Sparse Equisigned PCA: Algorithms and Performance Bounds in the Noisy Rank-1 Setting

Arvind Prasadan and Raj Rao Nadakuditi

Department of EECS
University of Michigan
1301 Beal Avenue
Ann Arbor, MI 48109
e-mail: prasadan@umich.edu; rajnrao@umich.edu

Debashis Paul

Department of Statistics
University of California
One Shields Avenue
Davis, CA 95616 e-mail: debpaul@ucdavis.edu

Abstract: Singular value decomposition (SVD) based principal component analysis (PCA) breaks down in the high-dimensional and limited sample size regime below a certain critical eigen-SNR that depends on the dimensionality of the system and the number of samples. Below this critical eigen-SNR, the estimates returned by the SVD are asymptotically uncorrelated with the latent principal components. We consider a setting where the left singular vector of the underlying rank one signal matrix is assumed to be sparse and the right singular vector is assumed to be equisigned, that is, having either only nonnegative or only nonpositive entries. We consider six different algorithms for estimating the sparse principal component based on different statistical criteria and prove that by exploiting sparsity, we recover consistent estimates in the low eigen-SNR regime where the SVD fails. Our analysis reveals conditions under which a coordinate selection scheme based on a sum-type decision statistic outperforms schemes that utilize the $\ell_1$ and $\ell_2$ norm-based statistics. We derive lower bounds on the size of detectable coordinates of the principal left singular vector and utilize these lower bounds to derive lower bounds on the worst-case risk. Finally, we verify our findings with numerical simulations and illustrate the performance with a video data example, where the interest is in identifying objects.

MSC 2010 subject classifications: Primary 65F50, 62H25, 62H15; secondary 62F03, 60G35, 15A18.

Keywords and phrases: Sparse PCA, Random Matrices, FDR, Sparsity.

Contents

1 Introduction .................................................. 2

*This work was partially supported by the NSF and the ONR.
## 1. Introduction

It is well-understood that singular value decomposition (SVD) based principal component analysis (PCA) breaks down in the high-dimensional and limited sample size regime below a certain critical eigen-SNR (eigenvalue signal-to-noise ratio) threshold. In this paper, we propose a new method called Sparse Equisigned PCA (SEPCA) that addresses this issue by using a sparsity assumption on the principal components. The method is based on a novel formulation of the PCA problem that takes into account the sign of the eigenvalues, which can be used to improve the performance of the PCA in high-dimensional settings. The proposed algorithms are designed to control the False Discovery Rate (FDR) and to estimate the smallest detectable coordinate. Simulations and theoretical results are presented to demonstrate the effectiveness of the proposed methods.
ratio) that depends on the dimensionality of the system and the number of samples [17, 4]. Several sparse PCA algorithms have been proposed in the literature (see [17, 4, 8, 21, 33, 3]) and have been shown to successfully estimate the principal components in the low eigen-SNR regime where the SVD fails.

Prior work in this area primarily considers the Gaussian signal-plus-noise model with random effects, where the signal matrix is assumed to have sparse left singular vectors, normally distributed right singular vectors, and the noise matrix is assumed to have normally distributed \( i.i.d. \) entries. Here, we consider the setting where the left singular vector of the rank one signal matrix is sparse \textit{and} the right singular vector is assumed to be equisigned. We say that a vector is equisigned if its entries are all non-negative or all non-positive. This is motivated by applications such as diffusion imaging in MRI where the right singular vector represents a physical quantity (e.g. intensity as the diffusion agent is absorbed by a tissue) that is non-negative, by imaging problems such as foreground-background separation in video data [25, 31] and object detection in astronomy [27], where the data are naturally non-negative, and by problems in bioinformatics where the data are (non-negative) counts of genes [30]. When analyzing data that are non-negative, it is logical to take advantage of this property, and investigate how we may use this knowledge to do better than the (generic) alternatives. Additionally, we motivate the rank-1 assumption by noting that for a video with a static background, the foreground is a perturbation of a rank-1 background [22, 12]. Finally, even though we do not pursue this angle here, our framework can be extended to deal with the scenario where the signal can be viewed of a rank 1 tensor with all but one of the representors in the Kroneker product representation of the tensor is an equisigned vector.

A natural question at this juncture is the following: how does our problem differ from that solved by Non-Negative Matrix Factorization (NNMF)? NNMF takes a given matrix \( X \) and looks for non-negative matrices \( F \) and \( G \) such that \( X = F G^T \) [15, 32]. Ordinary NNMF has no sparsity constraints. We might impose such constraints, as is done in [14] and [20], but except in special cases, these solutions have no known theoretical guarantee of statistical performance. This problem partly stems from the fact that solutions to the corresponding optimization problems may not be unique. In contrast, our problem only constrains the right singular vectors, while the left singular vectors are free to take any sign. The work in [9] extends the NNMF framework to one wherein only one of the factors is non-negative; nevertheless, the rest of the constraints we impose are not included. Hence, NNMF is not an answer to the problem we consider herein.

