Tight Lower Bound of Sparse Covariance Matrix Estimation in the Local Differential Privacy Model[☆,☆☆]

Di Wang[a,*], Jinhui Xu[a]

[a]*Department of Computer Science and Engineering*
State University of New York at Buffalo
338 Davis Hall, Buffalo, 14260

Abstract

In this paper, we study the sparse covariance matrix estimation problem in the local differential privacy model, and give a lower bound of $\Omega(\frac{s^2 \log p}{n\epsilon^2})$ on the ϵ non-interactive private minimax risk in the metric of squared spectral norm, where s is the row sparsity of the underlying covariance matrix, n is the sample size, and p is the dimensionality of the data. We show that the lower bound is actually tight, as it matches a previous upper bound. Our main technique for achieving this lower bound is a general framework, called **General Private Assouad Lemma**, which is a considerable generalization of the previous private Assouad lemma and can be used as a general method for bounding the private minimax risk of matrix-related estimation problems.

Keywords: Local Differential Privacy, Covariance Matrix Estimation

1. Introduction

Machine Learning and Statistical Estimation have made profound impacts in recent years to many applied domains such as social sciences, genomics, and medicine. A frequently encountered challenge in such applications is how to deal with the high dimensionality of the datasets, especially for those in genomics, educational and psychological research. A commonly adopted strategy is to assume that the underlying structure of the parameter space is sparse.

Another often encountered challenge is how to handle sensitive data, such as those in social science, biomedicine and genomics. A promising approach is to use some private mechanisms for the statistical inference and learning tasks. Differential Privacy (DP) and its distributed version, Local Differential Privacy (LDP) [1], are widely-accepted

[☆]A preliminary version appeared in Proceedings of The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019).

[☆☆]This research was supported in part by the National Science Foundation (NSF) through grants CCF-1422324 and CCF-1716400.

[*]Corresponding author

Email addresses: dwang45@buffalo.edu (Di Wang), jinhui@buffalo.edu (Jinhui Xu)

models that provide provable protection against identification and are resilient to arbitrary auxiliary information that might be available to attackers. Since its introduction over a decade ago, a rich line of works are now available, which have made (local) differential privacy compelling privacy enhancing technologies for many organizations, such as Uber [2], Google [3], Apple [4].

While differentially private high dimensional estimation is quite promising, such as sparse linear regression [5] and selection problem [6], estimating high dimensional datasets in a locally differentially private manner could be quite challenging for many problems, such as sparse linear regression [7], sparse mean estimation [8] and selection problem [9]. Fortunately, recent research has shown that the loss of some problems caused by the local differential privacy constraints can be quite small compared to their non-private counterparts. Examples of this phenomenon include high dimensional sparse PCA [10]. Recently, [11] studied the locally differentially private high dimensional sparse covariance estimation problem and proposed an algorithm which achieves an upper bound of $O(\frac{s^2 \log p}{n\epsilon^2})$ measured by the squared spectral norm, *i.e.,* $\|\Sigma^{\text{priv}} - \Sigma^*\|_2^2$, where s is the row sparsity of the underlying covariance matrix, p is the dimensionality, and n is the sample size. With the above upper bound, a natural question is the follows.

Is the upper bound of $O(\frac{s^2 \log p}{n\epsilon^2})$ in the LDP high dimensional sparse covariance estimation tight?

In this paper, we give an affirmative answer to the above question. Specifically, we have the following contributions.

1. We show that in the non-interactive local differnetial privacy model, the private minimax risk (in the metric of squared spectral norm) of high dimensional sparse covariance matrix estimation is lower bounded by $\Omega(\frac{s^2 \log p}{n\epsilon^2})$. Moreover, we show that the same lower bound also holds, even if the metric is generalized from the squared spectral norm to the general squared ℓ_w norm for any $w \in [1, \infty]$. Combining these with previous upper bounds, it indicates that these lower bounds are tight.

2. To prove the above lower bounds, we propose a framework, called **General Private Assouad Lemma**, for lower bounding the private minimax risk in the non-interactive or sequential differential privacy model. Our lemma is a generalization of the private Assoud lemma in [12], and can be viewed as a general method for locally differentially private matrix estimation problems. We believe that it has the potential to be used in other matrix-related estimation problems.

2. Related Work

Recently, there are several papers studying the private covariance matrix estimation problem [13, 14, 15, 16, 17, 11]. For covariance matrix estimation in the central differential privacy model, [15] considered the 1-dimensional Gaussian distribution estimation with (un)known variance. [13] studied the problem of privately learning a multivariate Gaussian and product distributions in the total variation distance and showed that it is privacy-free for these problems. [17] recently also investigated the low dimentional case of the problem in Frobenious norm and proposed an iterative eigenvector

sampling method. The work that is the most related to ours is probably the one in [11], where the authors studied the problem in the high dimensional sparse case and proposed a method based on the idea of thresholding the private empirical covariance matrix. A missing ingredient in all the above works is that no lower bound is given, which makes it difficult to tell how far their solutions are away from the optimal.

Covariance matrix estimation under local differential privacy has been studied in [16, 14, 11]. Specifically, [16, 14] comprehensively studied the 1-dimensional Gaussian distribution estimation and provided several lower bounds. However, none of these works can be extended to general distributions and to the high dimensional sparse case. For the high dimensional case of the problem, [11] proposed a general method which achieves an upper bound of $O(\frac{s^2 \log p}{n\epsilon^2})$ in the squared spectral norm. In this paper, we provide a lower bound which matches this upper bound.

Using information-theoretic techniques to prove lower bounds in the local differential privacy model has also been studied in many papers, such as [8, 12, 18, 14, 19, 20]. [8, 12, 18] proposed several general frameworks for bounding the private minimax risk, such as the private versions of Le Cam lemma, Fano lemma, and Assouad lemma. However, none of these methods can be applied to our problem since all the previous lemmas can only be used in the one-directional case (*i.e.,* the underlying parameter is a vector), while it is a two-directional case (*i.e.,* the underlying parameter is a matrix) in our problem. Moreover, all of the previous methods need to obtain some upper bounds of some hard distribution instances under the total variation distance (or KL-divergence) while in our problem we use χ^2-divergence, which makes our method quite different from the previous ones. The method that is the most related to ours is the private Assouad lemma proposed in [12] which can be seen as a special case of our general private Assoud lemma. Recently, [20] revisited the private Assouad lemma and proposed a general theorem with tighter lower bounds via some results in the theory of communication complexity. However, our theorems are incomparable with theirs since we cannot use their theorem directly to our problem (see Conclusion section for details) .

3. Preliminaries

In this section, we introduce some definitions that will be used throughout the paper. More details can be found in [12].

Notation. In this paper, we will always assume (except for Corollary 2) that $\Phi(x) = x^2$ and $\rho(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_2$ is the spectral norm between two matrices Σ_1 and Σ_2.

3.1. Classical Minimax Risk

Since all of our lower bounds are in the form of private minimax risk, we first introduce the classical statistical minimax risk before discussing its locally differentially private version.

Let $\mathcal{P}$ be a class of distributions over a data universe $\mathcal{X}$. For each distribution $p \in \mathcal{P}$, there is a deterministic function $\theta(p) \in \Theta$, where Θ is the parameter space. Let $\rho : \Theta \times \Theta :\mapsto \mathbb{R}_+$ be a semi-metric function on the space Θ and $\Phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a non-decreasing function with $\Phi(0) = 0$. We further assume that $\{X_i\}_{i=1}^n$ are n i.i.d

observations drawn according to some distribution $p \in \mathcal{P}$, and $\hat{\theta} : \mathcal{X}^n \mapsto \Theta$ be some estimator. Then the minimax risk in metric $\Phi \circ \rho$ is defined by the following saddle point problem:

$$\mathcal{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) := \inf_{\hat{\theta}} \sup_{p \in \mathcal{P}} \mathbb{E}_p[\Phi(\rho(\hat{\theta}(X_1, \cdots, X_n), \theta(p))],$$

where the supremum is taken over distributions $p \in \mathcal{P}$ and the infimum over all estimators $\hat{\theta}$.

