(\bar{y}_i,X_i,X_i')}{f_{M_*'}(\bar{y}_i,X_i,X_i')}\right]\\&=\mathbb{E}_{X_i,X_i'}\mathbb{E}_{f_{M_*}(\bar{y}_i|X_i,X_i')}\left[-\frac{\left(\bar{y}_i-\left\langle\begin{pmatrix}0&M\\M^\top&0\end{pmatrix},\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\right)^2}{4\sigma_\xi^2}\right]\\&\quad+\mathbb{E}_{X_i,X_i'}\mathbb{E}_{f_{M_*}(\bar{y}_i|X_i,X_i')}\left[-\frac{\left(\bar{y}_i-\left\langle\begin{pmatrix}0&M'\\M'^\top&0\end{pmatrix},\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\right)^2}{4\sigma_\xi^2}\right]\\&=\mathbb{E}_{X_i,X_i'}\left[\frac{\left\langle M_*-M_*',\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\cdot\left\langle M_*-M_*',\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle}{4\sigma_\xi^2}\right]\\&\lesssim\frac{1}{\sigma_\xi^2}C_u\|M_*-M_*'\|_F^2\lesssim\frac{1}{\sigma_\xi^2}C_u\sigma^2r\varepsilon_0^2,\end{aligned}$$

and further by Pinsker's inequality, we have

$$\mathrm{TV}\left(f_{M_*}(\bar{y}_i,X_i,X_i'),f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\lesssim\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0.$$

For any probability measures $P_{M_*}\neq P_{M_*'}\in\mathcal{P}_\sigma$, we have

$$\mathrm{KL}\left(\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i,X_i,X_i')\|\bigotimes_{i=1}^n f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\lesssim\frac{n}{\sigma_\xi^2}C_u\sigma^2r\varepsilon_0^2.$$

and

$$\sum_{k\in[n]}\mathrm{TV}\left(f_{M_*}(\bar{y}_i,X_i,X_i'),f_{M_*'}(\bar{y}_i,(X_i^{\perp\!\!\!\perp})_*)\right)\lesssim n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0,$$

高斯分布，然后

$$\begin{aligned}&\mathrm{KL}\left(\mathcal{N}\left(\mu_1,\Sigma_1\right)\|\mathcal{N}\left(\mu_2,\Sigma_2\right)\right)\\&=\frac{1}{2}\left(\mathrm{Tr}\left(\Sigma_2^{-1}\Sigma_1-I_p\right)+\left(\mu_2-\mu_1\right)^\top\Sigma_2^{-1}\left(\mu_2-\mu_1\right)+\log\left(\frac{\det\Sigma_2}{\det\Sigma_1}\right)\right).\end{aligned}$$

设 $M_*=\sigma HH^\top$ 和 $M_*'=\sigma H'H'^\top$，其中 $H\neq H'\in\mathcal{S}_q^{(d_1+d_2)}$，则

$$\begin{aligned}&\mathrm{KL}\left(f_{M_*}(\bar{y}_i,X_i,X_i')\|f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\\&=\mathbb{E}_{f_{M_*}(\bar{y}_i,X_i,X_i')}\left[\log\frac{f_{M_*}(\bar{y}_i,X_i,X_i')}{f_{M_*'}(\bar{y}_i,X_i,X_i')}\right]\\&=\mathbb{E}_{X_i,X_i'}\mathbb{E}_{f_{M_*}(\bar{y}_i|X_i,X_i')}\left[-\frac{\left(\bar{y}_i-\left\langle\begin{pmatrix}0&M\\M^\top&0\end{pmatrix},\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\right)^2}{4\sigma_\xi^2}\right]\\&\quad+\mathbb{E}_{X_i,X_i'}\mathbb{E}_{f_{M_*}(\bar{y}_i|X_i,X_i')}\left[-\frac{\left(\bar{y}_i-\left\langle\begin{pmatrix}0&M'\\M'^\top&0\end{pmatrix},\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\right)^2}{4\sigma_\xi^2}\right]\\&=\mathbb{E}_{X_i,X_i'}\left[\frac{\left\langle M_*-M_*',\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle\cdot\left\langle M_*-M_*',\begin{pmatrix}0&X_i\\X_i'^\top&0\end{pmatrix}\right\rangle}{4\sigma_\xi^2}\right]\\&\lesssim\frac{1}{\sigma_\xi^2}C_u\|M_*-M_*'\|_F^2\lesssim\frac{1}{\sigma_\xi^2}C_u\sigma^2r\varepsilon_0^2,\end{aligned}$$

并且根据 Pinsker 不等式，我们有

$$\mathrm{TV}\left(f_{M_*}(\bar{y}_i,X_i,X_i'),f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\lesssim\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0.$$

对于任何概率测度 $P_{M_*}=P_{M_*'}\in\mathcal{P}_\sigma$，我们有

$$\mathrm{KL}\left(\bigotimes_{i=1}^n f_{M_*}(\bar{y}_i,X_i,X_i')\|\bigotimes_{i=1}^n f_{M_*'}(\bar{y}_i,X_i,X_i')\right)\lesssim\frac{n}{\sigma_\xi^2}C_u\sigma^2r\varepsilon_0^2.$$

and

$$\sum_{k\in[n]}\mathrm{TV}\left(f_{M_*}(\bar{y}_i,X_i,X_i'),f_{M_*'}(\bar{y}_i,(X_i^{\perp\!\!\!\perp})_*)\right)\lesssim n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0,$$

The next Lemma 6 states that there exists a sufficiently large subsect of $\mathbb{O}_{d_1+d_2,2r}$ such that the elements in the subsets are well separated.

Lemma 6. *For any $r \le d$, there exists a subset $\mathcal{S}_q^{(d)} \subset \mathbb{O}_{d,r}$ with cardinality $\left|\mathcal{S}_q^{(d)}\right| \ge 2^{r(d-r)}$ such that for any $U_i \ne U_j \in \mathcal{S}_q^{(d)}$,*

$$\|U_iU_i^\top - U_jU_j^\top\|_q \ge \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_i - V_j\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_iV_i^\top - V_jV_j^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}r^{1/q}$$

where $\|\cdot\|_q$ denotes the Schatten-q norm and, meanwhile,

$$\|U_iU_i^\top - U_jU_j^\top\|_{\mathrm{F}} \lesssim \|U_i - U_j\|_{\mathrm{F}} \le \varepsilon_0\|V_i - V_j\|_{\mathrm{F}} \le \sqrt{2r}\varepsilon_0.$$

By Lemma 6, there exists a subset $\mathcal{S}_q^{(d_1+d_2)} \subset \mathbb{O}_{d_1+d_2,2r}$ with cardinality $\left|\mathcal{S}_q^{(d_1+d_2)}\right| \ge 2^{2r(d_1+d_2-2r)}$ such that for any $H \ne H' \in \mathcal{S}_q^{(d_1+d_2)}$,

$$\|HH^\top - H'H'^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}(2r)^{1/q},$$

and meanwhile,

$$\|HH^\top - H'H'^\top\|_{\mathrm{F}} \lesssim 2\sqrt{r}\varepsilon_0.$$

To invoke Lemma 3, we define the metric $\rho : \mathbb{O}_{d_1+d_2,2r} \times \mathbb{O}_{d_1+d_2,2r} \mapsto \mathbb{R}^+$ as $\rho(H,H') := \|HH^\top - H'H'^\top\|_q$ for any $q \in [1,\infty]$ and take $\rho_0 \asymp \tau\varepsilon_0 r^{1/q}$,

$$l_0 = c_0\frac{n}{\sigma_\xi^2}C_u\sigma^2 r\varepsilon_0^2 \quad \text{and} \quad t_0 = c_0 n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0,$$

for some small absolute constant $c_0, \tau > 0$. Then, by Lemma 3, for any (ε,δ)-DP estimator $\widetilde{H}$,

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q$$
$$\geqslant \max\left\{\frac{\tau\varepsilon_0 r^{1/q}}{2}\left(1 - \frac{c_0\frac{n}{\sigma_\xi^2}C_u\sigma^2 r\varepsilon_0^2 + \log 2}{\log N}\right), \frac{\tau\varepsilon_0 r^{1/q}}{4}\left(1 \wedge \frac{N-1}{\exp\left(4\varepsilon c_0 n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0\right)}\right)\right\}.$$

Recall that $N \ge 2^{2r(d_1+d_2)/2}$ if $d_1 + d_2 \ge 4r$. We can take

$$\varepsilon_0 \asymp \frac{\sigma_\xi}{\sqrt{C_u}\sigma}\left(\sqrt{\frac{d_1+d_2}{n}} + r^{\frac{1}{2}}\frac{d_1+d_2}{n\varepsilon}\right),$$

to get

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}\sigma}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}} + r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right) \bigwedge r^{1/q},$$

下一个引理 6 指出，存在一个足够大的子集 $\mathbb{O}_{d_1+d_2\ 2r}$,使得子集中的元素相互分离。

引理 6。对于任何 $r \le d$，存在一个子集 $\mathcal{S}_q^{(d)} \subset \mathbb{O}_{d\,r}$,其基数 $\left|\mathcal{S}_q^{(d)}\right| \ge 2^{r(d-r)}$使得对于任何 $U_i = U_j \in \mathcal{S}_q^{(d)}$, /

$$\|U_iU_i^\top - U_jU_j^\top\|_q \ge \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_i - V_j\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_iV_i^\top - V_jV_j^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}r^{1/q}$$

其中 $\|\cdot\|_q$ 表示 Schatten-q 范数，同时，

$$\|U_iU_i^\top - U_jU_j^\top\|_{\mathrm{F}} \lesssim \|U_i - U_j\|_{\mathrm{F}} \le \varepsilon_0\|V_i - V_j\|_{\mathrm{F}} \le \sqrt{2r}\varepsilon_0.$$

根据引理 6，存在一个子集 $\mathcal{S}_q^{(d_1+d_2)} \subset \mathbb{O}_{d_1+d_2\ 2r}$,其基数 $\left|\mathcal{S}_q^{(d_1+d_2)}\right| \ge 2^{2r(d_1+d_2-2r)}$使得对于任何 $H = H' \in \mathcal{S}_q^{(d_1+d_2)}$ /,

$$\|HH^\top - H'H'^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}(2r)^{1/q},$$

与此同时，

$$\|HH^\top - H'H'^\top\|_{\mathrm{F}} \lesssim 2\sqrt{r}\varepsilon_0.$$

要调用引理 3，我们定义度量 ρ： $\mathbb{O}_{d_1+d_2,2r} \times \mathbb{O}_{d_1+d_2,2r} \to \mathbb{R}^+$ 为 $\rho(H,H') := \|HH^\top - H'H'^\top\|_q$对任何 $q \in [1,\infty]$ 并取 $\rho_0 \asymp \tau\varepsilon_0 r^{1/q}$,

$$l_0 = c_0\frac{n}{\sigma_\xi^2}C_u\sigma^2 r\varepsilon_0^2 \quad \text{and} \quad t_0 = c_0 n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0,$$

对于某个小的绝对常数 $c_0, \tau > 0$。然后，根据引理 3，对于任何 (ε,δ)-DP估计器 $\widetilde{H}$,

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q$$
$$\geqslant \max\left\{\frac{\tau\varepsilon_0 r^{1/q}}{2}\left(1 - \frac{c_0\frac{n}{\sigma_\xi^2}C_u\sigma^2 r\varepsilon_0^2 + \log 2}{\log N}\right), \frac{\tau\varepsilon_0 r^{1/q}}{4}\left(1 \wedge \frac{N-1}{\exp\left(4\varepsilon c_0 n\frac{\sqrt{C_u}\sigma}{\sigma_\xi}\sqrt{r}\varepsilon_0\right)}\right)\right\}.$$

回想一下 $N \ge 2^{2r(d_1+d_2)/2}$ 如果 $d_1 + d_2 \ge 4r$。我们可以取

$$\varepsilon_0 \asymp \frac{\sigma_\xi}{\sqrt{C_u}\sigma}\left(\sqrt{\frac{d_1+d_2}{n}} + r^{\frac{1}{2}}\frac{d_1+d_2}{n\varepsilon}\right),$$

以得到

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}\sigma}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}} + r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right) \bigwedge r^{1/q},$$

where the last term is due to a trivial upper of $\|\widetilde{H}\widetilde{H}^\top - HH^\top\|_q \le (4r)^{1/q}$. Since σ is already known, it suffices to estimate $HH^\top$ differentially privately by the estimator $\widetilde{H}\widetilde{H}^\top$, and an estimator $(\widetilde{M})_*$ for the matrix M_* is given by $\sigma\widetilde{H}\widetilde{H}^\top$. Therefore,

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \ge \sigma\cdot\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q.$$

Due to $\mathcal{P}_\sigma\subset\mathcal{P}_\Sigma$, we have

$$\sup_{P\in\mathcal{P}_\Sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \ge \sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}}+r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right)\bigwedge r^{1/q}\sigma.$$

There is a one-to-one mapping between $\widetilde{M}$ and $(\widetilde{M})_*$. Let $\widetilde{M}-M=U_\Delta\Sigma_\Delta V_\Delta^\top$, then $\|\widetilde{M}-M\|_q^q=\|\Sigma_\Delta\|_q^q$. Note that

$$(\widetilde{M})_*-M_*=\begin{pmatrix}0 & \widetilde{M}-M\\ \widetilde{M}^\top-M^\top & 0\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix}U_\Delta & U_\Delta\\ V_\Delta & -V_\Delta\end{pmatrix}\begin{pmatrix}\Sigma_\Delta & 0\\ 0 & \Sigma_\Delta\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}U_\Delta & U_\Delta\\ V_\Delta & -V_\Delta\end{pmatrix}^\top,$$

and

$$\left\|(\widetilde{M})_*-M_*\right\|_q^q=\left\|\begin{pmatrix}\Sigma_A & 0\\ 0 & \Sigma_a\end{pmatrix}\right\|_q^q=2\left\|\Sigma_\Delta\right\|_q^q=2\|\widetilde{M}-M\|_q^q.$$

Therefore, $\left\|(\widetilde{M})_*-M*\right\|_q=2^{1/q}\|\widetilde{M}-M\|_q$ and

$$\inf_{\widetilde{M}}\sup_{P\in\mathcal{P}_\Sigma}\mathbb{E}\left\|\widetilde{M}-M\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}}+r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right)\bigwedge r^{1/q}\sigma, \tag{S.11}$$

where we use the fact $d_1+d_2\lesssim d_1\vee d_2$ and infimum is taken over all possible (ε,δ)-DP algorithms.

C Proof of Theorem 3

In Appendix C, we aim to prove Theorem 3. The proof is composed of three Parts. In Part C.1, we characterize the sensitivity Δ_l for iterations $l=1,\cdots,l^*$ and bound $\|\mathcal{P}_{\mathbb{T}_l}N_l\|_F$. In Part C.2, take mathematical induction to prove that if the RIP-condition holds and both Δ_l and $\|M_l-M\|_F$ are bounded with high probability, then we also have Δ_{l+1} and $\|M_{l+1}-M\|_F$ are bounded with high probability. In Part C.3, we choose an appropriate l^* as the total number of iterations and give the convergence rate of $\|\widetilde{M}_{l^*}-M\|_F$.

最后一个项是由于 $\|\widetilde{H}\widetilde{H}^\top - HH^\top\|_q \le (4r)^{1/q}$ 的平凡上界。由于 σ 已经已知，只需通过估计器 $\widetilde{H}\widetilde{H}^\top$ 差分隐私地估计 $HH^\top$ ，并且给出了矩阵 M_* 的估计器（$\widetilde{M}$）$_*$，因此，

$$\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \ge \sigma\cdot\sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{H}\widetilde{H}^\top - HH^\top\right\|_q.$$

由于 $\mathcal{P}_\sigma\subset\mathcal{P}_\Sigma$，我们有

$$\sup_{P\in\mathcal{P}_\Sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \ge \sup_{P\in\mathcal{P}_\sigma}\mathbb{E}\left\|\widetilde{M}_*-M_*\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}}+r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right)\bigwedge r^{1/q}\sigma.$$

在 $\widetilde{M}$ 和（$\widetilde{M}$）$_*$之间存在一一对应关系。令 $\widetilde{M}-M=U_\Delta\Sigma_\Delta V_\Delta^\top$，则 $\|\widetilde{M}-M\|_q^q=\|\Sigma_\Delta\|_q^q$。注意

$$(\widetilde{M})_*-M_*=\begin{pmatrix}0 & \widetilde{M}-M\\ \widetilde{M}^\top-M^\top & 0\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix}U_\Delta & U_\Delta\\ V_\Delta & -V_\Delta\end{pmatrix}\begin{pmatrix}\Sigma_\Delta & 0\\ 0 & \Sigma_\Delta\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}U_\Delta & U_\Delta\\ V_\Delta & -V_\Delta\end{pmatrix}^\top,$$

and

$$\left\|(\widetilde{M})_*-M_*\right\|_q^q=\left\|\begin{pmatrix}\Sigma_A & 0\\ 0 & \Sigma_a\end{pmatrix}\right\|_q^q=2\left\|\Sigma_\Delta\right\|_q^q=2\|\widetilde{M}-M\|_q^q.$$

因此，$\left\|(\widetilde{M})_*-M*\right\|_q=2^{1/q}\|\widetilde{M}-M\|_q$ 和

$$\inf_{\widetilde{M}}\sup_{P\in\mathcal{P}_\Sigma}\mathbb{E}\left\|\widetilde{M}-M\right\|_q \gtrsim \frac{\sigma_\xi}{\sqrt{C_u}}\left(r^{1/q}\sqrt{\frac{d_1+d_2}{n}}+r^{\frac{1}{2}+\frac{1}{q}}\frac{d_1+d_2}{n\varepsilon}\right)\bigwedge r^{1/q}\sigma, \tag{S.11}$$

其中我们使用了事实 $d_1+d_2\lesssim d_1\vee d_2$，并且下确界是在所有可能的 (ε,δ)-DP 算法上取的。

C 定理证明3

在附录 C，我们旨在证明定理 3。证明由三个部分组成。在部分C.1，我们刻画了敏感性 Δ_l 对于迭代 $l=1\cdots l^*$， ，并界定了 $\|\mathcal{P}_{\mathbb{T}_l}N_l\|_F$。在部分C.2，使用数学归纳法证明，如果 RIP 条件成立且 和 Δ_l 和$\|M_l-M\|_F$以高概率有界，那么我们也有 和 Δ_{l+1}和 $\|M_{l+1}-M\|_F$ 以高概率有界。在部分 C.3，我们选择一个合适的 l^* 作为迭代总数，并给出 的收敛速度。

C.1 Part 1

In Part C.1, we focus on upper bounding $\|\mathcal{P}_{\mathbb{T}_l} N_l\|_F$. The first step is to characterize the sensitivity of the l-th iteration for $l \in [l^*]$. Let

$$G_l^{(i)} := \frac{1}{n}\sum_{j\neq i}\left(\langle X_j, M_l\rangle - y_j\right) X_j + \frac{1}{n}\left[\langle X_i', M_l\rangle - y_i'\right] X_i',$$

which is the gradient of l-th iteration obtained by the dataset differes with the original one only by the i-th pair of observation. The sensitivity of the gradient descent on the tagent space $\mathbb{T}_l$ is

$$\begin{aligned}\Delta_l &:= \max_{\text{neighbouring}(Z,Z')} \|M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z)) - [M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z'))]\|_F \\ &= \max_{i\in[n]} \left\| M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) - \left[M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l^{(i)})\right]\right\|_F.\end{aligned}$$

By the definition of G_l and $G_l^{(i)}$,

$$\Delta_l \le \frac{\eta}{n}\max_{i\in[n]}\left[\left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i, M_l\rangle - y_i\right) X_i\right\|_F + \left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i', M_l\rangle - y_i'\right) X_i'\right\|_F\right],$$

where for all $i \in [n]$ and $l+1 \in [l^*]$

$$\begin{aligned}\left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i, M_l\rangle - y_i\right) X_i\right\|_F &\le \left|\langle X_i, M_l - M\rangle - \xi_i\right| \left\|\mathcal{P}_{\mathbb{T}_l} X_i\right\|_F \\ &\le \left(|\xi_i| + |\langle X_i, M_l - M\rangle|\right)\sqrt{2r}\,\|X_i\|.\end{aligned} \tag{S.12}$$

Here, the last inequality uses the fact that for any $B \in \mathbb{R}^{d_1\times d_2}$, the matrix $\mathcal{P}_{\mathbb{T}_l} B$ is of rank at most $2r$. Since both ξ_i and $\langle M_l - M, X_i\rangle$ are sub-Gaussians with $\|\xi_i\|_{\psi_2} = \sigma_\xi$ and $\|\langle M_l - M, X_i\rangle\|_{\psi_2} \le \|M_l - M\|_F\sqrt{C_u}$, we turn to Lemma 7 to upper bound $|\xi_i| + |\langle X_i, M_l - M\rangle|$.