The main contribution of this paper is a rigorous sparsistency analysis of the various algorithms that brings into focus the various very-low eigen-SNR regimes where the new algorithms work and the SVD based methods provably fail. Additionally, a major novelty of this work is the integration of FDR-controlling (False Discovery Rate) hypothesis testing to the Sparse PCA problem.

Our analysis illustrates the situations where the sum based coordinate selection scheme dramatically outperforms the \( \ell_1 \) and \( \ell_2 \) [17, 4] based sparse PCA schemes. Additionally, our proposed algorithms are non-iterative, do not require
the computation of the sample covariance matrix, and do not require knowledge of the sparsity level. We separate our algorithms into two groups: one where the Family-Wise Error Rate (FWER) is controlled, and another where the False Discovery Rate (FDR) is controlled. We utilize sharp tail probability bounds for relevant statistics to derive our FWER-controlling estimators [6]. For the FDR controlling estimators, we relate the problem at hand to that of the sparse normal means problem [10].

This paper is organized as follows. In Section 3, we describe three algorithms for estimating the sparse principal component that utilize a coordinate selection scheme based on the sum, \( \ell_1 \), and \( \ell_2 \) norm-based statistics respectively. We call our family of algorithms \( \text{SEPCA} \), an abbreviation for Sparse Equisigned PCA. Section 4 proposes three FDR-controlling refinements of the sum- and \( \ell_2 \)-based algorithms in Section 3 by relating coordinate detection to the sparse normal means estimation problem. In Section 5 we show how the estimation performance is governed by the size of the smallest detectable coordinate, which we analyze in Section 6 and validate using numerical simulations in Section 7. In Section 8, we provide some geometric intuitions about the relative performance of three of our algorithms. We show that the sum statistic is potentially the most powerful, while the \( \ell_1 \) is the least powerful. We provide some concluding remarks in Section 9.

2. Problem Formulation

Let \( \mathbf{X} \in \mathbb{R}^{p \times n} \) be a real-valued signal-plus-noise data matrix of the form

\[
\mathbf{X} = \theta \mathbf{u} \mathbf{v}^T + \sigma \mathbf{G}.
\]  

(2.1)

The columns of the \( p \times n \) data matrix \( \mathbf{X} \) represent \( p \)-dimensional observations. In (2.1), \( \mathbf{u} \) and \( \mathbf{v} \) are the left and right singular vectors of the rank-one latent signal matrix, and have entries \( u_i \) and \( v_j \), respectively. The entries of \( \mathbf{G} \), the noise matrix, are assumed to be \( i.i.d. \) Gaussian random variables with mean 0 and variance \( 1/n \). We assume that \( \mathbf{u} \in \mathbb{R}^p \) has unit norm and is sparse in the sense of small \( \ell_0 \) norm, with \( s \ll p \) non-zero entries, where \( s/n \to 0 \). That is, for a set \( I = \{i_1, \cdots, i_s\} \subset \{1, \cdots, p\} \),

\[
\begin{align*}
u_i &\neq 0 \quad \text{for } i \in I, \\
u_i &= 0 \quad \text{for } i \in I^C,
\end{align*}
\]  

(2.2)

where \( I^C \) denotes the complement of \( I \). We further assume \( \mathbf{v} \in \mathbb{R}^n \) to be of unit norm, deterministic, and equisigned. Given \( \mathbf{X} \), our goal is to recover \( \mathbf{u} \) and \( \mathbf{v} \).

Note that the \((i, k)\) entry of \( \mathbf{X} \), \( X_{ik} \), is a Gaussian random variable with mean \( \theta u_i v_k \) and variance \( \sigma^2/n \). Moreover, it follows that

\[
\mathbb{E}(\mathbf{X} \mathbf{X}^T) = \theta^2 \mathbf{u} \mathbf{u}^T + \sigma^2 \mathbf{I}_p,
\]

where \( \mathbf{I}_p \) denotes the \( p \times p \) identity matrix. The quantity \( (\theta/\sigma)^2 \) is, for this model, the eigen-SNR (signal-to-noise ratio).
2.1. Motivation: Breakdown of PCA / SVD

From [2], we have the following result: let \( \hat{\mathbf{u}} \) be the estimate of \( \mathbf{u} \) given by the Singular Value Decomposition (SVD) of \( \mathbf{X} \), and let \( p(n)/n \) have limit \( c \in [0, \infty] \) as \( n \) grows, with \( \theta \) fixed and \( \sigma = 1 \). Then, with probability 1,

\[
|\langle \hat{\mathbf{u}}, \mathbf{u} \rangle|_2 \to \begin{cases} 
1 - \frac{c(1+\theta^2)}{\sigma^2(c+\theta^2)} & \text{if } \theta \geq c^{1/4}, \\
0 & \text{otherwise.}
\end{cases}
\]  

(2.3)

For general \( \sigma \), we replace \( \theta \) by \( \theta/\sigma \) in (2.3). Hence, SVD based PCA leads to inconsistent estimates of \( \mathbf{u} \) (and also for \( \mathbf{v} \), which can be deduced from (2.3)) when the dimension \( p \) is comparable to or larger than the sample size \( n \). Moreover, in the low eigen-SNR regime, the estimates break down completely. SVD does not exploit any assumed structure in \( \mathbf{u} \) and \( \mathbf{v} \). Consequently, (2.3) holds for arbitrary \( \mathbf{u} \) and \( \mathbf{v} \), including our setting where \( \mathbf{u} \) is sparse and/or \( \mathbf{v} \) is equisigned. Our goal, in what follows, is to derive consistent estimators for \( \mathbf{u} \) and \( \mathbf{v} \) that outperform the SVD by exploiting the sparsity of \( \mathbf{u} \) and the equisigned nature of \( \mathbf{v} \).

