Optimal Estimation and Rank Detection for Sparse Spiked Covariance Matrices

Tony Cai, **Zongming Ma**, and **Yihong Wu**

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

Tony Cai: tcai@wharton.upenn.edu; Zongming Ma: zongming@wharton.upenn.edu; Yihong Wu: yihongwu@wharton.upenn.edu

Abstract

This paper considers a sparse spiked covariancematrix model in the high-dimensional setting and studies the minimax estimation of the covariance matrix and the principal subspace as well as the minimax rank detection. The optimal rate of convergence for estimating the spiked covariance matrix under the spectral norm is established, which requires significantly different techniques from those for estimating other structured covariance matrices such as bandable or sparse covariance matrices. We also establish the minimax rate under the spectral norm for estimating the principal subspace, the primary object of interest in principal component analysis. In addition, the optimal rate for the rank detection boundary is obtained. This result also resolves the gap in a recent paper by Berthet and Rigollet [2] where the special case of rank one is considered.

Keywords

Covariance matrix; Group sparsity; Low-rank matrix; Minimax rate of convergence; Sparse principal component analysis; Principal subspace; Rank detection

1 Introduction

Covariance matrix plays a fundamental role in multivariate analysis. Many methodologies, including discriminant analysis, principal component analysis and clustering analysis, rely critically on the knowledge of the covariance structure. Driven by a wide range of contemporary applications in many fields including genomics, signal processing, and financial econometrics, estimation of covariance matrices in the high-dimensional setting is of particular interest.

There have been significant recent advances on the estimation of a large covariance matrix and its inverse, the precision matrix. A variety of regularization methods, including banding, tapering, thresholding and penalization, have been introduced for estimating several classes of covariance and precision matrices with different structures. See, for example, [3, 4, 8, 12, 13, 17, 22, 26, 36, 41], among many others.

1.1 Sparse spiked covariance matrix model

In the present paper, we consider spiked covariance matrix models in the high-dimensional setting, which arise naturally from *factor models* with homoscedastic noise. To be concrete, suppose that we observe an $n \times p$ data matrix $\mathbf{X}$ with the rows $\mathbf{X}_{1*}, \ldots, \mathbf{X}_{n*}$ i.i.d. following a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ, denoted by by $N(\mathbf{0}, \Sigma)$, where the covariance matrix Σ is given by

$$\sum = Cov(\mathbf{X}_{i*}) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' + \sigma^2 \mathbf{I}_p, \quad (1)$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ with $\lambda_1 \quad \cdots \quad \lambda_r > 0$, and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ is $p \times r$ with orthonormal columns. The r largest eigenvalues of Σ are $\lambda_i + \sigma^2$, $i = 1, \ldots, r$, and the rest are all equal to σ^2. The r leading eigenvectors of Σ are given by the column vectors of $\mathbf{V}$. Since the spectrum of Σ has r spikes, (1) is termed by [19] as the spiked covariance matrix model. This covariance structure and its variations have been widely used in signal processing, chemometrics, econometrics, population genetics, and many other fields. See, for instance, [16, 24, 32, 34]. In the high-dimensional setting, various aspects of this model have been studied by several recent papers, including but not limited to [2, 5, 10, 20, 21, 31, 33, 35]. For simplicity, we assume σ is known. Since σ can always be factored out by scaling $\mathbf{X}$, without loss of generality, we assume $\sigma = 1$. Data-based estimation of σ will be discussed in Section 6.

The primary focus of this paper is on the setting where $\mathbf{V}$ and Σ are sparse, and our goal is threefold. First, we consider the minimax estimation of the spiked covariance matrix Σ under the spectral norm. The method as well as the optimal rates of convergence in this problem are considerably different from those for estimating other recently studied structured covariance matrices, such as bandable and sparse covariance matrices. Second, we are interested in rank detection. The rank r plays an important role in principal component analysis (PCA) and is also of significant interest in signal processing and other applications. Last but not least, we consider optimal estimation of the principal subspace span($\mathbf{V}$) under the spectral norm, which is the main object of interest in PCA. Each of these three problems is important in its own right.

We now explain the sparsity model of $\mathbf{V}$ and Σ. The difficulty of estimation and rank detection depends on the joint sparsity of the columns of $\mathbf{V}$. Let $\mathbf{V}_{j*}$ denote the jth row of $\mathbf{V}$. The row support of $\mathbf{V}$ is defined by

$$supp(\mathbf{V}) = \{j \in [p] : \mathbf{V}_{j*} \neq \mathbf{0}\}, \quad (2)$$

whose cardinality is denoted by $|supp(\mathbf{V})|$. Let the collection of $p \times r$ matrices with orthonormal columns be $O(p, r) = \{\mathbf{V} \in \mathbb{R}^{p \times r} : \mathbf{V}'\mathbf{V} = \mathbf{I}_r\}$. Define the following parameter spaces for Σ,

$$\Theta_0(k, p, r, \lambda, \tau) = \{\sum = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' + \mathbf{I}_p \quad : \lambda/\tau \leq \lambda_r \leq \cdots \leq \lambda_1 \leq \lambda, \\ \mathbf{V} \in O(p, r), |supp(\mathbf{V})| \leq k\}, \quad (3)$$

where $\tau \ge 1$ is a constant and $r \le k \le p$ is assumed throughout the paper. Note that the condition number of Λ is at most τ. Moreover, for each covariance matrix in Θ_0 (k, p, r, λ, τ), the leading r singular vectors (columns of $\mathbf{V}$) are *jointly k-sparse* in the sense that the row support size of $\mathbf{V}$ is upper bounded by k. The structure of group sparsity has proved useful for high-dimensional regression; See, for example, [30]. In addition to (3), we also define the following parameter spaces by dropping the dependence on τ and r, respectively:

$$\Theta_1(k, p, r, \lambda) \quad = \mathbf{cl} \bigcup_{\tau \ge 1} \Theta_0(k, p, r, \lambda, \tau)$$
$$= \{ \textstyle\sum = \mathbf{V}\Lambda\mathbf{V}' + \mathbf{I}_p : \begin{array}{l} 0 \le \lambda_r \le \cdots \le \lambda_1 \le \lambda, \\ \mathbf{V} \in O(p, r), |supp(\mathbf{V})| \le k \} \end{array} \quad (4)$$

and

$$\Theta_2(k, p, \lambda, \tau) = \bigcup_{r=0}^{k} \Theta_0(k, p, r, \lambda, \tau). \quad (5)$$

As a consequence of the group sparsity in $\mathbf{V}$, a covariance matrix Σ in any of the above parameter spaces has at most k rows and k columns containing nonzero off-diagonal entries. We note that the matrix is more structured than the so-called "*k-sparse*" matrices considered in [3, 8, 13], where each row (or column) has at most k nonzero off-diagonals.

1.2 Main contributions

In statistical decision theory, the minimax rate quantifies the difficulty of an inference problem and is frequently used as a benchmark for the performance of inference procedures. The main contributions of this paper include the *sharp non-asymptotic* minimax rates for estimating the covariance matrix Σ and the principal subspace span($\mathbf{V}$) under the squared spectral norm loss, as well as for detecting the rank r of the principal subspace. In addition, we also establish the minimax rates for estimating the precision matrix $\Omega = \Sigma^{-1}$ as well as the eigenvalues of Σ under the spiked covariance matrix model (1).

We establish the minimax rate for estimating the spiked covariance matrix Σ in (1) under the spectral norm

$$L(\textstyle\sum, \hat{\textstyle\sum}) = \|\textstyle\sum - \hat{\textstyle\sum}\|^2, \quad (6)$$

where for a matrix $\mathbf{A}$ its spectral norm is defined as $\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2$ with $\|\cdot\|_2$ the vector ℓ_2 norm. The minimax upper and lower bounds developed in Sections 3 and 4 yield the following optimal rate for estimating sparse spiked covariance matrices under the spectral norm

$$\inf_{\hat{\sum}} \sup_{\sum \in \Theta_1(k,p,r,\lambda)} \mathrm{E}\|\hat{\textstyle\sum} - \textstyle\sum\|^2 \asymp \left[\frac{(\lambda+1)k}{n} \log \frac{ep}{k} + \frac{\lambda^2 r}{n} \right] \wedge \lambda^2, \quad (7)$$

subject to certain mild regularity conditions.[1] The two terms in the squared bracket are contributed by the estimation error of the eigenvectors $\mathbf{V}$ and the eigenvalues, respectively. Note that the second term can be dominant if λ is large.

An important quantity of the spiked model is the rank r of the principal subspace span($\mathbf{V}$), or equivalently, the number of spikes in the spectrum of $\mathbf{\Sigma}$, which is of significant interest in chemometrics [24], signal array processing [25], and other applications. Our second goal is the minimax estimation of the rank r under *the zero–one loss*, or equivalently, the minimax detection of the rank r. It is intuitively clear that the difficulty in estimating the rank r depends crucially on the magnitude of the minimum spike λ_r. Results in Sections 3 and 4 show that the optimal rank detection boundary over the parameter space $\Theta_2(k, p, \lambda, \tau)$ is of

order $\sqrt{\frac{k}{n} \log \frac{ep}{k}}$. Equivalently, the rank r can be exactly recovered with high probability if

$\lambda_r \geq \beta \sqrt{\frac{k}{n} \log \frac{ep}{k}}$ for a sufficiently large constant β; On the other hand, reliable detection

becomes impossible by any method if $\lambda_r \leq \beta_0 \sqrt{\frac{k}{n} \log \frac{ep}{k}}$ for some positive constant β_0. Lying in the heart of the arguments is a careful analysis of the moment generating function of a squared symmetric random walk stopped at a hypergeometrically distributed time, which is summarized in Lemma 1. It is worth noting that the optimal rate for rank detection obtained in the current paper resolves a gap left open in a recent paper by Berthet and

Rigollet [2], where the authors obtained the optimal detection rate $\sqrt{\frac{k}{n} \log \frac{ep}{k}}$ for the rank-

one case in the regime of $k \ll \sqrt{p}$, but the lower bound deteriorates to $\sqrt{\frac{p}{kn}}$ when $k \gg \sqrt{p}$ which is strictly suboptimal.

In many statistical applications, instead of the covariance matrix itself, the object of direct interest is often a lower dimensional functional of the covariance matrix, e.g., the principal subspace span($\mathbf{V}$). This problem is known in the literature as sparse PCA [5, 10, 20, 31]. The third goal of the paper is the minimax estimation of the principal subspace span($\mathbf{V}$). To this end, we note that the principal subspace can be uniquely identified with the associated projection matrix $\mathbf{V}\mathbf{V}'$. Moreover, any estimator can be identified with a projection matrix $\hat{\mathbf{V}}\hat{\mathbf{V}}'$, where the columns of $\hat{\mathbf{V}}$ constitute an orthonormal basis for the subspace estimator. Thus, estimating span($\mathbf{V}$) is equivalent to estimating $\mathbf{V}\mathbf{V}'$. We aim to optimally estimate span($\mathbf{V}$) under the loss [37, Section II.4]

$$L(\mathbf{V}, \hat{\mathbf{V}}) = \left\| \mathbf{V}\mathbf{V}' - \hat{\mathbf{V}}\hat{\mathbf{V}}' \right\|^2, \quad (8)$$

---

[1]Here and after, $a \wedge b \triangleq \min(a, b)$ and $a_n \asymp b_n$ means that $\frac{a_n}{b_n}$ is bounded from both below and above by constants independent of n and all model parameters.

which equals the squared sine of the largest canonical angle between the respective linear spans. In the sparse PCA literature, the loss (8) was first used in [31] for multi-dimensional subspaces. For this problem, we shall show that, under certain regularity conditions, the minimax rate of convergence is

$$\inf_{\hat{\mathbf{V}}} \sup_{\Sigma \in \Theta_0(k,p,r,\lambda,\tau)} \mathrm{E}\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\|^2 \asymp \frac{(\lambda+1)k}{n\lambda^2} \log \frac{ep}{k} \wedge 1. \quad (9)$$

In the present paper we considered estimation of the principal subspace span($\mathbf{V}$) under the spectral norm loss (8). It is interesting to compare the results with those for optimal estimation under the Frobenius norm loss [10, 40] $\|\mathbf{V}\mathbf{V}' - \hat{\mathbf{V}}\hat{\mathbf{V}}'\|_F^2$, whose ratio to (8) is between 1 and $2r$. The optimal rate under the spectral norm loss given in (9) does not depend on the rank r, whereas the optimal rate under the Frobenius norm loss has an extra term $\frac{\lambda+1}{n\lambda^2} r(k-r)$, which depends on the rank r quadratically through $r(k-r)$ [10].

Therefore the rate under the Frobenius norm far exceeds (9) when $r \gg \log \frac{ep}{k}$. When $r = 1$, both norms lead to the same rate and the result in (9) recovers earlier results on estimating the leading eigenvector obtained in [5, 39, 29].

In addition to the optimal rates for estimating the covariance matrix Σ, the rank r and the principal subspace span($\mathbf{V}$), the minimax rates for estimating the precision matrix $\Omega = \Sigma^{-1}$ as well as the eigenvalues of Σ are also established.

1.3 Other related work

Apart from the spiked covariance matrix model studied in this paper, other covariance matrix models have been considered in the literature. The most commonly imposed structural assumptions include "Toeplitz", where each descending diagonal from left to right is constant, "bandable", where the entries of the covariance matrix decay as they move away from the diagonal, and "sparse", where only a small number of entries in each row/column are nonzero. The optimal rates of convergence were established in [11], [12] and [13] for estimating Toeplitz, bandable, and sparse covariance matrices, respectively. Estimation of sparse precision matrices has also been actively studied due to its close connection to Gaussian graphical models [9, 36, 41]. In addition, our work is also connected to the estimation of effective low-rank covariance matrices. See, for example, [28, 7] and the reference therein.

1.4 Organization

The rest of the paper is organized as the following. Section 2 introduces basic notation and then gives a detailed description of the procedures for estimating the spiked covariance matrix Σ, the rank r and the principal subspace span($\mathbf{V}$). The rates of convergence of these estimators are given in Section 3. Section 4 presents the minimax lower bounds that match the upper bounds in Section 3 in terms of the convergence rates, thereby establishing the minimax rates of convergence and rate-optimality of the estimators constructed in Section 2.

The minimax rates for estimating the eigenvalues and the precision matrix are given in Section 5. Section 6 discusses computational and other related issues. The proofs are given in Section 7.

2 Estimation Procedure

We give a detailed description of the estimation procedure in this section and study its properties in Section 3. Throughout, we shall focus on minimax estimation and assume the sparsity k is known, while the rank r will be selected based on data. Adaptation to k will be discussed in Section 6.