3.2. Local Differential Privacy and Private Minimax Risk

Since we will consider the sequential interactive and non-interactive local models in this paper, we follow the definitions in [18].

We assume that $\{Z_i\}_{i=1}^n$ are the private observations transformed from $\{X_i\}_{i=1}^n$ through some privacy mechanisms. We say that the mechanism is sequentially interactive, when it has the following conditional independence structure:

$$\{X_i, Z_1, \cdots, Z_{i-1}\} \mapsto Z_i, Z_i \perp\!\!\!\perp X_j \mid \{X_i, Z_1, \cdots, Z_{i-1}\}$$

for all $j \neq i$ and $i \in [n]$, where $\perp\!\!\!\perp$ means independent relation. The full conditional distribution can be specified in terms of conditionals $Q_i(Z_i \mid X_i = x_i, Z_{1:i} = z_{i:i})$. The full privacy mechanism can be specified by a collection $Q = \{Q_i\}_{i=1}^n$.

When Z_i is depending only on X_i, the mechanism is called non-interactive and in this case we have a simpler form for the conditional distributions $Q_i(Z_i \mid X_i = x_i)$. We now define local differential privacy by restricting the conditional distribution Q_i.

Definition 1 ([18])**.** *For a given privacy parameter $\epsilon > 0$, the random variable Z_i is an ϵ sequentially locally differentially private view of X_i if for all $z_1, z_2, \cdots, z_{i-1}$ and $x, x' \in \mathcal{X}$ we have the following for all the events S:*

$$\frac{Q_i(Z_i \in S \mid X_i = x_i, Z_{1:i-1} = z_{1:i-1})}{Q_i(Z_i \in S \mid X_i = x_i', Z_{1:i-1} = z_{1:i-1})} \leq e^{\epsilon}.$$

We say that the random variable Z_i is an ϵ non-interactively locally differentially private view of X_i if

$$\frac{Q_i(Z_i \in S \mid X_i = x_i)}{Q_i(Z_i \in S \mid X_i = x_i')} \leq e^{\epsilon}.$$

We say that the privacy mechanism $Q = \{Q_i\}_{i=1}^n$ is ϵ-sequentially (non-interactively) locally differentially private (LDP) if each Z_i is a sequentially (non-interactively) locally differentially private view.

For a given privacy parameter $\epsilon > 0$, let $\mathcal{Q}_\epsilon$ be the set of conditional distributions that have the ϵ-LDP property. For a given set of samples $\{X_i\}_{i=1}^n$, let $\{Z_i\}_{i=1}^n$ be the set of observations produced by any distribution $Q \in \mathcal{Q}_\epsilon$. Then, our estimator will be based on $\{Z_i\}_{i=1}^n$, that is, $\hat{\theta}(Z_1, \cdots, Z_n)$. This yields a modified version of the minimax risk:

$$\mathcal{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q) = \inf_{\hat{\theta}} \sup_{p \in \mathcal{P}} \mathbb{E}_p \Phi(\rho(\hat{\theta}(Z_1, \cdots, Z_n), \theta(p)).$$

From the above definition, it is natural for us to seek the mechanism $Q \in \mathcal{Q}_\epsilon$ that has the smallest value for the minimax risk. This allows us to define functions that characterize the optimal rate of estimation in terms of privacy parameter ϵ.

Definition 2. *Given a family of distributions $\theta(\mathcal{P})$ and a privacy parameter $\epsilon > 0$, the ϵ sequential private minimax risk in the metric $\Phi\circ\rho$ is:*

$$\mathcal{M}_n^{Int}(\theta(\mathcal{P}), \Phi\circ\rho, \epsilon) := \inf_{Q\in\mathcal{Q}_\epsilon} \mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho, Q),$$

where $\mathcal{Q}_\epsilon$ is the set of all ϵ sequentially locally differentially private mechanisms. Moreover, the ϵ non-interactive private minimax risk in the metric $\Phi\circ\rho$ is:

$$\mathcal{M}_n^{Nint}(\theta(\mathcal{P}), \Phi\circ\rho, \epsilon) := \inf_{Q\in\mathcal{Q}_\epsilon} \mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho, Q),$$

where $\mathcal{Q}_\epsilon$ is the set of all ϵ non-interactively locally differentially private mechanisms.

4. General Private Assouad Lemma

In this section we introduce our general framework for lower bounding. Before that, we first review the classical Assouad lemma [21] and its two-directional generalization [22].

Assouad's method works with a hypercube $\mathcal{V} = \{-1, +1\}^r$ for some $r \in \mathbb{N}$. It transforms an estimation problem into multiple hypothesis testing problems using the structure of the problem in an essential way. Let $\{P_v\}_{v\in\mathcal{V}} \in \mathcal{P}$ be a family of distributions with its corresponding parameters $\{\theta_v\}_{v\in\mathcal{V}}$ indexed by the hypercube. Similar to the standard reduction from estimation to testing, we consider the following random process. Let V be a random vector uniformly chosen from the hypercube $\{-1, +1\}^r$. After that, the samples $X_1, X_2, \cdots, X_n$ are drawn from the distribution P_v conditioned on $V = v$. For each $j \in [r]$, we define the mixture of distributions

$$P_{j,+1}^n = \frac{1}{2^{r-1}} \sum_{v:\, v_j=1} P_v^n, P_{j,-1}^n = \frac{1}{2^{r-1}} \sum_{v:\, v_j=-1} P_v^n, \tag{1}$$

where P_v^n is the product distribution of $X_1, \cdots, X_n$. Then, Assouad lemma can be stated as follows.

Lemma 1 (Assouad Lemma)**.** *Under the conditions stated in the above paragraph,*

$$\mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho) \geq \frac{\alpha}{4} \sum_{j=1}^{r} [1 - \|P_{j,+1}^n - P_{j,-1}^n\|_{TV}], \tag{2}$$

where $\|\cdot\|_{TV}$ is the total variation distance, $\alpha = \min_{H(v,v')\geq 1, v,v'\in\mathcal{V}} \frac{\Phi(\rho(\theta_v,\theta_{v'}))}{2H(v,v')}$, and $H(v, v')$ is the hamming distance between θ and θ', i.e., $H(v, v') = \sum_{j=1}^r \mathbb{1}\{v_j \neq v'_j\}$.

Instead of restricting to a hypercube $\mathcal{V}$, the general Assouad lemma in [22] works with the Cartesian product of a hypercube and the r-th power of a finite set of vectors. Specifically, for a given $r \in \mathbb{N}$ and a finite set of p-dimensional vectors $B \subset \mathbb{R}^p \backslash \{0_{1\times p}\}$, let $\mathcal{V} = \{-1, +1\}^r$ and $\Lambda \subseteq B^r$. Define $T = \mathcal{V} \otimes \Lambda = \{\tau = (v, \lambda) : v \in \mathcal{V} \text{ and } \lambda \in \Lambda\}$. This means that one can view an element $\lambda \in \Lambda$ as an $r\times p$ matrix with each row coming from set B, and $\mathcal{V}$ as a set of parameters with each row indicating whether a given row

of λ is present or not. Similar to Assouad lemma, we assume that there is a family of distributions in the class $\mathcal{P}$, $\{P_\tau\}_{\tau\in T}$ indexed by T and its corresponding parameters $\{\theta_\tau\}_{\tau\in T}$.