3. Proposed Algorithms

We propose six different two-stage algorithms for estimating \( \mathbf{u} \). The first three algorithms are designed to control the family-wise error rate (FWER), or, the probability of obtaining a false positive in the coordinate selection. The last three algorithms aim to control the false discovery rate (FDR), or, the proportion of false discoveries (coordinate detections) among all discoveries. We defer discussion of the FDR-based algorithms to Section 4.

All of the algorithms have the same basic form given in Algorithm 1. Given \( \mathbf{X} \), we associate a test statistic \( T_i \) to each row of \( \mathbf{X} \). The sparsity of \( \mathbf{u} \) implies that the majority of the rows of \( \mathbf{X} \) are purely noise, so that the majority of the \( T_i \) come from the null, noise-only distribution. Hence, based on the statistics \( \{T_i\} \), we perform a form of multiple hypotheses testing procedure, and select the set \( \hat{I} \) of indices that are non-null. In this way, we can estimate the support of \( \mathbf{u} \), thereby isolating the the rows of \( \mathbf{X} \) that contain the signal. Then, taking the SVD of this submatrix (comprised of only the selected rows of \( \mathbf{X} \)) yields a better estimate of the non-zero coordinates in \( \mathbf{u} \), as well as \( \mathbf{v} \).

We begin by discussing the FWER-controlling algorithms. The work in [17] proposed a covariance thresholding method for Sparse PCA called DT-SPCA; this is equivalent to a coordinate selection scheme based on the \( \ell_2 \) norm-based statistic. In our terminology and with our choice of thresholds, we label it as \( \ell_2\text{-SEPCA} \). We label the coordinate selection scheme based on the \( \ell_1 \) norm-based statistic \( \ell_1\text{-SEPCA} \). Finally, the \( \text{sum-SEPCA} \) algorithm utilizes row sums of the data matrix.

We shall choose the thresholds \( \tau_{n,p} \) for the coordinate selection scheme so
Algorithm 1 Variable Selection and Estimation Algorithm

Require: Threshold \( \tau_{n,p} \) and form of Test Statistic \( T_i \) from Table 1

Let \( \hat{I} \) be an empty list

for all Rows \( i \) of \( X \), \( 1 \leq i \leq p \) do

Form test statistic \( T_i \) from row \( i \) of \( X \)

if \( T_i \geq \tau_{n,p} \) then

Add \( i \) to \( \hat{I} \)

end if

end for

Let \( \tilde{u}, \tilde{\theta}, \hat{v} \) = SVD(\( X_{\hat{I}} \)) be the rank-1 SVD of \( X \) restricted to rows in \( \hat{I} = [i_1, \ldots, i_{|\hat{I}|}] \)

For \( i_k \in \hat{I} \), let \( \hat{u}_{i_k} = \tilde{u}_{i_k} \); the other entries of \( \hat{u} \) are set to 0.

that in the noise-only case,

\[
P \left( \max_{1 \leq i \leq p} T_i \geq \tau_{n,p} \right) \leq \frac{1}{ep} \to 0, \quad (3.1)
\]

where \( e \) is Euler’s number, or the base of the natural logarithm. This choice ensures that the probability of a false positive tends to zero as \( p \to \infty \). That is, the FWER is asymptotically zero and is bounded by \( 1/ep \) in the finite-dimensional case. Note that the constraint used to control the FWER is simply that the distribution of the noise is log-concave. In the Gaussian case, we obtain the specific expressions given summarized in Table 1; however, with knowledge of the moments \( E T_i \) and Var \( T_i \), we can repeat our analysis and find thresholds for the \( \ell_1 \) and \( \ell_2 \)-SEPCA algorithms with any log-concave noise distribution. The thresholds are summarized in Table 1.

Table 1: Test Statistics and Thresholds for Algorithm (1)

| Algorithm | Statistic \( T_i \) | Threshold \( \tau_{n,p} \) |
|-----------|-------------------|-----------------|
| \( \ell_1 \)-SEPCA | \( \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |X_{i,k}| \) | \( \sigma \left( \frac{\sqrt{2}}{\pi} + C_1 \frac{\log ep}{\sqrt{n}} \right) \) |
| \( \ell_2 \)-SEPCA | \( \sum_{k=1}^{n} X_{i,k}^2 \) | \( \sigma^2 \left( 1 + C_2 \frac{\log ep}{\sqrt{n}} \right) \) |
| sum-SEPCA | \( \frac{1}{\sqrt{n}} \left| \sum_{k=1}^{n} X_{i,k} \right| \) | \( \sigma C_U \sqrt{\frac{\log p}{n}} \) |

See (3.2) and (3.6) for definitions of the constants \( C_2 \), \( C_1 \), and \( C_U \).