Lemma 7 (Vershynin (2018)). *For any sub-Gaussian random variable $B \in \mathbb{R}$,*

$$\mathbb{P}(|B| \ge t) \le 2\exp\left(-ct^2/\|B\|_{\psi_2}\right), \quad \forall t \ge 0.$$

According to the tail probability of sub-Gaussian random variable stated in Lemma 7, we have with probability at least $1 - n^{-10}$,

$$|\xi_i| + |\langle X_i, M_l - M\rangle| \le C_1(\sigma_\xi + \|M_l - M\|_F\sqrt{C_u})\log^{1/2} n, \tag{S.13}$$

for some absolute constant $C_1 > 0$. Shen et al. (2023) offeres the folowing result on $\|X_i\|$.

Lemma 8 (Shen et al. (2023)). *Suppose the vectorization of $X \in \mathbb{R}^{d_1\times d_2}$ follows mean zero multivariate Gaussian distribution $N(\mathbf{0}, \Lambda)$ where $\Lambda \in \mathbb{R}^{d_1d_2\times d_1d_2}$ satisfies $\lambda_{\max}(\Lambda) \le C_u$. Then, for some constant $c > 0$*

$$\mathbb{P}\left(\|X\| \ge t + c\sqrt{C_u(d_1+d_2)}\right) \le \exp\left(-t^2\right).$$

C.1 第一部分

在第一部分 C.1，我们关注上界 $\|\mathcal{P}_{\mathbb{T}_l} N_l\|_F$。第一步是描述第 l-次迭代的敏感性对于 $l \in [l^*]$。令

$$G_l^{(i)} := \frac{1}{n}\sum_{j\neq i}\left(\langle X_j, M_l\rangle - y_j\right) X_j + \frac{1}{n}\left[\langle X_i', M_l\rangle - y_i'\right] X_i',$$

这是通过数据集获得的第 l-次迭代的梯度，与原始梯度仅在第 i-对观测值上不同。梯度下降在切空间 $\mathbb{T}_l$ 上的敏感性是

$$\begin{aligned}\Delta_l &:= \max_{\text{neighbouring}(Z,Z')} \|M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z)) - [M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l(Z'))]\|_F \\ &= \max_{i\in[n]} \left\| M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) - \left[M_l - \eta\mathcal{P}_{\mathbb{T}_l}(G_l^{(i)})\right]\right\|_F.\end{aligned}$$

根据 G_l 和 $G_l^{(i)}$ 的定义，

$$\Delta_l \le \frac{\eta}{n}\max_{i\in[n]}\left[\left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i, M_l\rangle - y_i\right) X_i\right\|_F + \left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i', M_l\rangle - y_i'\right) X_i'\right\|_F\right],$$

其中对于所有 $i \in [n]$ 和 $l+1 \in [l^*]$

$$\begin{aligned}\left\|\mathcal{P}_{\mathbb{T}_l}\left(\langle X_i, M_l\rangle - y_i\right) X_i\right\|_F &\le \left|\langle X_i, M_l - M\rangle - \xi_i\right| \left\|\mathcal{P}_{\mathbb{T}_l} X_i\right\|_F \\ &\le \left(|\xi_i| + |\langle X_i, M_l - M\rangle|\right)\sqrt{2r}\,\|X_i\|.\end{aligned} \tag{S.12}$$

这里，最后一个不等式使用了这样的事实：对于任何 $B \in \mathbb{R}^{d_1\times d_2}$，矩阵 $\mathcal{P}_{\mathbb{T}_l} B$的秩至多为$2r$。由于 ξ_i和 $\langle M_l - M\ X_i\rangle$,都是具有 $\|\xi_i\|_{\psi_2} = \sigma_\xi$和 $\|\langle M_l - M\ X_i\rangle\|_{\psi_2} \le$的次高斯分布，$\|M_l - M\|_F\sqrt{C_u}$，我们转向引理 7来上界 $|\xi_i| + |\langle X_i\ M_l - M\rangle|$。。

引理 7 (Vershynin (2018)). 对于任何次高斯随机变量 $B \in \mathbb{R}$,

$$\mathbb{P}(|B| \ge t) \le 2\exp\left(-ct^2/\|B\|_{\psi_2}\right), \quad \forall t \ge 0.$$

根据引理 7 中关于次高斯随机变量尾部概率的陈述，我们得到以至少 7的概率， $1 - n^{-10}$,

$$|\xi_i| + |\langle X_i, M_l - M\rangle| \le C_1(\sigma_\xi + \|M_l - M\|_F\sqrt{C_u})\log^{1/2} n, \tag{S.13}$$

对于某个绝对常数 $C_1 > 0$. Shen 等人 (2023) 在 $\|X_i\|$上提出了以下结果。

引理 8 (Shen 等人 (2023)). 假设 $X \in \mathbb{R}^{d_1\times d_2}$ 的向量化服从均值为零的多变量高斯分布 $N(\mathbf{0}, \Lambda)$，其中 $\Lambda \in \mathbb{R}^{d_1d_2\times d_1d_2}$满足 $\lambda_{\max}(\Lambda) \le C_u$。那么，对于某个常数 $c > 0$

$$\mathbb{P}\left(\|X\| \ge t + c\sqrt{C_u(d_1+d_2)}\right) \le \exp\left(-t^2\right).$$

It implies $\|\|\|X\|\|\|_{\psi_2} \leq c_1\sqrt{C_u(d_1+d_2)}$ *and* $\|\|\|X\|\|\|_{\psi_1} \leq c_2\sqrt{C_u(d_1+d_2)}$ *for some constants* $c_1, c_2 > 0$.

Thus, for some absolute constant $C_2 > 0$, with probability at least $1-\exp\left(-10C_u(d_1+d_2)\right)n^{-10}$

$$\|X_i\| \leq C_2\sqrt{C_u(d_1+d_2)+\log n}. \tag{S.14}$$

Combining (S.12), (S.13) and (S.14) and taking maximum over n, for some constant $C_3 > 0$, we have the event

$$\mathcal{E}'_{\Delta_l} := \left\{\Delta_l \leq C_3\frac{\eta}{n}(\sigma_\xi + \|M_l - M\|_F\sqrt{C_u})\sqrt{C_u r(d_1+d_2+\log n)\log n}\right\}, \tag{S.15}$$

happens with probability at least $1-n^{-9}-\exp\left(-10C_u(d_1+d_2)\right)n^{-9}$. In the event $\mathcal{E}'_{\Delta_l}$ stated in (S.15), the sensitivity Δ_l still relies on $\|M_l - M\|_F$. To get an upper bound irrelevant with l, we take condition on the event

$$\mathcal{E}_l = \left\{\|M_l - M\|_F \leq c_0\sigma_r\right\},$$

and obtain that for some absolute constant $\widetilde{C}_3 > 0$, the event

$$\mathcal{E}_{\Delta_l} := \left\{\Delta_l \leq \widetilde{C}_3\frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2+\log n)\log n}\right\}, \tag{S.16}$$

happens with the probability $\mathbb{P}(\mathcal{E}_{\Delta_l}) \geq 1-n^{-9}-\exp\left(-10C_u(d_1+d_2)\right)n^{-9}$.

In the l+1-th iteration of Algorithm 2, the matrix M_l and operator $\mathcal{P}_{\mathbb{T}_l}$ are known. Moreover, the rank r approximation SVD_r is irrelevant with the data set $Z = \{(X_i, y_i)\}_{i=1}^n$. Thanks to the post-processing property and composition property of differential privacy, we only need to guarantee that $M_l - \eta_l\mathcal{P}_{\mathbb{T}_l}(G_l)$ is $(\varepsilon/l^*, \delta/l^*)$-DP where the gradient

$$G_l = \frac{1}{n}\sum_{i=1}^n\left(\langle X_i, M_l\rangle - y_i\right)X_i,$$

is the only component depends on the data set Z. Let N_l be a $d_1 \times d_2$ matrix with entries i.i.d. to normal distribution with varicance $\frac{l^{*2}\Delta_l^2}{\varepsilon^2}\log\left(\frac{1.25l^*}{\delta}\right)$. Under the condition that $d_1+d_2 \gtrsim \log n$ and conditioned on the event $\mathcal{E}_{\Delta_l}\cap\mathcal{E}_l$, we have for some constant $C_4 > 0$

$$\|\mathcal{P}_{\mathbb{T}_l}N_l\|_F \leq C_4\eta l^*\frac{r(d_1+d_2)}{n\varepsilon}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u\log n}\log^{1/2}\left(\frac{1.25l^*}{\delta}\right),$$

with probability at least $1-\exp(-(d_1+d_2))$.

这意味着对于某些常数c_1 $c_2 > 0$， $\|\|\|X\|\|\|_{\psi_2} \leq c_1\sqrt{C_u(d_1+d_2)}$ 和

$\|\|\|X\|\|\|_{\psi_1} \leq c_2\sqrt{C_u(d_1+d_2)}$。

因此，对于某些绝对常数 $C_2 > 0$，以至少 $1-\exp\left(-10C_u(d_1+d_2)\right)n^{-10}$的概率

$$\|X_i\| \leq C_2\sqrt{C_u(d_1+d_2)+\log n}. \tag{S.14}$$

结合（S.12），（S.13）和（S.14）并取最大值在 n，对于某个常数 $C_3 > 0$，我们有事件

$$\mathcal{E}'_{\Delta_l} := \left\{\Delta_l \leq C_3\frac{\eta}{n}(\sigma_\xi + \|M_l - M\|_F\sqrt{C_u})\sqrt{C_u r(d_1+d_2+\log n)\log n}\right\}, \tag{S.15}$$

以至少 $1-n^{-9}-\exp\left(-10C_u(d_1+d_2)\right)n^{-9}$。在（S.15）中陈述的事件 $\mathcal{E}'_{\Delta_l}$，敏感性 Δ_l仍然依赖于 $\|M_l - M\|_F$。为了得到与l无关的上界，我们在事件上施加条件

$$\mathcal{E}_l = \left\{\|M_l - M\|_F \leq c_0\sigma_r\right\},$$

并得到对于某个绝对常数 $\widetilde{C}_3 > 0$，事件

$$\mathcal{E}_{\Delta_l} := \left\{\Delta_l \leq \widetilde{C}_3\frac{\eta}{n}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u r(d_1+d_2+\log n)\log n}\right\}, \tag{S.16}$$

以概率 $\mathbb{P}(\mathcal{E}_{\Delta_l}) \geq 1-n^{-9}-\exp\left(-10C_u(d_1+d_2)\right)n^{-9}$. 发生

在 l+1-次迭代中，算法 2，矩阵 M_l 和算子 $\mathcal{P}_{\mathbb{T}_l}$ 是已知的。此外，秩 r 近似 SVD_r 与数据集 $Z = \{(X_i, y_i)\}_{i=1}^n$无关。归功于差分隐私的后处理特性和组合特性，我们只需要保证 $M_l - \eta_l\mathcal{P}_{\mathbb{T}_l}(G_l)$是 $(\varepsilon/l^*, \delta/l^*)$-DP，其中梯度

$$G_l = \frac{1}{n}\sum_{i=1}^n\left(\langle X_i, M_l\rangle - y_i\right)X_i,$$

是唯一依赖于数据集 Z 的分量。设 N_l 是一个元素独立同分布且服从方差为 $\frac{l^{*2}\Delta_l^2}{\varepsilon^2}\log\left(\frac{1.25l^*}{\delta}\right)$ 的正态分布的 $d_1 \times d_2$ 矩阵。在 $d_1+d_2 \gtrsim \log n$ 的条件下，且在事件 $\mathcal{E}_{\Delta_l}\cap\mathcal{E}_l$ 的条件下，对于某个常数 $C_4 > 0$，我们有

$$\|\mathcal{P}_{\mathbb{T}_l}N_l\|_F \leq C_4\eta l^*\frac{r(d_1+d_2)}{n\varepsilon}(\sigma_\xi + \sigma_r\sqrt{C_u})\sqrt{C_u\log n}\log^{1/2}\left(\frac{1.25l^*}{\delta}\right),$$

以至少 $1-\exp(-(d_1+d_2))$的概率。

C.2 Part 2

In Part C.2, we take mathematical induction to prove that when the events

$$\mathcal{E}_l^* := \mathcal{E}_{\text{RIP}} \cap \mathcal{E}_l \cap \mathcal{E}_{\Delta_l},$$

occurs with high probability, the event $\mathcal{E}_{l+1}^*$ occurs with high probability as well. Here, $\mathcal{E}_{\text{RIP}}$ is defined as the event where the RIP condition of $\{X_i\}_{i\in[n]}$ holds, See (S.17).

Step 1: $\mathcal{E}_0^*$ is true with high probability.

We first consider the RIP condition. According to Lemma 1, for any $B \in \mathbb{R}^{d_1\times d_2}$ of rank r, there exist constants $c_1, c_2, c_3 > 0$ and $0 < c_4 < c_5$ such that when $n \geq c_1 r(d_1+d_2)$, with probability at least $1 - c_2\exp\left(-c_3 r(d_1+d_2)\right)$,

$$(1-R_r)\|B\|_{\text{F}}^2 \leq \frac{1}{n}\sum_{i=1}^{n}\langle X_i, B\rangle^2 \leq (1+R_r)\|B\|_{\text{F}}^2,$$

where $R_r := \left(1 - c_4\sqrt{C_u C_l}\right) \vee \left(c_5\sqrt{C_u C_l} - 1\right)$. The values of R_{2r}, R_{3r} and R_{4r} are defined similarly. Therefore, under the condition that $n \geq \widetilde{c}_1 r(d_1+d_2)$, for some constants $\widetilde{c}_1, \widetilde{c}_2, \widetilde{c}_3, \widetilde{c}_4, \widetilde{c}_5 > 0$,

$$\mathcal{E}_{\text{RIP}} := \left\{R_r \vee R_{2r} \vee R_{3r} \vee R_{4r} \leq \left(1 - \widetilde{c}_4\sqrt{C_u C_l}\right) \vee \left(\widetilde{c}_5\sqrt{C_u C_l} - 1\right)\right\}, \tag{S.17}$$

happens with probability at least $1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)$.