Notation

We first introduce some notation. For any matrix $\mathbf{X} = (x_{ij})$ and any vector $\mathbf{u}$, denote by $\|\mathbf{X}\|$ the spectral norm, $\|\mathbf{X}\|_F$ the Frobenius norm, and $\|\mathbf{u}\|$ the vector ℓ_2 norm. Moreover, the ith row of $\mathbf{X}$ is denoted by $\mathbf{X}_{i*}$ and the jth column by $\mathbf{X}_{*j}$. Let $\text{supp}(\mathbf{X}) = \{i : \mathbf{X}_{i*} \neq 0\}$ denote the row support of $\mathbf{X}$. For a positive integer p, $[p]$ denotes the index set $\{1, 2, \ldots, p\}$. For any set A, $|A|$ denotes its cardinality, and A^c its complement. For two subsets I and J of indices, denote by $\mathbf{X}_{IJ}$ the $|I| \times |J|$ submatrices formed by x_{ij} with $(i, j) \in I \times J$. Let $\mathbf{X}_{I*} = \mathbf{X}_{I[n]}$ and $\mathbf{X}_{*J} = \mathbf{X}_{[p]J}$. For any square matrix $\mathbf{A} = (a_{ij})$, we let $\text{Tr}(\mathbf{A}) = \Sigma_i a_{ii}$ be its trace. Define the inner product of any two matrices $\mathbf{B}$ and $\mathbf{C}$ of the same size by $\langle \mathbf{B}, \mathbf{C} \rangle = \text{Tr}(\mathbf{B}'\mathbf{C})$. For any matrix $\mathbf{A}$, we use $\sigma_i(\mathbf{A})$ to denote its ith largest singular value. When $\mathbf{A}$ is positive semi-definite, $\sigma_i(\mathbf{A})$ is also the ith largest eigenvalue of $\mathbf{A}$. For any real number a and b, set $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Let S^{p-1} denote the unit sphere in $\mathbb{R}^p$. For any event E, we write $\mathbf{1}_{\{E\}}$ as its indicator function.

For any set $B \subset [p]$, let B^c be its complement. For any symmetric matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$, we use $\mathbf{A}_B$ to denote the $p \times p$ matrix whose $B \times B$ block is $\mathbf{A}_{BB}$, the remaining diagonal elements are all ones and the remaining off-diagonal elements are all zeros, i.e.,

$$(\mathbf{A}_B)_{ij} = a_{ij}\mathbf{1}_{\{i \in B, j \in B\}} + \mathbf{1}_{\{i = j \in B^c\}}. \quad (10)$$

In other word, after proper reordering of rows and columns, we have

$$\mathbf{A}_B = \left[\begin{array}{cc} \mathbf{A}_{BB} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{B^c B^c} \end{array} \right].$$

Let $P \otimes Q$ denote the product measure of P and Q and $P^{\otimes n}$ the n-fold product of P. For random variables X and Y, we write $X \overset{(d)}{=} Y$ if they follow the same distribution, and $X \overset{\text{s.t.}}{\leq} Y$ if $P(X > t) \leq P(Y > t)$ for all $t \in \mathbb{R}$. Throughout the paper, we use C to denote a generic positive absolute constant, whose actual value may depend on the context. For any two sequences $\{a_n\}$ and $\{b_n\}$ of positive numbers, we write $a_n \lesssim b_n$ when $a_n \leq Cb_n$ for some numeric constant C, and $a_n \gtrsim b_n$ when $b_n \lesssim a_n$, and $a_n \asymp b_n$ when both $a_n \gtrsim b_n$ and $a_n \lesssim b_n$ hold.

Estimators

We are now ready to present the procedure for estimating the spiked covariance matrix Σ, the rank r, and the principal subspace span($\mathbf{V}$).

Let

$$A = supp(\mathbf{V}) \quad (11)$$

be the row support of $\mathbf{V}$. For any $m \in [p]$, let

$$\eta(m, n, p, \gamma_1) = 2\left(\sqrt{\frac{m}{n}} + \sqrt{\frac{\gamma_1}{n} m \, log \, \frac{ep}{m}}\right) + \left(\sqrt{\frac{m}{n}} + \sqrt{\frac{\gamma_1}{n} m \, log \, \frac{ep}{m}}\right)^2. \quad (12)$$

Recall that the observed matrix $\mathbf{X}$ has i.i.d. rows $\mathbf{X}_{i*} \sim N(\mathbf{0}, \Sigma)$. We define $\mathbf{S} = \frac{1}{n}\mathbf{X}'\mathbf{X}$ as the sample covariance matrix. Also recall that we assume knowledge of the sparsity level k which is an upper bound for the support size $|A|$. The first step in the estimation scheme is to select a subset $\hat{A}$ of k features based on the data. To this end, let

$$\mathrm{B}_k = \{B \subset [p] : |B| = k, \quad and \; for \; all \; D \subset B^c \; with \; |D| \leq k,$$
$$\|\mathbf{S}_D - \mathbf{I}\| \leq \eta(|D|, n, p, \gamma_1), \|\mathbf{S}_{DB}\| \leq 2\sqrt{\|\mathbf{S}_B\|\eta(k, n, p, \gamma_1)}\}. \quad (13)$$

The appropriate value of γ_1 will be specified later in the statement of the theorems. Intuitively speaking, the requirements in (13) aim to ensure that for any $B \in \mathrm{B}_k$, there is no evidence in data suggesting that B^c overlaps with the row support A of $\mathbf{V}$. If $\mathrm{B}_k \neq \emptyset$, denote by $\hat{A}$ an arbitrary element of B_k (or we can let $\hat{A} = \mathrm{argmax}_{B \in \mathrm{B}_k} \mathrm{Tr}(\mathbf{S}_{BB})$ for concreteness). As we will show later, B_k is non-empty with high probability; See Lemma 3 in Section 7.1. The set $\hat{A}$ represents the collection of selected features, which turns out to be instrumental in constructing optimal estimators for the three objects we are interested in: the covariance matrix, the rank of the spiked model, and the principle subspace. The estimator $\hat{A}$ of the support set A is obtained through searching over all subsets of size k. Such a global search, though computationally expensive, appears to be necessary in order for our procedure to optimally estimate Σ and $\mathbf{V}$ under the spectral norm. For example, estimating row-by-row is not guaranteed to always yield optimal results. Whether there exist computationally efficient procedures attaining the optimal rate is currently unknown. See Section 6 for more discussions.

Given $\hat{A}$, the estimators for the above three objects are defined as follows. Recalling the notation in (10), we define the covariance matrix estimator as

$$\hat{\Sigma} = \mathbf{S}_{\hat{A}} \mathbf{1}_{\{\mathrm{B}_k \neq \emptyset\}} + \mathbf{I}_p \mathbf{1}_{\{\mathrm{B}_k = \emptyset\}}, \quad (14)$$

The estimator for the rank is

$$\hat{r} = max\ \left\{ l : \sigma_l(\mathbf{S}_{\hat{A}\hat{A}}) \geq 1 + \gamma_2 \sqrt{\|\mathbf{S}_{\hat{A}}\|} \eta(k, n, p, \gamma_1) \right\}. \quad (15)$$

The appropriate value of γ_2 will be specified later in the statement of the theorems. Last but not least, the estimator for the principle subspace is span($\hat{\mathbf{V}}$), Where

$$\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_{\hat{r}}] \in O(p, \hat{r}) \quad (16)$$

with $\hat{\mathbf{v}}_l$ the lth eigenvector of $\hat{\Sigma}$. When $B_k = 0$, we set $\hat{r} = 0$ and $\hat{\mathbf{V}} = 0$ since $\hat{\Sigma} = \mathbf{I}_p$. Note that the estimator $\hat{\mathbf{V}}$ is based on the estimated rank r. Whenever $\hat{r} \neq r$, the value of the loss function (8) equals 1.

For a brief discussion on the comparison between the foregoing estimation procedure and that in [40], see Section 6.

3 Minimax Upper Bounds

We now investigate the properties of the estimation procedure introduced in Section 2. Rates of convergence for estimating the whole covariance matrix and its principal subspace under the spectral norm as well as for rank detection are established. The minimax lower bounds given in Section 4 will show that these rates are optimal and together they yield the minimax rates of convergence.

We begin with the estimation error of the covariance matrix estimator (14). Note that here we do not require the ratio λ_1/λ_r to be bounded.

Theorem 1

Let $\hat{\Sigma}$ be defined in (14) with B_k given by (13) for some $\gamma_1 \geq 10$. If $\frac{k}{n} log \frac{ep}{k} \leq c_0$ for a sufficiently small constant $c_0 > 0$, then

$$\sup_{\sum \in \Theta_1(k,p,r,\lambda)} \mathrm{E} \|\hat{\sum} - \sum\|^2 \lesssim \frac{(\lambda+1)k}{n} log \frac{ep}{k} + \frac{\lambda^2 r}{n}, \quad (17)$$

where the parameter space $\Theta_1(k, p, r, \lambda)$ is defined in (4).

As we shall show later in Theorem 4, the rates in (17) are optimal with respect to all the model parameters, namely k, p, r and λ.

Remark 1—Since the parameter space Θ_1 in (4) is contained in the set of k-sparse covariance matrices, it is of interest to compare the minimax rates for covariance matrix estimation in these two nested parameter spaces. For simplicity, only consider the case where the spectral norms of covariance matrices are uniformly bounded by a constant. Cai and Zhou [13] showed that, under certain regularity conditions, the minimax rate of convergence for estimating k-sparse matrices is $k^2 \frac{log\ p}{n}$, while the rate over Θ_1 in (7)

reduces to $\frac{k}{n} log \frac{ep}{k}$ when the spectral norm of the matrix and hence λ is bounded. Thus, ignoring the logarithmic terms, the rate over the smaller parameter space Θ_1 is faster by a factor of k. This faster rate can be achieved because the group k-sparsity considered in our parameter space imposes much more structure than the row-wise k-sparsity does for the general k-sparse matrices.

The next result concerns with the detection rate of the rank estimator (15) under the extra assumption that the ratio of the largest spike to the smallest, i.e., λ_1/λ_r, is bounded.

Theorem 2

Let $r \hat{=} r(\hat{\gamma}_1, \gamma_2)$ be defined in (15) for some constants $\gamma_1 \quad 10$ and $\gamma_2 \geq 8\sqrt{\gamma_1} + 34$. Assume that

$$\frac{\tau k}{n} log \frac{ep}{k} \leq c_0 \quad (18)$$

for a sufficiently small constant $c_0 \in (0, 1)$ which depends on γ_1. If $\lambda \geq \beta \sqrt{\frac{k}{n} log \frac{ep}{k}}$ for some sufficiently large β depending only on γ_1, γ_2 and τ, then

$$\sup_{\sum \in \Theta_2(k,p,\lambda,\tau)} P\{\hat{r} \neq rank(\boldsymbol{\Lambda})\} \leq Cp^{1-\gamma_1/2} \quad (19)$$

where the parameter space $\Theta_2(k, p, \lambda, \tau)$ is defined in (5).

By Theorem 5 to be introduced later, the detection rate of $\sqrt{\frac{k}{n} log \frac{ep}{k}}$ is optimal. For more details, see the discussion in Section 4.

Finally, we turn to the risk of the principal subspace estimator. As in Theorem 2, we require λ_1/λ_r to be bounded.

Theorem 3

Suppose

$$M_1 log p \geq log n \geq M_0 log \lambda \quad (20)$$

holds for some absolute constants $M_0, M_1 > 0$. Let $\hat{\mathbf{V}}$ be defined in (16) with constants $\gamma_1 \geq 10 \vee \left(1 + \frac{2}{M_0}\right) M_1$ and $\gamma_2 \geq 8\sqrt{\gamma_1} + 34$ in (13) and (15). If (18) holds for a sufficiently small constant c_0 depending on γ_1, then

$$\sup_{\sum \in \Theta_0(k,p,r,\lambda,\tau)} E\left\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\right\|^2 \lesssim \frac{k(\lambda+1)}{n\lambda^2} log \frac{ep}{k} \wedge 1, \quad (21)$$

where the parameter space $\Theta_0(k, p, r, \lambda, \tau)$ is defined in (3).

Remark 2—To ensure that the choice of γ_1 for achieving (21) is data-driven, we only need an under-estimate for $M_0 = \log n / \log \lambda$, or equivalently an overestimate for λ. (Note that $M_1 = \log n / \log p$ can be obtained directly given the data matrix.) To this end, we first estimate Σ by $\hat{\Sigma}$ in (14) with an initial $\gamma_1 = 10$ in (13). Then we control λ by $2\|\hat{\Sigma}\| - 1$. By the proof of Theorem 2, and in particular (75), this is an over-estimate of λ with high probability. The upper bound in (21) remains valid if we compute $\hat{\mathbf{V}}$ with a (possibly) new $\gamma_1 = 10 \wedge (1 + 2/\hat{M_0})M_1$ in (13), where $\hat{M_0} = \log n / \log(2\|\hat{\Sigma}\| - 1)$.

It is worth noting that the rate in (21) does not depend on r, and is optimal, by the lower bound given in Theorem 4 later.

The problems of estimating Σ and $\mathbf{V}$ are clearly related, but they are also distinct from each other. To discuss their relationship, we first note the following result (proved in Appendix 7.5) which is a variation of the well-known sin-theta theorem [15]:

Proposition 1

Let Σ and $\hat{\Sigma}$ be $p \times p$ symmetric matrices. Let $r \in [p - 1]$ be arbitrary and let $\mathbf{V}, \hat{\mathbf{V}} \in O(p, r)$ be formed by the r leading singular vectors of Σ and $\hat{\Sigma}$, respectively. Then

$$\|\hat{\textstyle\sum} - {\textstyle\sum}\| \geq \frac{1}{2}(\sigma_r\left({\textstyle\sum}\right) - \sigma_{r+1}({\textstyle\sum}))\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\|. \quad (22)$$

In view of Proposition 1, the minimax risks for estimating the spiked co-variance matrix Σ and the principle subspace $\mathbf{V}$ under the spectral norm can be tied as follows:

$$\underset{\hat{\sum}}{\inf}\,\underset{\sum \in \Theta}{\sup}\, \mathrm{E}\|\hat{\textstyle\sum} - {\textstyle\sum}\|^2 \gtrsim \lambda_r^2\, \underset{\hat{\mathbf{V}}}{\inf}\,\underset{\sum \in \Theta}{\sup}\, \mathrm{E}\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\|^2, \quad (23)$$

where $\Theta = \Theta_0(k, p, r, \lambda, \tau)$.