Let $D_\Lambda = |\Lambda|$. For a given $a \in \{-1, +1\}$ and $j \in [r]$, we let $T_{i,a} = \{\tau : \upsilon_i(\tau) = a\}$, where $\upsilon_i(\tau)$ is the i-th coordinate of the first component of τ. It is easy to see that $|T_{i,a}| = 2^{r-1} D_\Lambda$. We have the following mixture of distributions

$$P^n_{j,a} = \frac{1}{2^{r-1} D_\Lambda} \sum_{\tau\in T_{j,a}} P^n_\tau, P_{j,a} = \frac{1}{2^{r-1} D_\Lambda} \sum_{\tau\in T_{j,a}} P_\tau. \tag{3}$$

Lemma 2 (General Assouad's Lemma [23])**.** *Under the conditions stated in above paragraph, we have the following*

$$\mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho) \geq \frac{\alpha}{4} \sum_{j=1}^{r} [1 - \|P^n_{j,+1} - P^n_{j,-1}\|_{TV}],$$

where α satisfies

$$\alpha = \min_{H(\upsilon(\tau),\upsilon(\tau'))>1, \upsilon(\tau),\upsilon(\tau')\in\mathcal{V}} \frac{\Phi(\rho(\theta_\tau, \theta_{\tau'}))}{2H(\upsilon(\tau), \upsilon(\tau'))},$$

and $\upsilon(\tau)$ is the first component of τ.

Now, we present the locally private version of Lemma 2. Suppose that we draw samples $Z_1, \cdots, Z_n$ according to ϵ-LDP channel $Q(\cdot|X_{1:n})$. Then, conditioned on $V = \tau$, the private sample is distributed according to the marginal distribution M^n_τ:

$$M^n_\tau(S) = \int Q^n(S|x_1, x_2, \cdots, x_n) d P^n_\tau(x_1, x_2, \cdots, x_n). \tag{4}$$

Specifically, when Q is non-interactive, we have $M^n_\tau = (\int Q(\cdot|x) d P_\tau(x))^{\otimes n}$. Similarly to (3), we can define $M^n_{j,a}$ and $M_{j,a}$ for $a \in \{-1, +1\}$ and $j \in [r]$. Thus, combining the above with Lemma 2, we have the following theorem:

Theorem 1. *Under the conditions given in Lemma 2, the ϵ private minimax risk satisfies:*

$$\mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho, \epsilon) \geq \frac{\alpha}{4} \sum_{j=1}^{r} [1 - \|M^n_{j,+1} - M^n_{j,-1}\|_{TV}]. \tag{5}$$

For the sequential private minimax risk, we have the following general lower bound.

Theorem 2. *Under the conditions given in Theorem 1 and further assuming that $\epsilon \in (0, \frac{1}{2}]$, the ϵ sequential private minimax risk in the metric $\Phi\circ\rho$ satisfies*

$$\mathcal{M}^{Int}_n(\theta(\mathcal{P}, \Phi\circ\rho, \epsilon) \geq \frac{\alpha r}{4} [1 - \big(\frac{n\epsilon^2}{2r} \sup_{\gamma\in\mathbb{B}_\infty(\mathcal{X})} \sum_{j=1}^{r} \big(\int_{\mathcal{X}} \gamma(x)(d P_{j,+1} - d P_{j,-1})\big)^2\big)^{\frac{1}{2}}], \tag{6}$$

where $\mathbb{B}_\infty$ is the 1-ball of supremum norm $\mathbb{B}_\infty = \{\gamma \in L^\infty(\mathcal{X}) \mid \|\gamma\|_\infty \leq 1\}$, and $L^\infty(\mathcal{X}) = \{f : \mathcal{X} \mapsto \mathbb{R} \mid \|f\|_\infty < \infty\}$ is the space of uniformly bounded functions with the supremum norm $\|f\|_\infty = \sup_x |f(x)|$.

Proof of Theorem 2. The proof follows the proof of Theorem 3 in [12]. We will mainly prove the following lemma

Lemma 3. *[Theorem 3 in [12]] Under the condition in Theorem 1, for any ϵ sequential interactive private channel Q we have*

$$\sum_{j=1}^{r}[D_{kl}(M_{j,+1}^n\|M_{j,-1}^n)+D_{kl}(M_{j,-1}^n\|M_{j,+1}^n)] \leq (e^\epsilon-1)^2 n \sup_{\gamma\in\mathbb{B}_\infty(\mathcal{X})}\sum_{j=1}^{r}(\int_{\mathcal{X}}\gamma(x)(dP_{j,+1}-dP_{j,-1}))^2$$

By Lemma 3 we can easily get Theorem 1, which is due to the Pinsker's inequality and Cauchy-Schwartz:

$$\sum_{j=1}^{r}\|M_{j,+1}^n-M_{j,-1}^n\|_{TV}\leq\frac{1}{2}\sqrt{r}\Big(\sum_{j=1}^{r}D_{kl}(M_{j,+1}^n\|M_{j,-1}^n)+D_{kl}(M_{j,-1}^n\|M_{j,+1}^n)\Big)^{\frac{1}{2}}.$$

□

Note that the lower bound in Theorem 2 reduces to the same one in the private Assouad lemma [12] when Λ contains only one matrix which every row is non-zero. Thus, we call Theorem 1 as the **General Private Assouad Lemma**. Particularly, if we restrict our attention only to the non-interactive LDP mechanisms, we have the following theorem bounding the private minimax risk, which will be used to prove our lower bounds in this paper.

Theorem 3. *Under the conditions given in Theorem 1 and further assuming that $\epsilon\in(0,\frac{\ln 2}{2}]$, the ϵ non-interactive private minimax risk in the metric $\Phi\circ\rho$ satisfies*

$$\mathcal{M}_n^{Nint}(\theta,\Phi\circ\rho,\epsilon)\geq\frac{r\alpha}{4}\times\min_{1\leq j\leq r}(1-\sqrt{\frac{1}{2}(\epsilon^2 D_{\chi^2}(P_{j,+1}\|P_{j,-1}))^n}), \tag{7}$$

where $D_{\chi^2}(\cdot\|\cdot)$ is the χ^2-divergence, that is, $D_{\chi^2}(P\|Q)=\int\frac{(dP-dQ)^2}{dQ}$ for distributions P and Q.

Proof. By Theorem 1, we have

$$\mathcal{M}_n^{Nint}(\theta(\mathcal{P}),\Phi\circ\rho,\epsilon)\geq\frac{r\alpha}{4}\min_{j\in[r]}(1-\|M_{j,+1}^n-M_{j,-1}^n\|_{TV}).$$

By the non-interactivity, we have $M_{j,a}^n=(\int Q(\cdot|x)dP_{j,a})^{\otimes n}$. Let $M_{j,a}=\int Q(\cdot|x)dP_{j,a}$.

By Pinsker inequality, we have the following

$$|M_{j,+1}^n - M_{j,-1}^n\|_{TV}^2 \leq \frac{1}{2} D_{kl}(M_{j,+1}^n \| M_{j,-1}^n) \tag{8}$$

$$\leq \frac{1}{2} D_{\chi^2}(M_{j,+1}^n \| M_{j,-1}^n) \tag{9}$$

$$= \frac{1}{2}(D_{\chi^2}(M_{j,+1} \| M_{j,-1}))^n \tag{10}$$

$$\leq \frac{1}{2}(\min\{4, e^{2\epsilon}\}(e^\epsilon - 1)^2 \|P_{j,+1} - P_{j,-1}\|_{TV}^2)^n \tag{11}$$

$$\leq \frac{1}{2}(\min\{2, \frac{e^{2\epsilon}}{2}\}\epsilon^2 D_{\chi^2}(P_{j,+1} \| P_{j,-1})))^n, \tag{12}$$

where (8) is due to Pinsker inequality, (9) is by the relation between KL-divergence and χ^2-divergence $D_{kl}(P\|Q) \leq \log(1 + D_{\chi^2}(P\|Q)) \leq D_{\chi^2}(P\|Q)$ [21], (10) is due to the non-interactivity, (12) is by Pinsker inequality and inequalities $(e^\epsilon - 1)^2 \leq 2\epsilon^2$ and $e^{2\epsilon} \leq 2$. Next, we prove (11).

