Lower Bound of Locally Differentially Private Sparse Covariance Matrix Estimation\textsuperscript{*}

Di Wang\textsuperscript{†}, Jinhui Xu

Department of Computer Science and Engineering
State University of New York at Buffalo, NY, USA.
\{dwang45,jinhui\}@buffalo.edu.

Abstract

In this paper, we study the sparse covariance matrix estimation problem in the local differential privacy model, and give a non-trivial lower bound on the non-interactive private minimax risk in the metric of squared spectral norm. 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.

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) [Dwork et al., 2006], are widely-accepted 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 [Near, 2018], Google [Erlingsson et al., 2014], Apple [Tang et al., 2017].

\textsuperscript{*}The research of this work was supported in part by NSF through grant CCF-1716400.

\textsuperscript{†}Contact Author; Part of the work was done when Di Wang was visiting the Simons Institute of the Theory for Computing.

While differentially private high dimensional estimation is quite promising, such as sparse linear regression [Cai et al., 2019] and selection problem [Steinke and Ullman, 2017], estimating high dimensional datasets in a locally differentially private manner could be quite challenging for many problems, such as sparse linear regression [Wang and Xu, 2019b], sparse mean estimation [Duchi and Ruan, 2018] and selection problem [Ullman, 2018]. 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 [Ge et al., 2018]. Recently, [Wang and Xu, 2019a] studied the locally differentially private high dimensional sparse covariance estimation problem and proposed an algorithm which achieves an upper bound of $O(\frac{s^2 \log p}{n})$ measured by the squared spectral norm, i.e., $\|\Sigma^\text{priv} - \Sigma^\ast\|_2^2$, 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 $O(\frac{s^2 \log p}{n})$ 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 differential privacy model, the private minimax risk (in the metric of squared spectral norm) of high dimensional sparse covariance matrix estimation is lower bounded by $\Omega(\frac{s^2 \log p}{n^2})$. 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 $\ell_w$ 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 Assouad lemma in [Duchi et al., 2018], 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 [Kamath et al., 2018; Joseph et al., 2018; Karwa and Vadhan, 2017; Gaboardi et al., 2018; Amin et al., 2018; Wang and Xu, 2019a]. For covariance matrix estimation in the central differential privacy model, [Karwa and Vadhan, 2017] considered the 1-dimensional Gaussian distribution estimation with (un)known variance. [Kamath et al., 2018] 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. [Amin et al., 2018] recently also investigated the low dimensional case of the problem in Frobenius norm and proposed an iterative eigenvector sampling method. The work that is the most related to ours is probably the one in [Wang and Xu, 2019a], 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 [Gaboardi et al., 2018; Joseph et al., 2018; Wang and Xu, 2019a]. Specifically, [Gaboardi et al., 2018; Joseph et al., 2018] 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, [Wang and Xu, 2019a] proposed a general method which achieves an upper bound of $O(d^2 \log n)$ 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 [Duchi and Ruan, 2018; Duchi et al., 2018; 2013; Joseph et al., 2018; Duchi and Rogers, 2019]. [Duchi and Ruan, 2018; Duchi et al., 2018; 2013] 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 $\chi^2$-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 [Duchi et al., 2018] which can be seen as a special case of our general private Assouad lemma.

3 Preliminaries

In this section, we introduce some definitions that will be used throughout the paper. More details can be found in [Duchi et al., 2018].

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 $\Sigma_1$ and $\Sigma_2$.

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 $P$ be a class of distributions over a data universe $\mathcal{X}$. For each distribution $p \in P$, there is a deterministic function $\theta(p) \in \Theta$, where $\Theta$ is the parameter space. Let $p : \Theta \times \Theta \mapsto \mathbb{R}_+$ be a semi-metric function on the space $\Theta$ 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 $p \in P$, and $\hat{\theta} : \mathcal{X}^n \mapsto \Theta$ be some estimator. Then the minimax risk in metric $\Phi(p)$ is defined by the following saddle point problem:

$$M_p(\theta(P), \Phi(p)) := \inf_{\hat{\theta}} \sup_{p \in P} \mathbb{E}_p[\Phi(p(\hat{\theta}(X_1, \ldots, X_n), \theta(p)))],$$

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 [Duchi et al., 2013].