The results of Theorems 1 and 3 suggest, however, that the above inequality is not tight when λ is large. The optimal rate for estimating $\mathbf{V}$ is not equivalent to that for estimating Σ divided by λ^2 when $\frac{\lambda+1}{\lambda^2} \log \frac{ep}{k} \ll \frac{r}{k}$. Consequently, Theorem 3 cannot be directly deduced from Theorem 1 but requires a different analysis by introducing an intermediate matrix $\mathbf{S}_0$ defined later in (50). This is because the estimation of Σ needs to take into account the extra error in estimating the eigenvalues in addition to those in estimating $\mathbf{V}$. On the other hand, in proving Theorem 1 we need to contend with the difficulty that the loss function is unbounded.

4 Minimax Lower Bounds and Optimal Rates of Convergence

In this section we derive minimax lower bounds for estimating the spiked covariance matrix Σ and the principal subspace span($\mathbf{V}$) as well as for the rank detection. These lower bounds hold for all parameters and are non-asymptotic. The lower bounds together with the upper bounds given in Section 3 establish the optimal rates of convergence for the three problems.

The technical analysis heavily relies on a careful study of a rank-one testing problem and analyzing the moment generating function of a squared symmetric random walk stopped at a hypergeometrically distributed time. This lower bound technique is of independent interest and can be useful for other related matrix estimation and testing problems.

4.1 Lower bounds and minimax rates for matrix and subspace estimation

We first consider the lower bounds for estimating the spiked covariance matrix Σ and the principal subspace $\mathbf{V}$ under the spectral norm.

Theorem 4—*For any* $1 \le r \le k \le p$ *and* $n \in \mathbb{N}$,

$$\inf_{\tilde{\Sigma}} \sup_{\Sigma \in \Theta_1(k,p,r,\lambda)} \mathrm{E} \left\| \tilde{\Sigma} - \Sigma \right\|^2 \gtrsim \left(\frac{(\lambda+1)k}{n} \, log \, \frac{ep}{k} + \frac{\lambda^2 r}{n} \right) \wedge \lambda^2, \quad (24)$$

and

$$\inf_{\tilde{\mathbf{V}}} \sup_{\Sigma \in \Theta_0(k,p,r,\lambda,\tau)} \mathrm{E} \left\| \tilde{\mathbf{V}}\tilde{\mathbf{V}}' - \mathbf{V}\mathbf{V}' \right\|^2 \gtrsim \frac{(\lambda+1)(k-r)}{n\lambda^2} \, log \, \frac{e(p-r)}{k-r} \wedge 1, \quad (25)$$

where the parameter spaces $\Theta_1 (k, p, r, \lambda)$ *and* $\Theta_0 (k, p, r, \lambda, \tau)$ *are defined in* (4) *and* (3), *respectively.*

To better understand the lower bound (24), it is helpful to write it equivalently as

$$\left(\frac{(\lambda+1)k}{n} \, log \, \frac{ep}{k} \right) \wedge \lambda^2 + \frac{\lambda^2 r}{n},$$

which can be proved by showing that the minimax risk is lower bounded by each of these two terms. The first term does not depend on *r* and is the same as the lower bound in the rank-one case. The second term is the oracle risk when the true support of $\mathbf{V}$ is known. The key to the proof is the analysis of the rank-one case which will be discussed in more detail in Section 4.3. The proof of (25) is relatively straightforward by using known results on rank-one estimation.

In view of the upper bounds given in Theorems 1 and 3 and the lower bounds given in Theorem 4, we establish the following minimax rates of convergence for estimating the spiked covariance matrix Σ and the principal subspace $\mathbf{V}$, subject to the constraints on the parameters given in Theorems 1 and 3:

$$\inf_{\tilde{\Sigma}} \sup_{\Sigma \in \Theta_1(k,p,r,\lambda,\tau)} \mathrm{E} \left\| \tilde{\Sigma} - \Sigma \right\|^2 \asymp \left(\frac{(\lambda+1)k}{n} \, log \, \frac{ep}{k} + \frac{\lambda^2 r}{n} \right) \wedge \lambda^2, \quad (26)$$

$$\inf_{\tilde{\mathbf{V}}} \sup_{\Sigma \in \Theta_0(k,p,r,\lambda,\tau)} \mathrm{E} \left\| \tilde{\mathbf{V}}\tilde{\mathbf{V}}' - \mathbf{V}\mathbf{V}' \right\|^2 \asymp \frac{(\lambda+1)k}{n\lambda^2} \, log \, \frac{ep}{k} \wedge 1, \quad (27)$$

where (27) holds under the addition condition that $k - r \gtrsim k$. Therefore, the estimators of Σ and $\mathbf{V}$ given in Section 2 are rate optimal. In (26), the trivial upper bound of λ^2 can always be achieved by using the identity matrix as the estimator.

4.2 Lower bound and minimax rate for rank detection

We now turn to the lower bound and minimax rate for the rank detection problem.

Theorem 5—*Let $\beta_0 < \dfrac{1}{36}$ be a constant. For any $1 \leq r \leq k \leq p$ and $n \in \mathbb{N}$, if*

$$\lambda \leq 1 \wedge \sqrt{\frac{\beta_0 k}{n} \log \frac{ep}{k}}, \text{ then}$$

$$\inf_{\tilde{r}} \sup_{\sum \in \Theta_2(k, p, \lambda, \tau)} \mathrm{P}\left\{\tilde{r} \neq rank(\textstyle\sum)\right\} \geq w(\beta_0), \quad (28)$$

where the function $w : (0, \dfrac{1}{36}) \to (0, 1)$ satisfies $w(0+) = 1$.

The upper and lower bounds given in Theorems 2 and 5 show that the optimal detection

boundary for the rank r is $\sqrt{\dfrac{k}{n} \log \dfrac{ep}{k}}$. That is, the rank r can be estimated with an arbitrarily

small error probability when $\lambda_r \geq \beta \sqrt{\dfrac{k}{n} \log \dfrac{ep}{k}}$ for a sufficiently large constant β, whereas

this is impossible to achieve by any method if $\lambda_r \leq \sqrt{\dfrac{\beta_0 k}{n} \log \dfrac{ep}{k}}$ for some small positive constant β_0. Note that Theorem 5 applies to the full range of sparsity including the non-

sparse case $k = p$, which requires $\lambda_r \geq \sqrt{\dfrac{p}{n}}$. This observation turns out to be useful in proving the "parametric term" in the minimax lower bound for estimating Σ in Theorem 4.

The rank detection lower bound in Theorem 5 is in fact a direct consequence of the next proposition concerning testing the identity covariance matrix against rank-one alternatives,

$$H_0 : \textstyle\sum = \mathbf{I}_p, \quad versus \quad H_1 : \textstyle\sum = \mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}', \quad \mathbf{v} \in \mathrm{S}^{p-1} \cap B_0(k), \quad (29)$$

where $B_0(k), \triangleq \{\mathbf{x} \in \mathbb{R}^p : |\mathrm{supp}(\mathbf{x})| \leq k\}$. Note that Σ is in the parameter space Θ_2 under both the null and the alternative hypotheses. The rank-one testing problem (29) has been studied in [2], where there is a gap between the lower and upper bounds when $k \gtrsim \sqrt{p}$. The following result show that their lower bound is in fact sub-optimal in this case. We shall give below a dimension-free lower bound for the optimal probability of error and determine the optimal rate of separation. The proof is deferred to Section 7.2.2.

Proposition 2—*Let $\beta_0 < \frac{1}{36}$ be a constant. Let $\mathbf{X}$ be an $n \times p$ random matrix whose rows are independently drawn from $N(0, \boldsymbol{\Sigma})$. For any $k \in [p]$ and $n \in \mathbb{N}$, if $\lambda \leq 1 \wedge \sqrt{\frac{\beta_0 k}{n} \log \frac{ep}{k}}$, the minimax sum of Type-I and Type-II error probabilities for the testing problem* (29) *satisfies*

$$\varepsilon_n\,(k, p, \lambda) \triangleq \inf_{\phi : \mathbb{R}^{n \times p} \to \{0,1\}} \left(\mathrm{P}_{\mathbf{I}_p}\, \{\phi(\mathbf{X}) = 1\} + \sup_{\sum \in H_1} \mathrm{P}_{\sum}\, \{\phi(\mathbf{X}) = 0\} \right) \geq w\,(\beta_0)$$

where the function $w : (0, \frac{1}{36}) \to (0, 1)$ satisfies $w(0+) = 1$.

Proposition 2 shows that testing independence in the rank-one spiked co-variance model can be achieved reliably only if the *effective signal-to-noise-ratio*

$$\beta_0 = \frac{\lambda^2}{\frac{k}{n} \log \frac{ep}{k}} \to \infty.$$

Furthermore, the lower bound in Proposition 2 also captures the following phenomenon: if β_0 vanishes, then the optimal probability of error converges to one. In fact, the lower bound in Proposition 2 is optimal in the sense that the following test succeeds with vanishing probability of error if $\beta_0 \to \infty$:

$$reject\ H_0\ if\ and\ only\ if \max_{\substack{J \subset [p] \\ |J| = k}} \|\mathbf{S}_{JJ} - \mathbf{I}_{JJ}\| \geq c \sqrt{\frac{k}{n} \log \frac{ep}{k}}$$

for some c only depends on β_0. See, e.g., [2, Section 4]. However, the above test has high computational complexity since one needs to enumerate all $k \times k$ submatrices of $\mathbf{S}$. It remains an open problem to construct tests that are both computationally feasible and minimax rate-optimal.

4.3 Testing rank-one spiked model

As mentioned earlier, a careful study of the rank-one testing problem (29) provides a major tool for the lower bound arguments. A key step in this study is the analysis of the moment generating function of a squared symmetric random walk stopped at a hypergeometrically distributed time. We present the main ideas in this section as the techniques can also be useful for other related matrix estimation and testing problems.

It is well-known that the minimax risk is given by the least-favorable Bayesian risk under mild regularity conditions on the model [27]. For the composite testing problem (29), it turns out that the rate-optimal least-favorable prior for $\mathbf{v}$ is given by the distribution of the following random vector:

$$\mathbf{u} = \frac{1}{\sqrt{k}} \mathbf{J}_I \, \mathbf{w}, \quad (30)$$

where $\mathbf{w} = (w_1, \ldots, w_p)$ consists of iid Rademacher entries, and $\mathbf{J}_I$ is a diagonal matrix given by $(\mathbf{J}_I)_{ii} = \mathbf{1}_{\{i \in I\}}$ with I uniformly chosen from all subsets of $[p]$ of size k. In other words, $\mathbf{u}$ is uniformly distributed on the collection of k-sparse vectors of unit length with equal-magnitude non-zeros. Hence $\mathbf{u} \in S^{p-1} \cap B_0(k)$. We set

$$\lambda^2 = \frac{\beta_0 k}{n} \log \frac{ep}{k},$$

where $\beta_0 > 0$ is a sufficiently small absolute constant.[2] The desired lower bound then follows if we establish that the following (Bayesian) hypotheses

$$H_0 : \sum = \mathbf{I}_p \quad \text{v.s.} \quad H_1 : \sum = \mathbf{I}_p + \lambda \mathbf{u}\mathbf{u}' \quad (31)$$

cannot be separated with vanishing probability of error.

Remark 3—The composite testing problem (31) has also been considered in [2]. In particular, the following suboptimal lower bound is given in [2, Theorem 5.1]: If

$$\lambda \leq 1 \wedge \sqrt{\frac{\upsilon k}{n} \log \left(1 + \frac{p}{k^2}\right)} \quad (32)$$

then the optimal error probability satisfies $\varepsilon_n(k, p, \lambda) \quad C(\upsilon)$, where $C(\upsilon) \to 1$ as $\upsilon \to 0$. This result is established based on the following prior:

$$\mathbf{u} = \frac{1}{\sqrt{k}} \mathbf{J}_I \, \mathbf{1}, \quad (33)$$

which is a binary sparse vector with uniformly chosen support.

Compared to the result in Proposition 2, (32) is rate-optimal in the *very sparse* regime of $k = o(\sqrt{p})$. However, since $\log(1 + x) \asymp x$ when $x \lesssim 1$, in the *moderately sparse* regime of $k \gtrsim \sqrt{p}$, $\log(1 + \frac{p}{k^2}) \asymp \frac{p}{k^2}$ and so the lower bound in (32) is substantially smaller than the optimal rate in Proposition 2 by a factor of $\sqrt{\frac{k^2}{p} \log \frac{ep}{k}}$, which is a *polynomial* factor in k when $k \gtrsim p^a$ for any $a > 1/2$. In fact, by strengthening the proof in [2], one can show that the optimal separation for discriminating (31) using the binary prior (33) is

$1 \wedge \sqrt{\frac{k}{n} \log \frac{ep}{k}} \wedge \frac{p}{k\sqrt{n}}$. Therefore the prior (33) is rate-optimal *only* in the regime of

---

[2]Here β_0 can be chosen to be any constant smaller than $\frac{1}{36}$. See Proposition 2. The number $\frac{1}{36}$ is certainly not optimized.

$k=o(p^{\frac{2}{3}})$, while (30) is rate-optimal for all k. Examining the role of the prior (30) in the proof of Theorem 5, we see that it is necessary to randomize the signs of the singular vector in order to take advantage of the central limit theorem and Gaussian approximation. When

$k \gtrsim p^{\frac{2}{3}}$, the fact that the singular vector **u** is positive componentwise reduces the difficulty of the testing problem.

The main technical tool for establishing the rank-detection lower bound in Proposition 2 is the following lemma, which can be of independent interest. It deals with the behavior of a symmetric random walk stopped after a hypergeometrically distributed number of steps. Moreover, note that Lemma 1 also incorporates the non-sparse case ($k = p$ and $H = k$), which proves to be useful in establishing the minimax lower bound for estimating Σ in Theorem 4. The proof of Lemma 1 is deferred to Section 7.6.

Lemma 1: *Let $p \in \mathbb{N}$ and $k \in [p]$. Let $B_1, \dots, B_k$ be independently Rademacher distributed. Denote the symmetric random walk on $\mathbb{Z}$ stopped at the mth step by*

$$G_m \triangleq \sum_{i=1}^{m} B_i. \quad (34)$$

Let $H \sim$ Hypergeometric(p, k, k) with
$$\mathrm{P}\{H=i\}=\frac{\dbinom{k}{i}\dbinom{p-k}{k-i}}{\dbinom{p}{k}}$$
, $i = 0, \dots, k$. Then there exists a function $g : (0, \frac{1}{36}) \to (1, \infty)$ with $g(0+) = 1$, such that for any $a<\frac{1}{36}$,

$$\mathrm{E}\, exp\,(tG_H^2) \leq g(a), \quad (35)$$

where $t=\dfrac{a}{k}\,log\,\dfrac{ep}{k}$.