3.1. FWER Thresholds

3.1.1. \( \ell_2 \)- and \( \ell_1 \)-SEPCA

In the noise-only cases, the statistics for \( \ell_2 \)- and \( \ell_1 \)-SEPCA are distributed as scaled \( \chi_n^2 \) and sums of half-normal, respectively. Both of these quantities are log-concave random variables, so we may apply the result in [19] to set the threshold \( \tau_{n,p} \) in both cases.
Defining $K$ to be some absolute constant (we may use $K = e$, as in [5]), we define the constants

$$C_2 = \sqrt{2}K \text{ and } C_1 = K\sqrt{(1 - 2/\pi)}.$$  \hfill (3.2)

### 3.1.2. Sum-SEPCA

From Proposition 4.4 of [7], we obtain that the threshold for sum-SEPCA is given by

$$\tau_{n,p} = \frac{\sigma}{\sqrt{n}} \left( \sqrt{2\log p} + \frac{1}{U(p)} \left( \frac{1}{3} \log ep + \sqrt{\log ep} \right) + \delta_p \right).$$  \hfill (3.3)

In (3.3), we have that

$$U(p) = \sqrt{2} \text{ Erf}^{-1} \left( 1 - \frac{1}{p} \right) \text{ and } \delta_p \approx \frac{\pi^2}{12} (\log p)^{-3/2},$$  \hfill (3.4)

where Erf denotes the error function, or alternatively, the cumulative distribution function of a standard Gaussian random variable is given by

$$\Phi(x) = \frac{1}{2} \left( 1 + \text{Erf} \left( \frac{x}{\sqrt{2}} \right) \right).$$  \hfill (3.5)

Moreover, $\tau_{n,p} \leq \sigma C_U \sqrt{\frac{\log p}{n}}$ for some constant $C_U$. For a fixed value of $p$, choosing

$$\kappa_U \geq \sqrt{\frac{2}{U(p)}} \left( 3 + \sqrt{\log p} \right) > 1 \text{ and } C_U = \sqrt{2} + \frac{\kappa_U}{3\sqrt{2}}$$  \hfill (3.6)

is sufficient. The choice of $1/ep$ is the largest bound justified by Proposition 4.4 of [7], so we have calibrated all of our algorithms to the same constant factor times $1/p$. The thresholds are summarized in Table 1.

### 4. Controlling the False Discovery Rate

So far, we have controlled the probability of a false alarms when detecting coordinates. However, there are two relevant observations to make. First, under the Gaussian noise, rank-1, and equisigned assumptions, the vector of test statistics $\{T_i\}$ in the sum-SEPCA algorithm looks like a sparse vector plus Gaussian noise (or a vector of $\chi^2_n$-variates with varying non-centralities, in the $\ell_2$-SEPCA algorithm). Secondly, controlling the false discovery rate, that is, the proportion of rejected nulls that are false positives, can lead to increased detection power relative to controlling the false positive rate. We hence look at FDR-controlling tests for the Sparse Normal Means problem.

That is, given a vector of test statistics (as before), we replace the thresholding and selection in Algorithm 1 with an FDR-controlling selection procedure.
We summarize this change in Algorithm 2. There are three procedures we consider. The first two are known as Higher Criticism, and directly extend the sum- and $\ell_2$-SEPCA algorithms [10, 11]. The third is a method for detection in the sparse normal means problem that comes out of complexity-penalized estimation theory for linear inverse problems [18].

**Algorithm 2**

FDR-Controlling Variable Selection and Estimation Algorithm

**Require:** Test Statistic $T_i$ from Table 1 and Selection Procedure

Let $\hat{I}$ be an empty list

for all Rows $i$ of $X$, $1 \leq i \leq p$ do

Form test statistic $T_i$ from row $i$ of $X$

end for

Perform an FDR-Controlling selection procedure, and add the selected indices to $\hat{I}$

Let $[\tilde{u}, \tilde{\theta}, \hat{v}] = \text{SVD}(X_{\hat{I}}, :)$ be the rank-1 SVD of $X$ restricted to rows in $\hat{I} = [i_1, \ldots, i_{\hat{I}}]$

For $i_k \in \hat{I}$, let $\hat{u}_{i_k} = \tilde{u}_{i_k}$; the other entries of $\hat{u}$ are set to 0.

### 4.1. Higher Criticism

Assume we have $p$ independent tests of the form

\begin{align*}
H_{0,i} : \quad & W_i \sim \mathcal{N}(0, 1), \\
H_{1,i} : \quad & W_i \sim \mathcal{N}(\mu_i, 1),
\end{align*}

and assume that at most $p^{1-\beta}$ of the $p$ hypotheses are truly non-null, for some $\beta \in (1/2, 1)$. Further assume that the non-null means have magnitude

$$\mu_i = \mu_p = \sqrt{2r \log p},$$

for $r \in (0, 1)$. Here, the means will correspond to the coordinate size. Note that the expected maximum of $p$ standard Gaussian random variables is upper bounded by $\sqrt{2 \log p}$, with the bound being asymptotically sharp.