Lemma 4.

$$D_{\chi^2}(M_{j,+1} \| M_{j,-1}) \leq \min\{4, e^{2\epsilon}\}(e^\epsilon - 1)^2 \|P_{j,+1} - P_{j,-1}\|_{TV}^2.$$

Proof. W.l.o.g, we can assume that the density function of $M_{j,a}$ is $m_{j,a}(z) = \int q(z|x) d P_{j,a}$ and $q(\cdot|x)$ is the density function of $Q(\cdot|x)$. By the definition, we have

$$D_{\chi^2}(M_{j,+1} \| M_{j,-1}) = \int \frac{(m_{j,+1}(z) - m_{j,-1}(z))^2}{m_{j,-1}(z)} dz \tag{13}$$

$$\leq \int \frac{c_\epsilon^2 \inf_x q^2(z|x)(e^\epsilon - 1)^2 \|P_{j,+1} - P_{j,-1}\|_{TV}^2}{\int q(z|x) d P_{j,a}} dz \tag{14}$$

$$\leq c_\epsilon^2 (e^\epsilon - 1)^2 \|P_{j,+1} - P_{j,-1}\|_{TV}^2 \int \inf_x q(z|x) dz$$

$$\leq c_\epsilon^2 (e^\epsilon - 1)^2 \|P_{j,+1} - P_{j,-1}\|_{TV}^2, \tag{15}$$

where $c_\epsilon = \min\{2, e^\epsilon\}$, (13) is by the definition of χ^2-divergence, (14) is by Lemma 3 in [12] and (15) is due to the fact that $\int \inf_x q(z|x) dz \leq 1$.

□

□

The inequality in Lemma 4 is weaker than the one in Theorem 1 of [12] in the sense that it becomes the later one if combining the inequality of $\|M_{j,+1} - M_{j,-1}\|_{TV} \leq D_{\chi^2}(M_{j,+1} \| M_{j,-1})$.

Remark 1. *We note that comparing to existing general lower bounding methods on the private minimax risk, such as [12, 8, 19], Theorem 3 is quite different. Firstly, while all previous lower bounds depend only linearly on the sample size n, the lower bound in Theorem 3 depends exponentially on n. Secondly, due to the special structure of our*

indexing set T, Theorem 3 is more suitable for matrix estimation problems, while previous methods are more suitable for vector estimation problems. Thirdly, previous lower bounds are measured by (or derived from) the mutual information, the total variation distance, or the KL-divergence between the hard distribution instances, while in Theorem 3, the lower bound is measured by the χ^2- divergence between distributions. This indicates that although Theorem 3 is stronger than the previous ones, as it can be seen later in the sparse covariance estimation problem, it is easier to obtain a lower bound on the χ^2-divergence of the hard instances than other measurements. This is also the reason that existing methods cannot be applied to our problem.

From (7), we can see that, to obtain the lower bound, one needs to bound the terms of $D_{\chi^2}(P_{j,+1}\|P_{j,-1})$ for all j, which are quite complicated since they are mixture distributions. To simplify the task, we fix all the other terms and consider only the j-th term, which can be seen as an $r \times p$ matrix with all other rows fixed, except for the j-th one. Formally, for an element $\tau \in T$, we define the projection $\upsilon_A(\tau) = (\upsilon_i(\tau))_{i\in A}$ for a set $A \subseteq \{1, 2\cdots, r\}$, and the set $\{-j\} = [r]\backslash\{j\}$. $\lambda_A(\tau)$ and $\lambda_{-i}(\tau)(\lambda_i(\tau))_{i\in A}$ can be defined similarly, where $\lambda_i(\tau)$ is the i-th coordinate of the second component of τ. Denote by Λ_A the set $\Lambda_A = \{\lambda_A(\tau) : \tau \in T\}$. For $a \in \{+1, -1\}$, $b \in \{-1, +1\}^{r-1}$ and $c \in \Lambda_{-j} \subseteq B^{r-1}$, we let

$$T_{\Lambda_j(a,b,c)} = \{\tau \in T : \upsilon_j(\tau) = a, \upsilon_{-j}(\tau) = b, \lambda_{-j}(\tau) = c\}$$

and $D_{\Lambda_j(a,b,c)} = |T_{\Lambda_j(a,b,c)}|$. Let $\bar{P}^n_{j,a,b,c}$ denote the mixture distribution

$$\bar{P}^n_{j,a,b,c} = \frac{1}{D_{\Lambda_j(a,b,c)}} \sum_{\tau\in T_{\Lambda_j(a,b,c)}} P^n_\tau, \tag{16}$$

and $\bar{M}^n_{j,a,b,c}$ be its corresponding marginal distribution. Similar to Theorem 3, we have the following corollary.

Corollary 1. *Under the conditions given in Theorem 1 and further assuming that $\epsilon \in (0, \frac{\ln 2}{2}]$, the ϵ non-interactive private minimax risk in the metric $\Phi\circ\rho$ satisfies*

$$\mathcal{M}^{Nint}_n(\psi(\theta), \Phi\circ\rho, \epsilon) \geq \frac{r\alpha}{4} \times \min_{1\leq j\leq r} (1 - \sqrt{\frac{\epsilon^{2n}}{2} Average_{\upsilon_{-j},\lambda_{-j}}(D_{\chi^2}(\bar{P}_{j,+1,\upsilon_{-j},\lambda_{-j}}\|\bar{P}_{j,-1,\upsilon_{-j},\lambda_{-j}}))^n}), \tag{17}$$

where the average over $\upsilon_{-j}, \lambda_{-j}$ is induced by the uniform distribution over T.

Proof. The key observation is that the distributions $P^n_{j,a}$ can be represented by a linear combination of $\{\bar{P}^n_{j,a,b,c}\}_{b,c\in T_{-j}}$, where the set T_{-j} is

$$\begin{aligned} T_{-j} &= \{0, 1\}^{r-1} \otimes \Lambda_{-i} \\ &= \{(b, c) : \exists\tau \in T \text{ s.t } \upsilon_{-i}(\tau) = b \text{ and } \lambda_{-i}(\tau) = c\}. \end{aligned}$$

That is, $P^n_{j,a} = \sum_{(b,c)\in T_{-j}} w_{b,c}\bar{P}^n_{j,a,b,c}$, where $w_{b,c} = \frac{D_{\Lambda_j(a,b,c)}}{2^{r-1}D_\Lambda}$ (note that since $D_{\Lambda_j(a,b,c)}$ is independent of a, we omit it). Also, $\sum_{(b,c)\in T_{-j}} w_{b,c} = 1$. Thus, $P^n_{j,a}$ can be seen as an average over (b,c). The same also holds for $M^n_{j,a}$.

By the convexity of total variation norm and Lemma 4, we have

$$\begin{aligned}\|M^n_{j,+1} - M^n_{j,-1}\|_{TV} &\le \sum_{(b,c)\in T_{-j}} w_{b,c}\|\bar{M}^n_{j,+1,b,c} - \bar{M}^n_{j,-1,b,c}\|_{TV} \\ &= \text{Average}_{b,c}\|\bar{M}^n_{j,+1,b,c} - \bar{M}^n_{j,-1,b,c}\|_{TV}.\end{aligned}$$

By a similar argument given in the proof of Theorem 3, we get

$$\begin{aligned}\|\bar{M}^n_{j,+1,b,c} - \bar{M}^n_{j,-1,b,c}\|^2_{TV} &\le D_{\chi^2}(\bar{M}_{j,+1,\upsilon_{-j},\lambda_{-j}}\|\bar{M}_{j,-1,\upsilon_{-j},\lambda_{-j}})^n \\ &\le \frac{1}{2}(\min\{2, \frac{e^{2\epsilon}}{2}\}\epsilon^2 D_{\chi^2}(\bar{P}_{j,+1,b,c}\|\bar{P}_{j,-1,b,c})))^n \\ &\le \frac{1}{2}(\epsilon^2 D_{\chi^2}(\bar{P}_{j,+1,b,c}\|\bar{P}_{j,-1,b,c})))^n.\end{aligned}$$

Thus, by the inequality $\text{Average}_{b,c}\|\bar{M}^n_{j,+1,b,c} - \bar{M}^n_{j,-1,b,c}\|_{TV})^2 \le \text{Average}_{b,c}\|\bar{M}^n_{j,+1,b,c} - \bar{M}^n_{j,-1,b,c}\|^2_{TV}$, we have the proof. □