Remark 4 (Tightness of Lemma 1)—The purpose of Lemma 1 is to seek the largest t, as a function of (p, k), such that $\mathrm{E} \exp (tG_H^2)$ is upper bounded by a constant non-asymptotically. The condition that $t \asymp \dfrac{1}{k}\,log\,\dfrac{ep}{k}$ is in fact both necessary and sufficient. To see the necessity, note that $\mathrm{P}\{G_H = H | H = i\} = 2^{-i}$. Therefore

$$\mathrm{E}\, exp\,(tG_H^2) \geq \mathrm{E}\, exp\,(tH^2)2^{-H} \geq exp\,(tk^2)2^{-k}\, \mathrm{P}\{H=k\} \geq exp\left(tk^2 - k\,log\,\frac{2p}{k}\right),$$

which cannot be upper bounded by an absolutely constant unless $t \lesssim \dfrac{1}{k}\,log\,\dfrac{ep}{k}$.

5 Estimation of Precision Matrix and Eigenvalues

We have so far focused on the optimal rates for estimating the spiked covariance matrix Σ, the rank r and the principal subspace span($\mathbf{V}$). The technical results and tools developed in the earlier sections turn out to be readily useful for establishing the optimal rates of convergence for estimating the precision matrix $\Omega = \Sigma^{-1}$ as well as the eigenvalues of Σ under the spiked covariance matrix model (1).

Besides the covariance matrix Σ, it is often of significant interest to estimate the precision matrix Ω, which is closely connected to the Gaussian graphical model as the support of the precision matrix Ω coincides with the edges in the corresponding Gaussian graph. Let $\hat{\Sigma}$ be defined in (14) and let $\sigma_i(\hat{\Sigma})$ denote its ith largest eigenvalue value for all $i \in [p]$. Define the precision matrix estimator as

$$\hat{\Omega} = \begin{cases} (\hat{\Sigma})^{-1}, & \sigma_p(\hat{\Sigma}) \geq \frac{1}{2}, \\ \mathbf{I}_p, & \sigma_p(\hat{\Sigma}) < \frac{1}{2}. \end{cases} \quad (36)$$

The following result gives the optimal rates for estimating Ω under the spectral norm.

Proposition 3 (Precision matrix estimation)

Assume that $\lambda \asymp 1$. If $\dfrac{k}{n} \log \dfrac{ep}{k} \leq c_0$ for a sufficiently small constant $c_0 > 0$, then

$$\inf_{\hat{\Omega}} \sup_{\sum \in \Theta_1 (k,p,r,\lambda)} \mathrm{E} \left\| \hat{\Omega} - \Omega \right\|^2 \asymp \frac{k}{n} \log \frac{ep}{k}, \quad (37)$$

where the upper bound is attained by the estimator (36) *with γ_1 10 in obtaining $\hat{\Sigma}$.*

The upper bound follows from the lines in the proof of Theorem 1 after we control the smallest eigenvalue of $\hat{\Sigma}$ as in (36). Proposition 2 can be readily applied to yield the desired lower bound.

Note that the optimal rate in (37) is quite different from the minimax rate of convergence

$M^2 k^2 \dfrac{\log p}{n}$ for estimating k-sparse precision matrices where each row/column has at most k nonzero entries. Here M is the ℓ_1 norm bound for the precision matrices. See [9]. So the sparsity in the principal eigenvectors and the sparsity in the precision matrix itself have significantly different implications for estimation of Ω under the spectral norm.

We now turn to estimation of the eigenvalues. Since σ is assumed to be equal to one, it suffices to estimate the eigenvalue matrix $\mathbf{E} = \mathrm{diag}(\lambda_i)$ where $\lambda_i \triangleq 0$ for $i > r$. For any estimator $\tilde{\mathbf{E}} = \mathrm{diag}(\tilde{\lambda_i})$, we quantify the estimator error by the loss function $\|\tilde{\mathbf{E}} - \mathbf{E}\|^2 = \max_{i \in [p]} |\tilde{\lambda_i} - \lambda_i|^2$. The following result gives the optimal rate of convergence for this estimation problem.

Proposition 4 (Uniform estimation of spectra)

Under the same conditions of Proposition 3,

$$\inf_{\hat{\mathbf{E}}} \sup_{\Sigma \in \Theta_1 (k,p,r,\lambda)} \mathrm{E} \left\| \hat{\mathbf{E}} - \mathbf{E} \right\|^2 \asymp \frac{k}{n} log \frac{ep}{k}, \quad (38)$$

where the upper bound is attained by the estimator $\hat{\mathbf{E}} = diag(\sigma_i(\hat{\Sigma})) - \mathbf{I}_p$ *with* $\hat{\Sigma}$ *defined in* (14) *with* γ_1 10.

Hence, the spikes and the eigenvalues can be estimated uniformly at the rate of $\dfrac{k}{n} log \dfrac{ep}{k}$ when $k \ll n/ log\ p$. Proposition 4 is a direct consequence of Theorem 1 and Proposition 2. The proofs of Propositions 3 and 4 are deferred to Section 7.4.

6 Discussions

We have assumed the knowledge of the noise level $\sigma = 1$ and the support size k. For a given value of σ^2, one can always rescale the data and reduce to the case of unit variance. As a consequence of rescaling, the results in this paper remain valid for a general σ^2 by replacing each λ with λ/σ^2 in both the expressions of the rates and the definitions of the parameter spaces. When σ^2 is unknown, it can be easily estimated based on the data. Under the sparsity models (3)-(5), when $k < p/2$, σ^2 can be well estimated by $\hat{\sigma^2} = \mathrm{median}(s_{jj})$ as suggested in [20], where s_{jj} is jth diagonal element of the sample covariance matrix.

The knowledge of the support size k is much more crucial for our procedure. An interesting topic for future research is the construction of adaptive estimators which could achieve the minimax rates in Theorems 1–3 without knowing k. One possibility is to define B_k in (13) for all $k \in [p]$, find the smallest k such that B_k is non-empty, and then define estimators for that particular k similar to those in Section 2 with necessary adjustments accounting for the extra multiplicity in the support selection procedure.

For ease of exposition, we have assumed that the data are normally distributed and so various non-asymptotic tail bounds used in the proof follow. Since these bounds typically only rely on sub-Gaussianity assumptions on the data, we expect the results in the present paper readily extendable to data generated from distributions with appropriate sub-Gaussian tail conditions. The estimation procedure in Section 2 is different from that in [40]. Although both are based on enumerating all possible support sets of size k, Vu and Lei [40] proposed to pick the support set which maximizes a quadratic form, while ours is based on picking the set satisfying certain deviation bounds.

A more important issue is the computational complexity required to obtain the minimax rate optimal estimators. The procedure described in Section 2 entails a global search for the support set A, which can be computationally intensive. In many cases, this seems unavoidable since the spectral norm is not separable in terms of the entries/rows/columns. However, in some other cases, there are estimators that are computationally more efficient and can attain the same rates of convergence. For instance, in the low rank cases where

$r \lesssim log \frac{ep}{k}$, the minimax rates for estimating span(**V**) under the spectral norm and under the Frobenius norm coincide with each other. See the discussion following Theorem 3 in Section 3. Therefore the procedure introduced in [10, Section 3] attains the optimal rates under both norms simultaneously. As shown in [10], this procedure is not only computationally efficient, but also adaptive to the sparsity k. Finding the general complexity-theoretic limits for attaining the minimax rates under the spectral norm is an interesting and challenging topic for future research. Following the initial post of the current paper, Berthet and Rigollet [1] showed in a closely related sparse principal component detection problem that the minimax detection rate cannot be achieved by any computationally efficient algorithm in the highly sparse regime.

7 Proofs

We first collect a few useful technical lemmas in Section 7.1 before proving the main theorems in Section 7.2 in the order of Theorems 1 - 4. We then give the proofs of the propositions in the order of Propositions 2, 3, 4, and 1. As mentioned in Section 4, Theorem 5 on the rank detection lower bound is a direct consequence of Proposition 2. We complete this section with the proof of Lemma 1.

Recall that the row vector of **X** are i.i.d. samples from the $N(\mathbf{0}, \boldsymbol{\Sigma})$ distribution with $\boldsymbol{\Sigma}$ specified by (1). Equivalently, one can think of **X** as an $n \times p$ data matrix generated as

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}' + \boldsymbol{Z}, \quad (39)$$

where **U** is an $n \times r$ random effects matrix with iid $N(0, 1)$ entries, $\mathbf{D} = diag(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$, **V** is $p \times r$ orthonormal, and **Z** has iid $N(0, 1)$ entries which are independent of **U**.

7.1 Technical Lemmas

Lemma 2—*Let* $\mathbf{S} = \frac{1}{n}\mathbf{X}'\mathbf{X}$ *be the sample covariance matrix, then*

$$\mathrm{P}\left(\max_{|B|=k}\|\mathbf{S}_{BB}\| \leq (\lambda_1+1)\left(1+\sqrt{\frac{k}{n}}+\frac{t}{\sqrt{n}}\right)^2\right) \leq \binom{p}{k} \mathrm{e}^{-t^2/2}.$$

Proof: Note that

$$\max_{|B|=k}\|\mathbf{S}_{BB}\| \leq \max_{|B|=k}\|\textstyle\sum_{BB}\| \cdot \|\textstyle\sum_{BB}^{-1/2}\mathbf{S}_{BB}\textstyle\sum_{BB}^{-1/2}\|$$

$$\leq (\lambda_1+1)\max_{|B|=k}\|\textstyle\sum_{BB}^{-1/2}\mathbf{S}_{BB}\textstyle\sum_{BB}^{-1/2}\| \overset{\text{(d)}}{=} (\lambda_1+1)\max_{|B|=k}\|\boldsymbol{Z}_{*B}\|^2.$$

The result follows from the Davidson–Szarek bound [14, Theorem II.7] and the union bound.

Lemma 3—*Let* B_k *be defined as in* (13) *with* $\gamma_1 \geq 3$. *Then* $P(A \notin B_k) \leq 5(ep)^{1-\gamma_1/2}$.

Proof: Note that by union bound

$$P(A \notin B_k) \leq \quad P(\exists D \subset A^c, |D| \leq k, \|\mathbf{S}_D - \mathbf{I}\| > \eta(|D|, n, p, \gamma_1))$$
$$+ P(\exists D \subset A^c, |D| \leq k, \|\mathbf{S}_{DA}\| > 2\sqrt{\|\mathbf{S}_A\|}\eta(k, n, p, \gamma_1)).$$

We now bound the two terms on the right-hand side separately.

For the first term, note that $A = \mathrm{supp}(\mathbf{V})$. Then for any $D \subset A^c$, $\mathbf{S}_D = \frac{1}{n}\mathbf{Z}'_{*D}\mathbf{Z}_{*D}$. Hence

$$P(\exists D \subset A^c, |D| \leq k, \|\mathbf{S}_D - \mathbf{I}\| > \eta(|D|, n, p, \gamma_1))$$
$$\leq \sum_{D \subset A^c, |D| \leq k} P(\|\mathbf{S}_D - \mathbf{I}\| > \eta(|D|, n, p, \gamma_1))$$
$$\leq \sum_{l=1}^{k} \binom{p-k}{l} 2\, exp\left(-\tfrac{\gamma_1}{2}l\, log\, \tfrac{ep}{l}\right)$$
$$\leq 2\sum_{l=1}^{k} \left(\tfrac{ep}{l}\right)^{l(1-\gamma_1/2)} \leq 4(ep)^{1-\gamma_1/2},$$

where the second inequality is [10, Proposition 4], and the last inequality holds for all $\gamma_1 \geq 3$ and $p \geq 2$.

For the second term, note that for any fixed $D \subset A^c$, $\mathbf{S}_{DA} = \frac{1}{n}\mathbf{Z}'_{*D}\mathbf{X}_{*A}$, where $\mathbf{Z}_{*D}$ and $\mathbf{X}_{*A}$ are independent. Thus, let $\mathbf{W}$ be the left singular vector matrix of $\mathbf{X}_{*A}$, we obtain that[3]

$$\|\mathbf{S}_{DA}\| \leq \frac{1}{n}\|\mathbf{Z}'_{*D}\mathbf{W}\|\|\mathbf{X}_{*A}\| \overset{(d)}{=} \frac{1}{\sqrt{n}}\|\mathbf{Y}\|\sqrt{\|\mathbf{S}_A\|},$$

where $\mathbf{Y}$ is a $|D| \times k$ matrix with i.i.d. $N(0, 1)$ entries. Thus, we have

$$P(\exists D \subset A^c, |D| \leq k, \|\mathbf{S}_{DA}\| > 2\sqrt{\|\mathbf{S}_A\|}\eta(k, n, p, \gamma_1))$$
$$\leq \sum_{D \subset A^c, |D| \leq k} P(\|\mathbf{S}_{DA}\| > 2\sqrt{\|\mathbf{S}_A\|}\eta(k, n, p, \gamma_1))$$
$$\leq \sum_{D \subset A^c, |D| \leq k} P(\|\mathbf{Y}\| > 2\sqrt{n}\eta(k, n, p, \gamma_1))$$
$$\leq \sum_{l=1}^{k} \binom{p-k}{l} exp\left(-2\gamma_1 k\, log\, \tfrac{ep}{k}\right)$$
$$\leq k\left(e\tfrac{p}{k}\right)^{k(1-2\gamma_1)} \leq (ep)^{1-\gamma_1/2},$$

where the third inequality is due to the Davidson-Szarek bound. Combining the two bounds completes the proof.

---

[3]If rank$(X_{*A}) < k$, then $\overset{(d)}{=}$ is changed to $\overset{s.t.}{\leq}$, and the subsequent arguments continue to hold verbatim.

Lemma 4—*Let $\gamma_1 \geq 3$. Suppose that $\frac{k}{n} \log \frac{ep}{k} \leq C_0$ for a sufficiently small constant $c_0 > 0$. Then with probability at least $1 - 12(ep)^{1-\gamma_1/2}$, $B_k \neq 0$, and*

$$\|\mathbf{S}_{\hat{A}} - \mathbf{S}_A\| \leq C_0(\gamma_1) \sqrt{\lambda_1 + 1} \eta(k, n, p, \gamma_1), \quad (40)$$

where $C_0(\gamma_1) = 14 + 4\sqrt{\gamma_1}$.