As for $\mathcal{E}_0$, we refer to Corollary 1, which shows that as the sample size

$$n \geq \widetilde{O}\left((\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2)\right),$$

the event $\mathcal{E}_0$ happens with probability at least $1-(d_1+d_2)^{-10}-n^{-9}-\exp(-d_1)-\exp(-d_2)-10^{-20r}$. Conditioned on $\mathcal{E}_0$, plugging $l = 0$ to the event $\mathcal{E}'_{\Delta_l}$ defined in (S.15), we have the event $\mathcal{E}_{\Delta_l}$ defined in (S.16) happens with probability at least $1 - n^{-9} - \exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$. To this end, we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_0^*\right) = \mathbb{P}\left(\mathcal{E}_{\text{RIP}} \cap \mathcal{E}_0 \cap \mathcal{E}_{\Delta_0}\right) &\geq 1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right) \\ &- (d_1+d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r} \\ &- n^{-9} - \exp\left(-10C_u(d_1+d_2)\right)n^{-9}.\end{aligned}$$

Step 2: induction. The following analysis is conditioned on the event $\mathcal{E}_l^*$. Let $\mathcal{X} : \mathbb{R}^{d_1\times d_2} \to \mathbb{R}^n$ be an operator defined by $\mathcal{X}(B) = \left(\langle X_1, B\rangle, \cdots, \langle X_n, B\rangle\right)^\top \in \mathbb{R}^n$, for all $B \in \mathbb{R}^{d_1\times d_2}$. It is easy to check that the adjoint operator of $\mathcal{X}$ is $\mathcal{X}^* : \mathbb{R}^n \to \mathbb{R}^{d_1\times d_2}$ which is defined by $\mathcal{X}^*(b) :=$

C.2 第二部分

在部分 C.2 中，我们使用数学归纳法证明当事件

$$\mathcal{E}_l^* := \mathcal{E}_{\text{RIP}} \cap \mathcal{E}_l \cap \mathcal{E}_{\Delta_l},$$

以高概率发生，事件 $\mathcal{E}_{l+1}^*$ 以高概率也会发生。这里， $\mathcal{E}_{\text{RIP}}$ 定义为满足 RIP 条件的事件 $\{X_i\}_{i\in[n]}$ ，参见 (S.17)。

步骤 1： $\mathcal{E}_0^*$ 以高概率为真。

我们首先考虑RIP条件。根据引理1，对于任何秩为r的 $B \in \mathbb{R}^{d_1\times d_2}$ ，存在常数 c_1 c_2 $c_3 > 0$、{v16}和 $0 < c_4 < c_5$，使得当 $n \geq c_1 r(d_1+d_2)$时，以至少 $1 - c_2\exp\left(-c_3 r(d_1+d_2)\right)$的概率，

$$(1-R_r)\|B\|_{\text{F}}^2 \leq \frac{1}{n}\sum_{i=1}^{n}\langle X_i, B\rangle^2 \leq (1+R_r)\|B\|_{\text{F}}^2,$$

在 $R_r := \left(1 - c_4\sqrt{C_u C_l}\right) \vee \left(c_5\sqrt{C_u C_l} - 1\right)$ 处。 R_{2r}、 R_{3r} 和 R_{4r} 的值定义类似。因此，在 $n \geq \widetilde{c}_1 r(d_1+d_2)$ 的条件下，对于某些常数 $\widetilde{c}_1$ $\widetilde{c}_2$ $\widetilde{c}_3$ $\widetilde{c}_4$ $\widetilde{c}_5 > 0$、

$$\mathcal{E}_{\text{RIP}} := \left\{R_r \vee R_{2r} \vee R_{3r} \vee R_{4r} \leq \left(1 - \widetilde{c}_4\sqrt{C_u C_l}\right) \vee \left(\widetilde{c}_5\sqrt{C_u C_l} - 1\right)\right\}, \tag{S.17}$$

以至少 $1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)$ 的概率发生。

至于 $\mathcal{E}_0$，我们参考推论 1，它表明随着样本量

$$n \geq \widetilde{O}\left((\kappa^4 r^2 + \kappa^2\kappa_\xi^2 r)(d_1 \vee d_2)\right),$$

该事件 $\mathcal{E}_0$ 以至少 $1-(d_1+d_2)^{-10}-n^{-9}-\exp(-d_1)-\exp(-d_2)-10^{-20r}$的概率发生。在 $\mathcal{E}_0$下，将 $l = 0$ 代入事件 $\mathcal{E}'_{\Delta_l}$ 中定义的（S.15），我们得到事件 $\mathcal{E}_{\Delta_l}$中定义的（S.16）以至少 $1 - n^{-9} - \exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$的概率发生。为此，我们有

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_0^*\right) = \mathbb{P}\left(\mathcal{E}_{\text{RIP}} \cap \mathcal{E}_0 \cap \mathcal{E}_{\Delta_0}\right) &\geq 1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right) \\ &- (d_1+d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r} \\ &- n^{-9} - \exp\left(-10C_u(d_1+d_2)\right)n^{-9}.\end{aligned}$$

步骤 2：归纳。以下分析基于事件 $\mathcal{E}_l^*$。令 $\mathcal{X} : \mathbb{R}^{d_1\times d_2} \to \mathbb{R}^n$ 是一个由 $\mathcal{X}(B) = \left(\langle X_1, B\rangle, \cdots, \langle X_n, B\rangle\right)^\top \in \mathbb{R}^n$ 定义的算子，对所有 $B \in \mathbb{R}^{d_1\times d_2}$。容易验证 $\mathcal{X}$ 的伴随算子是 $\mathcal{X}^* : \mathbb{R}^n \to \mathbb{R}^{d_1\times d_2}$，它由 $\mathcal{X}^*(b) :=$ 定义

$\sum_{i=1}^n b_i X_i$, for all $b \in \mathbb{R}^n$. Therefore, $\mathcal{X}^*\mathcal{X}(M_l) = \sum_{i=1}^n \langle X_i, M_l\rangle X_i$ and $\mathcal{X}^*(\xi) = \sum_{i=1}^n \xi_i X_i$, where $\xi := (\xi_1, \cdots, \xi_n) \in \mathbb{R}^n$ and accordingly,

$$P_{\mathbb{T}_l} G_l = \frac{1}{n}\left[\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\left(M_l - M\right) - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}(\xi)\right].$$

Our first goal is to upper bound

$$\begin{aligned}
&\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)\|_F = \|M_l - M - \frac{\eta}{n}\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\left(M_l - M\right) - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}(\xi)\|_F \\
&\le \left(1 - \frac{\eta}{n}\right)\|M_l - M\|_F + \underbrace{\frac{\eta}{n}\left\|\left(\mathcal{I} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\right)\left(M_l - M\right)\right\|_F}_{D_1} \\
&+ \underbrace{\frac{\eta}{n}\left\|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\right\|_F}_{D_2} + \underbrace{\frac{\eta}{n}\left\|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\left(\xi\right)\right\|_F}_{D_3}.
\end{aligned} \tag{S.18}$$

Lemma 9 characterizes the operators $\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}$ and $\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}$, which are critical to upper bound D_1 and D_2 in (S.18).

Lemma 9 (Wei et al. (2016), Luo and Zhang (2022))**.** *Suppose the event $\mathcal{E}_{\text{RIP}}$ happens, then the following conclusions hold*

1. $\|\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\| \le R_{2r}$.
2. $\|\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F = R_{4r}\left\|\mathcal{P}_{\mathbb{T}_l^\perp}(M_l - M)\right\|_F$,

where $\|\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F \le \frac{1}{\sigma_r}\|M_l - M\|\|M_l - M\|_F$ according to Wei et al. (2016).

According to Lemma 9, conditioned on the event $\mathcal{E}_l^*$

$$\begin{aligned}
&D_1 = \frac{\eta}{n}\left\|\mathcal{I} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\right\|_F \\
&\le \frac{\eta}{n}\left[\|\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\|_F + \|\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F\right] \le \frac{\eta}{n}\left(R_{2r} + c_0\right)\|M_l - M\|_F,
\end{aligned}$$

and $D_2 = \frac{\eta}{n}\|\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\|_F \le \frac{\eta}{n}R_{4r}c_0\|M_l - M\|_F$. To this end, the only term unknown in (S.18) is

$$D_3 = \|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\left(\xi\right)\|_F = \left\|\mathcal{P}_{\mathbb{T}_l}\sum_{i=1}^n \xi_i X_i\right\|_F \le \sqrt{2r}\left\|\sum_{i=1}^n \xi_i X_i\right\|,$$

where $\|\sum_{i=1}^n \xi_i X_i\|$ is the spectral norm of a summation of n i.i.d. mean zero sub-exponential random matrices. We upper bound $\|\sum_{i=1}^n \xi_i X_i\|$ by Theorem 4. Let $B_i := \xi_i X_i$ for all $i = 1, \cdots, n$, then

$$K := \max_{i\in[n]}\|\|B_i\|\|_{\psi_1} = \max_{i\in[n]}\|\|\xi_i X_i\|\|_{\psi_1} \le \max_{i\in[n]}\|\xi_i\|_{\psi_2}\|\|X_i\|\|_{\psi_2} \le \sqrt{C_u(d_1+d_2)\sigma_\xi^2}.$$

$\sum_{i=1}^n b_i X_i$, 对所有 $b \in \mathbb{R}^n$。因此， $\mathcal{X}^*\mathcal{X}(M_l) = \sum_{i=1}^n \langle X_i\ M_l\rangle X_i$， 以及 $\mathcal{X}^*(\xi) = \sum_{i=1}^n \xi_i X_i$, 其中 $\xi := (\xi_1, \cdots, \xi_n) \in \mathbb{R}^n$， 并相应地，

$$P_{\mathbb{T}_l} G_l = \frac{1}{n}\left[\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\left(M_l - M\right) - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}(\xi)\right].$$

我们的第一个目标是上界

$$\begin{aligned}
&\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)\|_F = \|M_l - M - \frac{\eta}{n}\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\left(M_l - M\right) - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}(\xi)\|_F \\
&\le \left(1 - \frac{\eta}{n}\right)\|M_l - M\|_F + \underbrace{\frac{\eta}{n}\left\|\left(\mathcal{I} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\right)\left(M_l - M\right)\right\|_F}_{D_1} \\
&+ \underbrace{\frac{\eta}{n}\left\|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\right\|_F}_{D_2} + \underbrace{\frac{\eta}{n}\left\|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\left(\xi\right)\right\|_F}_{D_3}.
\end{aligned} \tag{S.18}$$

引理 9 描述了算子 $\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}$和 $\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}$， 它们对上界至关重要 D_1 和 D_2 在（S.18）。

引理 9 (魏等人 (2016)， 罗和张 (2022))。 假设事件 $\mathcal{E}_{\text{RIP}}$ 发生，则以下结论成立

1. $\|\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\| \le R_{2r}$.
2. $\|\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F = R_{4r}\left\|\mathcal{P}_{\mathbb{T}_l^\perp}(M_l - M)\right\|_F$,

其中 $\|\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F \le \frac{1}{\sigma_r}\|M_l - M\|\|M_l - M\|_F$根据 魏等人 (2016)。

根据引理 9，在事件 $\mathcal{E}_l^*$

$$\begin{aligned}
&D_1 = \frac{\eta}{n}\left\|\mathcal{I} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\right\|_F \\
&\le \frac{\eta}{n}\left[\|\mathcal{P}_{\mathbb{T}_l} - \mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\|_F + \|\mathcal{P}_{\mathbb{T}_l^\perp}\left(M_l - M\right)\|_F\right] \le \frac{\eta}{n}\left(R_{2r} + c_0\right)\|M_l - M\|_F,
\end{aligned}$$

和 $D_2 = \frac{\eta}{n}\|\mathcal{P}_{T_l}\mathcal{X}^*\mathcal{X}\mathcal{P}_{\mathbb{T}_l}\left(M_l - M\right)\|_F \le \frac{\eta}{n}R_{4r}c_0\|M_l - M\|_F$。 为此，(S.18)中唯一的未知项是

$$D_3 = \|\mathcal{P}_{\mathbb{T}_l}\mathcal{X}^*\left(\xi\right)\|_F = \left\|\mathcal{P}_{\mathbb{T}_l}\sum_{i=1}^n \xi_i X_i\right\|_F \le \sqrt{2r}\left\|\sum_{i=1}^n \xi_i X_i\right\|,$$

其中 $\|\sum_{i=1}^n \xi_i X_i\|$ 是 n 独立同分布、均值为零的次指数随机矩阵和的谱范数。我们用定理4上界估计 $\|\sum_{i=1}^n \xi_i X_i\|$。令 $B_i := \xi_i X_i$ 对所有 i =1 $\cdots$ n， ， ， 则

$$K := \max_{i\in[n]}\|\|B_i\|\|_{\psi_1} = \max_{i\in[n]}\|\|\xi_i X_i\|\|_{\psi_1} \le \max_{i\in[n]}\|\xi_i\|_{\psi_2}\|\|X_i\|\|_{\psi_2} \le \sqrt{C_u(d_1+d_2)\sigma_\xi^2}.$$

Since $\mathbb{E}\xi_i^4 \le 3\sigma_\xi^4$ and $\mathbb{E}\|X_i\|^4 \le 3C_u(d_1+d_2)^2$,

$$\|\mathbb{E}B_iB_i^\top\| = \mathbb{E}\xi_i^2\mathbb{E}\|X_i\|^2 \le \frac{C_u(d_1+d_2)}{2\sigma_\xi^2}\mathbb{E}\xi_i^4 + \frac{\sigma_\xi^2}{2C_u(d_1+d_2)}\mathbb{E}\|X_i\|^4 \le 3\sigma_\xi^2 C_u(d_1+d_2),$$

where the first inequality uses the fact $ab \le \frac{a^2}{2c} + \frac{cb^2}{2}$. Similarly, $\|\mathbb{E}B_iB_i^\top\| \le 3\sigma_\xi^2 C_u(d_1+d_2)$.

Therefore,

$$S^2 := \|\mathbb{E}B_iB_i^\top\| \vee \|\mathbb{E}B_i^\top B_i\| \le 3\sigma_\xi^2 C_u(d_1+d_2).$$

Applying Thorem 4 with $\alpha = 1$, $K = c_1\sqrt{C_u(d_1+d_2)\sigma_\xi^2}$ and $S = \sqrt{3\sigma_\xi^2 C_u(d_1+d_2)}$, we have

$$\mathbb{P}\left(\frac{1}{n}\left\|\sum_{i=1}^n \xi_i X_i\right\| \ge C_5\sqrt{C_u(d_1+d_2)\sigma_\xi^2}\sqrt{\frac{\log(d_1+d_2)}{n}}\right) \le (d_1+d_2)^{-10}.$$

In conclusion, with probability at least $1-(d_1+d_2)^{-10}$

$$D_3 \le C_1\sqrt{r}\left\|\sum_{i=1}^n \xi_i X_i\right\|_F \le C_1\sqrt{C_u\sigma_\xi^2 nr(d_1+d_2)\log(d_1+d_2)},$$

for some constant $C_1 > 0$.

Conditioned on $\mathcal{E}_l^*$, we plug D_1, D_2 and D_3 into (S.18) and obtain that for some small constant $0 < c_0 < 1$ and absolute constant $C_2 > 0$

$$\begin{aligned}
&\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\|_F \le \|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)\|_F + \|\mathcal{P}_{\mathbb{T}_l}N_l\|_F \\
&\le (1-\rho_0)\|M_l - M\|_F + C_2\eta\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) \\
&\quad + C_2\eta l^*\sqrt{C_u}(\sigma_\xi + \sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2} n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right),
\end{aligned}$$

with probability at least $1-(d_1+d_2)^{-10} - \exp(-(d_1+d_2))$ where we define

$$\rho_0 := \frac{\eta}{n}\left(1 - R_{2r} - c_0 - R_{4r}c_0\right).$$

Suppose that $c_0 \lesssim \frac{1}{R_{2r}(1+R_{4r})} \wedge \frac{1}{8}$, the step size $\eta \le n$ being a small constant, then we have $0 \le \rho_0 < 1$. Further, as for some absolute constant $C_3 > 0$, the sample size satisfies

$$\begin{aligned}
n \ge C_3\max\Bigg\{&\eta^2\left(\frac{\sigma_\xi}{\sigma_r}\right)^2 C_u r(d_1+d_2)\log(d_1+d_2), \\
&\eta l^*\sqrt{C_u}\left(\frac{\sigma_\xi + \sigma_r\sqrt{C_u}}{\sigma_r}\right)\frac{r(d_1+d_2)}{\varepsilon}\log^{1/2} n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right)\Bigg\},
\end{aligned}$$

we have for some small constant $0 < \rho_1 < 1$,

$$\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\|_F \le (1-\rho_1)\|M_l - M\|_F \le (1-\rho_1)c_0\sigma_r.$$

自 $\mathbb{E}\xi_i^4 \le 3\sigma_\xi^4$ 和 $\mathbb{E}\|X_i\|^4 \le 3C_u(d_1+d_2)^2$ 以来

$$\|\mathbb{E}B_iB_i^\top\| = \mathbb{E}\xi_i^2\mathbb{E}\|X_i\|^2 \le \frac{C_u(d_1+d_2)}{2\sigma_\xi^2}\mathbb{E}\xi_i^4 + \frac{\sigma_\xi^2}{2C_u(d_1+d_2)}\mathbb{E}\|X_i\|^4 \le 3\sigma_\xi^2 C_u(d_1+d_2),$$

在第一个不等式中使用了 $ab \le \frac{a^2}{2c} + \frac{cb^2}{2}$ 的事实。类似地，$\|\mathbb{E}B_iB_i^\top\| \le 3\sigma_\xi^2 C_u(d_1+d_2)$。

因此，

$$S^2 := \|\mathbb{E}B_iB_i^\top\| \vee \|\mathbb{E}B_i^\top B_i\| \le 3\sigma_\xi^2 C_u(d_1+d_2).$$

应用定理，4，与 $\alpha = 1$，$K = c_1 \quad C_u$，$(d_1 + d_2)$，σ_ξ^2，和 $S = \quad 3\sigma_\xi^2 C_u$，$(d_1 + d_2)$，我们得到

$$\mathbb{P}\left(\frac{1}{n}\left\|\sum_{i=1}^n \xi_i X_i\right\| \ge C_5\sqrt{C_u(d_1+d_2)\sigma_\xi^2}\sqrt{\frac{\log(d_1+d_2)}{n}}\right) \le (d_1+d_2)^{-10}.$$

总之，至少以 $1-(d_1+d_2)^{-10}$ 的概率

$$D_3 \le C_1\sqrt{r}\left\|\sum_{i=1}^n \xi_i X_i\right\|_F \le C_1\sqrt{C_u\sigma_\xi^2 nr(d_1+d_2)\log(d_1+d_2)},$$

对于某个常数 $C_1 > 0$。

给定 $\mathcal{E}_l^*$，我们代入 D_1， D_2 和 D_3 到（S.18）并得到对于某个小的常数 $0 < c_0 < 1$ 和绝对常数 $C_2 > 0$

$$\begin{aligned}
&\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\|_F \le \|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)\|_F + \|\mathcal{P}_{\mathbb{T}_l}N_l\|_F \\
&\le (1-\rho_0)\|M_l - M\|_F + C_2\eta\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) \\
&\quad + C_2\eta l^*\sqrt{C_u}(\sigma_\xi + \sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2} n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right),
\end{aligned}$$

至少以概率 $1-(d_1+d_2)^{-10} - \exp(-(d_1+d_2))$，我们定义

$$\rho_0 := \frac{\eta}{n}\left(1 - R_{2r} - c_0 - R_{4r}c_0\right).$$

假设 $c_0 \lesssim \frac{1}{R_{2r}(1+R_{4r})} \wedge \frac{1}{8}$，步长 $\eta \le n$ 是一个小的常数，那么我们有$0 \le \rho_0 < 1$。此外，对于某个绝对常数 $C_3 > 0$，样本量满足

$$\begin{aligned}
n \ge C_3\max\Bigg\{&\eta^2\left(\frac{\sigma_\xi}{\sigma_r}\right)^2 C_u r(d_1+d_2)\log(d_1+d_2), \\
&\eta l^*\sqrt{C_u}\left(\frac{\sigma_\xi + \sigma_r\sqrt{C_u}}{\sigma_r}\right)\frac{r(d_1+d_2)}{\varepsilon}\log^{1/2} n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right)\Bigg\},
\end{aligned}$$

对于某个小的常数 $0 < \rho_1 < 1$，我们有

$$\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l) + \mathcal{P}_{\mathbb{T}_l}N_l\|_F \le (1-\rho_1)\|M_l - M\|_F \le (1-\rho_1)c_0\sigma_r.$$

Applying Lemma 4, we obtain that under the condition $40c_0 < \frac{\rho_1}{2}$,

$$\begin{aligned}\left\|M_{l+1}-M\right\|_F &\leq (1+40c_0)\left\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)+\mathcal{P}_{\mathbb{T}_l}N_l\right\|_F \\ &\leq (1+40c_0)\,(1-\rho_1)\|M_l-M\|_F \\ &\leq \left(1-\frac{\rho_1}{2}\right)\|M_l-M\|_F \leq \left(1-\frac{\rho_1}{2}\right)c_0\sigma_r.\end{aligned} \tag{S.19}$$

In summary, conditioned on the event $\mathcal{E}_l^*$, the event

$$\mathcal{E}_{l+1} := \left\{\|M_{l+1}-M\|_F \leq c_0\sigma_r\right\},$$

occurs with probability at least $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))$. Besides, according to $\mathcal{E}'_{\Delta_l}$ defined in (S.15), the event

$$\mathcal{E}_{\Delta_{l+1}} := \left\{\Delta_{l+1} \leq \widetilde{C}_3\frac{\eta}{n}\left(\sigma_\xi+\sigma_r\sqrt{C_u}\right)\sqrt{C_u r\,(d_1+d_2+\log n)\log n}\right\},$$

occurs with probability $1-n^{-9}-\exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$. Therefore, conditioned on $\mathcal{E}_l^*$, the event $\mathcal{E}_{l+1}^*$ happens with probability at least $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))-n^{-9}-\exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$.