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:

$$\{X_i, Z_1, \ldots, Z_{i-1}\} \Rightarrow Z_i, Z_i \perp X_i \mid \{X_i, Z_1, \ldots, Z_{i-1}\}$$

for all $j \neq i$ and $i \in [n]$, where $\perp$ means independent relation.

The full conditional distribution can be specified in terms of conditionals $Q_i(Z_i \mid X_i = x_i, Z_{i:1} = z_{i:1})$. 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 ([Duchi et al., 2013]). For a given privacy parameter $\epsilon > 0$, the random variable $Z_i$ is an $\epsilon$ sequentially locally differentially private view of $X_i$ if for all $z_1, z_2, \ldots, z_{i-1}$ and $x, x' \in \mathcal{X}$ we have the following for all the events $S$:

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

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

$$\frac{Q(Z_i \in S \mid X_i = x_i)}{Q(Z_i \in S \mid X_i = x_i')} \leq e^\epsilon.$$
We say that the privacy mechanism \( Q = \{Q_i\}_{i=1}^{n} \) is \( \epsilon \)-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 \( Q_\epsilon \) be the set of conditional distributions that have the \( \epsilon \)-LDP property. For a given set of samples \( \{X_j\}_{j=1}^{n} \), let \( \{Z_i\}_{i=1}^{n} \) be the set of observations produced by any distribution \( Q \in Q_\epsilon \). Then, our estimator will be based on \( \{Z_i\}_{i=1}^{n} \), that is, \( \hat{\theta}(Z_1, \ldots, Z_n) \). This yields a modified version of the minimax risk:

\[
\mathcal{M}_n(\theta(P), \Phi, Q) = \inf_{\theta \in \mathcal{P}} \inf_{Q \in Q_\epsilon} \mathbb{E}_q(\hat{\theta}(Z_1, \ldots, Z_n), \theta(p)).
\]

Lemma 1 (Assouad Lemma). Under the conditions stated in above paragraph, we have the following.

\[
\mathcal{M}_n(\theta(P), \Phi, Q) \geq \frac{\alpha}{4} \sum_{j=1}^{r} \mathbb{E}_q(1 - \|P^n_{j+1} - P^n_{j-1}\|_{TV}).
\]

where \( \alpha \) satisfies

\[
\alpha = \min_{\mathbb{H}(v, v') \geq 1, v, v' \in \mathbb{V}} \mathbb{E}_q(\hat{\theta}(Z_1, \ldots, Z_n), \theta(p)).
\]

Theorem 1. Under the conditions given in Lemma 1, the \( \epsilon \) private minimax risk satisfies:

\[
\mathcal{M}_n(\theta(P), \Phi, \epsilon, Q) \geq \frac{\alpha}{4} \sum_{j=1}^{r} \mathbb{E}_q(1 - \|M^n_{j+1} - M^n_{j-1}\|_{TV}).
\]

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 [Tsypakov, 2008] and its two-directional generalization [Cai et al., 2012].

Assouad’s method works with a hypercube \( \mathcal{V} = \{-1, +1\}^r \) for some \( r \in \mathbb{N} \). It transforms an estimation problem into 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 \( X_1, X_2, \ldots, X_n \) are drawn from the distribution \( P \) conditioned on \( V = v \). For each \( j \in [r] \), we define the product of distributions

\[
P^n_{j+1} = \frac{1}{2r-1} \sum_{v', v_j = \pm 1} P^n_{v', v_j = +1} P^n_{v', v_j = -1}.
\]

where \( P^n_j \) is the product distribution of \( X_1, \ldots, X_n \). Then, Assouad lemma can be stated as follows.

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

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

where \( \| \cdot \|_{TV} \) is the total variation distance, \( \alpha = \min_{\mathbb{H}(v, v') \geq 1, v, v' \in \mathbb{V}} \mathbb{E}_q(\hat{\theta}(Z_1, \ldots, Z_n), \theta(p)) \), and \( H(v, v') \) is the hamming distance between \( \theta \) and \( \theta' \), i.e., \( H(v, v') = \sum_{j=1}^{r} 1 \{v_j \neq v'_j\} \).