Proof: We focus on the event $B_k \neq 0$. Define the sets

$$G = A \cap \hat{A}, \quad M = A \cap \hat{A}^c, \quad O = A^c \cap \hat{A}, \quad (41)$$

which correspond to the sets of correctly identified, missing, and overly identified features by the selected support set $\hat{A}$, respectively. By the triangle inequality, we have

$$\|\mathbf{S}_{\hat{A}} - \mathbf{S}_A\| \leq \|\mathbf{S}_{MM} - \mathbf{I}_{MM}\| + \|\mathbf{S}_{OO} - \mathbf{I}_{OO}\| + 2\|\mathbf{S}_{GM}\| + 2\|\mathbf{S}_{GO}\|. \quad (42)$$

We now provide high probability bounds for each term on the right-hand side.

For the first term, recall that $\hat{A} \in B_k$ which is defined in (13). Since $M \subset \hat{A}^c$, we have

$$\|\mathbf{S}_{MM} - \mathbf{I}_{MM}\| \leq \eta(|M|, n, p, \gamma_1) \leq \eta(k, n, p, \gamma_1). \quad (43)$$

For the second term, by similar calculation to that in the proof of Lemma 3, we have that when $\gamma_1 \geq 3$,

$$\|\mathbf{S}_{OO} - \mathbf{I}_{OO}\| \leq \eta(k, n, p, \gamma_1) \quad (44)$$

with probability at least $1 - 4(ep)^{1-\gamma_1/2}$. For the third term, we turn to the definition of B_k in (13) again to obtain

$$\|\mathbf{S}_{GM}\| \leq \|\mathbf{S}_{\hat{A}M}\| \leq 2\sqrt{\|\mathbf{S}_{\hat{A}\hat{A}}\|} \eta(k, n, p, \gamma_1).$$

By Lemma 2, with probability at least $1 - (ep)^{1-\gamma_1/2}$,

$$\begin{aligned}
\|\mathbf{S}_{\hat{A}\hat{A}}\|, \|\mathbf{S}_{AA}\| &\leq \max_{|B|=k} \|S_{BB}\| \leq (\lambda_1+1)\left(1 + \sqrt{\frac{k}{n}} + \sqrt{\frac{\gamma_1 k}{n} \log \frac{ep}{k}}\right)^2 \\
&\leq (1 + (\sqrt{\gamma_1}+1)\sqrt{c_0})^2(\lambda_1+1) \\
&\leq \left(\frac{1}{2}\sqrt{\gamma_1} + \frac{3}{2}\right)^2(\lambda_1+1),
\end{aligned} \quad (45)$$

where the second inequality holds when $\frac{k}{n} \log \frac{ep}{k} \leq c_0$ and the last inequality holds for sufficiently small constant c_0. Moreover, the last two displays jointly imply

$$\|\mathbf{S}_{GM}\| \leq (\sqrt{\gamma_1} + 3)\sqrt{\lambda_1 + 1} \eta(k, n, p, \gamma_1). \quad (46)$$

For the fourth term, we obtain by similar arguments that with probability at least $1 - (ep)^{1 - \gamma_1/2}$,

$$\|\mathbf{S}_{GO}\| \le \|\mathbf{S}_{AO}\| \le (\sqrt{\gamma_1}+3)\sqrt{\lambda_1+1}\eta(k,n,p,\gamma_1). \quad (47)$$

Note that $P(B_k = 0)' \quad P(A \notin B_k)$, and by Lemma 3, the latter is bounded above by $5(ep)^{1 - \gamma_1/2}$. So the union bound implies that the intersection of the event $\{B_k \quad 0\}'$ and the event that (43)-(47) all hold has probability at least $1 - 12(ep)^{1 - \gamma_1/2}$. On this event, we assemble (42)-(47) to obtain

$$
\begin{aligned}
\|\mathbf{S}_{\hat{A}} - \mathbf{S}_A\| &\le [2+4(\sqrt{\gamma_1}+3)\sqrt{\lambda_1+1}]\eta(k,n,p,\gamma_1) \\
&\le [2+4(\sqrt{\gamma_1}+3)]\sqrt{\lambda_1+1}\,\eta(k,n,p,\gamma_1).
\end{aligned}
$$

This completes the proof.

Lemma 5—*Let $\mathbf{Y} \in \mathbb{R}^{n \times k}$ and $\mathbf{Z} \in \mathbb{R}^{n \times m}$ be two independent matrices with i.i.d. $N(0,1)$ entries. Then there exists an absolute constant $C > 0$, such that*

$$\mathrm{E}\left\|\frac{1}{n}\mathbf{Y}'\mathbf{Y} - \mathbf{I}_k\right\|^2 \le C\left(\frac{k}{n}+\frac{k^2}{n^2}\right), \quad and \quad (48)$$

$$\mathrm{E}\|\mathbf{Y}'\mathbf{Z}\|^2 \le C(n(k+m)+km). \quad (49)$$

Proof: The inequality (48) follows directly from integrating the high-probability upper bound in [10, Proposition 4].

Let the SVD of $\mathbf{Y}$ be $\mathbf{Y} = \mathbf{ACB}'$, where $\mathbf{A},\mathbf{B}$ are $n \times (n \wedge k)$ and $k \times (n \wedge k)$ uniformly Haar distributed, and $\mathbf{C}$ is an $(n \wedge k) \times (n \wedge k)$ diagonal matrix with $\|\mathbf{C}\| = \|\mathbf{Y}\|$. Since $\mathbf{A}$ and $\mathbf{Z}$ are independent, $\mathbf{A}'\mathbf{Z}$ has the same law as a $(n \wedge k) \times m$ iid Gaussian matrix. Therefore

$$\mathrm{E}\|\mathbf{Y}'\mathbf{Z}\|^2 \le \mathrm{E}\|\mathbf{Y}\|^2\mathrm{E}\|\mathbf{A}'\mathbf{Z}\|^2 \le C_0(n+k)(n \wedge k+m).$$

for some absolute constant C_0, where the last inequality follows from the Davidson-Szarek theorem [14, Theorem II.7]. Exchanging the role of $\mathbf{Y}$ and $\mathbf{Z}$, we have $\mathrm{E}\|\mathbf{Y}'\mathbf{Z}\|^2 \quad C_0(n+m)(n \wedge m + k)$. Consequently,

$$\mathrm{E}\|\mathbf{Y}'\mathbf{Z}\|^2 \le C_0((n+k)(n\wedge k+m)\wedge(n+m)(n\wedge m+k)) \le 2C_0(n(k+m)+km).$$

This completes the proof.

In the proofs, the following *intermediate matrix*

$$\mathbf{S}_0 = \frac{1}{n}\mathbf{VDU}'\mathbf{UDV}' + \mathbf{I}_p \quad (50)$$

plays a key role. In particular, the following results on $\mathbf{S}_0$ will be used repeatedly.

Lemma 6—*Suppose $\lambda_1/\lambda_r \leq \tau$ for some constant $\tau \geq 1$. If $\|\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\| < 1/\tau$, then*

$$\sigma_l(\mathbf{S}_0) > 1, \quad \forall l \in [r], \quad and \quad \sigma_l(\mathbf{S}_0) = 1, \quad \forall l > r. \quad (51)$$

Moreover, $\mathbf{VV}'$ is the projection matrix onto the rank r principal subspace of $\mathbf{S}_0$.

Proof: It is straightforward to verify that

$$\mathbf{S}_0 - \sum = \mathbf{VD}\left(\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\right)\mathbf{DV}'. \quad (52)$$

When $\|\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\| < 1, \|\mathbf{S}_0 - \sum\| \leq \|\mathbf{D}\|^2\|\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\| < \lambda_1/\tau \leq \lambda_r$. So Weyl's inequality [18, Theorem 4.3.1] leads to

$$\sigma_r(\mathbf{S}_0) \geq \sigma_r\left(\sum\right) - \|\mathbf{S}_0 - \sum\| > \lambda_r + 1 - \lambda_r = 1.$$

Note that $\mathbf{S}_0$ always has 1 as its eigenvalue with multiplicity at least $p - r$. We thus obtain (51).

When (51) holds, (50) shows that the rank r principal subspace of $\mathbf{S}_0$ is equal to that of

$\frac{1}{n}\mathbf{VDU}'\mathbf{UDV}'$. Therefore, the subspace is spanned by the column vectors of $\mathbf{V}$, and $\mathbf{VV}'$ is the projection matrix onto it since $\mathbf{V} \in O(p, r)$.

To prove the lower bound for rank detection, we need the following lemma concerning the the χ^2-divergences in covariance models. Recall that the χ^2-divergence between two probability measures is defined as

$$\chi^2(P\|Q) \triangleq \int \left(\frac{\mathrm{d}P}{\mathrm{d}Q} - 1\right)^2 \mathrm{d}Q.$$

For a distribution F, we use $F^{\otimes n}$ to denote the product distribution of n independent copies of F.

Lemma 7—*Let ν be a probability distribution on the space of $p \times p$ symmetric random matrix $\mathbf{M}$ such that $\|\mathbf{M}\| \leq 1$ almost surely. Consider the scale mixture distribution $\mathrm{E}[N(0, \mathbf{I}_p + \mathbf{M})^{\otimes n}] = \int N(0, \mathbf{I}_p + \mathbf{M})^{\otimes n} \nu(\mathrm{d}\mathbf{M})$. Then*

$$\chi^2 \left(\mathrm{E}\left[N(0, \mathbf{I}_p + \mathbf{M})^{\otimes n} \right] \| N(0, \mathbf{I}_p)^{\otimes n} \right) + 1 = \mathrm{E} \, det \left(\mathbf{I}_p - \mathbf{M}_1 \mathbf{M}_2 \right)^{-\frac{n}{2}} \quad (53)$$

$$\geq \mathrm{E} \, exp \left(\frac{n}{2} \langle \mathbf{M}_1, \mathbf{M}_2 \rangle \right), \quad (54)$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are independently drawn from v. Moreover, if rank(**M**) = 1 *a.s., then* (54) *holds with equality.*

Proof: Denote by g_i the probability density function of $N(0, \Sigma_i)$ for $i = 0, 1$ and 2, respectively. Then it is straightforward to verify that

$$\int \frac{g_1 g_2}{g_0} = det \left(\sum\nolimits_0 \right)^{-1} \left[det \left(\sum\nolimits_0 (\sum\nolimits_1 + \sum\nolimits_2) - \sum\nolimits_1 \sum\nolimits_2 \right) \right]^{-\frac{1}{2}} \quad (55)$$

if $\Sigma_0(\Sigma_1 + \Sigma_2)$ $\Sigma_1\Sigma_2$; otherwise, the integral on the left-hand side of (55) is infinite. Applying (55) to $\Sigma_0 = \mathbf{I}_p$ and $\Sigma_i = \mathbf{I}_p + \mathbf{M}_i$ and using Fubini's theorem, we have

$$\chi^2(\mathrm{E}\left[N(0, \mathbf{I}_p + \mathbf{M})^{\otimes n} \right] \| N(0, \mathbf{I}_p)^{\otimes n}) + 1$$
$$= \mathrm{E} \, det \left(\mathbf{I}_p - \mathbf{M}_1 \mathbf{M}_2 \right)^{-\frac{n}{2}} \quad (56)$$

$$= \mathrm{E} \, exp \left(-\frac{n}{2} \, log \, det \left(\mathbf{I}_p - \mathbf{M}_1 \mathbf{M}_2 \right) \right) \quad (57)$$

$$\geq \mathrm{E} \, exp \left(\frac{n}{2} Tr \left(\mathbf{M}_1 \mathbf{M}_2 \right) \right), \quad (58)$$

where (56) is due to $\|\mathbf{M}_1 \mathbf{M}_2\|$ $\|\mathbf{M}_1\| \|\mathbf{M}_2\|$ 1 a.s., and (58) is due to log det($\mathbf{I} + \mathbf{A}$) Tr(**A**), with equality if and only if rank(**A**) = 1.

7.2 Proofs of Main Results

7.2.1 Proofs of the Upper Bounds

Proof of Theorem 1: Let γ_1 10 be a constant. Denote by E the event that (40) holds, which, in particular, contains the event $\{ B_k$ 0$\}$. By triangle inequality,

$$\begin{aligned} \mathrm{E} \| \mathbf{S}_{\hat{A}} - \sum \|^2 \mathbf{1}_E &\leq 2\mathrm{E} \| \mathbf{S}_{\hat{A}} - \mathbf{S}_A \|^2 \mathbf{1}_E + 2\mathrm{E} \| \mathbf{S}_A - \sum \|^2 \mathbf{1}_E \\ &\lesssim 2(\lambda_1 + 1)\eta^2(k, n, p, \gamma_1) + 2\mathrm{E} \| \mathbf{S}_A - \sum \|^2 \quad (59) \\ &\asymp (\lambda_1 + 1)\frac{k}{n} \, log \, \frac{ep}{k} + \mathrm{E} \| \mathbf{S}_A - \sum \|^2, \end{aligned}$$

where the last step holds because $\frac{k}{n} \, log \, \frac{ep}{k} \leq c_0$. In view of the definition of η in (12), we have $\eta^2(k, n, p, \gamma_1) \asymp \frac{k}{n} \, log \, \frac{ep}{k}$.