5. Lower Bound of Private Sparse Covariance Estimation

We follow the settings in [22, 11]. Let $X_1, \cdots, X_n$ be random samples from a zero-mean p-variate distribution with covariance matrix $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$. The goal of sparse covariance matrix estimation is to estimate the unknown matrix Σ based on samples $\{X_1, \cdots, X_n\}$, and the locally private version is to determine a locally differentially private estimator. In this paper, we focus on the high dimensional case, that is, $c_1 n^\beta \le p \le \exp(c_2 n)$ for some $\beta > 1, c_1, c_2 > 0$. We assume that the underlying covariance is sparse. That is, $\Sigma \in \mathcal{G}(s)$ with

$$\mathcal{G}(s) = \{\Sigma = (\sigma_{ij})_{1\le i,j\le p} : \|\sigma_{-j,j}\|_0 \le s, \forall j \in [p]\}, \tag{18}$$

where $\sigma_{-j,j}$ is the j-th column of Σ with $\sigma_{j,j}$ removed, i.e., a matrix in $\mathcal{G}(s)$ has at most s-nonzero off-diagonal elements on each column.

Moreover, we assume that each X_i is sampled from a ρ-sub-Gaussian distribution. That is, for all $t > 0$ and $\|v\|_2 = 1$,

$$\mathbb{P}\{|\langle v, X\rangle| > t\} \le \exp(\frac{-t^2}{2\rho}), \tag{19}$$

which means that all the one-dimensional marginals of X have sub-Gaussian tails.

Additionally, in private matrix-related estimation problems, it is always assumed that the ℓ_2 norm of each X_i are bounded by 1 [24, 10, 25, 11]. In this paper, we relax the bounded norm assumption in the following way; for the random vector $X \in \mathbb{R}^p$, we

assume that $\|X\|_2 \leq 1$ with probability at least $1-e^{-\Omega(p)}$. This leads us to the following class of distributions $\mathcal{P}(\tau, s)$.

$$\mathcal{P}(\rho, s) = \{P : X \sim P \text{ satisfies (19) and } \|X\|_2 \leq 1 \text{ w.p at least } 1-e^{-\Omega(p)}, \mathbb{E}X = 0, \Sigma = \mathbb{E}[XX^T] \in \mathcal{G}(s)\}. \quad (20)$$

Before showing the lower bound, we first describe our construction of the hard indexing set T with their distributions $\{P_\tau\}_{\tau \in T}$ instances, which is motivated by the ones in [22].

We first construct the parameter set, which is the same as in [22]. Let $r = \lfloor \frac{p}{2} \rfloor$ and B be the collection of all row vectors $b = (v_j)_{1 \leq j \leq p}$ such that $v_j = 0$ for all $1 \leq j \leq p-r$ and $v_j = 0$ or 1 for $p-r+1 \leq j \leq p$ under the constraint that $\|b\|_0 = k$ (where the value of k will be specified later). We can view each $(b_1, \cdots, b_r)$ as an $r \times p$ matrix with the i-th row being b_i.

Then, we define the set T and its corresponding distributions. Define $\Lambda \subset B^r$ to be the set of all elements in B^r such that each column is less than or equal to 2k. For each matrix $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_r) \in \Lambda$, define a $p \times p$ matrix $A_m(\lambda_m)$ by making the m-th row and column of $A_m(\lambda_m)$ be λ_m and the rest of entries be 0.

Next, we construct the distributions. Let $T = \mathcal{V} \otimes \Lambda$. For each $\tau = (v, \lambda)$, we define a matrix $P_\tau = \mathcal{N}(0, \Sigma(\tau))$ with the matrix $\Sigma(\tau)$ having the following form

$$\Sigma(\tau) = cI_p + c\alpha_{n,p,\epsilon} \sum_{j=1}^{n} v_j A_j(\lambda_j), \quad (21)$$

where $c > 0$ is some constant to be specified later and $\alpha_{n,p,\epsilon} = \gamma\sqrt{\frac{\log p}{n\epsilon^2}}$ for some universal small enough constant γ.

We first choose c, γ and k to make the Gaussian distribution $\mathcal{N}(0, \Sigma(\tau))$ contained in the class (20).

Lemma 5. *Under the assumption of $n \geq C\frac{s^2 \log p}{\epsilon^2}$, if let $c \leq \min\{\frac{\rho}{2}, \frac{1}{10p}\}$ and $k = \max\{\lceil \frac{s}{2} \rceil - 1, 0\}$, then there is a γ, which depends only on C, such that $\mathcal{N}(0, \Sigma(\tau)) \in \mathcal{P}(\rho, s)$ for every $\tau \in T$, where T is the set defined in the above paragraph.*

Proof. We first bound the term of $\|\Sigma(\tau)\|_2$. Note that since $\Sigma(\tau)$ is symmetric, we have $\|\Sigma(\tau)\|_2 \leq \|\Sigma(\tau)\|_1$. By the construction of $\Sigma(\tau)$, we can see that the ℓ_1 norm of each column in $\Sigma(\tau)$ is less than $1 + 2k\alpha_{n,p,\epsilon} \leq 1 + s\gamma\sqrt{\frac{\log p}{n\epsilon^2}}$. Thus, we have $\|\Sigma(\tau)\|_2 \leq c + cs\gamma\sqrt{\frac{\log p}{n\epsilon^2}}$.

We need $\mathcal{N}(0, \Sigma(\tau))$ satisfying (19). By [26], we know that it is sufficient to have $\|\Sigma(\tau)\|_2 \leq \rho$.

Let $\Sigma(\tau) = V^T Q V$ be the SVD decomposition and $Q = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$. Then, for $X \sim \mathcal{N}(0, \Sigma(\tau))$, we have $VX \sim \mathcal{N}(0, Q)$. Thus, $\|X\|_2^2 = \|VX\|_2^2 \leq \|\Sigma(\tau)\|_2 Y$, where Y is a χ_p^2 random variable. For the χ^2-distribution, we have the following concentration bound.

Lemma 6 ([27])**.** *If $z \sim \chi_n^2$, then*

$$\mathbb{P}[z - n \geq 2\sqrt{nx} + 2x] \leq \exp(-x).$$

Thus, with probability at least $1 - \exp(-p)$, we have $Y \leq 5p$. This means that, to ensure $\|X\|_2 \leq 1$, it is sufficient to have $5p\|\Sigma(\tau)\|_2 \leq 1$. Thus, we need that

$$c + cs\gamma\sqrt{\frac{\log p}{n\epsilon^2}} \leq \min\{\rho, \frac{1}{5p}\}. \tag{22}$$

Taking $c = \min\{\rho/2, \frac{1}{10p}\}$ and choosing a small enough $\gamma \leq \frac{\sqrt{C}}{2}$, we can get the proof. □

In order to use Theorem 4, we need to bound the term

$$\alpha = \min_{H(v(\tau),v(\tau'))>1, v(\tau),v(\tau')\in\mathcal{V}} \frac{\|\Sigma(\tau) - \Sigma(\tau')\|_2^2}{2H(v(\tau), v(\tau'))},$$

which is due to the following Lemma in [22].

Lemma 7. *Under the conditions given in Lemma 5, we have* $\alpha \geq \frac{(k\alpha_{n,p,\epsilon})^2}{p}$.