Instead of restricting to a hypercube \( \mathcal{V} \), the general Assouad lemma in [Cai et al., 2012] 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 \( B \subseteq \mathbb{R}^p \), let \( \mathcal{V} = (\{-1, +1\}^r)^p \). Define \( T = \mathcal{V} \otimes \Lambda = \{ (v, \lambda) : v \in \mathcal{V} \text{ and } \lambda \in \Lambda \} \). Let \( \mathcal{D}_\lambda = |\Lambda| \). For a given \( a \in \{-1, +1\}^r \) and \( j \in [r] \), we let \( T_{a,j} = \{ v : v_j = a \} \), where \( v_j \) is the \( j \)-th coordinate of the first component of \( v \). It is easy to see that \( |T_{a,j}| = 2^{r-1} \mathcal{D}_\lambda \).

Let \( P_{a,j} = \frac{1}{2r-1} \sum_{v \in T_{a,j}} P^n_v \), \( P_{a,j} = \frac{1}{2r-1} \sum_{v \in T_{a,j}} P^n_v \).

Lemma 2 (General Assouad’s Lemma [Cai et al., 2016]). Under the conditions stated in above paragraph, we have the following.

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

where \( \alpha \) satisfies

\[
\alpha = \min_{H(v, v') \geq 1, v, v' \in \mathbb{V}} \mathbb{E}_q(\hat{\theta}(Z_1, \ldots, Z_n), \theta(p)).
\]

and \( v(a) \) is the first component of \( v \).

Now, we present the locally private version of Lemma 2. Suppose that we draw samples \( Z_1, \ldots, Z_n \) according to \( \epsilon \)-LDP channel \( Q(\cdot | X_1, n) \). Then, conditioned on \( V = v \), the private sample is distributed according to the marginal distribution \( M^n_v \):

\[
M^n_v(S) = \int Q^n(S|x_1, x_2, \ldots, x_n) dP^n_v(x_1, x_2, \ldots, x_n).
\]

Specifically, when \( Q \) is non-interactive, we have \( M^n_v = (\int Q^n(\cdot | x) dP^n_v(x))^{\otimes n} \). Similarly to (3), we can define \( M^n_{a,j} \) and \( M^n_{a,j} \) for \( a \in \{-1, +1\}^r \) 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 \( \epsilon \) private minimax risk satisfies:

\[
\mathcal{M}_n(\theta(P), \Phi, \epsilon) \geq \frac{\alpha}{4} \sum_{j=1}^{r} \mathbb{E}_q(1 - \|M^n_{j+1} - M^n_{j-1}\|_{TV}).
\]
Theorem 2. Under the conditions given in Theorem 1 and further assuming that $e \in (0, \frac{1}{2}]$, the $\epsilon$ sequential private minimax risk in the metric $\Phi_\rho$ satisfies

\[
M_n^{\text{seq}}(\theta, \Phi_\rho, e) \geq \frac{ar}{4} \left[ 1 - \frac{e^2}{2} \sup_{y \in \mathbb{B}_\infty(x)} \sum_{j=1}^{r} \left( \int y(x) (dP_{j+1} - dP_{j-1})^2 \right)^{\frac{1}{2}} \right],
\]

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

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

\[
D_{\chi^2}(M_{j+1}, M_{j-1}) = \int \frac{(m_{j+1}(z) - m_{j-1}(z))^2}{m_{j-1}(z)} dz \quad (13)
\]

\[
\leq \int c_e^2 (\inf_x q^2(z|x)(e^e - 1)^2 ||P_{j+1} - P_{j-1}||_{TV}^2 dz \quad (14)
\]

\[
\leq c_e^2 (e^e - 1)^2 ||P_{j+1} - P_{j-1}||_{TV}^2 \quad (15)
\]

where $c_e = \min\{2, e^e\}$, (13) is by the definition of $\chi^2$ divergence, (14) is by Lemma 3 in [Duchi et al., 2018] and (15) is due to the fact that $\int \inf_x q(z|x) dz \leq 1$.