To this end, we has finished the induction and conclude Part C.2 by

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=0}^{l}\mathcal{E}_i^*\right) \geq 1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)-(d_1+d_2)^{-10}-n^{-9}-e^{-d_1}-e^{-d_2}-10^{-20r} \\ -l\left((d_1+d_2)^{-10}+\exp(-(d_1+d_2))\right)-(l+1)\left(n^{-9}+\exp\left(-10C_u(d_1+d_2)\right)n^{-9}\right).\end{aligned}$$

C.3 Part 3

In Part C.3, we derive the convergence rate of $\|M_{l^*}-M\|_F$ and choose an appropriate value for l^*. Conditioned on the event $\bigcap_{i=0}^{l^*-1}\mathcal{E}_i^*$, according to (S.19), with probability at least $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))$

$$\begin{aligned}&\left\|\widetilde{M}_{l^*}-M\right\|_F = \|M_{l^*}-M\|_F \\ &\leq \left(1-\rho_0\right)^{l^*}\|M_0-M\|_F+\left(\sum_{l=0}^{l^*-1}\left(1-\rho_0\right)^{l^*-l-1}\right)C_2\eta\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) \\ &\quad+\left(\sum_{l=0}^{l^*-1}\left(1-\rho_0\right)^{l^*-l-1}\right)C_2\eta l^*\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2}n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right).\end{aligned}$$

Let $\|M_0-M^*\|_F = c_0^*$ and $l^* := \log\left(c_0^* n\right)/\rho_0$, then we have $\left(1-\rho_0\right)^{l^*}\|M_0-M\|_F \asymp \frac{1}{n}$, indicating that there is little reason to run the algorithm further than $O(\log n)$ iterations.

应用引理 4，我们得到在条件 $40c_0 < \frac{\rho_1}{2}$ 下，

$$\begin{aligned}\left\|M_{l+1}-M\right\|_F &\leq (1+40c_0)\left\|M_l - M - \eta\mathcal{P}_{\mathbb{T}_l}(G_l)+\mathcal{P}_{\mathbb{T}_l}N_l\right\|_F \\ &\leq (1+40c_0)\,(1-\rho_1)\|M_l-M\|_F \\ &\leq \left(1-\frac{\rho_1}{2}\right)\|M_l-M\|_F \leq \left(1-\frac{\rho_1}{2}\right)c_0\sigma_r.\end{aligned} \tag{S.19}$$

总之，在事件 $\mathcal{E}_l^*$ 的条件下，事件

$$\mathcal{E}_{l+1} := \left\{\|M_{l+1}-M\|_F \leq c_0\sigma_r\right\},$$

以至少 $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))$的概率发生。此外，根据 $\mathcal{E}'_{\Delta_l}$在（S.15）中定义的，事件

$$\mathcal{E}_{\Delta_{l+1}} := \left\{\Delta_{l+1} \leq \widetilde{C}_3\frac{\eta}{n}\left(\sigma_\xi+\sigma_r\sqrt{C_u}\right)\sqrt{C_u r\,(d_1+d_2+\log n)\log n}\right\},$$

以概率 $1-n^{-9}-\exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$ 发生。因此，在 $\mathcal{E}_l^*$ 的条件下，事件 $\mathcal{E}_{l+1}^*$ 以至少 $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))-n^{-9}-\exp\left(-10C_u\left(d_1+d_2\right)\right)n^{-9}$. 的概率发生

为此，我们已经完成了归纳并得出第 C.2 部分，

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=0}^{l}\mathcal{E}_i^*\right) \geq 1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)-(d_1+d_2)^{-10}-n^{-9}-e^{-d_1}-e^{-d_2}-10^{-20r} \\ -l\left((d_1+d_2)^{-10}+\exp(-(d_1+d_2))\right)-(l+1)\left(n^{-9}+\exp\left(-10C_u(d_1+d_2)\right)n^{-9}\right).\end{aligned}$$

第 C.3 部分 3

在部分 C.3，我们推导出收敛速度 $\|M_{l^*}-M\|_F$并选择一个合适的值用于 l^*。在事件 $\bigcap_{i=0}^{l^*-1}\mathcal{E}_i^*$ 下，根据（S.19），以至少 $1-(d_1+d_2)^{-10}-\exp(-(d_1+d_2))$

$$\begin{aligned}&\left\|\widetilde{M}_{l^*}-M\right\|_F = \|M_{l^*}-M\|_F \\ &\leq \left(1-\rho_0\right)^{l^*}\|M_0-M\|_F+\left(\sum_{l=0}^{l^*-1}\left(1-\rho_0\right)^{l^*-l-1}\right)C_2\eta\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) \\ &\quad+\left(\sum_{l=0}^{l^*-1}\left(1-\rho_0\right)^{l^*-l-1}\right)C_2\eta l^*\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2}n\log^{1/2}\left(\frac{1.25l^*}{\delta}\right).\end{aligned}$$

令 $\|M_0-M^*\|_F = c_0^*$和 $l^* := \log\left(c_0^* n\right)/\rho_0$，则我们有 $\left(1-\rho_0\right)^{l^*}\|M_0-M\|_F \asymp \frac{1}{n}$，表明几乎没有理由运行算法超过 $O(\log n)$ 次迭代。

In conclusion,

$$\left\|\widetilde{M}_{l^*}-M\right\|_F \leq C_3\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) + C_3\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2}n\log^{3/2}\left(\frac{1.25l^*}{\delta}\right).$$

with probability at least

$$1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)-(d_1+d_2)^{-10}-n^{-9}-e^{-d_1}-e^{-d_2}-10^{-20r} - \left((d_1+d_2)^{-10}+\exp(-(d_1+d_2))+n^{-9}+\exp\left(-10C_u(d_1+d_2)\right)n^{-9}\right)\log n.$$

D The lower bound derived by score attack argument

Let $\mathbb{M}_r := \{M\in\mathbb{R}^{d_1\times d_2}:\mathrm{rank}(M)=r\}$. This section establishes the minimax lower bound of differentially privately estimating the matrix $M\in\bigcup_{k=1}^r\mathbb{M}_k$, within the trace regression model based on an alternative approach, score attack argument Cai et al. (2023).

The score attack argument involves designing a test statistic and establishing the lower bound of the statistic with the help of a prior distribution of the parameters to estimate. It is unclear, however, how to construct a prior distribution for the low-rank matrix M such that the prior complies with the parameter space $\mathbb{M}_r$ and the *score attack* is easy to compute at the same time. Compared to DP-fano's Lemma (See Lemma 3) which requires $\delta\lesssim e^{-n}$, the score attack argument is valid for a wider range of $\delta\lesssim n^{1+\gamma}$ where $\gamma>0$ is a constant. We first define some necessary notations for the elaboration of score attack argument. For any matirx $B,C\in\mathbb{R}^{d_1\times d_2}$, we denote $\mathrm{supp}(B):=\{(i,j)\in[d_1]\times[d_2]:B_{ij}\neq 0\}$ as the support of B and the matrix C restricted on supp B is $[C]_{\mathrm{supp}(B)}=\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}C_{ij}e_ie_j^\top\mathbb{I}(B_{ij}\neq 0)$ where e_i is the i-th canonical basis in $\mathbb{R}^{d_1}$ and e_j is the j-th canonical basis in $\mathbb{R}^{d_2}$.

To apply score attack argument, we relax the problem to deriving minimax lower bounds over $\mathbb{M}_{r,d_1}:=\{M\in\mathbb{R}^{d_1\times d_2}:\mathrm{supp}(M)\subset[d_1]\times[r]\}\subset\bigcup_{k=1}^r\mathbb{M}_k$. The benefit is that there exists a trivial prior of $M\in\mathbb{M}_{r,d_1}$ such that $M_{ij}\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$ for $(i,j)\in[d_1]\times[r]$ and $M_{ij}=0$ otherwise. Similarly, we may consider establish minimax lower bound over $\mathbb{M}_{r,d_2}:=\{M\in\mathbb{R}^{d_1\times d_2}:\mathrm{supp}(M)\subset[r]\times[d_2]\}\subset\bigcup_{k=1}^r\mathbb{M}_k$. For any $M\in\mathbb{M}_{r,d_2}$, there is a trivial prior as well. Let A be a randomized algorithm mapping a dataset Z to a $d_1\times d_2$ matrix. We define the DP-constrained minimax risk over $\bigcup_{k=1}^r\mathbb{M}_k$ as

$$\mathrm{risk}(\bigcup_{k=1}^r\mathbb{M}_k):=\inf_A\sup_{M\in\mathbb{M}_r}\mathbb{E}\left\|A(Z)-M\right\|_F^2,$$

总之，

$$\left\|\widetilde{M}_{l^*}-M\right\|_F \leq C_3\sigma_\xi\sqrt{C_u}\sqrt{\frac{r(d_1+d_2)}{n}}\log^{1/2}(d_1+d_2) + C_3\sqrt{C_u}(\sigma_\xi+\sigma_r\sqrt{C_u})\frac{r(d_1+d_2)}{n\varepsilon}\log^{1/2}n\log^{3/2}\left(\frac{1.25l^*}{\delta}\right).$$

至少以概率

$$1-\widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1+d_2)\right)-(d_1+d_2)^{-10}-n^{-9}-e^{-d_1}-e^{-d_2}-10^{-20r} - \left((d_1+d_2)^{-10}+\exp(-(d_1+d_2))+n^{-9}+\exp\left(-10C_u(d_1+d_2)\right)n^{-9}\right)\log n.$$

D 推导出的下界基于评分攻击论证

让 $\mathbb{M}_r:=\{M\in\mathbb{R}^{d_1\times d_2}:\ \mathrm{rank}(M)=r\}$。本节建立了在替代方法、评分攻击论证 Cai 等人 (2023) 的基础上，对矩阵 $M\in\bigcup_{k=1}^r\mathbb{M}_k$，，在迹回归模型中，以差分隐私估计的最小最大下界

评分攻击论证涉及设计一个检验统计量，并借助参数的先验分布来建立统计量的下界。然而，如何构造一个低秩矩阵 M 的先验分布，使得先验分布符合参数空间 $\mathbb{M}_r$，同时评分攻击 易于计算，这一点尚不清楚。与需要 $\delta\lesssim e^{-n}$的 DP-fano 引理（参见引理 3）相比，评分攻击论证对更广泛的 $\delta\lesssim n^{1+\gamma}$ 有效，其中 $\gamma>0$ 是一个常数。我们首先定义一些必要的符号，以便详细阐述评分攻击论证。对于任何 matirx$B$ $C\in\mathbb{R}^{d_1\times d_2}$，，我们记 supp $(B):=\{(i,j)\in[d_1]\times[d_2]:B_{ij}=0\}$ 为 B的支持，且在 supp B上限制的矩阵 C是 $[C]_{\mathrm{supp}(B)}=\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}C_{ij}e_ie_j^\top\mathbb{I}(B_{ij}=0)$，其中 e_i是 i-th 标准基，$\mathbb{R}^{d_1}$，且 e_j 是 j-th 标准基，$\mathbb{R}^{d_2}$。

/

/

要应用得分攻击论证，我们将问题放宽到在 $\mathbb{M}_{r,d_1}:=\{M\in\mathbb{R}^{d_1\times d_2}:\mathrm{supp}(M)\subset[d_1]\times[r]\}\subset\bigcup_{k=1}^r\mathbb{M}_k$ 上导出 minimax 下界。优点在于存在一个平凡的先验 $M\in\mathbb{M}_{r\,d_1}$，使得 $M_{ij}\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$ 当 $(i,j)\in[d_1]\times[r]$ 时，$M_{ij}=0$否则。类似地，我们也可以考虑在 $\mathbb{M}_{r\,d_2}:=\{M\in\mathbb{R}^{d_1\times d_2}:\mathrm{supp}(M)\subset[r]\times[d_2]\}\subset\bigcup_{k=1}^r\mathbb{M}_k$. 上建立 minimax 下界。对于任何 $M\in\mathbb{M}_{r\,d_2}$，也存在一个平凡的先验。令 A 是一个随机算法，将数据集 Z 映射到一个 $d_1\times d_2$ 矩阵。我们将 $\bigcup_{k=1}^r\mathbb{M}_k$ 上的 DP-约束 minimax 风险定义为

$$\mathrm{risk}(\bigcup_{k=1}^r\mathbb{M}_k):=\inf_A\sup_{M\in\mathbb{M}_r}\mathbb{E}\left\|A(Z)-M\right\|_F^2,$$

where A is taken over all (ϵ, δ)-DP algorithms. Similarly, we define $\text{risk}(\mathbb{M}_{r,d_1})$ and $\text{risk}(\mathbb{M}_{r,d_2})$. Since $\mathbb{M}_{r,d_1} \subset \bigcup_{k=1}^{r} \mathbb{M}_k$ and $\mathbb{M}_{r,d_2} \subset \bigcup_{k=1}^{r} \mathbb{M}_k$, we have

$$\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k) \geq \text{risk}(\mathbb{M}_{r,d_1}) \bigvee \text{risk}(\mathbb{M}_{r,d_2}), \tag{S.20}$$

which indicates that the lower bound of $\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k)$ will be an immediate result once we successfully lower bound $\text{risk}(\mathbb{M}_{r,d_1})$ and $\text{risk}(\mathbb{M}_{r,d_2})$.

Next, we construct *score attacks* to derive the lower bounds of $\text{risk}(\mathbb{M}_{r,d_1})$ and $\text{risk}(\mathbb{M}_{r,d_2})$. Let $z = (X, y)$ be the pair of a measurement matrix and its corresponding response variable, drawn independently from (3). The score function is defined by

$$S_M(z) := \nabla_M \log f_M(z) = \nabla_M \log f_M(y|X),$$

and the score attack is defined by

$$\mathcal{A}_M^{(1)}(z, A(Z)) := \left\langle [A(Z) - M]_{[d_1]\times[r]}, S_M(z) \right\rangle,$$

where A is an (ε, δ)-DP algorithm to estimate $M \in \mathbb{M}_r$; $z = (X, y)$ is a piece of datum that we want to test whether it belongs to $Z = \{(X_i, y_i)\}_{i=1}^n$; the quantity $[A(Z) - M]_{[d_1]\times[r]}$ is obtained by restricting $A(Z) - M \in \mathbb{R}^{d_1 \times d_2}$ to the index set $[d_1] \times [r]$. Under some regularity conditions, the score attack $\mathcal{A}_M^{(1)}(z, A(Z))$ will lead to the lower bound of $\text{risk}(\mathbb{M}_r, d_1)$. Similarly, we derive the lower bound of $\text{risk}(\mathbb{M}_r, d_2)$ with the help of the attack

$$\mathcal{A}_M^{(2)}(z, A(Z)) := \left\langle [A(Z) - M]_{[r]\times[d_2]}, S_M(z) \right\rangle.$$

Finally, Theorem 5 establishes the lower bound for estimating the low-rank matrix M. The proof of Theorem 5.

Theorem 5. *Consider i.i.d. observations $Z = \{z_1, \cdots, z_n\}$ drawn from the trace regression model defined in (1), where $z_i := (X_i, y_i)$ for $i = 1, \cdots, n$. We assume that $\{X_i\}_{i\in[n]}$ satisfy the Assumption 1, $r(d_1 \vee d_2) \lesssim n\varepsilon$, $0 < \varepsilon < 1$ and $\delta \lesssim n^{-(1+\gamma)}$ for some $\gamma > 0$, then*

$$\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k) = \inf_A \sup_{M \in \bigcup_{k=1}^{r} \mathbb{M}_k} \mathbb{E} \|A(Z) - M\|_F^2 \gtrsim \underbrace{\sigma_\xi^2 \frac{r(d_1 \vee d_2)}{n}}_{a_1} + \underbrace{\sigma_\xi^2 \frac{r^2(d_1 \vee d_2)^2}{n^2\varepsilon^2}}_{a_2}. \tag{S.21}$$

By Theorem 5, the lower bound of $\text{risk}(\mathbb{M}_r)$ consists of two terms where the first term a_1 accounts for the statistical error and the second term a_2 is the cost of privacy. The proof for a_1 can be found in Rohde and Tsybakov (2011) and the *cost of privacy* is deduced in the following proof.