To bound the second term in (59), define $\mathbf{J}_B$ as the diagonal matrix given by

$$(\mathbf{J}_B)_{ii}=\mathbf{1}_{\{i\in B\}}. \quad (60)$$

Then, for $\mathbf{S}_0$ in (50),

$$
\begin{aligned}
\mathbf{S}_A &= \mathbf{J}_A\left(\mathbf{S}-\mathbf{I}\right)\mathbf{J}_A+\mathbf{I}=\mathbf{J}_A\,\mathbf{S}\mathbf{J}_A-\mathbf{I}_{AA}+\mathbf{I}\\
&= \mathbf{S}_0+\tfrac{1}{n}\mathbf{VDU}'\mathbf{Z}\mathbf{J}_A+\tfrac{1}{n}\mathbf{J}_A\mathbf{Z}'\mathbf{UDV}'+\mathbf{J}_A\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}-\mathbf{I}\right)\mathbf{J}_A.
\end{aligned}\quad (61)
$$

Therefore

$$\left\|\mathbf{S}_A-\sum\right\|\le\left\|\mathbf{S}_0-\sum\right\|+\frac{2}{n}\|\mathbf{VDU}'\mathbf{Z}\mathbf{J}_A\|+\frac{1}{n}\left\|\mathbf{J}_A\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}-\mathbf{I}\right)\mathbf{J}_A\right\| \quad (62)$$

In view of (52) and Lemma 5, we have

$$\mathrm{E}\left\|\mathbf{S}_0-\sum\right\|^2\le\|\mathbf{D}\|^4\mathrm{E}\left\|\frac{1}{n}\mathbf{U}'\mathbf{U}-\mathbf{I}_r\right\|^2\lesssim\lambda_1^2\left(\frac{r}{n}+\frac{r^2}{n^2}\right)\asymp\frac{\lambda_1^2 r}{n}, \quad (63)$$

where the last step is due to $n\ge c_0 k\,log\,\dfrac{ep}{k}\ge c_0 k$ by assumption. Similarly,

$$\mathrm{E}\left\|\mathbf{J}_A\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}-\mathbf{I}\right)\mathbf{J}_A\right\|^2\lesssim\frac{k}{n}, \quad (64)$$

Again by (49) in Lemma 5,

$$\frac{1}{n^2}\mathrm{E}\|\mathbf{VDU}'\mathbf{Z}\mathbf{J}_A\|^2\le\frac{\lambda_1}{n^2}\mathrm{E}\|\mathbf{U}'\mathbf{Z}\mathbf{J}_A\|^2\lesssim\frac{\lambda_1(n\,(k+r)+kr)}{n^2}\asymp\frac{\lambda_1 k}{n}. \quad (65)$$

Assembling (62) - (65), we have

$$\mathrm{E}\left\|\mathbf{S}_A-\sum\right\|^2\lesssim\frac{\lambda_1^2 r}{n}+\frac{(\lambda_1+1)\,k}{n}. \quad (66)$$

Combining (59) and (66) yields

$$\mathrm{E}\left\|\mathbf{S}_{\hat A}-\sum\right\|^2\mathbf{1}_E\lesssim(\lambda_1+1)\frac{k}{n}\,log\,\frac{ep}{k}+\frac{\lambda_1^2 r}{n}. \quad (67)$$

Next we control the estimation error conditioned on the complement of the event E. First note that $\mathbf{S}_{\hat A}-\sum=\begin{bmatrix}\mathbf{S}_{\hat A\hat A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{\hat A^c\hat A^c}\end{bmatrix}-\sum.$ Then $\|\mathbf{S}_{\hat A}-\Sigma\|$ $\|\mathbf{S}\|+\|\Sigma\|+1$ $\|\Sigma\|(\|\mathbf{W}_p\|$ $+1)$, where $\mathbf{W}_p$ is equal to $\dfrac{1}{n}$ times a $p\times p$ Wishart matrix with n degrees of freedom. Also,

$\|\Sigma - \mathbf{I}\| = \lambda_1$. In view of (14), we have $\|\hat{\Sigma} - \Sigma\| \quad (1 + \lambda_1)(\|\mathbf{W}_p\| + 2)$. Using Cauchy-Schwartz inequality, we have

$$
\begin{aligned}
\mathrm{E}\|\hat{\Sigma} - \Sigma\|^2 \mathbf{1}_{E^c} &\leq \quad (1+\lambda_1)^2 \mathrm{E}(\|\mathbf{W}_p\|+2)^2 \mathbf{1}_{E^c} \\
&\leq \quad (1+\lambda_1)^2 \sqrt{\mathrm{E}(\|\mathbf{W}_p\|+2)^4} \sqrt{\mathrm{P}\{\mathbf{1}_{E^c}\}}.
\end{aligned}
$$

Note that $\|\mathbf{W}_p\| \overset{(d)}{=} \frac{1}{n}\sigma_1^2(\mathbf{Z})$. By [14, Theorem II.7],

$\mathrm{P}\{\sigma_1(\mathbf{Z}) \geq 1 + \sqrt{\frac{p}{n}} + t\} \leq exp(-nt^2/2)$. Then $\mathrm{E}\|\mathbf{W}_p\|^4 \lesssim \frac{1}{n^4}(1+\frac{p^4}{n^4})$. By Lemma 4, we have $\mathrm{P}\{E^c\} \quad 12(ep)^{1-\gamma_1/2}$. Therefore

$$
\mathrm{E}\|\hat{\Sigma} - \Sigma\|^2 \mathbf{1}_{E^c} \lesssim (1+\lambda_1)^2 \frac{1}{n^2}\left(1+\frac{p^2}{n^2}\right) p^{\frac{1}{2}-\frac{\gamma_1}{4}} \quad (68)
$$

Choose a fixed $\gamma_1 \quad 10$. Assembling (67) and (68), we have

$$
\begin{aligned}
\mathrm{E}\|\hat{\Sigma} - \Sigma\|^2 &= \quad \mathrm{E}\|\mathbf{S}_{\hat{A}} - \Sigma\|^2 \mathbf{1}_E + \mathrm{E}\|\mathbf{S}_{\hat{A}} - \Sigma\|^2 \mathbf{1}_{E^c} \\
&\lesssim \quad (1+\lambda_1)\frac{k}{n} log \frac{ep}{k} + \frac{\lambda_1^2 r}{n} + (1+\lambda_1)^2 \frac{1}{n^2} \asymp (1+\lambda_1)\frac{k}{n} log \frac{ep}{k} + \frac{\lambda_1^2 r}{n}.
\end{aligned}
$$

<u>Proof of Theorem 2:</u> To prove the theorem, it suffices to show that with the desired probability,

$$
\sigma_r(\mathbf{S}_{\hat{A}}) > 1 + \gamma_2 \sqrt{\|\mathbf{S}_{\hat{A}}\|} \eta(k,n,p,\gamma_1) > \sigma_{r+1}(\mathbf{S}_{\hat{A}}). \quad (69)
$$

Recall (50) and (52). By [10, Proposition 4], with probability at least $1 - 2(ep)^{-\gamma_1/2}$,

$\|\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\| \leq \eta(k,n,p,\gamma_1)$. Under (18), $\sqrt{\frac{k}{n} log \frac{ep}{k}} \asymp \eta(k,n,p,\gamma_1) \leq 1/(2\tau)$ when c_0 is sufficiently small, and so $\eta(k,n,p,\gamma_1) \quad 1/(2\tau)$. Thus,

$$
\|\mathbf{S}_0 - \Sigma\| \leq \|\mathbf{D}\|^2 \|\frac{1}{n}\mathbf{U}'\mathbf{U} - \mathbf{I}_r\| \leq \lambda_r/2. \quad (70)
$$

Therefore, Lemma 6 leads to (51). Moreover, Weyl's inequality leads to

$$
\sigma_r(\mathbf{S}_0) \geq \sigma_r(\Sigma) - |\sigma_r(\mathbf{S}_0) - \sigma_r(\Sigma)| \geq \lambda_r + 1 - \lambda_1\eta(k,n,p,\gamma_1) \geq \frac{1}{2}\lambda_r + 1 > 1. \quad (71)
$$

Next, we consider $\mathbf{S}_A - \mathbf{S}_0$. By (61), we have

$$
\|\mathbf{S}_A - \mathbf{S}_0\| \leq \frac{2}{n}\|\mathbf{V}\mathbf{D}\mathbf{U}'\mathbf{Z}\mathbf{J}_A\| + \frac{1}{n}\|\mathbf{J}_A(\mathbf{Z}'\mathbf{Z} - \mathbf{I})\mathbf{J}_A\|.
$$

By [10, Proposition 4] and (18), when c_0 is sufficiently small, with probability at least $1 - (ep + 1)(ep)^{-\gamma_1/2}$,

$$
\begin{aligned}
\|\mathbf{V}\mathbf{D}\mathbf{U}'\mathbf{Z}\mathbf{J}_A\| &\leq \|\mathbf{D}\|\|\mathbf{U}'\mathbf{Z}\mathbf{J}_A\| \\
&\leq \sqrt{\lambda_1}n\sqrt{1+\tfrac{8\gamma_1\,log\,p}{3n}}\left(\sqrt{\tfrac{r}{n}}+\sqrt{\tfrac{k}{n}}+\sqrt{\gamma_1\,log\,\tfrac{ep}{n}}\right) \\
&\leq n\sqrt{\lambda_1}\eta(k,p,n,\gamma_1).
\end{aligned}
$$

Moreover, [10, Proposition 3] implies that with probability at least $1-2(ep)^{-\gamma_1/2}$,

$$
\frac{1}{n}\|\mathbf{J}_A(\mathbf{Z}'\mathbf{Z}-\mathbf{I})\mathbf{J}_A\| \leq \eta(k,n,p,\gamma_1).
$$

Assembling the last three displays, we obtain that with probability at least $1 - (ep + 3)(ep)^{-\gamma_1/2}$,

$$
\|\mathbf{S}_A - \mathbf{S}_0\| \leq (2\sqrt{\lambda_1}+1)\eta(k,n,p,\gamma_1). \quad (72)
$$

Last but not least, Lemma 4 implies that, under the condition of Theorem 2, with probability at least $1 - 12(ep)^{1-\gamma_1/2}$, (40) holds. Together with (72), it implies that

$$
\|\mathbf{S}_{\hat{A}} - \mathbf{S}_0\| \leq [C_0(\gamma_1)+3]\sqrt{\lambda_1+1}\eta(k,n,p,\gamma_1), \quad (73)
$$

where $C_0(\gamma_1)=14+4\sqrt{\gamma_1}$. By (18), we could further upper bound the right-hand side by $\lambda_1/4 \vee 1/2$. When $\lambda_1 \geq 1$, $\sqrt{\lambda_1+1} \leq \sqrt{2\lambda_1} < \sqrt{2}\lambda_1$, so for sufficiently small c_0, (18) implies that the right side of (73) is further bounded by $\lambda_1/4$. When $\lambda_1 \leq 1$, $\sqrt{\lambda_1+1} \leq \sqrt{2}$, and so the right side of (73) is further bounded by $1/2$ for sufficiently small c_0.

Thus, the last display, together with (70), implies

$$
\begin{aligned}
\|\mathbf{S}_{\hat{A}} - \textstyle\sum\| &\leq \lambda_1\eta(k,n,p,\gamma_1)+[C_0(\gamma_1)+3]\sqrt{\lambda_1+1}\eta(k,n,p,\gamma_1) \\
&< \tfrac{1}{2}(\lambda_1+1).
\end{aligned} \quad (74)
$$

Here, the last inequality comes from the above discussion, and the fact that $\eta(k, n, p, \gamma_1) < 1/4$ for small c_0. The triangle inequality further leads to

$$
\frac{1}{2}(\lambda_1+1)<\|\textstyle\sum\|-\|\mathbf{S}_{\hat{A}}-\textstyle\sum\| \leq \|\mathbf{S}_{\hat{A}}\| \leq \|\textstyle\sum\|+\|\mathbf{S}_{\hat{A}}-\textstyle\sum\|<\frac{3}{2}(\lambda_1+1). \quad (75)
$$

Set $\gamma_2 \geq 2[C_0(\gamma_1)+3]=8\sqrt{\gamma_1}+34$. Then (51), (73) and (75) jointly imply that, with probability at least $1-(13ep+5)(ep)^{-\gamma_1/2}$, the second inequality in (69) holds. Moreover, (74) and the triangle inequality implies that, with the same probability,

$$
\begin{aligned}
\sigma_r(\mathbf{S}_{\hat{A}}) &\geq & \sigma_r(\Sigma) - \|\mathbf{S}_{\hat{A}} - \Sigma\| \\
&\geq & 1 + \lambda_r - \lambda_1\,\eta(k,n,p,\gamma_1) - [C_0(\gamma_1) + 3]\sqrt{\lambda_1 + 1}\,\eta(k,n,p,\gamma_1) \\
&\geq & 1 + \lambda_r \left[1 - \tau\eta(k,n,p,\gamma_1) - \tfrac{\sqrt{\lambda_1+1}}{2\lambda_1}\gamma_2\tau\eta(k,n,p,\gamma_1) \right] \\
&\geq & 1 + \tfrac{3}{2}\gamma_2(\lambda_1 + 1)\eta(k,n,p,\gamma_1).
\end{aligned}
$$

Here, the last inequality holds when $\lambda_r \geq \beta\sqrt{\frac{k}{n}\,log\,\frac{ep}{k}}$ for a sufficiently large β which depends only on γ_1, γ_2 and τ. In view of (75), the last display implies the first inequality of (69). This completes the proof of the upper bound.

Proof of Theorem 3: Let E be the event such that Lemma 3, Lemma 4, the upper bound in Theorem 2, and (71) hold. Then $P(E^c) \quad C(ep)^{1-\gamma_1/2}$.

On the event E, $\hat{\Sigma} = \mathbf{S}_{\hat{A}}$. Moreover, Lemma 6 shows that $\mathbf{VV}'$ is the projection matrix onto the principal subspace of $\mathbf{S}_0$, and Theorem 2 ensures $\hat{\mathbf{V}} \in \mathbb{R}^{p \times r}$. Thus, Proposition 1 leads to

$$
\begin{aligned}
\|\hat{\Sigma} - \mathbf{S}_0\|\mathbf{1}_{\{E\}} &\geq & \tfrac{1}{2}(\sigma_r(\mathbf{S}_0) - \sigma_{r+1}(\mathbf{S}_0))\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\|\mathbf{1}_{\{E\}} \\
&\geq & \tfrac{\lambda_r}{4}\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\|\mathbf{1}_{\{E\}},
\end{aligned}
$$

and so

$$
E\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\|^2\mathbf{1}_{\{E\}} \leq \frac{16}{\lambda_r^2}E\|\hat{\Sigma} - \mathbf{S}_0\|^2\mathbf{1}_{\{E\}}.
$$

To further bound the right-hand side of the last display, we apply (61), (64), (65), Lemma 3 and Lemma 4 to obtain

$$
\begin{aligned}
E\|\hat{\Sigma} - \mathbf{S}_0\|^2\mathbf{1}_{\{E\}} &\leq & E\|\mathbf{S}_{\hat{A}} - \mathbf{S}_0\|^2\mathbf{1}_{\{E\}} \\
&\lesssim & E\|\mathbf{S}_{\hat{A}} - \mathbf{S}_A\|^2\mathbf{1}_{\{E\}} + E\|\mathbf{S}_A - \mathbf{S}_0\|^2 \lesssim (\lambda_1 + 1)\tfrac{k}{n}\,log\,\tfrac{ep}{k}.
\end{aligned}
$$

Together with the second last display, this implies

$$
E\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\|^2\mathbf{1}_{\{E\}} \leq C_{\tau^2}\frac{k(\lambda_1 + 1)}{n\lambda_1^2}\,log\,\frac{ep}{k}. \quad (76)
$$

Now consider the event E^c. Note that $\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\| \quad 1$ always holds. Thus,

$$
E\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{VV}'\|^2\mathbf{1}_{\{E^c\}} \leq P(E^c) \leq C(ep)^{1-\gamma_1/2} \lesssim \frac{\lambda_1 + 1}{n\lambda_1^2}, \quad (77)
$$

where the last inequality holds under condition (20) for all $\gamma_1 \geq (1 + \frac{2}{M_o}) M_1$. Assembling (76) and (77), we obtain the upper bounds.