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

The inequality in Lemma 3 is weaker than the one in Theorem 1 of [Duchi et al., 2018] in the sense that it becomes the lower 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 [Duchi et al., 2018; Duchi and Ruan, 2018; Duchi and Rogers, 2019], 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 $\chi^2$-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 $\chi^2$-divergence of the hard instances than other measurements. This is also the reason that existing methods cannot be applied to our problem.
\( \Lambda_A \) the set \( \Lambda_A = \{ \lambda_A (r) : r \in T \} \). For \( a \in \{+1, -1\}, \ b \in \{-1, +1\} \) and \( c \in \Lambda_{-j} \subseteq B^{-1} \), we let

\[
T_{\lambda_A(a,b,c)} = \{ r \in T : u_j(r) = a, \ u_{-j}(r) = b, \ \lambda_{-j}(r) = c \}
\]

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

\[
\tilde{P}^n_{j,a,b,c} = \frac{1}{D_{\lambda_A(a,b,c)}} \sum_{r \in T_{\lambda_A(a,b,c)}} P^n_r,
\]

and \( \tilde{M}^n_{j,a,b,c} \) 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 \( e \in (0, \ln 2) \), the \( \epsilon \)-non-private interactive privacy minimax risk in the metric \( \Phi_{\infty} \) satisfies

\[
M_n N(\psi(\Theta), \Phi_{\infty}, c) \geq \frac{6a}{4 \epsilon} \min_{1 \leq j \leq r} \left( 1 + \sqrt{\frac{2 \epsilon}{n}} \right)
\]

where the average over \( u_{-j}, \lambda_{-j} \) 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 \( \{ P^n_{j,a,b,c} \}_{b,c \in T_{\lambda_A}} \), where the set \( T_{\lambda_A} \) is

\[
T_{\lambda_A} = [0, 1]^{r-1} \otimes \lambda_{-j} = \{(b, c) : \exists r \in T \text{ s.t. } u_{-j}(r) = b \text{ and } \lambda_{-j}(r) = c \}.
\]

That is, \( P^n_{j,a} = \sum_{(b,c) \in T_{\lambda_A}} w_{b,c} P^n_{j,a,b,c} \) where \( w_{b,c} = \frac{D_{\lambda_A(b,c)}}{2^{r-1} D_{\lambda_A}} \) (note that since \( D_{\lambda_A(b,c)} \) is independent of \( a \), we omit it). Also, \( \sum_{(b,c) \in T_{\lambda_A}} w_{b,c} = 1 \). 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 3, we have

\[
\| M^n_{j+1} - M^n_{j-1} \|_{TV} \leq \sum_{(b,c) \in T_{\lambda_A}} w_{b,c} \| M^n_{j+1,b,c} - M^n_{j-1,b,c} \|_{TV}
\]

\[
= \text{Average}_{b,c} \| M^n_{j+1,b,c} - M^n_{j-1,b,c} \|_{TV}.
\]

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

\[
\| M^n_{j+1,b,c} - M^n_{j-1,b,c} \|_{TV}^2 \leq D_{\lambda_A}^2 (\| \tilde{M}^n_{j+1,b,c} - \tilde{M}^n_{j-1,b,c} \|_{TV})^2 \leq \frac{1}{2} (\min \{ \frac{2 \epsilon^2}{2} \} e^2 D_{\lambda_A}^2 (\| \tilde{M}^n_{j+1,b,c} - \tilde{M}^n_{j-1,b,c} \|_{TV}))^2.
\]

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

\[ \Box \]

### 5 Lower Bound of Private Sparse Covariance Estimation

We follow the settings in [Cai et al., 2012; Wang and Xu, 2019a]. Let \( X_1, \ldots, X_p \) be random samples from a zero-mean \( p \)-variate distribution with covariance matrix \( \Sigma = (\sigma_{ij})_{1 \leq i,j \leq p} \).