If we let $p(1) \leq p(2) \leq \cdots \leq p(p)$ be the sorted p-values of the individual tests, we may define the Higher Criticism statistic:

$$HC_p = \max_{i : 1/p \leq p(i) \leq 1/2} \sqrt{p(i/p - p(i))} \sqrt{p(i)(1 - p(i))}.$$  \hfill (4.2)

Rejecting the global null hypothesis (that there are no non-null coordinates) when $HC_p > \sqrt{2 \log \log p(1 + o(1))}$ leads to asymptotically full power when $r$ is greater than some decision boundary $\rho$, and that under the global null,

$$\frac{HC_p}{\sqrt{2 \log \log p}} \to 1$$

in probability as $n, p \to \infty$. The function $\rho$ depends on the sparsity index $\beta$, and as [11] indicate:

$$\rho(\beta) = \begin{cases} 
\beta - 1/2 & \text{when } \beta \in (1/2, 3/4), \\
(1 - \sqrt{1 - \beta})^2 & \text{when } \beta \in (3/4, 1). 
\end{cases} \hfill (4.4)$$
If we replace the normal distribution with a $\chi^2_n$ distribution, the same results hold for tests of the form

$$
H_{0,i} : \quad W_i \sim \chi^2_n,
H_{1,i} : \quad W_i \sim \chi^2_n(\delta),
$$

(4.5)

where $\delta$ is a non-centrality parameter and we consider $r \in (0, 1)$ such that $\delta = 2r \log p$. That is to say, we form the Higher Criticism statistic in the same manner as for the sum statistic, perform the test with the same threshold, and the form of the decision boundary $\rho$ is identical [11].

To summarize, taking sums across the rows of $X$, we obtain a vector $y$ where

$$
y_i = \mu_i + \sigma z_i,
$$

with $\mu_i = (\theta u_i) \| v \|_1$; this situation is exactly that of a sparse mean vector embedded in Gaussian noise. Similarly, taking sums of squares across the rows of $X$ yields scaled $\chi^2_n$ distributed random variables, of which only a few have non-zero non-centrality parameters.

As a point of interest, the test in (4.1) can be extended to (and potentially strengthened in) the case where the $p$ tests are correlated, i.e., when the additive Gaussian noise has a non-identity covariance [13].

**Remark** While Higher Criticism is typically formulated for the case of identical non-null means or parameters (all of the non-zero $\mu_i$ are identical), this constraint is not mandatory [1, 13]. Indeed, the results hold without modification for the Gaussian model with non-null means of size $\mu_i = \alpha_i \sqrt{2 \log p}$, where $\alpha_i$ is a non-negative random variable with the property that $P(\alpha_i \leq \sqrt{r}) = 1$ and $P(\alpha_i > \sqrt{r} - \epsilon) > 0$ for all $\epsilon > 0$ [13]. The case of a $\chi^2_n$ distribution is similar.

### 4.2. FDR-SEPCA

In this section, we give an summary of the algorithm for uncorrelated noise and defer the general case and details to Appendix D. We continue in the same vein as in the previous section on Higher Criticism.

We note that in the equisigned, rank-1 setting, coordinate selection is equivalent to the estimation of a sparse mean vector. Let $y_i = \mu_i + \sigma z_i$, where $i \in \{1, \ldots, p\}$ and the vector $z$ of the $z_i$ is normally distributed with mean 0 and covariance $I_p$. The mean vector $\mu$ of the $\mu_i$ is assumed to be sparse; the goal is to estimate $\mu$. Taking sums across the rows of $X$, we obtain a vector $y$ where $y_i = \mu_i + \sigma z_i$, with $\mu_i = (\theta u_i) \| v \|_1$. Hence, we are in the same setting as in the previous section.

The following penalized least squares formulation, taken from [18], yields an estimator for $\mu$:

$$
\hat{\mu} = \arg \min_{\mu} \| y - \mu \|_2^2 + \sigma^2 \text{pen}(\| \mu \|_0),
$$

(4.6)

where pen($k$) is defined as

$$
\text{pen}(k) = \zeta k \left( 1 + \sqrt{2 \log(p/k)} \right)^2,
$$

(4.7)
with $\zeta > 1$; we may take $\zeta = 1 + o(1)$. The parameter $\nu$ is no smaller than $e$.

We define $\|\mu\|_0$ to be the number of non-zero coordinates of $\mu$.

The solution to (4.6) is given by hard-thresholding. Let $|y_{(i)}|$ be the $i^{th}$ order statistic of $|y_i|$, namely $|y_{(1)}| \geq \cdots \geq |y_{(p)}|$. Then if

$$\hat{k} = \arg\min_{k \geq 0} \sum_{i > k} |y_{(i)}|^2 + \sigma^2 \text{pen}(k),$$

defining

$$t_k^2 = \text{pen}(k) - \text{pen}(k - 1),$$

the solution is to hard threshold at $t_{\hat{k}}$.