Proof of Lemma 7. Let the vector $v = (v_i)_{1\leq i\leq p}$ be a p-vector with $v_i = 0$ for $1 \leq i \leq p - r$ and $v_i = 1$ for $p - r + 1 \leq i \leq p$. Denote $w = (w_i)_{1\leq i\leq p} = (\Sigma(\tau) - \Sigma(\tau'))v$. Note that for each i, if $|v_i(\tau) - v_i(\tau')| = 1$, then we have $|w_i| = k\alpha_{n,p,\epsilon}$. Then there are at least $H(v_i(\tau), v_i(\tau')$ number of elements w_i with $|w_i| = k\alpha_{n,p,\epsilon}$, which implies

$$\|(\Sigma(\tau) - \Sigma(\tau'))v\|_2^2 \geq H(v_i(\tau), v_i(\tau')(k\alpha_{n,p,\epsilon})^2.$$

Since $\|v\|_2^2 \leq p$, we have

$$\begin{aligned}\|\Sigma(\tau) - \Sigma(\tau')\|_2^2 &\geq \frac{\|(\Sigma(\tau) - \Sigma(\tau'))v\|_2^2}{\|v\|_2^2} \\ &\geq \frac{H(v_i(\tau), v_i(\tau')(k\alpha_{n,p,\epsilon})^2}{p}\end{aligned}$$

Thus, $\alpha \geq \frac{(k\alpha_{n,p,\epsilon})^2}{p}$. □

The following key lemma gives a lower bound on the term

$$\text{Average}_{v_{-j},\lambda_{-j}}(D_{\chi^2}(\bar{P}_{j,+1,v_{-j},\lambda_{-j}} \| \bar{P}_{j,-1,v_{-j},\lambda_{-j}}))^n.$$

Lemma 8. *Under the conditions on T, $\Sigma(\tau)$ and the conditions of given in Lemma 5, the following holds for every $j \in [r]$, when γ is sufficiently small and p is sufficiently large*

$$Average_{v_{-j},\lambda_{-j}}(D_{\chi^2}(\bar{P}_{j,+1,v_{-j},\lambda_{-j}} \| \bar{P}_{j,-1,v_{-j},\lambda_{-j}}))^n \leq \frac{3}{4}\frac{1}{\epsilon^{2n}}.$$

Proof of Lemma 8. Our proof is similar to the proof of Lemma 6 in [22] with difference parameters. Here we only give a sketch of the proof.

Without loss of generality, we only consider the case where $j = 1$. And we denote the density function of $\bar{P}_{1,a,\upsilon_{-1},\lambda_{-1}}$ be $\bar{p}_{1,a,\upsilon_{-1},\lambda_{-1}}$. Also, we have

$$D_{\chi^2}(\bar{P}_{1,+1,\upsilon_{-1},\lambda_{-1}} \| \bar{P}_{1,-1,\upsilon_{-1},\lambda_{-1}}) = \int \frac{\bar{p}^2_{1,1,\upsilon_{-1},\lambda_{-1}}(x)}{\bar{p}_{1,-1,\upsilon_{-1},\lambda_{-1}}(x)} dx - 1.$$

By the definition, we know that the covariance matrix of the distribution $\bar{P}_{1,-1,\upsilon_{-1},\lambda_{-1}}$ has the form

$$\Sigma_0 = \begin{pmatrix} c & 0_{1\times(p-1)} \\ 0_{(p-1)\times 1} & S_{(p-1)\times(p-1)} \end{pmatrix} \tag{23}$$

Here $S_{(p-1)\times(p-1)} = (s_{ij})_{2\le i,j\le p}$ is a symmetric matrix uniquely determined by $(\upsilon_{-1}, \lambda_{-1})$ where for $i \le j$,

$$s_{ij} = \begin{cases} 1, i = 1 \\ c\alpha_{n,p,\epsilon}, \upsilon_i = \lambda_i(j) = 1 \\ 0, \text{otherwise} \end{cases} \tag{24}$$

Let

$$\Lambda_1(m) = \{a \in B : \exists \tau \in T \text{ s.t. } \lambda_1(\tau) = a, \lambda_{-1} = m\}$$

which gives the rest of all possible values of the first row with the rest of the rows fixed, that $\lambda_{-1}(\tau) = m$. Let $n_{\lambda_{-1}}$ be the number of columns of λ_{-1} with the column sum equal to $2k$ for which the first row has no choice but to take value 0 in this column. Set $p_{\lambda_{-1}} = r - n_{\lambda_{-1}}$. We have $p_{\lambda_{-1}} \ge \frac{p}{4} - 1$. Since $2kn_{\lambda_{-1}} \le rk$, the total number of 1s in the upper triangular matrix by the construction of the parameter set, we thus have $n_{\lambda_{-1}} \le \frac{r}{2}$, thus $p_{\lambda_{-1}} = r - n_{\lambda_{-1}} \ge \frac{r}{2} \ge \frac{p}{4} - 1$. Thus we have $|\Lambda_1(\lambda_{-1})| = \binom{p_{\lambda_{-1}}}{k}$. Then from the definition, we have $\bar{P}_{1,1,\upsilon_{-1},\lambda_{-1}}$ is an average of $\binom{p_{\lambda_{-1}}}{k}$ multivariate normal distribution with the covariance matrix has the following form:

$$\begin{pmatrix} c & \mathbf{r}_{1\times(p-1)} \\ \mathbf{r}_{(p-1)\times 1} & S_{(p-1)\times(p-1)} \end{pmatrix} \tag{25}$$

With $\|\mathbf{r}\|_0 = k$ with non-zero elements of $\mathbf{r}$ equal $c\alpha_{n,p,\epsilon}$ and the submatrix $S_{(p-1)\times(p-1)}$ is the same as the ones in Σ_0 in (23).

We have the following lemma, given by [22]

Lemma 9. *Let g_i be the density function of $\mathcal{N}(0, \Sigma_i)$ for $i = 0, 1, 2$, then we have*

$$\int \frac{g_1 g_2}{g_0} = [det(I - \Sigma_0^{-2}(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)]^{-\frac{1}{2}}. \tag{26}$$

Let Σ_0 defined above and determined by $\upsilon_{-1}, \lambda_{-1}$. Let Σ_1 and Σ_2 be the form above with the first row λ_1, λ_1', respectively. Set

$$R^{\upsilon_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} = -\log\det(I - \Sigma_0^{-2}(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)). \tag{27}$$

Now we denote the average as the expectation, then we have

$$\mathbb{E}_{v_{-1},\lambda_{-1}}(D_{\chi^2}(\bar{P}_{1,+1,v_{-1},\lambda_{-1}} \| \bar{P}_{1,-1,v_{-1},\lambda_{-1}}))^n \tag{28}$$

$$\leq \mathbb{E}_{v_{-1},\lambda_{-1}}[\mathbb{E}_{(\lambda_1,\lambda_1')|\lambda_{-1}}[\exp(\frac{n}{2}(R_{\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}} - 1))] \tag{29}$$

$$\leq \mathbb{E}_{v_{-1},\lambda_{-1}}[\mathbb{E}_{(\lambda_1,\lambda_1')|\lambda_{-1}}[\exp(\frac{n}{2}(R_{\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}})) - 1] \tag{30}$$

$$= \mathbb{E}_{\lambda_1,\lambda_1'}[\mathbb{E}_{(v_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}[\exp(\frac{n}{2}(R_{\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}})) - 1] \tag{31}$$

where λ_1 and λ_1' are independent and uniformly distributed over $\Lambda_1(\lambda_{-1})$ for given λ_{-1}, and the distribution of (v_{-1}, λ_{-1}) given (λ_1, λ_1') is inform over $T_{-1}(\lambda_1, \lambda_{-1})$, where

$$T_{-1}(a_1, a_2) = \{-1, +1\}^{r-1} \otimes \{c \in \Lambda_{-1} : \exists \tau_i \in T, i = 1, 2 \text{ s.t. } \lambda_1(\tau_i) = a_i, \lambda_{-1}(\tau_i) = v\}$$

□

We now have the following lemma for the term $(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)$, which corresponds to the Lemma 10 in [22]:

Lemma 10. *Let $\Sigma_0, \Sigma_1, \Sigma_2$ be the same covariance matrices as above. Define J to be the number of overlapping $c\alpha_{n,p,\epsilon}$'s between Σ_1 and Σ_2 on the first row, and define the matrix Q as the following*

$$Q = (q_{ij})_{1 \leq i,j \leq p} = (\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0).$$

Then there are index subsets I_r and I_c in $\{2, \cdots, p\}$ with $|I_r| = |I_c| = k$ and $|I_r \bigcap I_c| = J$

$$q_{ij} = \begin{cases} Jc^2\alpha_{n,p,\epsilon}^2, i = j = 1 \\ c^2\alpha_{n,p,\epsilon}^2, i \in I_r \text{ and } j \in I_c \\ 0, \text{otherwise} \end{cases} \tag{32}$$

And the matrix $(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$ has rank 2 with two identical non-zero eigenvalues $Jc^2\alpha_{n,p,\epsilon}^2$.