7.2.2 Proofs of the Lower Bounds

Proof of Theorem 4: 1° The minimax lower bound for estimating span($\mathbf{V}$) follows straightforwardly from previous results on estimating the leading singular vector, i.e., the rank-one case (see, e.g., [5, 39]). The desired lower bound (25) can be found in [10, Eq. (58)] in the proof of [10, Proof of Theorem 3].

2° Next we establish the minimax lower bound for estimating the spiked covariance matrix Σ under the spectral norm, which is considerably more involved. Let $\Theta_1 = \Theta_1(k, p, r, \lambda)$. In view of the fact that $a \wedge (b + c) \quad a \wedge b + a \wedge c \quad 2(a \wedge (b + c))$ for all $a, b, c \quad 0$, it is convenient to prove the following equivalent lower bound

$$\inf_{\hat{\Sigma}} \sup_{\Sigma \in \Theta} \mathrm{E} \| \hat{\Sigma} - \Sigma \|^2 \gtrsim \lambda^2 \wedge \left(\frac{(\lambda+1)k}{n} \, log \, \frac{ep}{k} \right) + \lambda^2 \left(1 \wedge \frac{r}{n} \right). \quad (78)$$

To this end, we show that the minimax risk is lower bounded by the two terms on the right-hand side of (78) separately. In fact, the first term is the minimax rate in the rank-one case and the second term is the rate of the oracle risk when the estimator knows the true support of $\mathbf{V}$.

2.1° Consider the following rank-one subset of the parameter space

$$\Theta' = \{ \Sigma = \mathbf{I}_p + \lambda \mathbf{v} \mathbf{v}' : \mathbf{v} \in B_0(k) \cap \mathbb{S}^{p-1} \} \subset \Theta_1.$$

Then $\sigma_1(\Sigma) = 1 + \lambda$ and $\sigma_2(\Sigma) = 1$. For any estimator $\hat{\Sigma}$ denote by $\hat{\mathbf{v}}$ its leading singular vector. Applying Proposition 1 yields

$$\| \hat{\Sigma} - \Sigma \| \geq \frac{\lambda}{2} \| \hat{\mathbf{v}} \hat{\mathbf{v}}' - \mathbf{v} \mathbf{v}' \|.$$

Then

$$\begin{aligned} \inf_{\hat{\Sigma}} \sup_{\Sigma \in \Theta'} \mathrm{E} \| \hat{\Sigma} - \Sigma \|^2 &\geq \frac{\lambda^2}{4} \inf_{\hat{\mathbf{v}}} \sup_{\mathbf{v} \in B_0(k) \cap \mathbb{S}^{p-1}} \mathrm{E} \| \hat{\mathbf{v}} \hat{\mathbf{v}}' - \mathbf{v} \mathbf{v}' \|^2 \\ &\geq \frac{\lambda^2}{8} \inf_{\hat{\mathbf{v}}} \sup_{\mathbf{v} \in B_0(k) \cap \mathbb{S}^{p-1}} \mathrm{E} \| \hat{\mathbf{v}} \hat{\mathbf{v}}' - \mathbf{v} \mathbf{v}' \|_{\mathrm{F}}^2 \end{aligned} \quad (79)$$

$$\gtrsim \lambda^2 \left(1 \wedge \frac{(\lambda+1)k}{n\lambda^2} \, log \, \frac{ep}{k} \right), \quad (80)$$

where (79) follows from rank($\mathbf{v}\hat{\mathbf{v}}' - \mathbf{v}\mathbf{v}'$) 2 and (80) follows from the minimax lower bound in [39, Theorem 2.1] (see also [5, Theorem 2]) for estimating the leading singular vector.

2.2° To prove the lower bound $\lambda^2(1 \wedge \frac{r}{n})$, consider the following (composite) hypotheses testing problem:

$$H_0: \sum = \begin{bmatrix} \left(1+\frac{\lambda}{2}\right)\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-r} \end{bmatrix}, vs. H_1: \sum = \begin{bmatrix} \left(1+\frac{\lambda}{2}\right)(\mathbf{I}_r+\rho\mathbf{vv}') & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-r} \end{bmatrix}, \mathbf{v} \in S^{r-1}, \quad (81)$$

where $\rho = \frac{b\lambda}{\lambda+2}(1 \wedge \sqrt{\frac{r}{n}})$ with a sufficiently small absolute constant $b > 0$. Since $r \in [k]$ and $\rho\left(1+\frac{\lambda}{2}\right) \le \lambda$, both the null and the alternative hypotheses belong to the parameter set Θ_1 defined in (4). Following the Le Cam's twopoint argument [38, Section 2.3], next we show that the minimal sum of Type-I and Type-II error probabilities of testing (81) is non-vanishing. Since any pair of covariance matrices in H_0 and H_1 differ in operator norm by at least $\rho(1+\frac{\lambda}{2})=\frac{b\lambda}{2}\left(1 \wedge \sqrt{\frac{r}{n}}\right)$, we obtain a lower bound of rate $\lambda^2(1 \wedge \frac{r}{n})$.

To this end, let $\mathbf{X}$ consist of n iid rows drawn from $N(0, \Sigma)$, where Σ is either from H_0 or H_1. Since under both the null and the alternative, the last $p-r$ columns of $\mathbf{X}$ are standard normal and independent of the first r columns, we conclude that the first r columns form a sufficient statistic. Therefore the minimal Type-I+II error probability testing (81), denoted by ε_n, is equal to that of the following testing problem of dimension r and sample size n:

$$H_0: \sum = \mathbf{I}_r, \quad versus \quad H_1: \sum = \mathbf{I}_r+\rho\mathbf{vv}', \mathbf{v} \in S^{r-1}. \quad (82)$$

Recall that the minimax risk is lower bounded by the Bayesian risk. For any random vector $\mathbf{u}$ taking values in S^{r-1}, denote by the $E[N(0, \mathbf{I}_r + \rho\mathbf{uu}')^{\otimes n}]$ the mixture alternative distribution with a prior equal to the distribution of $\mathbf{u}$. Applying [38, Theorem 2.2 (iii)] we obtain the following lower bound in terms of the χ^2-divergence from the mixture alternative to the null:

$$\varepsilon_n \ge 1 - (\chi^2(E[N(0, \mathbf{I}_r+\rho\mathbf{uu}')^{\otimes n}]\|N(0, \mathbf{I}_r)^{\otimes n})/2)^{1/2} \quad (83)$$

Consider the unit random vector $\mathbf{u}$ with iid coordinates taking values in $\frac{1}{\sqrt{r}}\{\pm 1\}$ uniformly. Since $\rho \quad 1$, applying the equality case of Lemma 7 yields

$$1+\chi^2\left(E[N(0, \mathbf{I}_r+\rho\mathbf{uu}')^{\otimes n}]\|N(0, \mathbf{I}_r)^{\otimes n}\right) = E\,exp\,(n\rho^2\langle\mathbf{uu}, \widetilde{\mathbf{u}}\widetilde{\mathbf{u}}\rangle^2)$$
$$= E\,exp\,(n\rho^2 G_r^2),$$

where in the last step we recall that G_r is the symmetric random walk on $\mathbb{Z}$ at r^{th} step defined in (34). Since $n\rho^2 \le b^2\frac{r}{n}$, choosing $b^2 \le \frac{1}{36}$ as a fixed constant and applying Lemma 1 with $p = k = r$ (the non-sparse case), we conclude that

$$\chi^2 \left(\mathrm{E}[N(0, \mathbf{I}_p + \lambda \mathbf{u}\mathbf{u}')^{\otimes n}] \| N(0, \mathbf{I}_p)^{\otimes n}\right) \le g(b^2) - 1, \quad (84)$$

where g is given by in Lemma 1 and satisfies $g > 1$. Combining (83) - (84), we obtain the following lower bound for estimating $\mathbf{\Sigma}$:

$$\inf_{\hat{\Sigma}} \sup_{\Sigma \in \Theta} \mathrm{E} \| \hat{\Sigma} - \Sigma \|^2 \gtrsim \left(1 + \frac{\lambda}{2}\right)^2 \rho^2 \asymp \lambda^2 \left(1 \wedge \frac{r}{n}\right). \quad (85)$$

As we mentioned in Section 4, the rank-detection lower bound in Theorem 5 is a direct consequence of Proposition 2 concerning testing rank-zero versus rank-one perturbation, which we prove below.

7.3 Proof of Proposition 2

Proof—Analogous to (83), any random vector $\mathbf{u}$ taking vales in $S^{p-1} \cap B_0(k)$ gives the Bayesian lower bound

$$\varepsilon_n(k, p, \lambda) \ge 1 - (\chi^2(\mathrm{E}N(0, \mathbf{I}_p + \lambda \mathbf{u}\mathbf{u}')^{\otimes n} \| N(0, \mathbf{I}_p)^{\otimes n})/2). \quad (86)$$

Put

$$t \triangleq \frac{n\lambda^2}{k^2} = \frac{\beta_0}{k} \log \frac{ep}{k}.$$

Let the random sparse vector $\mathbf{u}$ be defined in (30). In view of Lemma 7 as well as the facts that $\mathrm{rank}(\lambda \mathbf{v}\mathbf{v}') = 1$ and $\|\lambda \mathbf{v}\mathbf{v}'\| = \lambda \quad 1$, we have

$$
\begin{aligned}
1 + \chi^2 \left(\mathrm{E}[N(0, \mathbf{I}_p + \lambda \mathbf{u}\mathbf{u}')^{\otimes n} \quad] \qquad\qquad \| N(0, \mathbf{I}_p)^{\otimes n}\right) \\
= \mathrm{E} exp \left(n\lambda^2 \langle \mathbf{u}\mathbf{u}', \tilde{\mathbf{u}}\tilde{\mathbf{u}}' \rangle^2\right) \\
= \mathrm{E} exp \left(n\lambda^2 \left\langle \tfrac{1}{\sqrt{k}} \mathbf{J}_I \mathbf{w}, \tfrac{1}{\sqrt{k}} \mathbf{J}_{\tilde{I}} \tilde{\mathbf{w}} \right\rangle^2\right) \\
= \mathrm{E} exp \left(t(\mathbf{w}' \mathbf{J}_{I \cap \tilde{I}} \tilde{\mathbf{w}}')^2\right) \\
= \mathrm{E} exp \left(t G_H^2\right),
\end{aligned}
$$

where in the last step we have defined $H \triangleq |I \cap \tilde{I}| \sim \text{Hypergeometric}(p, k, k)$ and $\{G_m\}$ is the symmetric random walk on $\mathbb{Z}$ defined in (34). Now applying Lemma 1, we conclude that

$$\chi^2 \left(\mathrm{E}[N(0, \mathbf{I}_p + \lambda \mathbf{u}\mathbf{u}')^{\otimes n}] \| N(0, \mathbf{I}_p)^{\otimes n}\right) \le g(\beta_0) - 1$$

where g is given by in Lemma 1 satisfying $g(0+) = 1$. In view of (86), we conclude that

$$\varepsilon_n(k, p, \lambda) \ge max\{0, 1 - \sqrt{(g(\beta_0) - 1)/2}\} \triangleq w(\beta_0).$$

Note that the function w satisfies $w(0+) = 1$.

7.4 Proof of Propositions 3 and 4

We give here a joint proof of Propositions 3 and 4.

Proof—1° (Upper bounds) By assumption, we have $\lambda \asymp 1$ and $\frac{k}{n} \log \frac{ep}{k} \leq c_0$. Since $r \in [k]$, applying Theorem 1 yields

$$\sup_{\Sigma \in \Theta_1(k,p,r,\lambda)} \mathrm{E} \|\hat{\Sigma} - \Sigma\|^2 \lesssim \frac{k}{n} \log \frac{ep}{k}. \quad (87)$$

Note that

$$\|\hat{\Sigma}^{-1} - \Sigma^{-1}\| \leq \|\hat{\Sigma}^{-1}\|\|\Sigma^{-1}\|\|\hat{\Sigma} - \Sigma\| \leq \frac{\|\hat{\Sigma} - \Sigma\|}{1 - \|\hat{\Sigma} - \Sigma\|}, \quad (88)$$

where the last inequality follows from $\sigma_p(\Sigma) = 1$ and Weyl's inequality. By Chebyshev's

inequality, $\mathrm{P}\left\{\|\hat{\Sigma} - \Sigma\| \geq t\right\} \leq \frac{1}{t^2} \mathrm{E}\|\hat{\Sigma} - \Sigma\|^2$. Let $E = \{\|\hat{\Sigma} - \Sigma\| \leq \frac{1}{2}\}$. Then

$$\mathrm{P}\{E^c\} \lesssim \frac{k}{n} \log \frac{ep}{k}. \quad (89)$$

Moreover, again by Weyl's inequality, we have $E \subset \{\sigma_p(\hat{\Sigma}) \geq \frac{1}{2}\}$ hence $\hat{\Omega}\mathbf{1}_{\{E\}} = (\hat{\Sigma})^{-1}\mathbf{1}_{\{E\}}$ in view of (36). On the other hand, by definition we always have $\|\hat{\Omega}\|$ 2. Therefore

$$\begin{aligned}
\mathrm{E}\|\hat{\Omega} - \Omega\|^2 &= \mathrm{E}\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|^2 \mathbf{1}_{\{E\}} + \mathrm{E}\|\hat{\Omega} - \Omega\|^2 \mathbf{1}_{\{E^c\}} \\
&\leq \mathrm{E}\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|^2 \mathbf{1}_{\{E\}} + 3\mathrm{P}\{E^c\} \\
&\leq 4\mathrm{E}\|\hat{\Sigma} - \Sigma\|^2 + 3\mathrm{P}\{E^c\}
\end{aligned} \quad (90)$$

$$\lesssim \frac{k}{n} \log \frac{ep}{k}. \quad (91)$$

where (90) follows from (88) and (91) is due to (87) and (89).

Next we prove upper bound for estimating **E**. Recall that $\mathbf{E} + \mathbf{I}_p$ and $\hat{\mathbf{E}} + \mathbf{I}_p$ give the diagonal matrices formed by the ordered singular values of Σ and $\hat{\Sigma}$, respectively. Similar to the proof of Proposition 4, Weyl's inequality implies that $\|\hat{\mathbf{E}} - \mathbf{E}\| = \|\hat{\mathbf{E}} + \mathbf{I}_p - (\mathbf{E} + \mathbf{I}_p)\| = \max_i |\sigma_i(\hat{\Sigma}) - \sigma_i(\Sigma)| \quad \|\hat{\Sigma} - \Sigma\|$, where for any $i \in [p]$, $\sigma_i(\Sigma)$ is the ith largest eigenvalue of Σ. In view of Theorem 1, we have $\mathrm{E}\|\hat{\mathbf{E}} - \mathbf{E}\|^2 \leq \frac{k}{n} \log \frac{ep}{k}$.