在所有 (ϵ, δ)-DP 算法中取 A。类似地，我们定义 $\text{risk}(\mathbb{M}_{r,d_1})$ 和 $\text{risk}(\mathbb{M}_{r,d_2})$。由于 $\mathbb{M}_{r\,d_1} \subset \bigcup_{k=1}^{r} \mathbb{M}_k$, 和 $\mathbb{M}_{r\,d_2} \subset \bigcup_{k=1}^{r} \mathbb{M}_k$, 我们得到

$$\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k) \geq \text{risk}(\mathbb{M}_{r,d_1}) \bigvee \text{risk}(\mathbb{M}_{r,d_2}), \tag{S.20}$$

这表明，一旦我们成功地对 $\text{risk}(\mathbb{M}_{r,d_1})$ 和 $\text{risk}(\mathbb{M}_{r,d_2})$ 进行下界估计，$\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k)$ 的下界将是一个直接结果。

接下来，我们构造 得分攻击 来推导 $\text{risk}(\mathbb{M}_{r,d_1})$ 和 $\text{risk}(\mathbb{M}_{r,d_2})$。令 $z = (X, y)$ 是测量矩阵及其对应响应变量的对，独立地从 (3) 中抽取。得分函数定义为

$$S_M(z) := \nabla_M \log f_M(z) = \nabla_M \log f_M(y|X),$$

得分攻击定义为

$$\mathcal{A}_M^{(1)}(z, A(Z)) := \left\langle [A(Z) - M]_{[d_1]\times[r]}, S_M(z) \right\rangle,$$

其中 A 是一个 (ε, δ)-DP 算法来估计 $M \in \mathbb{M}_r$; $z = (X, y)$ 是我们想要测试是否属于 $Z = \{(X_i, y_i)\}_{i=1}^n$; 的数据片段 $[A(Z) - M]_{[d_1]\times[r]}$ 是通过将 $A(Z) - M \in \mathbb{R}^{d_1 \times d_2}$ 限制在索引集 $[d_1] \times [r]$ 中获得的。在一些正则条件下，得分攻击 $\mathcal{A}_M^{(1)}(z, A(Z))$ 将导致 $\text{risk}(\mathbb{M}_r, d_1)$ 的下界。类似地，我们借助攻击

$$\mathcal{A}_M^{(2)}(z, A(Z)) := \left\langle [A(Z) - M]_{[r]\times[d_2]}, S_M(z) \right\rangle.$$

最后，定理 5 建立了低秩矩阵估计的下界 M。定理 5的证明。

Theorem 5. *Consider i.i.d. observations $Z = \{z_1, \cdots, z_n\}$ drawn from the trace regression model defined in (1), where $z_i := (X_i, y_i)$ for $i = 1, \cdots, n$. We assume that $\{X_i\}_{i\in[n]}$ satisfy the Assumption 1, $r(d_1 \vee d_2) \lesssim n\varepsilon$, $0 < \varepsilon < 1$ and $\delta \lesssim n^{-(1+\gamma)}$ for some $\gamma > 0$, then*

$$\text{risk}(\bigcup_{k=1}^{r} \mathbb{M}_k) = \inf_A \sup_{M \in \bigcup_{k=1}^{r} \mathbb{M}_k} \mathbb{E} \|A(Z) - M\|_F^2 \gtrsim \underbrace{\sigma_\xi^2 \frac{r(d_1 \vee d_2)}{n}}_{a_1} + \underbrace{\sigma_\xi^2 \frac{r^2(d_1 \vee d_2)^2}{n^2\varepsilon^2}}_{a_2}. \tag{S.21}$$

根据定理 5，风险($\mathbb{M}_r$)的下界包含两项，其中第一项 a_1用于统计误差，第二项 a_2 是隐私成本。的证明可以在 Rohde and Tsybakov (2011)中找到，隐私成本 在以下证明中推导。

Proof of Theorem 5. We now start proving Theorem 5 by score attack argument. Throughout the proof, we assume that $Z = \{z_1, \cdots, z_n\}$ is an i.i.d. sample drawn from f_M and Z_i' is a neighbouring data set of Z obtained by replacing z_i with an independent copy $z_i' \sim f_M$. Besides, we mainly focus on the case $M \in \mathbb{M}_{r,d_1}$ and states the result for the case $M \in \mathbb{M}_{r,d_2}$ in Remark 1. Let

$$\mathcal{A}_M^{(1)}(z, A(Z)) := \left\langle [A(Z) - M]_{[d_1]\times[r]}, S_M(z) \right\rangle.$$

We derive the lower bound of $\text{risk}(\mathbb{M}_{r,d_1}) := \inf_A \sup_{M\in\mathbb{M}_{r,d_1}} \mathbb{E}\,\|A(Z) - M\|_F^2$, in three steps. For ease of notation, we define

$$A_i' := \mathcal{A}_M(z_i, A(Z_i')) \quad \text{and} \quad A_i := \mathcal{A}_M(z_i, A(Z)).$$

Step 1: bounding the summation. The following Lemma 10 bounds $\mathbb{E}\,|A_i'|$; Lemma 11 develops the upper bound of $\sum_{i\in[n]} \mathbb{E}\,A_i$ based on $\mathbb{E}\,|A_i'|$ discussed in Lemma 10 and a tunning parameter T. The proof of Lemma 10 and 11 can be found in Appendix E.7 and E.8.

Lemma 10. *For $i \in [n]$, we have $\mathbb{E}A_i' = 0$ and $\mathbb{E}\,|A_i'| \le \sqrt{\mathbb{E}\,\|A(Z) - M\|_F^2}\sqrt{\frac{C_u}{\sigma_\xi^2}}$.*

Lemma 11. *Let A be an (ε, δ)-DP algorithm with $0 < \varepsilon < 1$ and $\delta \ge 0$, under model (1), by choosing $T = \sqrt{2/\sigma_\xi^2} r d_1 \sqrt{\log(\frac{1}{\delta})}$, we have*

$$\sum_{i\in[n]} \mathbb{E}A_i \le 2n\varepsilon\sqrt{\mathbb{E}\|A(Z) - M\|_F^2}\sqrt{C_u/\sigma_\xi^2} + 4\sqrt{2}\delta r d_1 \sqrt{\log(1/\delta)/\sigma_\xi^2}. \tag{S.22}$$

Step 2: lower bounding the summation. Under some regularity conditions, the following Lemma 12 characterize the quantity $\sum_{i\in[n]} A_i$ as a summation of functions of M. Lemma 13 lower bounds the summation of functions by assigning an appropriate prior distribution π to M. The proof of Lemma 12 and 13 can be found in Appendix E.8.

Lemma 12. *If for every $(i,j) \in [d_1]\times[r]$, $\log f_M(Z)$ is continuously differentiable with respect to M_{ij} and $\left|\frac{\partial}{\partial M_{ij}} \log f_M(Z)\right| < h_{ij}(Z)$ such that $\mathbb{E}\,|h_{ij}(Z)A(M)_{ij}| < \infty$, we have*

$$\sum_{i\in[n]} \mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z)) = \sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} \mathbb{E}A(Z)_{ij}.$$

Lemma 12 has its general form stated in Theorem 2.1, Cai et al. (2023). Let g_{ij} be a function defined by $g_{ij}(M) := \left(\mathbb{E}_{Z|M} A(Z)\right)_{ij}$ for all $(i,j) \in [d_1]\times[r]$, then

$$\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} \mathbb{E}A(Z)_{ij} = \mathbb{E}_\pi \left(\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} g_{ij} \right).$$

定理的证明 5. 我们现在开始证明定理 5 通过评分攻击论证。在证明过程中，我们假设 $Z = \{z_1, \cdots, z_n\}$ 是一个独立同分布的样本，从 f_M 中抽取，而 Z_i' 是一个 Z 的邻近数据集，通过用独立副本 $z_i' \sim f_M$ 替换 z_i 获得<style id='31'>。此外，我们主要关注 $M \in \mathbb{M}_{r\,d_1}$ 的情况，并在备注1中陈述 $M \in \mathbb{M}_{r\,d_2}$ 的情况，。令

$$\mathcal{A}_M^{(1)}(z, A(Z)) := \left\langle [A(Z) - M]_{[d_1]\times[r]}, S_M(z) \right\rangle.$$

我们分三步推导出风险$(\mathbb{M}_{r,d_1}) := \inf_A \sup_{M\in\mathbb{M}_{r\,d_1}} \mathbb{E}\,\|A(Z) - M\|_{F\text{的界}}^2$，。为简洁起见，我们定义

$$A_i' := \mathcal{A}_M(z_i, A(Z_i')) \quad \text{and} \quad A_i := \mathcal{A}_M(z_i, A(Z)).$$

步骤1：估计求和。以下引理 10 估计 $\mathbb{E}\,|A_i'|$；引理 11 基于 $\sum_{i\in[n]} \mathbb{E}\,A_i$ 开发了 $\mathbb{E}\,|A_i'|$ 的界，该界在引理 10 中讨论过，并有一个调整参数T。引理 10 和 11 的证明可以在附录 E.7 和 E.8中找到<style id='46'>。

Lemma 10. *For $i \in [n]$, we have $\mathbb{E}A_i' = 0$ and $\mathbb{E}\,|A_i'| \le \quad \mathbb{E}\,\|A(Z) - M\|_F^2 \quad \frac{C_u}{\sigma_\xi^2}$.*

引理 11。设 A 是一个 (ε, δ)-DP 算法，具有 $0 < \varepsilon < 1$ 和 $\delta \ge 0$，在模型（*1*）下，通过

$\sqrt{\quad}$ $\sqrt{}$ 选择 $T = \quad 2/\sigma_\xi^2 r d_1 \quad \log(\frac{1}{\delta})$，我们得到

$$\sum_{i\in[n]} \mathbb{E}A_i \le 2n\varepsilon\sqrt{\mathbb{E}\|A(Z) - M\|_F^2}\sqrt{C_u/\sigma_\xi^2} + 4\sqrt{2}\delta r d_1 \sqrt{\log(1/\delta)/\sigma_\xi^2}. \tag{S.22}$$

步骤 2：对求和下界。在一定的正则条件下，以下引理 12 刻画了量 $\sum_{i\in[n]} A_i$ 作为 M的函数求和的形式。引理 13 通过为 π 分配适当的先验分布 M，对函数求和进行下界估计。引理 12 和 13 的证明可以在附录 E.8中找到。

引理 12。如果对于每一个 $(i,j) \in [d_1]\times[r]$，$\log f_M(Z)$ 关于 M_{ij}和 $\left|\frac{\partial}{\partial M_{ij}} \log f_M(Z)\right| < h_{ij}(Z)$ 是连续可微的，并且满足 $\mathbb{E}\,|h_{ij}(Z)A(M)_{ij}| < \infty$，则我们有

$$\sum_{i\in[n]} \mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z)) = \sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} \mathbb{E}A(Z)_{ij}.$$

词条 12在其一般形式中陈述于定理2。1,Cai等人 (2023). 让 g_{ij}是一个定义函数 $g_{ij}(M) := \left(\mathbb{E}_{Z|M} A(Z)\right)_{ij}$对所有 $(i,j) \in [d_1]\times[r]$，则

$$\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} \mathbb{E}A(Z)_{ij} = \mathbb{E}_\pi \left(\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} g_{ij} \right).$$

Lemma 13 lower bounds this quantity by assigning the prior distribution π to M such that $M_{ij} \sim \mathcal{N}(0,1)$ for all $(i,j) \in [d_1] \times [r]$ and otherwise, $M_{ij} = 0$.

Lemma 13. *Let $M \in \mathbb{M}_{r,d_1}$ be distributed according to a density π whose marginal densities are $\{\pi_{ij}\}$ for $i = 1, \cdots, d_1$ and $j = 1, \cdots, d_2$ such that $\pi_{ij} \sim \mathcal{N}(0,1)$ for all $(i,j) \in [d_1] \times [r]$, and otherwise, π_{ij} be the density function such that $\mathbb{P}(M_{ij} = 0) = 1$. Then,*

$$\mathbb{E}_\pi \left(\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} g_{ij} \right) \geq \sum_{(i,j)\in[d_1]\times[r]} \mathbb{E}_{\pi_{ij}} M_{ij}^2 - \sqrt{C} \sqrt{\mathbb{E}_{\pi_{ij}} M_{ij}^2} = rd_1 - \sqrt{Crd_1} \gtrsim rd_1.$$

Combining Lemma 12 and 13, we obtain

$$\sum_{i\in[n]} \mathbb{E} A_i = \sum_{i\in[n]} \mathbb{E}\, \mathcal{A}_M\left(z_i, A(Z)\right) \gtrsim rd_1. \tag{S.23}$$

Step 3: combining the upper and lower bounds. Combining the lower bound (S.23) of $\sum_{i\in[n]} \mathbb{E} A_i$ and the upper bound (S.22) of $\sum_{i\in[n]} \mathbb{E} A_i$, we have

$$2n\varepsilon \sqrt{\mathbb{E}_\pi \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2} \sqrt{C_u/\sigma_\xi^2} \gtrsim rd_1 - 4\sqrt{2}\delta r d_1 \sqrt{\log(1/\delta)/\sigma_\xi^2}.$$

Under the assumption that $\delta < n^{-(1+\gamma)}$ for some $\gamma > 0$, we have $rd_1 - 4\sqrt{2}\delta r d_1 \sqrt{\log(1/\delta)/\sigma_\xi^2} \gtrsim rd_1$, and therefore $\mathbb{E}_\pi \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}$. Since the sup-risk is greater than the Bayesian risk,

$$\sup_{M\in\mathbb{M}_{r,d_1}} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}.$$

Furthermore, due to $\mathbb{M}_{r,d_1} \subset \bigcup_{k=1}^r \mathbb{M}_k$, we have $\sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}$ and

$$\inf_A \sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2},$$

where A is an (ε,δ)-DP algorithm that satisfies $\mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \lesssim 1$. This conclusion extends to all differentially private A if we assume that $rd_1 \lesssim n\varepsilon$ such that $\frac{r^2 d_1^2}{n^2\varepsilon^2} \lesssim 1$.

Remark 1. *Lemma 11 and 13 are also applicable to the case where the parameter space is M_{r,d_2}. For $M \in \mathbb{M}_{r,d_2}$, Lemma 11 implies that*

$$\sum_{i\in[n]} \mathbb{E} A_i \leq 2n\varepsilon \sqrt{\mathbb{E}\|A(Z) - M\|_F^2} \sqrt{C_u/\sigma_\xi^2} + 4\sqrt{2}\delta r d_2 \sqrt{\log(1/\delta)/\sigma_\xi^2};$$

and Lemma 13 results in $\mathbb{E}\left(\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} g_{ij}\right) \gtrsim rd_2$. Therefore, as $\delta < n^{-(1+\gamma)}$ for some $\gamma > 0$, the minimax lower bound

$$\inf_A \sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_2^2}{n^2\varepsilon^2},$$

引理 13 通过分配先验分布来下界这个量 π 给 M 使得$M_{ij} \sim \mathcal{N}(0,1)$ 对于所有 $(i,j) \in [d_1] \times [r]$，否则，$M_{ij} = 0$。

引理 13. 令 $M \in \mathbb{M}_{r\,d_1}$,按照密度 π 分布，其边缘密度为 $\{\pi_{ij}\}$ 对于 $i = 1, \cdots, d_1$ 和 $j = 1 \cdots d_2$，使得 $\pi_{ij} \sim \mathcal{N}(0,1)$ 对于所有 $(i,j) \in [d_1] \times [r]$，否则，$\pi_{ij}$ 是满足 $\mathbb{P}(M_{ij} = 0) = 1$的密度函数。然后，

$$\mathbb{E}_\pi \left(\sum_{(i,j)\in[d_1]\times[r]} \frac{\partial}{\partial M_{ij}} g_{ij} \right) \geq \sum_{(i,j)\in[d_1]\times[r]} \mathbb{E}_{\pi_{ij}} M_{ij}^2 - \sqrt{C} \sqrt{\mathbb{E}_{\pi_{ij}} M_{ij}^2} = rd_1 - \sqrt{Crd_1} \gtrsim rd_1.$$

结合引理 12 和 13，我们得到

$$\sum_{i\in[n]} \mathbb{E} A_i = \sum_{i\in[n]} \mathbb{E}\, \mathcal{A}_M\left(z_i, A(Z)\right) \gtrsim rd_1. \tag{S.23}$$

步骤 3：结合上下界。结合 $\sum^{i\in[n]} \mathbb{E} A_i$ 的下界（S.23）和 $\sum^{i\in[n]} \mathbb{E} A_i$的上界（S.22）的 $\sum_{i\in[n]} \mathbb{E} A_i$，我们有

$$2n\varepsilon \sqrt{\mathbb{E}_\pi \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2} \sqrt{C_u/\sigma_\xi^2} \gtrsim rd_1 - 4\sqrt{2}\delta r d_1 \sqrt{\log(1/\delta)/\sigma_\xi^2}.$$

在假设 $\delta < n^{-(1+\gamma)}$ 对于某些 $\gamma > 0$ 的前提下，我们有 $rd_1 - 4\ \ 2\delta r d_1 \sqrt{\ \log(1/\delta)/\sigma_\xi^2} \gtrsim rd_1$，因此 $\mathbb{E}_\pi \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}$. 由于超风险大于贝叶斯风险，

$$\sup_{M\in\mathbb{M}_{r,d_1}} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}.$$

此外，由于 $\mathbb{M}_{r\,d_1} \subset \ \ _{k=1}^r \mathbb{M}_k$，我们得到 $\sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2}$和

$$\inf_A \sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_1^2}{n^2\varepsilon^2},$$

其中 A 是一个 (ε,δ)-DP 算法，满足 $\mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \lesssim 1$。这一结论扩展到所有差分隐私的 A，如果我们假设 $rd_1 \lesssim n\varepsilon$ 使得 $\frac{r^2 d_1^2}{n^2\varepsilon^2} \lesssim 1$。

注意 1。引理 *11* 和 *13* 也适用于参数空间是 M_{r,d_2} 的情况。对于 $M \in \mathbb{M}_{r\,d_2}$，引理 *11* 意味着

$$\sum_{i\in[n]} \mathbb{E} A_i \leq 2n\varepsilon \sqrt{\mathbb{E}\|A(Z) - M\|_F^2} \sqrt{C_u/\sigma_\xi^2} + 4\sqrt{2}\delta r d_2 \sqrt{\log(1/\delta)/\sigma_\xi^2};$$

和引理 *13* 结果为 $\mathbb{E}\sum_{(i,j)\in\left([d_1]\times[r]\right.} \frac{\partial}{\partial M_{ij}} g_{ij} \quad \gtrsim rd_2$。因此，作为 $\delta < n^{-(1+\gamma)}$ 对于某些$\gamma > 0$，最小最大下界

$$\inf_A \sup_{M\in\bigcup_{k=1}^r \mathbb{M}_k} \mathbb{E}_{Z|M} \|A(Z) - M\|_F^2 \gtrsim \sigma_\xi^2 \cdot \frac{r^2 d_2^2}{n^2\varepsilon^2},$$

where A is an (ε, δ)-DP algorithm that satisfies $\mathbb{E}_{Z|M}\|A(Z)-M\|_F^2 \lesssim 1$. Similarly, this conclusion extends to all differentially private A if we assume that $rd_2 \lesssim n\varepsilon$ such that $\frac{r^2 d_2^2}{n^2\varepsilon^2} \lesssim 1$.