Thus by Lemma 10 and the Lemma 11 in [22] we have:

Lemma 11 (Lemma 11 in [22])**.** *Let $R_{1,\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}}$ satisfies*

$$R_{\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}} = -2\log(1 - Jc^2\alpha_{n,p,\epsilon}^2) + R_{1,\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}} \tag{33}$$

Then uniformly over J, we have

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}[\mathbb{E}_{(v_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}[\exp(\frac{n}{2}(R_{1,\lambda_1,\lambda_1'}^{v_{-1},\lambda_{-1}}))] \leq \frac{3}{2}.$$

Next we will prove our lemma. By (31) and Lemma 11 we now have

$$\begin{aligned}
&\mathbb{E}_{\lambda_1,\lambda_1'}[\mathbb{E}_{(\upsilon_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}[\exp(\frac{n}{2}(R^{\upsilon_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'}))-1]\\
&=\mathbb{E}_J\{\exp[-n\log(1-Jc^2\alpha^2_{n,p,\epsilon})]\times\\
&\mathbb{E}_{(\lambda_1,\lambda_1')|J}[\mathbb{E}_{(\upsilon_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}[\exp(\frac{n}{2}(R^{\upsilon_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}))]-1\}\\
&\le\mathbb{E}_J\{\frac{3}{2}\exp[-n\log(1-Jc^2\alpha^2_{n,p,\epsilon})]-1\}
\end{aligned}$$

Recall that J is the number of overlapping $c\alpha_{n,p,\epsilon}$'s between Σ_1 and Σ_2 on the first row. Thus J has the hypergeometric distribution as λ_1, λ_1' vary in B for each given λ_{-1}. For $0 \le j \le k$, the same as in [22], we have

$$\mathbb{E}(\mathbb{I}\{J=j\}) = \binom{k}{j}\binom{p_{\lambda_{-1}}-k}{k-j}/\binom{p_{\lambda_{-1}}}{k} \le (\frac{k^2}{p/4-1-k})^j.$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}_J\{\frac{3}{2}\exp[-n\log(1-Jc^2\alpha^2_{n,p,\epsilon})]-1\}\\
&\le\sum_{j=0}^{k}(\frac{k^2}{p/4-1-k})^j\{\frac{3}{2}\exp[-n\log(1-jc^2\alpha^2_{n,p,\epsilon})]-1\}\\
&\le\frac{1}{\epsilon^{2n}}\sum_{j=0}^{k}(\frac{k^2}{p/4-1-k})^j\{\frac{3}{2}\exp[-n\log(1-jc^2\gamma^2\frac{\log p}{n})]-1\} \qquad (34)\\
&\le\frac{1}{\epsilon^{2n}}\sum_{j=0}^{k}(\frac{k^2}{p/4-1-k})^j\{\frac{3}{2}\exp[2jc^2\gamma^2\log p]\}+\frac{1}{\epsilon^{2n}}\frac{1}{2}\\
&\le\frac{1}{\epsilon^{2n}}\frac{3}{2}\sum_{j\ge1}(p^{1-1/\beta}p^{-2\gamma^2c^2})^{-j}+\frac{1}{2}\frac{1}{\epsilon^{2n}} \qquad (35)\\
&\le C\frac{1}{\epsilon^{2n}}\sum_{j\ge1}(p^{1/2-1/2\beta})^{-j}+\frac{1}{2}\frac{1}{\epsilon^{2n}}\le\frac{3}{4}\frac{1}{\epsilon^{2n}} \qquad (36)
\end{aligned}$$

Where (34) is due to that, let $a = \frac{1}{\epsilon^2}$ and $b = jc^2\gamma^2\frac{\log p}{n}$, then it is sufficient to prove

$$\begin{aligned}
&-\log(1-ab)\le\log a-\log(1-b)\\
&\equiv\frac{1}{1-ab}\le\frac{a}{1-b}\\
&\equiv b(a+1)\le1
\end{aligned}$$

The final inequality is true due to that $b(a+1) \le 2ab \le 2kc^2\gamma^2\frac{\log p}{n\epsilon^2} \le 1$ when γ is small enough.

(35) is due to that $k^2 = O(\frac{n\epsilon^2}{\log p}) = O(\frac{n}{\log p}) = O(\frac{p^{1/\beta}}{\log p})$, and $\gamma^2 \le \frac{\beta-1}{54\beta}$ for sufficient large p. Combining Lemmas 5, 7 and 8 with $r = \lfloor\frac{p}{2}\rfloor$, by Corollary 1 we have the following lower bound theorem.

Theorem 4. *If $\epsilon \in (0, \frac{\ln 2}{2}]$, $n \geq C\frac{s^2 \log p}{\epsilon^2}$ and $p \geq c_1 n^\beta$ for $\beta > 1$, then the ϵ non-interactive private minimax risk in the metric of squared spectral norm satisfies the following inequality*

$$\mathcal{M}_n^{Nint}(\Sigma(\mathcal{P}(s,\rho)), \Phi\circ\rho, \epsilon) \geq \Omega(\frac{s^2 \log p}{n\epsilon^2}). \tag{37}$$

Proof of Theorem 4. By Corollary 1, Lemma 8 and 7 we have

$$\begin{aligned}
&\mathcal{M}_n^{Nint}(\psi(\theta), \Phi\circ\rho, \epsilon) \geq \frac{r\alpha}{4} \times \min_{1\leq j\leq r} \\
&(1 - \sqrt{\frac{\epsilon^{2n}}{2}\text{Average}_{v_{-j},\lambda_{-j}}(D_{\chi^2}(\bar{P}_{j,+1,v_{-j},\lambda_{-j}} \| \bar{P}_{j,-1,v_{-j},\lambda_{-j}}))^n}) \\
&\geq \frac{p}{2}\frac{k^2\alpha_{n,p,\epsilon}^2}{p}(1 - \sqrt{\frac{\epsilon^{2n}}{2}\frac{3}{4}\frac{1}{\epsilon^{2n}}}) \\
&\geq \Omega(\frac{s^2 \log p}{n\epsilon^2}).
\end{aligned}$$

□

For the upper bound, [11] recently showed that if each $\|X_i\|_2 \leq 1$ and $\{X_i\}_{i=1}^n \sim P$, where $P \in \mathcal{P}(s,\rho)$, then by using a thresholding method on the perturbed empirical covariance matirx with some well-defined threshold, the output $\tilde{\Sigma}$ satisfies $\|\tilde{\Sigma} - \Sigma\|_2^2 \leq O(\frac{s^2 \log p}{n\epsilon^2})$ with high probability. Combining this upper bound with Theorem 4, we can see that the bound $\Theta(\frac{s^2 \log p}{n\epsilon^2})$ is actually tight (i.e., optimal).

We note that for the non-private case, the optimal rate of minimax risk under the same measurement is $\Theta(\frac{s^2 \log p}{n})$ [22]. Thus, in this case, the impact of the local differential privacy is to change the number of efficient samples from n to $n\epsilon^2$. However, the collection of the considered distributions needs another assumption, which says that $\|X\|_2$ is bounded by 1 with high probability. This is not necessary in the non-private case [22], but needed for showing the upper bound.

Moreover, [11] also show that there is an (ϵ, δ) non-interactive LDP algorithm whose output $\tilde{\Sigma}$ satisfies $\|\tilde{\Sigma} - \Sigma\|_w^2 \leq O(\frac{s^2 \log p}{n\epsilon^2})$ for every $w \in [1, \infty]$ with high probability. One natural question is whether it is optimal. The following corollary provides an affirmative answer.