In this set-up, we have that $t_k \approx \sqrt{\zeta} (1 + \sqrt{2 \log(vp/k)})$.

We provide a precise quantification of $t_k$ in Appendix D.

Hence, by computing $t_k$ and performing hard thresholding of the row sums, we can perform coordinate selection. Once again, this procedure replaces the test statistic/thresholding in Algorithm 1.

5. Estimation Error and Smallest Detectable Coordinate

As we will see, our theorems discuss the “detectability” of the coordinates $u_i$ of $u$. However, it is common in the sparse PCA literature to discuss lower bounds for the risk (estimation error) [17, 4, 21]. In what follows, we will show that these two notions are equivalent.

We define the $L^2$ estimation error for a principal component estimator as

$$L(\hat{u}, u) = \|u - \text{sign}(\langle u, \hat{u} \rangle)\hat{u}\|_2^2.$$  (5.1)

The quantity in (5.1) is upper bounded by 2; this bound is attained when $u$ and $\hat{u}$ are unit norm and mutually orthogonal. Following [4], we want to compute a lower bound for the maximum expected loss for the $s$-sparse vectors $u$ (in the sense of $\ell_0$ sparsity) defined as

$$\sup_{u \in \mathbb{S}^{p-1} : \|u\|_0 \leq s} \mathbb{E} L(\hat{u}, u),$$  (5.2)

where $\mathbb{S}^{p-1}$ denotes the unit sphere in $\mathbb{R}^p$. Let $\hat{I}$ be some index set of coordinates selected by an algorithm of the form given in Algorithm (1). We may take $\langle u, \hat{u} \rangle$ to be non-negative, and decompose the loss as

$$\|u - \hat{u}\|^2 = \|u_{\hat{I}} - \hat{u}\|^2 + \|u_{\hat{I}^c}\|^2 \geq \|u_{\hat{I}^c}\|^2.$$  (5.3)
Equation (5.3) shows that the loss is lower-bounded by the squared sum of the missed coordinates. Indeed, it is a natural consequence of the result in [2] that if the sparsity $s$ grows slower than does $n$, and we have a consistent estimate of the support of $u$, the estimation error will asymptotically be small. Essentially, we are estimating the singular vectors of an $s \times n$ matrix instead of a $p \times n$ matrix, so that if the ratio $s/n$ has limit zero, our estimates will be consistent (see (2.3) and [2]). This suggests the following strategy for lower-bounding (5.2): we want to construct a non-trivial ‘worst-case’ sparse vector. That is, we want a vector $u$ that has a non-trivial loss (less than 2), is sparse (fewer than $s$ non-zero coordinates), and has maximal error from missed coordinates. To ensure a non-trivial loss, we set the first coordinate $u_1$ to be large, i.e., $u_1 = \sqrt{1 - r^2}$, where $r = o(1)$. To ensure sparsity, we set $u_2, \ldots, u_{m+1}$ to be non-zero for some $m \leq s - 1$, with the subsequent coordinates of $u$ set to 0. Then, the expected loss has the lower bound

$$E L(u, \hat{u}) \geq \sum_{k=1}^{p} |u_k|^2 P(\text{Not Selecting Coordinate } k)$$

$$\geq \sum_{k=2}^{m+1} |u_k|^2 P(\text{Not Selecting Coordinate } k),$$

(5.4)

since $u_1$ is detected with probability approaching 1 and $u_k$ is zero for $k > m+1$. Now, let $u_2$ through $u_{m+1}$ all have value $r/\sqrt{m}$, so that we may simplify the lower bound to

$$E L(u, \hat{u}) \geq r^2 P(\text{Not Selecting Coordinate } k).$$

(5.5)

If coordinates of size $r/\sqrt{m}$ are not detected with a probability approaching 1, $r^2$ is a lower-bound on the risk. This construction shows that specifying the sizes of coordinates that are not detected with probability approaching 1 is equivalent to specifying a worst-case risk lower bound. Consequently, in what follows we focus on the smallest detectable and largest undetectable coordinates because they directly shed light on the attainable estimation error. The details of the risk calculations and extensions to approximate sparsity are deferred to Appendix C, where we summarize our findings in Theorem 3.

6. Main Results

The following theorem characterizes consistent support recovery conditions. These results are the analogue of the ‘sparsistency’ guarantees found in the LASSO and $\ell_1$-norm minimization literature [26]. Throughout, $\hat{I}$ denotes the set of coordinates selected by the coordinate selection scheme.