□

E Proofs of Technical Lemmas

E.1 Proof of Lemma 1

Lemma 1 is a consequence of Proposition 10.4, Vershynin (2015) and Lemma 1, Chen et al. (2019) by setting $c_l^2 = C_l$ and $c_u^2 = C_u$ and $\tau^2 \asymp \frac{\sqrt{C_u}}{\sqrt{C_l}}$. See the definiton of c_l^2 and c_u^2 in Lemma 1, Chen et al. (2019).

E.2 Proof of Lemma 2

The proof of Lemma 2 involves applying the symmetric dilation trick to Theorem 1, Xia (2021).

Lemma 14 (Theorem 1, Xia (2021))**.** *Let $B \in d \times d$ be a rank-r symmetric matrix with eigendecomposition of the form $B = \Theta\Lambda\Theta^\top$ where $\Theta \in \mathbb{O}_{d,r}$ and the diagonal matrix $\Lambda = \{\lambda_1, \cdots, \lambda_r\}$ has the eigenvalues of B arranging in the non-increasing order. Let $\widehat{B} = B + \Delta_B$ be another $d \times d$ symmetric matrix and leading r eigen vector of $\widehat{B}$ is given by $\widehat{\Theta} \in \mathbb{O}_{r,d}$. Then,*

$$\widehat{\Theta}\widehat{\Theta}^\top - \Theta\Theta^\top = \sum_{k\geq 1} \mathcal{S}_{B,k}(\Delta_B),$$

where the k-th order term $\mathcal{S}_{M_*,k}(\Delta)$ is a summation of $\binom{2k}{k}$ terms defined by

$$\mathcal{S}_{B,k}(\Delta_B) = \sum_{\mathbf{s}: s_1+\ldots+s_{k+1}=k} (-1)^{1+\tau(\mathbf{s})} \cdot Q^{-s_1}\Delta_B Q^{-s_2} \ldots \Delta_B Q^{-s_{k+1}},$$

where $\mathbf{s} = (s_1, \ldots, s_{k+1})$ contains non-negative indices and $\tau(\mathbf{s}) = \sum_{j=1}^{k+1} \mathbb{I}(s_j > 0)$.

Lemma 14 provides an explicit representation formula for the spectral projector $\widehat{\Theta}\widehat{\Theta}^\top$ given that B is symmetric and of rank-r. Since we are interested in the asymmetric rank-r matrix $M = U\Sigma V^\top \in \mathbb{R}^{d_1\times d_2} \in \mathbb{R}^{d_1\times d_2}$, we apply the symmetric dilation trick to M and obtain the rank-$2r$ symmetric matrix M_* has eigendecomposition of the form

$$M^* = U_{M^*}\Sigma_{M^*}U_{M^*}^\top = \frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}\begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}^\top.$$

E 技术引理的证明

E.1 引理的证明1

引理 1 是命题10的推论。4, Vershynin (2015)和引理1， Chenet al.(2019)通过设置 $c_l^2 = C_l$ 和 $c_u^2 = C_u$ 和 $\tau^2 \asymp \frac{\sqrt{C_u}}{\sqrt{C_l}}$。参见 c_l^2和 c_u^2 在引理1， Chen et al.(2019)中的定义。

E.2 引理的证明2

引理 2 的证明涉及将对称膨胀技巧应用于定理1， Xia (2021)。

引理14（定理1， 夏 (2021)). 设 $B \in d \times d$ 是一个秩-r 对称矩阵，其特征分解形式为 $B = \Theta\Lambda\Theta^\top$，其中 $\Theta \in \mathbb{O}_{d\,r}$，对角矩阵 $\Lambda = \{\lambda_1, \cdots, \lambda_r\}$ 的特征值 B 按非递增顺序排列。设 $\widehat{B} = B + \Delta_B$ 是另一个$d \times d$ 对称矩阵， $\widehat{B}$ 的前导 r 特征向量由 $\widehat{\Theta} \in \mathbb{O}_{r\,d,}$ 给出。则，

$$\widehat{\Theta}\widehat{\Theta}^\top - \Theta\Theta^\top = \sum_{k\geq 1} \mathcal{S}_{B,k}(\Delta_B),$$

其中 k-阶项 $\mathcal{S}_{M_*\,k}(\Delta)$, 是由 $\binom{2k}{k}$ 项定义的和

$$\mathcal{S}_{B,k}(\Delta_B) = \sum_{\mathbf{s}: s_1+\ldots+s_{k+1}=k} (-1)^{1+\tau(\mathbf{s})} \cdot Q^{-s_1}\Delta_B Q^{-s_2} \ldots \Delta_B Q^{-s_{k+1}},$$

其中 $\mathbf{s} = (s_1, \ldots, s_{k+1})$ 包含非负索引和 $\tau(\mathbf{s}) = \sum_{j=1}^{k+1} \mathbb{I}(s_j > 0)$.

引理 14 提供了谱投影器 $\widehat{\Theta}\widehat{\Theta}^\top$ 的显式表示公式，前提是 B 是对称的且秩为 r。由于我们对非对称秩-r 矩阵$M = U\Sigma V^\top \in \mathbb{R}^{d_1\times d_2} \in \mathbb{R}^{d_1\times d_2}$感兴趣，我们将对称膨胀技巧应用于 M 并得到秩-$2r$ 对称矩阵 M_* 具有特征分解形式

$$M^* = U_{M^*}\Sigma_{M^*}U_{M^*}^\top = \frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}\begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}^\top.$$

The proof is finished by applying Lemma 14 with $B = M_*$, $\widehat{B} = M_* + \Delta_*$, $d = d_1 + d_2$, the rank be $2r$ and

$$\Theta = \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \quad \text{and} \quad \widehat{\Theta} = \frac{1}{\sqrt{2}} \begin{pmatrix} \widehat{U} & \widehat{U} \\ \widehat{V} & -\widehat{V} \end{pmatrix}.$$

E.3 Proof of Lemma 3 and 4

See the proof of Lemma 3 in Acharya et al. (2021) and Cai et al. (2024). See the proof of 4 in Shen et al. (2023).

E.4 Proof of Lemma 5

$$\begin{aligned} \max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq \frac{1}{n} \left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\left(\langle X_i, M\rangle + \xi_i\right)\right\| \\ + \max_{i\in[n]} \frac{1}{n} \left\|\mathrm{mat}((\Lambda_i')^{-1}\,\mathrm{vec}(X_i'))\left(\langle X_i', M\rangle + \xi_i\right)\right\|, \end{aligned}$$

where $\|\xi_i\|_{\psi_2} = \sigma_\xi$ and $\langle X_i, M\rangle \sim N\left(0, \mathrm{vec}\,(M)^\top \Lambda_i\,\mathrm{vec}\,(M)\right)$, $\quad \Lambda_i^{-1}\,\mathrm{vec}\,(X_i) \sim N\left(0, \Lambda_i^{-1}\right)$.

$$\begin{aligned} &\left\|\|\xi_i + \langle X_i, M\rangle\,\mathrm{mat}\left(\Lambda_i^{-1}\,\mathrm{vec}\,(X_i)\right) - M\|\right\|_{\Psi_1} \\ &\leq \|\xi_i + \langle X_i, M\rangle\|_{\Psi_2} \left\|\|\,\mathrm{mat}\left(\Lambda_i^{-1}\,\mathrm{vec}\,(X_i)\right)\|\right\|_{\Psi_2} \leq c_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right), \end{aligned}$$

for some absolute constant $c_0 > 0$. Therefore, for some absolute constant $C_3 > 0$, with probability at least $1 - n^{-10}$,

$$\left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\left(\langle X_i, M\rangle + \xi_i\right)\right\| \leq C_3\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\log n.$$

Taking maximum over n, with probability at least $1 - n^{-9}$

$$\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq C_3 \cdot n^{-1}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n.$$

In (S.1), we have already shown that that for some absolute constant $C_1 > 0$,

$$\|\Delta\| = \|\widehat{L} - M\| \leq C_1\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1 + d_2)}{n}}, \tag{S.24}$$

with probability at least $1 - (d_1 + d_2)^{-10}$. Note that for all $i \in [n]$, $\|\Delta^{(i)}\| = \|\Delta - (\Delta - \Delta^{(i)})\| \leq \|\Delta\| + \|\Delta - \Delta^{(i)}\|$, and thus

$$\|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq 2\|\Delta\| + \max_{i\in[n]} \|\Delta - \Delta^{(i)}\|.$$

通过应用引理 14，证明完成 $B = M_*$，$\widehat{B} = M_* + \Delta_*$，$d = d_1 + d_2$，秩为 $2r$ 和

$$\Theta = \frac{1}{\sqrt{2}} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \quad \text{and} \quad \widehat{\Theta} = \frac{1}{\sqrt{2}} \begin{pmatrix} \widehat{U} & \widehat{U} \\ \widehat{V} & -\widehat{V} \end{pmatrix}.$$

E.3 引理的证明3 和 4

参见引理 3 的证明 在 (2021)和 Cai等人 (2024)的证明，参见引理 4 的证明 在(2023)。

E.4 引理的证明5

$$\begin{aligned} \max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq \frac{1}{n} \left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\left(\langle X_i, M\rangle + \xi_i\right)\right\| \\ + \max_{i\in[n]} \frac{1}{n} \left\|\mathrm{mat}((\Lambda_i')^{-1}\,\mathrm{vec}(X_i'))\left(\langle X_i', M\rangle + \xi_i\right)\right\|, \end{aligned}$$

where $\|\xi_i\|_{\psi_2} = \sigma_\xi$ and $\langle X_i, M\rangle \sim N\left(0, \mathrm{vec}\,(M)^\top \Lambda_i\,\mathrm{vec}\,(M)\right)$, $\quad \Lambda_i^{-1}\,\mathrm{vec}\,(X_i) \sim N\left(0, \Lambda_i^{-1}\right)$.

$$\begin{aligned} &\left\|\|\xi_i + \langle X_i, M\rangle\,\mathrm{mat}\left(\Lambda_i^{-1}\,\mathrm{vec}\,(X_i)\right) - M\|\right\|_{\Psi_1} \\ &\leq \|\xi_i + \langle X_i, M\rangle\|_{\Psi_2} \left\|\|\,\mathrm{mat}\left(\Lambda_i^{-1}\,\mathrm{vec}\,(X_i)\right)\|\right\|_{\Psi_2} \leq c_0\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right), \end{aligned}$$

对于某个绝对常数 $c_0 > 0$。因此，对于某个绝对常数 $C_3 > 0$，以至少 $1 - n^{-10}$ 的概率，

$$\left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\left(\langle X_i, M\rangle + \xi_i\right)\right\| \leq C_3\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\log n.$$

在 n 上取最大值，以至少 $1 - n^{-9}$ 的概率

$$\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq C_3 \cdot n^{-1}\sqrt{C_l^{-1}}\left(\sqrt{C_u r}\sigma_1 + \sigma_\xi\right)\log n.$$

在（S.1）中，我们已经证明了对于某些绝对常数 $C_1 > 0$，

$$\|\Delta\| = \|\widehat{L} - M\| \leq C_1\sqrt{C_l^{-1}}\left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right)\sqrt{\frac{(d_1 \vee d_2)\log(d_1 + d_2)}{n}}, \tag{S.24}$$

以至少 $1 - (d_1 + d_2)^{-10}$ 的概率。请注意对于所有 $i \in [n]$ $\|\Delta^{(i)}\| = \|\Delta - (\Delta - \Delta^{(i)})\| \leq$，$\|\Delta\| + \|\Delta - \Delta^{(i)}\|$，因此

$$\|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq 2\|\Delta\| + \max_{i\in[n]} \|\Delta - \Delta^{(i)}\|.$$

As long as the sample size $n \geq \frac{\log^2 n}{(d_1 \vee d_2)\log(d_1+d_2)}$, there exists an absolute constant $C_0 > 0$ such that

$$\|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq C_0 \sqrt{C_l^{-1}} \left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right) \sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}},$$

with probability at least $1 - (d_1+d_2)^{-10} - n^{-9}$.

E.5 Proof of Lemma 6

By (Pajor, 1998, Proposition 8) and (Koltchinskii and Xia, 2015, Lemma 5), for any $q \in [1,\infty]$, there exists an absolute constant $c' > 0$ and a subset $\mathcal{S}_q^{(d-r)} \subset \mathbb{O}_{d-r,r}$ such that for any $V_i \neq V_j \in \mathcal{S}_q^{(d-r)}$, $\|V_iV_i^\top - V_jV_j^\top\|_q \geq c' r^{1/q}$, and the cardinality of $\mathcal{S}_q^{(d-r)}$ is at least $2^{r(d-r)}$. Here, $\|\cdot\|_q$ denotes the Schatten-q norm of a matrix. In particular, spectral norm is Schatten-∞ norm, Frobenius norm is Schatten-2 norm, and nuclear norm is Schatten-1 norm. Let $\varepsilon_0 > 0$ be a small number to be decided later. Now, for each $V \in \mathcal{S}_q^{(d-r)}$, we define

$$U = \begin{pmatrix} \sqrt{1-\varepsilon_0^2} I_r \\ \sqrt{\varepsilon_0^2} V \end{pmatrix}$$

such that $U \in \mathbb{R}^{d\times r}$ and $U^\top U = I_r$. This means that, for any $V \in \mathcal{S}_q^{(d-r)}$, we can construct a $U \in \mathbb{O}_{d,r}$. This defines a subset $\mathcal{S}_q^{(d)} \subset \mathbb{O}_{d,r}$ with $\mathrm{Card}\big(\mathcal{S}_q^{(d)}\big) \geq 2^{r(d-r)}$ such that for any $U_i \neq U_j \in \mathcal{S}_q^{(d)}$,

$$\|U_iU_i^\top - U_jU_j^\top\|_q \geq \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_i - V_j\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_iV_i^\top - V_jV_j^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)} r^{1/q}$$

and, meanwhile,

$$\|U_iU_i^\top - U_jU_j^\top\|_{\mathrm{F}} \lesssim \|U_i - U_j\|_{\mathrm{F}} \leq \varepsilon_0\|V_i - V_j\|_{\mathrm{F}} \leq \sqrt{2r}\varepsilon_0.$$

E.6 Proof of Lemma 7, 8 and 9

See the proof of Lemma 7 in Vershynin (2018), Lemma 8 in Shen et al. (2023) and Lemma 9 in Wei et al. (2016) and Luo and Zhang (2022).