2° (Lower bounds) The lower bound follows from the testing result in Proposition 2. Consider the testing problem (29), then both the null ($\Sigma = \mathbf{I}_p$) and alternatives ($\Sigma = \mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}'$) are contained in the parameter space Θ_1. By Proposition 2, they cannot be distinguished with probability 1 if $\lambda^2 \asymp \frac{k}{n} \log \frac{ep}{k}$. The spectra differs at least $|\sigma_1(\mathbf{I}_p) - \sigma_1(\mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}')| = \lambda$. By the Woodbury identity, $(\mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}')^{-1} = \mathbf{I}_p - \frac{\lambda}{\lambda+1} \mathbf{v}\mathbf{v}'$, hence $\|\mathbf{I}_p - (\mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}')^{-1}\| = \frac{\lambda}{1+\lambda} \asymp \lambda$. The lower bound $\frac{k}{n} \log \frac{ep}{k}$ is now completed by way of the usual two-point argument.

7.5 Proof of Proposition 1

Proof—Recall that $\sigma_1(\Sigma) \quad \ldots \quad \sigma_p(\Sigma) \quad 0$ denote the ordered singular values of Σ. If $\sigma_r\left(\hat{\Sigma}\right) \leq \frac{\sigma_r(\Sigma) + \sigma_{r+1}(\Sigma)}{2}$, then by Weyl's inequality, we have $\|\hat{\Sigma} - \Sigma\| \geq \frac{1}{2}\left(\sigma_r\left(\Sigma\right) - \sigma_{r+1}\left(\Sigma\right)\right)$. If $\sigma_r\left(\hat{\Sigma}\right) \geq \frac{\sigma_r(\Sigma) + \sigma_{r+1}(\Sigma)}{2}$, then by David–Kahn's sin-theta theorem [15] (see also [10, Theorem 10]) we have

$$\|\hat{\mathbf{V}}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\| \leq \frac{\|\hat{\Sigma} - \Sigma\|}{\left|\sigma_r(\hat{\Sigma}) - \sigma_{r+1}(\Sigma)\right|} \leq \frac{2\|\hat{\Sigma} - \Sigma\|}{\sigma_r(\Sigma) - \sigma_{r+1}(\Sigma)},$$

completing the proof of (22) in view of the fact that $\|\mathbf{V}\hat{\mathbf{V}}' - \mathbf{V}\mathbf{V}'\| \quad 1$.

7.6 Proof of Lemma 1

Proof—First of all, we can safely assume that $p \quad 5$, for otherwise the expectation on the right-hand side of (35) is obviously upper bounded by an absolutely constant. In the sequel we shall assume that

$$0 < a \leq \frac{1}{36}. \quad (92)$$

We use normal approximation of the random walk G_m for small m and use truncation argument to deal with large m. To this end, we divide $[p]$, the whole range of k, into three regimes.

Case I: Large k—Assume that $\frac{p}{2} \leq k \leq p$. Then $t \leq \frac{2a \log(2e)}{p}$. By the following non-asymptotic version of the Tusnády's coupling lemma (see, for example, [6, Lemma 4, p. 242]), for each m, there exists $Z_m \sim N(0, m)$, such that

$$|G_m| \overset{\text{s.t.}}{\leq} |Z_m| + 2. \quad (93)$$

Since $H \quad p$, in view of (92), we have

$$\begin{aligned}
\mathrm{E}\, exp\,(tG_H^2) &\leq \mathrm{E}\, exp\,(tG_p^2) \leq \mathrm{E}\, exp\,(t(|Z_p|+2)^2) \\
&\leq exp\,(8t)\mathrm{E}\, exp\,(2tZ_p^2)
\end{aligned} \tag{94}$$

$$= \frac{exp(8t)}{\sqrt{1-4pt}} \leq \frac{exp(7a)}{\sqrt{1-8log(2e)a}} \tag{95}$$

Case II: Small k—Assume that $1 \leq k \leq log\dfrac{ep}{k}$, which, in particular, implies that $k \leq p^{\frac{1}{3}} \wedge \dfrac{p}{2}$ since $p \quad 5$. Using $G_H^2 \leq H^2 \leq kH$, we have

$$\mathrm{E}\, exp\,(tG_H^2) \leq \mathrm{E}\left(\frac{ep}{k}\right)^{aH} \leq \left(1 + \frac{k}{p-k}\left(\left(\frac{ep}{k}\right)^a - 1\right)\right)^k \tag{96}$$

$$\leq exp\left\{\frac{2\sqrt{k}}{\sqrt{p}}\left(\left(\frac{ep}{k}\right)^a - 1\right)\right\} \tag{97}$$

$$\leq exp\,(8a), \tag{98}$$

where (96) follows from the stochastic dominance of hypergeometric distributions by binomial distributions (see, e.g., [23, Theorem 1.1(d)])

$$Hypergeometric(p, k, k) \overset{\text{s.t.}}{\leq} Bin\left(k, \frac{k}{p-k}\right), \quad k \leq \frac{p}{2} \tag{99}$$

and the moment generating function of binomial distributions, (97) is due to $(1+x)^k$ $exp(kx)$ for $x \quad 0$ and $\dfrac{k^2}{p} \leq \sqrt{\dfrac{k}{p}}$, (98) is due to

$$\sup_{y \geq 1} \frac{(ey)^a - 1}{\sqrt{y}} = 2\sqrt{e}a((1-2a))^{\frac{1}{2a}-1} \leq 4a, \quad \forall a \leq \frac{1}{2},$$

and

Case III: Moderate k—Assume that $log\dfrac{ep}{k} \leq k \leq \dfrac{p}{2}$. Define

$$A \triangleq \frac{k}{\sqrt{a}\, log\,\frac{ep}{k}} \wedge k \tag{100}$$

and write

$$\mathrm{E}\, exp\,(tG_H^2) = \mathrm{E}\, exp\,(tG_H^2)\mathbf{1}_{\{H \leq A\}} + \mathrm{E}\, exp\,(tG_H^2)\,\mathbf{1}_{\{H > A\}}. \tag{101}$$

Next we bound the two terms in (101) separately: For the first term, we use normal approximation of G_m. By (93) - (94), for each fixed $m \quad A$, we have

$$\mathrm{E}\,exp\left(tG_m^2\right) \leq \frac{exp(8t)}{\sqrt{1-4mt}} \leq \frac{exp(8a)}{\sqrt{1-4\sqrt{a}}}, \quad (102)$$

where the last inequality follows from $k \geq log\frac{ep}{k}$ and (100).

To control the second term in (101), without loss of generality, we assume that $A \quad k$, i.e.,

$A = \frac{k}{\sqrt{a}\,log\frac{ep}{k}} \geq \frac{1}{\sqrt{a}}$. we proceed similarly as in (96) - (98):

$$
\begin{aligned}
& \mathrm{E}\,exp\left(tG_H^2\right)\mathbf{1}_{\{H>A\}} \\
\leq \quad & \mathrm{E}\,exp\left(tkH\right)\mathbf{1}_{\{H>A\}} \\
\leq \quad & \sum_{m>A} exp(tkm)\left(\frac{k}{p-k}\right)^m\left(\frac{p-2k}{p-k}\right)^{k-m}\binom{k}{m}
\end{aligned}
\quad (103)
$$

$$\leq \sum_{m>A} exp\left(am\,log\frac{p}{k} - m\,log\frac{p}{2k} + m\,log\left(\sqrt{a}\mathrm{e}\,log\frac{ep}{k}\right)\right) \quad (104)$$

$$\leq \sum_{m>A} exp\left\{-\left(1-a-\frac{1}{\mathrm{e}}\right)m\,log\,2 + m\,log\left(2\sqrt{a}\mathrm{e}\right) + \frac{m}{\mathrm{e}}\right\} \quad (105)$$

$$\leq 7exp\left(-\frac{1}{\sqrt{a}}\right), \quad (106)$$

where (103) follows from (99), (104) follows from that $\binom{k}{m} \leq \left(\frac{\mathrm{e}k}{m}\right)^m$, $p \quad 2k$ and m A, (105) follows from that $\mathrm{e}\,log\,x \quad x$ for all $x \quad 1$, and (106) is due to (100) and our choice of a in (92). Plugging (102) and (106) into (101), we obtain

$$\mathrm{E}\,exp\left(tG_H^2\right) \leq \frac{exp(8a)}{\sqrt{1-4\sqrt{a}}} + 7exp\left(-\frac{1}{\sqrt{a}}\right). \quad (107)$$

We complete the proof of (35), with $g(a)$ defined as the maxima of the righthand sides of (95), (98) and (107).

Acknowledgments

References

1. Berthet Q, Rigollet P. Complexity theoretic lower bounds for sparse principal component detection. Journal of Machine Learning Research: Workshop and Conference Proceedings. 2013; 30:1–21.

2. Berthet Q, Rigollet P. Optimal detection of sparse principal components in high dimension. Annals of Statistics. 2013; 41(4):1780–1815.

3. Bickel P, Levina E. Covariance regularization by thresholding. The Annals of Statistics. 2008; 36(6):2577–2604.

4. Bickel P, Levina E. Regularized estimation of large covariance matrices. The Annals of Statistics. 2008; 36(1):199–227.

5. Birnbaum A, Johnstone I, Nadler B, Paul D. Minimax bounds for sparse PCA with noisy high-dimensional data. arXiv preprint arXiv:1203.0967. 2012

6. Bretagnolle J, Massart P. Hungarian constructions from the nonasymptotic viewpoint. The Annals of Probability. 1989; 17(1):239–256.

7. Bunea F, Xiao L. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. arXiv preprint arXiv:1212.5321. 2012

8. Cai T, Liu W. Adaptive thresholding for sparse covariance matrix estimation. Journal of the American Statistical Association. 2011; 106:672–684.

9. Cai T, Liu W, Zhou H. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. arXiv preprint arXiv:1212.2882. 2012

10. Cai, T.; Ma, Z.; Wu, Y. Sparse PCA: Optimal rates and adaptive estimation. 2012. URL http://arxiv.org/abs/1211.1309. Preprint

11. Cai T, Ren Z, Zhou H. Optimal rates of convergence for estimating toeplitz covariance matrices. Probability Theory and Related Fields. 2012:1–43.

12. Cai T, Zhang CH, Zhou H. Optimal rates of convergence for covariance matrix estimation. The Annals of Statistics. 2010; 38(4):2118–2144.

13. Cai T, Zhou H. Optimal rates of convergence for sparse covariance matrix estimation. The Annals of Statistics. 2012

14. Davidson, K.; Szarek, S. Handbook on the Geometry of Banach Spaces, chap Local operator theory, random matrices and Banach spaces. Vol. 1. Elsevier Science; 2001. p. 317-366.

15. Davis C, Kahan W. The rotation of eigenvectors by a perturbation. III. SIAM J Numer Anal. 1970; 7(1):1–46.

16. Fan J, Fan Y, Lv J. High dimensional covariance matrix estimation using a factor model. Journal of Econometrics. 2008; 147(1):186–197.

17. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9(3):432–441. [PubMed: 18079126]

18. Horn, RA.; Johnson, CR. Matrix Analysis. Cambridge University Press; 1990.

19. Johnstone I. On the distribution of the largest eigenvalue in principal component analysis. The Annals of Statistics. 2001; 29:295–327.

20. Johnstone I, Lu A. On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Association. 2009; 104:682–693. [PubMed: 20617121]

21. Jung S, Marron J. PCA consistency in high dimension, low sample size context. The Annals of Statistics. 2009; 37(6B):4104–4130.

22. Karoui N. Operator norm consistent estimation of large-dimensional sparse covariance matrices. The Annals of Statistics. 2008; 36:2717–2756.

23. Klenke A, Mattner L. Stochastic ordering of classical discrete distributions. Advances in Applied probability. 2010; 42(2):392–410.

24. Kritchman S, Nadler B. Determining the number of components in a factor model from limited noisy data. Chemometrics and Intelligent Laboratory Systems. 2008; 94(1):19–32.

25. Kritchman S, Nadler B. Non-parametric detection of the number of signals: hypothesis testing and random matrix theory. 2009; 57(10):3930–3941.

26. Lam C, Fan J. Sparsistency and rates of convergence in large covariance matrix estimation. The Annals of Statistics. 2009; 37:4254–4278. [PubMed: 21132082]

27. Le Cam, L. Asymptotic Methods in Statistical Theory. Springer-Verlag New York, Inc; 1986.

28. Lounici K. High-dimensional covariance matrix estimation with missing observations. arXiv preprint arXiv:1201.2577. 2012

29. Lounici K. Sparse principal component analysis with missing observations. arXiv preprint arXiv: 1205.7060. 2012

30. Lounici K, Pontil M, Van De Geer S, Tsybakov A. Oracle inequalities and optimal inference under group sparsity. The Annals of Statistics. 2011; 39(4):2164–2204.

31. Ma Z. Sparse principal component analysis and iterative thresholding. arxiv preprint arXiv: 1112.2432. 2011

32. Onatski A. Asymptotics of the principal components estimator of large factor models with weak factors. J Econometrics. 2012; 168:244–258.

33. Onatski A, Moreira M, Hallin M. Signal detection in high dimension: The multispiked case. arXiv preprint arXiv:1210.5663. 2012

34. Patterson N, Price A, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190.10.1371/journal.pgen.0020190 [PubMed: 17194218]

35. Paul D. Asymptotics of sample eigenstruture for a large dimensional spiked covariance model. Statistica Sinica. 2007; 17(4):1617–1642.

36. Ravikumar P, Wainwright M, Raskutti G, Yu B. High-dimensional covariance estimation by minimizing l_1-penalized log-determinant divergence. Electronic Journal of Statistics. 2011; 5:935–980.

37. Stewart, G.; Sun, JG. Computer science and scientific computing. Academic Press; 1990. Matrix Perturbation Theory.

38. Tsybakov, A. Introduction to Nonparametric Estimation. Springer Verlag; 2009.

39. Vu, V.; Lei, J. Minimax rates of estimation for sparse PCA in high dimensions. the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS'12); 2012. URL http://arxiv.org/abs/1202.0786

40. Vu V, Lei J. Minimax sparse principal subspace estimation in high dimensions. arXiv preprint arXiv:1211.0373. 2012

41. Yuan M. High dimensional inverse covariancematrix estimation via linear programming. The Journal of Machine Learning Research. 2010; 99:2261–2286.