The goal of sparse covariance matrix estimation is to estimate the unknown \( \Sigma \) based on samples \( \{ X_1, \ldots, X_n \} \), 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, \( c_1 n^\theta \leq r \leq \exp(c_2 n) \) for some \( \theta > 1, c_1, c_2 > 0 \). We assume that the underlying covariance is sparse. That is, \( \Sigma \in \mathcal{G}(s) \) with

\[
G(s) = \{ \Sigma = (\sigma_{ij})_{1 \leq i,j \leq p} : \| \sigma_{j,j} \|_0 \leq s, \forall j \in [p] \},
\]

where \( \sigma_{j,j} \) is the \( j \)-th column of \( \Sigma \) 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 \( p \)-sub-Gaussian distribution. That is, for all \( i > 0 \) and \( \| v \|_2 = 1 \),

\[
\mathbb{P} \{ |\langle v, X \rangle | > t \} \leq \exp \left( -\frac{t^2}{2p} \right),
\]

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 \( \ell_2 \) norm of each \( X_i \) are bounded by 1 [Dwork et al., 2014; Ge et al., 2018; Wang et al., 2018; Wang and Xu, 2019a]. 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 \( 1 - e^{-2k p} \). This leads us to the following class of distributions \( P(s) \),

\[
P(s) = \{ P : X \sim \mathcal{P} \text{ satisfies (19) and } \| X \|_2 \leq 1 \}
\]

w.p at least \( 1 - e^{-2k p} \). \( \mathbb{E} X = 0, \mathbb{E} [X X^T] \in \mathcal{G}(s) \).

Before showing the lower bound, we first describe our construction of the hard indexing set \( T \) with their distributions \( \{ P_j \}_{j \in T} \) instances, which is motivated by the ones in [Cai et al., 2012].

We first construct the parameter set, which is the same as in [Cai et al., 2012]. Let \( r = \lceil \frac{p}{2} \rceil \) 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 \( v_j = 1 \) or \( v_j = 0 \) 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 \( (b_1, \ldots, b_p) \) 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 \( A \subseteq B \) to be the set of all elements in \( B \) such that each column is less than or equal to 2. For each matrix \( A = (\lambda_1, \lambda_2, \ldots, \lambda_r) \in A \), define a \( p \times p \) matrix \( A_m(\lambda_m) \) by making the \( m \)-th row and column of \( A_m(\lambda_m) \) be \( e_{\lambda_m} \) 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 \( P_\tau = N(0, \Sigma(\tau)) \) with the matrix \( \Sigma(\tau) \) having the following form

\[
\Sigma(\tau) = c I_p + c a_n p \sum_{j=1}^n v_j A_j(\lambda_j).
\]
where \( c > 0 \) is some constant to be specified later and \( a_{n,p,c} = \gamma \sqrt{\log \frac{p}{n}} \) for some universal small enough constant \( \gamma \).

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

**Lemma 4.** Under the assumption of \( n \geq \frac{C^2 \log p}{c^2} \), if let \( c \leq \min \{ \frac{\rho}{2}, \frac{1}{10p} \} \) and \( k = \max \{ \lceil \frac{2}{\rho} \rceil - 1, 0 \} \), then there is a \( \gamma \), which depends only on \( C \), such that \( \mathcal{N}(0, \Sigma(\tau)) \in P(\rho, s) \) 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 \( \| \Sigma(\tau) \|_2 = \| \Sigma(\tau) \|_1 \). By the construction of \( \Sigma(\tau) \), we can see that the \( \ell_1 \) norm of each column in \( \Sigma(\tau) \) is less than 1 + \( 2k a_{n,p,c} \leq 1 + s \gamma \sqrt{\frac{\log p}{nc^2}} \). Thus, we have \( \| \Sigma(\tau) \|_2 \leq c + cs \gamma \sqrt{\frac{\log p}{nc^2}} \).

We need \( \mathcal{N}(0, \Sigma(\tau)) \) satisfying (19). By [Weller and others, 2013], 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 \( Q = \text{diag}(\lambda_1, \ldots, \lambda_p) \). Then, for \( X \sim \mathcal{N}(0, \Sigma(\tau)) \), we have \( X \sim \mathcal{N}(0, Q) \). Thus, \( \| X \|_2^2 = \| VX \|_2^2 \leq \| \Sigma(\tau) \|_2 Y \), where \( Y \) is a \( \chi^2_p \) random variable. For the \( \chi^2 \)-distribution, we have the following concentration bound.