E.7 Proof of Lemma 10

Since Z_i' is independent of z_i and $\mathbb{E}\,\mathcal{S}_M(z_i) = \mathbb{E}\,\nabla_M f_M(y_i|X_i) = 0$,

$$\mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z_i')) = \mathbb{E}\left\langle [A(Z_i') - M]_{[d_1]\times[r]}, \mathcal{S}_M(z_i)\right\rangle = \left\langle \mathbb{E}\,[A(Z_i') - M]_{[d_1]\times[r]}, \mathbb{E}\,\mathcal{S}_M(z_i)\right\rangle = 0,$$

只要样本量 $n \geq \frac{\log^2 n}{(d_1 \vee d_2)\log(d_1+d_2)}$，就存在一个绝对常数 $C_0 > 0$ 使得

$$\|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq C_0 \sqrt{C_l^{-1}} \left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right) \sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}},$$

以至少 $1 - (d_1+d_2)^{-10} - n^{-9}$ 的概率。

E.5 引理的证明6

由 (Pajor, 1998, 定理 8) 和 (Koltchinskii and Xia, 2015, 引理 5) 知，对于任何 $q \in [1,\infty]$，存在一个绝对常数 $c' > 0$ 和一个子集 $\mathcal{S}_q^{(d-r)} \subset \mathbb{O}_{d-r\,r}$,使得对于任何 $V_i = V_j \in \mathcal{S}_q^{(d-r)}$，$\|V_iV_i^\top - V_jV_j^\top\|_q \geq c' r^{1/q}$，以及 $\mathcal{S}_q^{(d-r)}$ 的基数至少为 $2^{r(d-r)}$。这里，$\|\cdot\|_q$ 表示矩阵的 Schatten-q 范数。特别是，谱范数是 Schatten-∞ 范数，Frobenius 范数是 Schatten-2 范数，核范数是 Schatten-1 范数。令 $\varepsilon_0 > 0$ 是一个稍后将要确定的数。现在，对于每个 $V \in \mathcal{S}_q^{(d-r)}$，我们定义

$$U = \begin{pmatrix} \sqrt{1-\varepsilon_0^2} I_r \\ \sqrt{\varepsilon_0^2} V \end{pmatrix}$$

使得 $U \in \mathbb{R}^{d\times r}$ 和 $U^\top U = I_r$。这意味着，对于任何 $V \in \mathcal{S}_q^{(d-r)}$，我们可以构造一个 $U \in \mathbb{O}_{d\,r,}$。这定义了一个子集 $\mathcal{S}_q^{(d)} \subset \mathbb{O}_{d,\,r}$，其$\mathrm{Card}\big(\mathcal{S}_q^{(d)}\big) \geq 2^{r(d-r)}$，使得对于任何 $U_i \neq U_j \in \mathcal{S}_q^{(d)}$,

$$\|U_iU_i^\top - U_jU_j^\top\|_q \geq \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_i - V_j\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)}\|V_iV_i^\top - V_jV_j^\top\|_q \gtrsim \sqrt{\varepsilon_0^2(1-\varepsilon_0^2)} r^{1/q}$$

同时,

$$\|U_iU_i^\top - U_jU_j^\top\|_{\mathrm{F}} \lesssim \|U_i - U_j\|_{\mathrm{F}} \leq \varepsilon_0\|V_i - V_j\|_{\mathrm{F}} \leq \sqrt{2r}\varepsilon_0.$$

E.6 引理的证明 7， 8 和9

参见引理 7 的证明 Vershynin (2018)，引理 8 在 Shen等人 (2023)和引理 9 在Wei等人。(2016)中，以及 Luo和Zhang (2022)。

E.7 引理的证明10

由于 Z_i'独立于 z_i和 $\mathbb{E}\,\mathcal{S}_M(z_i) = \mathbb{E}\,\nabla_M f_M(y_i|X_i) = 0$,

$$\mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z_i')) = \mathbb{E}\left\langle [A(Z_i') - M]_{[d_1]\times[r]}, \mathcal{S}_M(z_i)\right\rangle = \left\langle \mathbb{E}\,[A(Z_i') - M]_{[d_1]\times[r]}, \mathbb{E}\,\mathcal{S}_M(z_i)\right\rangle = 0,$$

As for $\mathbb{E}A_i = \mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z))$, we apply Jensen's inequality and have

$$\begin{aligned}
&\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right| \le \sqrt{\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right|^2} \\
&\le \sqrt{\mathbb{E}\operatorname{vec}\left([A(Z_i') - M]_{[d_1]\times[r]}\right)^\top \operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top \operatorname{vec}\left([A(Z_i') - M]_{[d_1]\times[r]}\right)} \qquad \text{(S.25)} \\
&\le \sqrt{\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\|} \cdot \sqrt{\mathbb{E}\left\|[A(Z_i') - M]_{[d_1]\times[r]}\right\|_F^2},
\end{aligned}$$

where the second line is due to $\langle B, C\rangle = \operatorname{vec}(B)^\top \operatorname{vec}(C)$ and the last inequality is because Z_i' is independent of z_i. By the definition of $\mathcal{S}_M(z_i) = \frac{1}{\sigma_\xi^2}(y_i - \langle X_i, M\rangle)X_i$ and the independence between ξ_i and X_i,

$$\left\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\right\| = \left\|\mathbb{E}\left(\frac{y_i - \langle X_i, M\rangle}{\sigma_\xi^2}\right)^2 \mathbb{E}\operatorname{vec}(X_i)\operatorname{vec}(X_i)^\top\right\| = \frac{1}{\sigma_\xi^2}\|\Lambda_i\| \le \frac{C_u}{\sigma_\xi^2}.$$

Plugging the upper bound of $\left\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\right\|$ into (S.25),

$$\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right| \le \sqrt{\mathbb{E}\left\|A(Z) - M\right\|_F^2}\sqrt{\frac{C_u}{\sigma_\xi^2}},$$

E.8 Proof of Lemma 11, Proof of Lemma 12, Lemma 13

Lemma 11 is a trivial consequence by setting $T = \sqrt{2/\sigma_\xi^2 r d_1}\sqrt{\log(\frac{1}{\delta})}$ to Proposition 2.1, Cai et al. (2023). Lemma 12 is a trivial consequence of Theorem 2.1, Cai et al. (2023) along with the definition of $\mathcal{A}_M^1(z_i, A(Z))$. Lemma 13 is a trivial consequence of Proposition 2.2, Cai et al. (2023) by taking $M_{ij} \sim \mathcal{N}(0,1)$ for $(i,j) \in [d_1]\times[r]$ and $M_{ij} = 0$ otherwise.

E.9 Proof of Lemma 14

See the proof of Lemma 14 in Xia (2021).

F Weak Differential privacy

This section proposes a weaker definition than differential privacy such that the sensitivities are *free of* $\{X_i\}_{i\in[n]}$.

Definition 2 (*weak* (ε,δ)-differential privacy)**.** *Let Z be a given data set and Z' be a weak neighbouring data set of Z, i.e., Z and Z' differs by at most one pair of observations $z \in Z$ and*

至于 $\mathbb{E}A_i = \mathbb{E}\,\mathcal{A}_M^1(z_i, A(Z))$，我们应用Jensen不等式，得到

$$\begin{aligned}
&\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right| \le \sqrt{\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right|^2} \\
&\le \sqrt{\mathbb{E}\operatorname{vec}\left([A(Z_i') - M]_{[d_1]\times[r]}\right)^\top \operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top \operatorname{vec}\left([A(Z_i') - M]_{[d_1]\times[r]}\right)} \qquad \text{(S.25)} \\
&\le \sqrt{\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\|} \cdot \sqrt{\mathbb{E}\left\|[A(Z_i') - M]_{[d_1]\times[r]}\right\|_F^2},
\end{aligned}$$

其中第二行是由于 $\langle B\ C\rangle = \operatorname{vec}(B)^\top \operatorname{vec}(C)$，最后的不等式是因为 Z_i'独立于 z_i。根据 $\mathcal{S}_M(z_i) = \frac{1}{\sigma_\xi^2}(y_i - \langle X_i, M\rangle)X_i$的定义以及 ξ_i 和 X_i的独立性，

$$\left\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\right\| = \left\|\mathbb{E}\left(\frac{y_i - \langle X_i, M\rangle}{\sigma_\xi^2}\right)^2 \mathbb{E}\operatorname{vec}(X_i)\operatorname{vec}(X_i)^\top\right\| = \frac{1}{\sigma_\xi^2}\|\Lambda_i\| \le \frac{C_u}{\sigma_\xi^2}.$$

将 $\left\|\mathbb{E}\operatorname{vec}(\mathcal{S}_M(z_i))\operatorname{vec}(\mathcal{S}_M(z_i))^\top\right\|$的上界代入（S.25），

$$\mathbb{E}\left|\mathcal{A}_M^1(z_i, A(Z_i'))\right| \le \sqrt{\mathbb{E}\left\|A(Z) - M\right\|_F^2}\sqrt{\frac{C_u}{\sigma_\xi^2}},$$

E.8 引理11的证明，引理12的证明，引理13

引理11是通过对 $T = \sqrt{2/\sigma_\xi^2 r d_1}\sqrt{\log(\frac{1}{\delta})}$设为命题2的平凡推论1，Cai等人(2023).引理12 是定理2.1的平凡推论，Cai等人(2023)以及 $\mathcal{A}_M^1(z_i, A(Z))$的定义的平凡推论。引理13 是命题2的平凡推论2，Cai等人(2023)通过对 $M_{ij} \sim \mathcal{N}(0,1)$ 设为 $(i,j) \in [d_1]\times[r]$，否则为 $M_{ij} = 0$。

E.9 引理的证明14

参见引理的证明14 在夏(2021).

F 弱差分隐私

本节提出一种比差分隐私更弱的定义，使得灵敏度不受 $\{X_i\}_{i\in[n]}$ 影响。

定义 2 (弱 (ε,δ)-差分隐私)。设 Z 是一个给定数据集，Z' 是 Z 的一个弱邻近数据集，即 Z 和 Z' 最多相差一对观测值 $z \in Z$ 和

$z' \in Z'$ *sharing the same measurement* X*. The algorithm* A *that maps* Z *into* $\mathbb{R}^{d_1 \times d_2}$ *is weak* (ε, δ)*-differentially private over the dataset* Z *if*

$$\mathbb{P}\big(A(Z) \in \mathcal{Q}\big) \leq e^{\varepsilon}\mathbb{P}\big(A(Z') \in \mathcal{Q}\big) + \delta, \tag{S.26}$$

for all weak neighbouring data set Z, Z' *and all subset* $\mathcal{Q} \subset \mathbb{R}^{d_1 \times d_2}$*.*

Compared to the standard (ε, δ)-DP, *weak* (ε, δ)-differential privacy is a less powerful constraint. Definition 2 only requires the algorithm A to preserve the property (S.26) over weak neighbouring datasets, i.e., datasets that differs by at most one pair of observations sharing the same measurement X. As we consider a pair of observations $z = (X, y)$ and $z' = (X, y')$ under the model (1), where $y = \langle X, M\rangle + \xi$ and $y' = \langle X, M\rangle + \xi'$, the difference $y - y' = \xi - \xi'$ is free of the measurement X.

Next, we list the Theorem 6, Corollary 2 and Theorem 7 as the analogues of Theorem 1, Corollary 1 and Theorem 3. All proofs for this section are deferred to the end of this section.

Theorem 6 (Weak DP and utility guarantees of the initialization $\widetilde{M}_0$)**.** *Consider i.i.d. observations* $Z = \{z_1, \cdots, z_n\}$ *drawn from the trace regression model stated in* (1) *where* $z_i := (X_i, y_i)$ *for* $i = 1, \cdots, n$*. Let the true low-rank regression coefficients matrix being* $M \in \mathbb{M}_r$*. Suppose that* $\{X_i\}_{i\in[n]}$ *satisfy the Assumption 1. Under the mild condition* $n \geq \frac{\sigma_\xi}{\sigma_\xi + \sqrt{C_u r}\sigma_1}$*, there exists absolute constants* $C_1, C_2, C_3 > 0$ *such that as the sample size* $n \geq n_0$*, the sensitivity for leading* r *left and right singular vectors takes the value*

$$\Delta^{(1)}_{weak} := \max_{i\in[n]} \left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F \vee \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right) = C_2\sqrt{C_l^{-1}}\frac{\sigma_\xi}{\sigma_r}\frac{\sqrt{r}}{n}\log n;$$

the sensitivity for the r *singular values takes the value*

$$\Delta^{(2)}_{weak} := \max_{i\in[n]} \left\|\widetilde{U}^\top\left(\widehat{L} - \widehat{L}^{(i)}\right)\widetilde{V}\right\|_F = C_2\sqrt{C_l^{-1}}\sigma_\xi\frac{\sqrt{r}}{n}\log n,$$

and Algorithm 1 is weak (ε, δ)*-differentially private. Moreover,*

$$\|\widetilde{M}_0^{weak} - M\| \bigvee \left(\|\widetilde{M}_0^{weak} - M\|_F/\sqrt{2r}\right) \leq e_1 + \underbrace{C_3\sqrt{C_l^{-1}}\sigma_\xi\left(\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon} + \frac{r}{n\varepsilon}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})}_{e_2^{weak}},$$

with probability at least $1 - (d_1 + d_2)^{-10} - n^{-9} - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$*.*

Theorem 6 requires the same sample size condition $n \geq n_0$ as Theorem 1, however, the sensitivities $\Delta^{(1)}_{weak}$ and $\Delta^{(2)}_{weak}$ derived under weak DP, differs with their DP counterpart $\Delta^{(1)}$ and $\Delta^{(2)}$ by the factor $\sqrt{C_u r}\sigma_1$. This leads to a smaller cost of privacy e_2^{weak} than the cost of privacy e_2 we obtained under stronger standard DP-constraints, as presented in Theorem 1.

$z' \in Z'$ 共享相同的测量 X。将 Z 映射到 $\mathbb{R}^{d_1 \times d_2}$ 的算法 A 在数据集 Z 上是弱(ε, δ)-差分隐私的，如果

$$\mathbb{P}\big(A(Z) \in \mathcal{Q}\big) \leq e^{\varepsilon}\mathbb{P}\big(A(Z') \in \mathcal{Q}\big) + \delta, \tag{S.26}$$

对于所有弱相邻数据集 Z, Z' 和所有子集 $\mathcal{Q} \subset \mathbb{R}^{d_1 \times d_2}$。

与标准(ε, δ)-DP，弱(ε, δ)-差分隐私是一种约束力较弱的约束。定义2 仅要求算法 A 在弱相邻数据集上保留属性(S.26)，即最多有一对共享相同测量的观测值不同的数据集 X。在模型(1)下，我们考虑一对观测值 $z = (X, y)$ 和 $z' = (X, y')$，其中 $y = \langle X, M\rangle + \xi$和 $y' = \langle X\ M\rangle + \xi'$，，，差异 $y - y' = \xi - \xi'$不受测量 X的影响。

接下来，我们列出定理 6，推论 2 和定理 7 作为定理 1，推论 1 和定理 3的类似物。本节的全部证明都推迟到本节末尾。

定理 6 (弱 DP 和初始化效用保证 $\widetilde{M}_0$)。考虑独立同分布观测 $Z = \{z_1, \cdots, z_n\}$ 从 (1) 中抽取的，其中 $z_i := (X_i, y_i)$对于 $i = 1, \cdots, n$。设真实低秩回归系数矩阵为 $M \in \mathbb{M}_r$。假设 $\{X_i\}_{i\in[n]}$ 满足假设 1。在温和条件 $n \geq \frac{\sigma_\xi}{\sigma_\xi + \sqrt{C_u r}\sigma_1}$下，存在绝对常数 $C_1\ C_2\ C_3 > 0$，，使得当样本大小 $n \geq n_0$，主左和右奇异向量的敏感性取值为

$$\Delta^{(1)}_{weak} := \max_{i\in[n]} \left(\|\widehat{U}\widehat{U}^\top - \widehat{U}^{(i)}\widehat{U}^{(i)\top}\|_F \vee \|\widehat{V}\widehat{V}^\top - \widehat{V}^{(i)}\widehat{V}^{(i)\top}\|_F\right) = C_2\sqrt{C_l^{-1}}\frac{\sigma_\xi}{\sigma_r}\frac{\sqrt{r}}{n}\log n;$$

对于 r 奇异值的灵敏度取值

$$\Delta^{(2)}_{weak} := \max_{i\in[n]} \left\|\widetilde{U}^\top\left(\widehat{L} - \widehat{L}^{(i)}\right)\widetilde{V}\right\|_F = C_2\sqrt{C_l^{-1}}\sigma_\xi\frac{\sqrt{r}}{n}\log n,$$

并且算法 1 较弱 (ε, δ)-差分隐私。此外，

$$\|\widetilde{M}_0^{weak} - M\| \bigvee \left(\|\widetilde{M}_0^{weak} - M\|_F/\sqrt{2r}\right) \leq e_1 + \underbrace{C_3\sqrt{C_l^{-1}}\sigma_\xi\left(\frac{\sigma_1}{\sigma_r}\frac{\sqrt{r(d_1 \vee d_2)}}{n\varepsilon} + \frac{r}{n\varepsilon}\right)\log n \log^{\frac{1}{2}}(\frac{3.75}{\delta})}_{e_2^{weak}},$$

以至少 $1 - (d_1 + d_2)^{-10} - n^{-9} - \exp(-d_1) - \exp(-d_2) - 10^{-20r}$的概率。

定理 6 要求相同的样本量条件 $n \geq n_0$ 与定理 1相同，然而，在弱DP下的灵敏度 $\Delta^{(1)}_{weak}$ 和 $\Delta^{(2)}_{weak}$ 与DP对应项 $\Delta^{(1)}$ 和 $\Delta^{(2)}$ 相差因子 $C_u r\sigma_1$。这导致隐私成本 e_2^{weak}低于在更强的标准DP约束下获得的隐私成本 e_2，如定理 1所述。

Corollary 2. *Under the conditions stated in Theorem 6, as the sample size is sufficiently large such that for some absolute constant $c_2 > 0$,*

$$n \geq C_1 \max\left\{n_1, \underbrace{\sqrt{C_l^{-1}}\left(\frac{\sigma_\xi}{\sigma_r}\right)\left(\kappa r\sqrt{d_1 \vee d_2} + r^{\frac{3}{2}}\right)\log n \frac{\log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)}{\varepsilon}}_{n_2^{weak}}\right\},$$

we have for some small constant $0 < c_0 < 1$, $\|\widetilde{M}_0 - M\|_F \leq \sqrt{2r}\|\widetilde{M}_0 - M\| \leq c_0\sigma_r$.

Compared with Corollary 1, Corollary 2 requires smaller sample size as $n_2^{weak} \leq n_2$.

Theorem 7. *Consider i.i.d. observations $Z = \{z_1, \cdots, z_n\}$ drawn from the trace regression model stated in (1) where the true low-rank regression coefficients matrix being $M \in \mathbb{M}_r$. Here, $z_i := (X_i, y_i)$ for $i = 1, \cdots, n$ and we assume that $\{X_i\}_{i\in[n]}$ satisfy the Assumption 1 and $(d_1 + d_2) > \log n$. Suppose the weak (ε, δ)-DP initialization satisfies 2, then Algorithm 2 is weak $(2\varepsilon, 2\delta)$-differentially private with the sensitivities*

$$\Delta_l = C_3\frac{\eta}{n}\sigma_\xi\sqrt{C_u r(d_1 + d_2)\log n},$$

for some absolute constant $C_3 > 0$. Moreover, as the sample size

$$n \geq c_4 \max\left\{n_3, n_4, \underbrace{\eta\sqrt{C_u}\kappa_\xi r(d_1 + d_2)\log^{3/2}(n)\frac{\log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}{\varepsilon}}_{n_5^{weak}}\right\},$$

for some small constant $0 < c_4 < 1$, number of iteration $l^* = O(\log n)$, and the step size $0 < \eta < 1$, we have the output of Algorithm 2 satisfies

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq u_1 + \underbrace{C_4\sqrt{C_u}\sigma_\xi\frac{r(d_1 + d_2)}{n\varepsilon}\log^{3/2} n \log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}_{u_2^{weak}}.$$

with probability at least

$$1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1 + d_2)\right) - (d_1 + d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r}$$
$$- \left((d_1 + d_2)^{-10} + \exp(-(d_1 + d_2)) + n^{-9} + \exp\left(-10C_u(d_1 + d_2)\right)n^{-9}\right)\log n.$$

Theorem 7 shows that as the sample size $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$, the estimator $\widetilde{M}_{l^*}$ given by Algorithm 2 attains the optimal convergence rate $\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}\right)$, in the sense of weak differential privacy.