**Theorem 1.** For the model specified in (2.1) and (2.2) and the algorithms specified in Table 1, assume that $p(n), n \to \infty$, $s(n)/n \to 0$, and $\log p(n) = o(n)$. Let $\epsilon \in (0, 1)$. We have that
a. For $i \in I^c$, 
\[ \max_{i \in I^c} \mathbb{P}\left(i \in \hat{I}\right) \to 0, \]

b. For $i \in I$, 
\[ \min_{i \in I : |\theta_{ui}| > \beta_{crit}(1+\epsilon)} \mathbb{P}\left(i \in \hat{I}\right) \to 1, \]
\[ \max_{i \in I : |\theta_{ui}| < \beta_{crit}(1-\epsilon)} \mathbb{P}\left(i \in \hat{I}\right) \to 0. \]

Here
\[ \beta_{crit} = \begin{cases} 
\sigma C_U \frac{\log p}{\sqrt{\sum_k v_k}} & \text{for sum-SEPCA,} \\
\sigma \sqrt{C_2} \frac{\log p}{\sqrt{n}} & \text{for } \ell_2\text{-SEPCA,} \\
\sigma t_{\ell_1} & \text{for } \ell_1\text{-SEPCA,} 
\end{cases} \] (6.1)

and $t_{\ell_1}$ satisfies the relation
\[ \left(\sqrt{\frac{2}{\pi}} + C_1 \frac{\log p}{\sqrt{n}}\right) = \frac{1}{n} \sqrt{\frac{2}{\pi}} \sum_k \exp \left(- \left(\sqrt{n} \frac{(t_{\ell_1}) v_k}{\sqrt{2}}\right)^2\right) + \sqrt{\pi} \sum_k \left(\frac{(t_{\ell_1}) v_k}{\sqrt{2}}\right) \text{Erf} \left(\sqrt{n} \frac{(t_{\ell_1}) v_k}{\sqrt{2}}\right). \]

We defer the proof to Appendix A.

Theorem 1 identifies a phase transition in the ability of the algorithms to accurately estimate the support of $u$. Note that the analysis brings into sharp focus the dependence of $\beta_{crit}$ on $v$ for the $\ell_1$- and sum-SEPCA algorithms, but not the $\ell_2$-SEPCA algorithm. Consequently, we can expect the algorithms to perform differently depending on the structure of the underlying $v$. It is important to note that the sparsity $s$ of $u$ is not a parameter in the thresholds and results.

It is also important to note that $\ell_2$-SEPCA and $\ell_1$-SEPCA do not rely on the equisigned character of $v$. However, it is clear that the sum-SEPCA algorithm explicitly depends on the equisigned assumption.

### 6.1. FDR-Based Algorithms

We may summarize the coordinate selection properties of the FDR refinements as follows:

**Theorem 2.** For the model specified in (2.1) and (2.2) and the three FDR-controlling algorithms summarized in Algorithm 2, assume that $p(n), n \to \infty$, $s(n)/n \to 0$, and $\log p(n) = o(n)$. Let $\epsilon \in (0, 1)$. We have that

a. For all three algorithms and $i \in I^c$, 
\[ \max_{i \in I^c} \mathbb{P}\left(i \in \hat{I}\right) \to 0, \]
b. For the Higher Criticism-based algorithms and \( i \in I \),
\[
\min_{i \in I : |\theta_u| > \beta_{\text{crit}}(1+\epsilon)} P(i \in \hat{I}) \to 1,
\]
\[
\max_{i \in I : |\theta_u| < \beta_{\text{crit}}(1-\epsilon)} P(i \in \hat{I}) \to 0.
\]

c. For the FDR-SEPCA algorithm, uniformly over \( i \in I \),
\[
\begin{align*}
\text{if } |\theta_u| > \beta_{\text{crit}}(1+\epsilon), & \text{ coordinate } i \text{ is selected;} \\
\text{if } |\theta_u| < \beta_{\text{crit}}(1-\epsilon), & \text{ coordinate } i \text{ is not selected}
\end{align*}
\]
with probability tending to 1.

Here
\[
\beta_{\text{crit}} = \begin{cases} 
\sigma \sqrt{\rho(\beta)} \frac{2 \log p}{\|v\|_1} & \text{for HC-sum-SEPCA,} \\
\sigma \rho(\beta) \frac{2 \log p}{\sqrt{n}} & \text{for HC-\ell_2-SEPCA,} \\
\sigma (1 - o(1)) \sqrt{\zeta^{1+\sqrt{2 \log (vp/k) / \|v\|_1}}} & \text{for FDR-SEPCA,}
\end{cases}
\]
where \( \zeta > 1 \), \( \nu > e \), and the FDR-SEPCA algorithm detects \( \hat{k} \) coordinates.

We defer the proof to Appendix B.

Once again, we see that the structure of the underlying \( v \) plays a role in the performance of the sum-based algorithms, but not for the \( \ell_2 \)-based HC-\( \ell_2 \)-SEPCA algorithm. Unlike in the FWER-controlling cases, the sparsity of \( u \) plays a (small) role here, via the constant \( \rho(\beta) \) for the Higher Criticism-based methods and via \( \hat{k} \) for FDR-SEPCA. Moreover, \( \ell_2 \)-HC-SEPCA, like \( \ell_2 \)-SEPCA, does not make use of the equisigned nature of \( v \).