**Lemma 5 ([Laurent and Massart, 2000]).** If \( z \sim \chi^2_n \), 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

\[
c + cs \gamma \sqrt{\frac{\log p}{nc^2}} \leq \min\{\rho, \frac{1}{5p}\}.
\]

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

In order to use Theorem 4, we need to bound the term \( a = \min_{H(\alpha(\tau), \alpha'(\tau)) \geq 1, \alpha(\tau), \alpha'(\tau) \in \overset{\circ}{V}} \frac{\| \Sigma(\tau) - \Sigma(\tau') \|^2}{2H(\alpha(\tau), \alpha'(\tau))} \), which is due to the following Lemma in [Cai et al., 2012].

**Lemma 6.** Under the conditions given in Lemma 4, we have

\[
a \geq \frac{(ka_{n,p,c})^2}{p}.
\]

The following key lemma gives a lower bound on the term \( \text{Average}_{\alpha_1, \ldots, \alpha_p}(D_X(\tilde{P}_{j+1, \ldots, j+p} \| \tilde{P}_{j-1, \ldots, j-p}))^p \).

**Lemma 7.** Under the conditions on \( T, \Sigma(\tau) \) and the conditions of given in Lemma 4, the following holds for every \( j \in [r] \), when \( \gamma \) is sufficiently small and \( p \) is sufficiently large

\[
\text{Average}_{\alpha_1, \ldots, \alpha_p}(D_X(\tilde{P}_{j+1, \ldots, j+p} \| \tilde{P}_{j-1, \ldots, j-p}))^p \leq \frac{3}{4} \frac{1}{e^{2n}}.
\]

Combining Lemmas 4, 6 and 7 with \( r = \lceil \frac{p}{2} \rceil \), by Corollary 1 we have the following lower bound theorem.

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

\[
\mathcal{M}_n^{Nint}(\mathcal{P}(\sigma, \rho), \mathcal{F}_n, \epsilon) \geq \Omega(\frac{s^2 \log p}{ne^{2n}}).
\]

For the upper bound, [Wang and Xu, 2019a] recently showed that if each \( \| X_i \|_2 \leq 1 \) and \( \{X_i\}_{i=1}^P \sim P \), where \( P \in \mathcal{P}(s, \rho) \), then by using a thresholding method on the perturbed empirical covariance matrix with some well-defined threshold, the output \( \Sigma \) satisfies \( \| \Sigma - \Sigma_0 \|_2 \leq O(\frac{s^2 \log p}{ne^{2n}}) \) with high probability. Combining this upper bound with Theorem 4, we can see that the bound \( \Theta(\frac{s^2 \log p}{ne^{2n}}) \) 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 \( \Theta(\frac{s^2 \log p}{n}) \) [Cai et al., 2012]. Thus, in this case, the impact of the local differential privacy is to change the number of efficient samples from \( n \) to \( ne^{2n} \). 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 [Cai et al., 2012], but needed for showing the upper bound.

Moreover, [Wang and Xu, 2019a] also show that there is an \( (\epsilon, \delta) \) non-interactive LDP algorithm whose output \( \Sigma \) satisfies \( \| \Sigma - \Sigma_0 \|_2 \leq O(\frac{s^2 \log p}{ne^{2n}}) \) 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 \( \epsilon \) non-interactive private minimax risk in the metric of squared \( \ell_w \)-norm satisfies the following

\[
\mathcal{M}_n^{Nint}(\mathcal{P}(\sigma, \rho), \mathcal{F}_n, \epsilon) \geq \Omega(\frac{s^2 \log p}{n e^{2n}}),
\]

where the \( \ell_w \)-norm of any matrix \( A \) is defined as \( \| A \|_w = \sup_{\| x \|_w = 1} |Ax| \).

6 Discussions

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 \( c \in (0, \frac{\ln 2}{2}] \). Thus, the second open question is whether the range of \( c \) can be enlarged, or whether better result can be achieved when \( c \) is larger, such as those in [Ye and Barg, 2018]? 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.

References