The proofs of Theorem 6, 7 and Corollary 2 will be a trivial concequence of replacing the first part of Lemma 5 by the following Lemma 15

推论 2。 在定理中所述的条件下 6，当样本量足够大，以至于对于某个绝对常数 $c_2 > 0$,

$$n \geq C_1 \max\left\{n_1, \underbrace{\sqrt{C_l^{-1}}\left(\frac{\sigma_\xi}{\sigma_r}\right)\left(\kappa r\sqrt{d_1 \vee d_2} + r^{\frac{3}{2}}\right)\log n}_{n_2^{weak}} \frac{\log^{\frac{1}{2}}\left(\frac{3.75}{\delta}\right)}{\varepsilon}\right\},$$

对于某个小的常数 $0 < c_0 < 1$ $\|\widetilde{M}_0 - M\|_F \leq \sqrt{2r}\|\widetilde{M}_0 - M\| \leq c_0\sigma_r$.,

与推论 1相比，推论 2 要求更小的样本量，因为 $n_2^{weak} \leq n_2$。

定理 7。考虑从（1）中所述的轨迹回归模型独立同分布抽取的观测值 $Z = \{z_1, \cdots, z_n\}$，其中真实的低秩回归系数矩阵为 $M \in \mathbb{M}_r$。这里,$z_i := (X_i, y_i)$ 对于 $i = 1, \cdots, n$，并且我们假设 $\{X_i\}_{i\in[n]}$ 满足假设 1 和 $(d_1 + d_2) > \log n$。假设弱 (ε, δ)-差分隐私初始化满足 2，则算法 2 是弱 $(2\varepsilon, 2\delta)$-差分隐私的，其敏感度为

$$\Delta_l = C_3\frac{\eta}{n}\sigma_\xi\sqrt{C_u r(d_1 + d_2)\log n},$$

对于某个绝对常数 $C_3 > 0$。此外，随着样本量

$$n \geq c_4 \max\left\{n_3, n_4, \underbrace{\eta\sqrt{C_u}\kappa_\xi r(d_1 + d_2)\log^{3/2}(n)}_{n_5^{weak}}\frac{\log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right)}{\varepsilon}\right\},$$

对于某个小常数 $0 < c_4 < 1$，迭代次数 $l^* = O(\log n)$，以及步长$0 < \eta < 1$，我们有算法 2 的输出满足

$$\left\|\widetilde{M}_{l^*} - M\right\|_F \leq u_1 + \underbrace{C_4\sqrt{C_u}\sigma_\xi\frac{r(d_1 + d_2)}{n\varepsilon}\log^{3/2} n}_{u_2^{weak}}\log^{1/2}\left(\frac{1.25\log(n)}{\delta}\right).$$

至少以概率

$$1 - \widetilde{c}_2\exp\left(-\widetilde{c}_3 r(d_1 + d_2)\right) - (d_1 + d_2)^{-10} - n^{-9} - e^{-d_1} - e^{-d_2} - 10^{-20r}$$
$$- \left((d_1 + d_2)^{-10} + \exp(-(d_1 + d_2)) + n^{-9} + \exp\left(-10C_u(d_1 + d_2)\right)n^{-9}\right)\log n.$$

定理 7 表明，随着样本量 $n \gtrsim \widetilde{O}\left(\left(\kappa_\xi^2 \vee \kappa_\xi\right) r(d_1 \vee d_2)\right)$， 估计量 $\widetilde{M}_{l^*}$ 由算法 2 达到最优收敛速度 $\widetilde{O}_p\left(\sigma_\xi\sqrt{\frac{r(d_1+d_2)}{n}} + \sigma_\xi\frac{r(d_1+d_2)}{n\varepsilon}\right)$， 在弱差分隐私的意义下。

定理 6， 7 以及推论 2 的证明将是替换引理 5 的第一部分为以下引理 15的平凡推论

Lemma 15. *Under model (1), Assumption 1, and the condition* $n \geq \frac{\sigma_\xi}{\sigma_\xi + \sqrt{C_u r}\sigma_1}$*, there exists some absolute constant* $C_0, C_1 > 0$ *such that the event*

$$\mathcal{E}_* := \left\{ \max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq C_0 \cdot n^{-1}\sqrt{C_l^{-1}}\sigma_\xi \log n \right\} \bigcap \left\{ \|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq C_0\sqrt{C_l^{-1}} \left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right) \sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}} \right\},$$

holds with probability at least $1 - (d_1+d_2)^{-10} - n^{-9}$.

Proof of Lemma 15. We only need to focus on $\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\|$ since the rest of the proof is the same as Lemma 5.

$$\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq \max_{i\in[n]} \frac{1}{n} \left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\,(\xi_i - \xi_i')\right\|,$$

where $\|\xi_i\|_{\psi_2} = \|\xi_i'\|_{\psi_2} = \sigma_\xi$ and $\Lambda_i^{-1}\,\mathrm{vec}\,(X_i) \sim N\left(0, \Lambda_i^{-1}\right)$ for all $i = 1, \cdots, n$. Therefore, for some absolute constant $c_0 > 0$,

$$\left\|\left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\,(\xi_i - \xi_i')\right\|\right\|_{\Psi_1} \leq c_0\sqrt{C_l^{-1}}\sigma_\xi.$$

We complete the proof by appying tail bound for sub-exponential random variable and taking a maximum over n. □

References

Abowd, J. M., I. M. Rodriguez, W. N. Sexton, P. E. Singer, and L. Vilhuber (2020). The modernization of statistical disclosure limitation at the us census bureau. *US Census Bureau*.

Absil, P.-A., R. Mahony, and R. Sepulchre (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.

Acharya, J., Z. Sun, and H. Zhang (2021). Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pp. 48–78. PMLR.

Adler, R. L., J. Dedieu, J. Y. Margulies, M. Martens, and M. Shub (2002, 07). Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis 22*(3), 359–390.

Apple Differential Privacy Team (2017). Learning with privacy at scale.

引理 15。 在模型（*1*）下，假设 *1*，以及条件 $n \geq \frac{\sigma_\xi}{\sigma_\xi + \sqrt{C_u r}\sigma_1}$，存在某个绝对常数 C_0 $C_1 > 0$,，使得事件

$$\mathcal{E}_* := \left\{ \max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq C_0 \cdot n^{-1}\sqrt{C_l^{-1}}\sigma_\xi \log n \right\} \bigcap \left\{ \|\Delta\| + \max_{i\in[n]} \|\Delta^{(i)}\| \leq C_0\sqrt{C_l^{-1}} \left(\sigma_\xi + \sqrt{C_u}\sqrt{r}\sigma_1\right) \sqrt{\frac{(d_1 \vee d_2)\log(d_1+d_2)}{n}} \right\},$$

以至少 $1 - (d_1+d_2)^{-10} - n^{-9}$ 的概率成立。

引理 15 的证明。 *15*。 我们只需要关注 $\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\|$，因为其余的证明与引理 5相同。

$$\max_{i\in[n]} \left\|\Delta - \Delta^{(i)}\right\| \leq \max_{i\in[n]} \frac{1}{n} \left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\,(\xi_i - \xi_i')\right\|,$$

其中 $\|\xi_i\|_{\psi_2} = \|\xi_i'\|_{\psi_2} = \sigma_\xi$和 $\Lambda_i^{-1}\mathrm{vec}\,(X_i) \sim N\left(0, \Lambda_i^{-1}\right)$ 对所有 $i = 1 \cdots n$, ,。因此，对于某个绝对常数 $c_0 > 0$,

$$\left\|\left\|\mathrm{mat}(\Lambda_i^{-1}\,\mathrm{vec}(X_i))\,(\xi_i - \xi_i')\right\|\right\|_{\Psi_1} \leq c_0\sqrt{C_l^{-1}}\sigma_\xi.$$

我们通过应用次指数随机变量的尾部界限并取 n 的最大值来完成证明。

参考文献

Abowd, J. M., I. M. Rodriguez, W. N. Sexton, P. E. Singer, and L. Vilhuber (2020). The modernization of statistical disclosure limitation at the us census bureau. *US Census Bureau*.

Absil, P.-A., R. Mahony, and R. Sepulchre (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.

Acharya, J., Z. Sun, and H. Zhang (2021). Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pp. 48–78. PMLR.

Adler, R. L., J. Dedieu, J. Y. Margulies, M. Martens, and M. Shub (2002, 07). Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis 22*(3), 359–390.

Apple Differential Privacy Team (2017). Learning with privacy at scale.

Blum, A., C. Dwork, F. McSherry, and K. Nissim (2005, 06). Practical privacy: The sulq framework. *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 128–138.

Brown, G., S. Hopkins, and A. Smith (2023, 12–15 Jul). Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. In G. Neu and L. Rosasco (Eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, Volume 195 of *Proceedings of Machine Learning Research*, pp. 5578–5579. PMLR.

Burer, S. and R. D. Monteiro (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming 95*(2), 329–357.

Cai, T. T., Y. Wang, and L. Zhang (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics 49*, 2825–2850.

Cai, T. T., Y. Wang, and L. Zhang (2023). Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*.

Cai, T. T., D. Xia, and M. Zha (2024). Optimal differentially private pca and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*.

Candes, E. J. and Y. Plan (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory 57*(4), 2342–2359.

Candes, E. J. and T. Tao (2005). Decoding by linear programming. *IEEE transactions on information theory 51*(12), 4203–4215.

Chaudhuri, K., A. Sarwate, and K. Sinha (2012). Near-optimal differentially private principal components. *Advances in Neural Information Processing Systems 25*.

Chen, H., G. Raskutti, and M. Yuan (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *The Journal of Machine Learning Research 20*(1), 172–208.

Chen, Y. and M. J. Wainwright (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees.

Blum, A., C. Dwork, F. McSherry, and K. Nissim (2005, 06). Practical privacy: The sulq framework. *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 128–138.

Brown, G., S. Hopkins, and A. Smith (2023, 12–15 Jul). Fast, sample-efficient, affine-invariant private mean and covariance estimation for subgaussian distributions. In G. Neu and L. Rosasco (Eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, Volume 195 of *Proceedings of Machine Learning Research*, pp. 5578–5579. PMLR.

Burer, S. and R. D. Monteiro (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming 95*(2), 329–357.

Cai, T. T., Y. Wang, and L. Zhang (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics 49*, 2825–2850.

Cai, T. T., Y. Wang, and L. Zhang (2023). Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*.

Cai, T. T., D. Xia, and M. Zha (2024). Optimal differentially private pca and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*.

Candes, E. J. and Y. Plan (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory 57*(4), 2342–2359.

Candes, E. J. and T. Tao (2005). Decoding by linear programming. *IEEE transactions on information theory 51*(12), 4203–4215.

Chaudhuri, K., A. Sarwate, and K. Sinha (2012). Near-optimal differentially private principal components. *Advances in Neural Information Processing Systems 25*.

Chen, H., G. Raskutti, and M. Yuan (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *The Journal of Machine Learning Research 20*(1), 172–208.

Chen, Y. and M. J. Wainwright (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees.

Chien, S., P. Jain, W. Krichene, S. Rendle, S. Song, A. Thakurta, and L. Zhang (2021). Private alternating least squares: Practical private matrix completion with tighter rates. pp. 1877–1887.

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference*, 265–284.

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.

Dwork, C., K. Talwar, A. Thakurta, and L. Zhang (2014). Analyze gauss: optimal bounds for privacy-preserving principal component analysis. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 11–20.

Edelman, A., T. A. Arias, and S. T. Smith (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications 20*(2), 303–353.

Erlingsson, U., V. Pihur, and A. Korolova (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, New York, NY, USA, pp. 1054–1067. Association for Computing Machinery.

Hamidi, N. and M. Bayati (2022). On low-rank trace regression under general sampling distribution. *Journal of Machine Learning Research 23*(321), 1–49.

Jain, P., O. D. Thakkar, and A. Thakurta (2018). Differentially private matrix completion revisited. pp. 2215–2224.

Kamath, G., J. Li, V. Singhal, and J. Ullman (2019). Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pp. 1853–1902. PMLR.

Kamath, G., V. Singhal, and J. Ullman (2020, 09–12 Jul). Private mean estimation of heavy-tailed distributions. In J. Abernethy and S. Agarwal (Eds.), *Proceedings of Thirty Third Conference on Learning Theory*, Volume 125 of *Proceedings of Machine Learning Research*, pp. 2204–2235. PMLR.

Koltchinskii, V. (2011). Von neumann entropy penalization and low-rank matrix estimation.

Chien, S., P. Jain, W. Krichene, S. Rendle, S. Song, A. Thakurta, and L. Zhang (2021). Private alternating least squares: Practical private matrix completion with tighter rates. pp. 1877–1887.

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference*, 265–284.

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.

Dwork, C., K. Talwar, A. Thakurta, and L. Zhang (2014). Analyze gauss: optimal bounds for privacy-preserving principal component analysis. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 11–20.

Edelman, A., T. A. Arias, and S. T. Smith (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications 20*(2), 303–353.

Erlingsson, U., V. Pihur, and A. Korolova (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, New York, NY, USA, pp. 1054–1067. Association for Computing Machinery.

Hamidi, N. and M. Bayati (2022). On low-rank trace regression under general sampling distribution. *Journal of Machine Learning Research 23*(321), 1–49.

Jain, P., O. D. Thakkar, and A. Thakurta (2018). Differentially private matrix completion revisited. pp. 2215–2224.

Kamath, G., J. Li, V. Singhal, and J. Ullman (2019). Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pp. 1853–1902. PMLR.

Kamath, G., V. Singhal, and J. Ullman (2020, 09–12 Jul). Private mean estimation of heavy-tailed distributions. In J. Abernethy and S. Agarwal (Eds.), *Proceedings of Thirty Third Conference on Learning Theory*, Volume 125 of *Proceedings of Machine Learning Research*, pp. 2204–2235. PMLR.

Koltchinskii, V. (2011). Von neumann entropy penalization and low-rank matrix estimation.

Koltchinskii, V., K. Lounici, and A. B. Tsybakov (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion.

Koltchinskii, V. and D. Xia (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res. 16*(53), 1757–1792.

Kuditipudi, R., J. Duchi, and S. Haque (2023, 12–15 Jul). A pretty fast algorithm for adaptive private mean estimation. In G. Neu and L. Rosasco (Eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, Volume 195 of *Proceedings of Machine Learning Research*, pp. 2511–2551. PMLR.

Liu, X., W. Kong, P. Jain, and S. Oh (2022). Dp-pca: Statistically optimal and differentially private pca. *Advances in Neural Information Processing Systems 35*, 29929–29943.

Liu, Z., Y.-X. Wang, and A. Smola (2015). Fast differentially private matrix factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 171–178.

Luo, Y. and A. R. Zhang (2022). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap, and their interplay. *arXiv preprint arXiv:2206.08756*.

McSherry, F. and I. Mironov (2009). Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 627–636.

Negahban, S. and M. J. Wainwright (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling.

Pajor, A. (1998). Metric entropy of the grassmann manifold. *Convex Geometric Analysis 34*(181-188), 0942–46013.

Rohde, A. and A. B. Tsybakov (2011). Estimation of high-dimensional low-rank matrices.

Shen, Y., J. Li, J.-F. Cai, and D. Xia (2023). Computationally efficient and statistically optimal robust high-dimensional linear regression. *arXiv preprint arXiv:2305.06199*.

Tropp, J. A. et al. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning 8*(1-2), 1–230.

Koltchinskii, V., K. Lounici, and A. B. Tsybakov (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion.

Koltchinskii, V. and D. Xia (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res. 16*(53), 1757–1792.

Kuditipudi, R., J. Duchi, and S. Haque (2023, 12–15 Jul). A pretty fast algorithm for adaptive private mean estimation. In G. Neu and L. Rosasco (Eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, Volume 195 of *Proceedings of Machine Learning Research*, pp. 2511–2551. PMLR.

Liu, X., W. Kong, P. Jain, and S. Oh (2022). Dp-pca: Statistically optimal and differentially private pca. *Advances in Neural Information Processing Systems 35*, 29929–29943.

Liu, Z., Y.-X. Wang, and A. Smola (2015). Fast differentially private matrix factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 171–178.

Luo, Y. and A. R. Zhang (2022). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap, and their interplay. *arXiv preprint arXiv:2206.08756*.

McSherry, F. and I. Mironov (2009). Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 627–636.

Negahban, S. and M. J. Wainwright (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling.

Pajor, A. (1998). Metric entropy of the grassmann manifold. *Convex Geometric Analysis 34*(181-188), 0942–46013.

Rohde, A. and A. B. Tsybakov (2011). Estimation of high-dimensional low-rank matrices.

Shen, Y., J. Li, J.-F. Cai, and D. Xia (2023). Computationally efficient and statistically optimal robust high-dimensional linear regression. *arXiv preprint arXiv:2305.06199*.

Tropp, J. A. et al. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning 8*(1-2), 1–230.

Vadhan, S. (2017). The complexity of differential privacy. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, 347–450.

Vandereycken, B. (2013). Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization 23*(2), 1214–1236.

Vershynin, R. (2015). Estimation in high dimensions: a geometric perspective. In *Sampling Theory, a Renaissance: Compressive Sensing and Other Developments*, pp. 3–66. Springer.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Wang, L., B. Zhao, and M. Kolar (2023, 25–27 Apr). Differentially private matrix completion through low-rank matrix factorization. *206*, 5731–5748.

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

Wei, K., J.-F. Cai, T. F. Chan, and S. Leung (2016). Guarantees of riemannian optimization for low rank matrix recovery. *SIAM Journal on Matrix Analysis and Applications 37*(3), 1198–1222.

Xia, D. (2021). Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics 15*(2), 3798–3851.

Zheng, Q. and J. Lafferty (2016). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements.

Vadhan, S. (2017). The complexity of differential privacy. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, 347–450.

Vandereycken, B. (2013). Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization 23*(2), 1214–1236.

Vershynin, R. (2015). Estimation in high dimensions: a geometric perspective. In *Sampling Theory, a Renaissance: Compressive Sensing and Other Developments*, pp. 3–66. Springer.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Wang, L., B. Zhao, and M. Kolar (2023, 25–27 Apr). Differentially private matrix completion through low-rank matrix factorization. *206*, 5731–5748.

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

Wei, K., J.-F. Cai, T. F. Chan, and S. Leung (2016). Guarantees of riemannian optimization for low rank matrix recovery. *SIAM Journal on Matrix Analysis and Applications 37*(3), 1198–1222.

Xia, D. (2021). Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics 15*(2), 3798–3851.

Zheng, Q. and J. Lafferty (2016). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements.

30