Corollary 2. *Under the assumptions given in Theorem 4, for each $w \in [1, \infty]$, the ϵ non-interactive private minimax risk in the metric of squared ℓ_w norm satisfies the following*

$$\mathcal{M}_n^{Nint}(\Sigma(\mathcal{P}(s,\rho)), \Phi\circ\rho, \epsilon) \geq \Omega(\frac{s^2 \log p}{n\epsilon^2}), \tag{38}$$

where the ℓ_w-norm of any matrix A is defined as $\|A\|_w = \sup \frac{\|Ax\|_w}{\|x\|_w}$.

Proof of Corollary 2. First, by the Riesz-Thorin Interpolation Theorem [23], we know that for every symmetric matrix M, $\|M\|_2 \leq \|M\|_w$ for all $w \in [1, \infty]$. Thus we have

under ℓ_w norm, by Lemma 6 we always have $\alpha_{n,p,\epsilon} \geq \frac{(k\alpha_{n,p,\epsilon})^2}{p}$, also since the term $\text{Average}_{v_{-j},\lambda_{-j}}(D_{\chi^2}(\bar{P}_{j,+1,v_{-j},\lambda_{-j}} \| \bar{P}_{j,-1,v_{-j},\lambda_{-j}})$ is independent on the norm, so we have the corollary. □

6. Conclusion and Discussion

In this paper we propose a general framework called General Assouad Lemma, which can be used to derive lower bounds for private mininax risk of matrix-related estimation problems in the local differential privacy model. The method generalizes the previous private Le Cam and private Assouad lemma in [12]. As an application of this lemma, we give the optimal lower bound on the non-interactive private mininax risk of the LDP sparse covariance matrix estimation problem in the metric of squared spectral norm.

There are still some open problems. Firstly, both Theorem 3 and 4 are restricted to non-interactive LDP protocols. The first open question is whether they can be extended to the sequential LDP model. Secondly, from Theorem 3 we can see that the lower bound holds under the assumption of $\epsilon \in (0, \frac{\ln 2}{2}]$. Thus, the second open question is whether the range of ϵ can be enlarged, or whether better result can be achieved when ϵ is larger, such as those in [28]? Recently, [20] extended the classical private Assouad lemma to the case where $\epsilon \in [0, \infty)$ via some results in the theory of communication complexity. However, their theorem cannot be used in our problem. The main reason is that, in their main results (Theorem 10 and Corollary 11 in [20]), they need the two distributions P_1 and P_{-1} satisfy strong data processing inequalities (SDPI), and also they should satisfy $|\log \frac{dP_1}{dP_{-1}}|$ is bounded by a constant under the assumption that the coordinates of X are independent. However, it is quite hard to bound the term or proof the SDPI property for our distributions in (21) due to the facts that the coordinates of the samples are dependent and the forms of our distributions are quite complicated. Thus, to extend to general $\epsilon \in (0, \infty)$ case we need new methods, which will be left for future work. The third open question is whether Theorem 2 and 3 can be used to other matrix-related estimation problems? We leave them for future research.

Acknowledgements

The research of this work was supported in part by NSF through grants CCF-1716400 and IIS-1910492. Part of this work was done while Di Wang was visiting the Simons Institute of the Theory for Computing.

References

[1] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in: TCC, Vol. 3876, Springer, 2006, pp. 265–284.

[2] J. Near, Differential privacy at scale: Uber and berkeley collaboration, in: Enigma 2018 (Enigma 2018), USENIX Association, Santa Clara, CA, 2018.

[3] Ú. Erlingsson, V. Pihur, A. Korolova, Rappor: Randomized aggregatable privacy-preserving ordinal response, in: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, ACM, 2014, pp. 1054–1067.

[4] J. Tang, A. Korolova, X. Bai, X. Wang, X. Wang, Privacy loss in apple's implementation of differential privacy on macos 10.12, CoRR abs/1709.02753. arXiv:1709.02753.

[5] T. T. Cai, Y. Wang, L. Zhang, The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy, arXiv preprint arXiv:1902.04495.

[6] T. Steinke, J. Ullman, Tight lower bounds for differentially private selection, in: 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 552–563.

[7] D. Wang, A. Smith, J. Xu, High dimensional sparse linear regression under local differential privacy: Power and limitations, 2018 NIPS workshop in Privacy-Preserving Machine Learning.

[8] J. C. Duchi, F. Ruan, The right complexity measure in locally private estimation: It is not the fisher information, arXiv preprint arXiv:1806.05756.

[9] J. Ullman, Tight lower bounds for locally differentially private selection, arXiv preprint arXiv:1802.02638.

[10] J. Ge, Z. Wang, M. Wang, H. Liu, Minimax-optimal privacy-preserving sparse pca in distributed systems, in: International Conference on Artificial Intelligence and Statistics, 2018, pp. 1589–1598.

[11] D. Wang, J. Xu, Differentially private high dimensional sparse covariance matrix estimation, CoRR abs/1901.06413.
URL http://arxiv.org/abs/1901.06413

[12] J. C. Duchi, M. I. Jordan, M. J. Wainwright, Minimax optimal procedures for locally private estimation, Journal of the American Statistical Association 113 (521) (2018) 182–201.

[13] G. Kamath, J. Li, V. Singhal, J. Ullman, Privately learning high-dimensional distributions, arXiv preprint arXiv:1805.00216.

[14] M. Joseph, J. Kulkarni, J. Mao, Z. S. Wu, Locally private gaussian estimation, arXiv preprint arXiv:1811.08382.

[15] V. Karwa, S. Vadhan, Finite sample differentially private confidence intervals, arXiv preprint arXiv:1711.03908.

[16] M. Gaboardi, R. Rogers, O. Sheffet, Locally private mean estimation: Z-test and tight confidence intervals, arXiv preprint arXiv:1810.08054.

[17] K. Amin, T. Dick, A. Kulesza, A. M. Medina, S. Vassilvitskii, Private covariance estimation via iterative eigenvector sampling, 2018 NIPS workshop in Privacy-Preserving Machine Learning.

[18] J. C. Duchi, M. I. Jordan, M. J. Wainwright, Local privacy and statistical minimax rates, in: Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on, IEEE, 2013, pp. 429–438.

[19] J. Duchi, R. Rogers, Lower bounds for locally private estimation via communication complexity, arXiv preprint arXiv:1902.00582.

[20] J. Duchi, R. Rogers, Lower bounds for locally private estimation via communication complexity, in: A. Beygelzimer, D. Hsu (Eds.), Proceedings of the Thirty-Second Conference on Learning Theory, Vol. 99 of Proceedings of Machine Learning Research, PMLR, Phoenix, USA, 2019, pp. 1161–1191.
URL http://proceedings.mlr.press/v99/duchi19a.html

[21] A. B. Tsybakov, Introduction to Nonparametric Estimation, 1st Edition, Springer Publishing Company, Incorporated, 2008.

[22] T. T. Cai, H. H. Zhou, et al., Optimal rates of convergence for sparse covariance matrix estimation, The Annals of Statistics 40 (5) (2012) 2389–2420.

[23] T. T. Cai, W. Liu, H. H. Zhou, et al., Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation, The Annals of Statistics 44 (2) (2016) 455–488.

[24] C. Dwork, K. Talwar, A. Thakurta, L. Zhang, Analyze gauss: optimal bounds for privacy-preserving principal component analysis, in: Proceedings of the forty-sixth annual ACM symposium on Theory of computing, ACM, 2014, pp. 11–20.

[25] D. Wang, M. Huai, J. Xu, Differentially private sparse inverse covariance estimation, in: 2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP, 2018, pp. 26–29.

[26] J. Wellner, et al., Weak convergence and empirical processes: with applications to statistics, Springer Science & Business Media, 2013.

[27] B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection, Annals of Statistics (2000) 1302–1338.

[28] M. Ye, A. Barg, Optimal schemes for discrete distribution estimation under locally differential privacy, IEEE Transactions on Information Theory 64 (8) (2018) 5662–5676.