7. Simulations

To illustrate the relative powers of the six algorithms, we compute the theoretical limits on the sizes of detectable coordinates as a function of \( n \). We use a unit-norm, equisigned \( v \) such that
\[
v_k \propto \exp\left(-5 \frac{k}{n}\right) \sin\left(4 \frac{k}{n}\right) \quad \text{for } 1 \leq k \leq n.
\]

This choice of \( v \) has a ‘rise and fall’ sort of behavior, and is motivated by physical signals, e.g., chemical reactions or nerve signals in the brain. The value of \( \beta_{\text{crit}} \) is shown in Figure 1; for this choice of \( v \), it is clear that the sum-SEPCA dramatically outperforms the other SEPCA variants in terms of size of the smallest detectable component. The FDR-SEPCA algorithm has similar performance to sum-SEPCA, and the HC-sum-SEPCA algorithm has the strongest performance.

In Figure 2, we plot the estimation error as a function of \( n \) and \( \theta \) for all six algorithms. We also include results for the SVD and competing algorithms.
Fig 1: This plot shows $\beta_{crit}$ for all six algorithms for the $v$ described in (7.1).

TPower [33] and ITSPCA [21]. In the simulations, we fix $p = 1000$ and vary $n$, since the dependence in $p$ in the thresholds is logarithmic, whereas that in $n$ is not. The left singular vector $u$ is chosen to be the vector with 1 in the first coordinate and 0 elsewhere. We fix the noise variance $\sigma^2$ at 1, so that $\theta^2$ is the eigen-SNR. The results should be interpreted as follows. For the particular $v$ chosen here, we expect HC-sum-SEPCA to have the lowest detectable limit, and $\ell_1$-SEPCA to have the largest. This behavior is confirmed. Moreover, the sum-based algorithms offer a slight strengthening of both ITSPCA and TPower.

7.1. Comments on the FDR-controlling procedures

The Higher Criticism for the $\chi^2_n$-variates ‘pushes back’ the phase transition between detecting nothing and something to a lower value of $\theta$ relative to the $\ell_2$-SEPCA algorithm, but is still less powerful than any of the sum-based algorithms. Moreover, even above the phase transition, the $\ell_2$-SEPCA algorithm may be preferable, as the error is increased by unacceptably many false positives.

The Higher Criticism procedure for the sum statistic has the lowest phase transition point and hence the highest power. Its transition is more gradual than the penalized FDR thresholding procedure and sum-SEPCA, which have roughly the same performance in this simulation.
7.2. An example where $\ell_2$-based algorithms outperform sum-based algorithms

Sum-SEPCA has a $\beta_{\text{crit}}$ that depends on $v$. Looking at the form in (6.1), if $\|v\|_1$ is smaller than $n^{1/4}$, we would expect $\ell_2$-SEPCA to detect a smaller coordinate size. Vectors with smaller coordinates have a smaller $\ell_1$-norm, i.e., one that is closer to their $\ell_2$-norm. Hence, if we choose

$$v_k \propto \frac{1}{k^2} \text{ for } 1 \leq k \leq n,$$

(7.2)

we expect sum-SEPCA to have worse performance relative to $\ell_2$-SEPCA. Figures 3 and 4 confirm this expectation. The FDR refinements perform poorly. It should be noted, however, that TPower and ITSPCA retain their performance. This choice of $v$ effectively corresponds to a very small value of $n$: the majority of coordinates are tiny in size and buried beneath noise regardless of the value of $\theta$. If we ‘corrected’ the scenario and used a smaller $n$ and a subset of $v$, we would be in a situation closer to that given in (7.1).

7.3. A video data example

We conclude our sequence of examples with a real data study. This example is motivated by the problem of foreground-background separation in videos. Consider a grayscale video of stars twinkling against a black background [28].
Our goal is to estimate the locations of the stars: by reshaping the video, we may treat each frame as a vector and hence treat the video as a sparse matrix. Only a few locations have a star and are hence non-zero. The scale of the video pixels is between 0 and 255. We examine the top-left 72 × 64 pixels for 89 frames, as shown in Figure 2a. In Figure 2b, we plot the singular values of the video matrix. The first singular value stands out strongly against the rest, and at most two more singular values are well-separated from the bulk. This structure suggests that our rank-1 based approach is well suited to this problem.

We add Gaussian noise of variance $\sigma^2$ and study the True Positive Rates (TPR) and False Discovery Rates (FDR) across all algorithms and across different values of $\sigma$. In Figure (5), we show the results of our simulations. In terms of the TPR, everything other than the SVD has a similar performance, while the test-statistic SEPCA-based algorithms enjoy the best performance in terms of the FDR. In Figure 6 we zoom in on the top-right three stars and show how the algorithms perform as noise increases. Here, we see that the behavior alluded to in the TPR/FDR results actually occurs in the video.

### 7.4. Estimation of the Noise Variance, $\sigma^2$

In general, estimation of $\sigma^2$ may not be straightforward [23]. However, in most applications, including the video example we consider, one can obtain a relatively sparse representation of the object in a multiscale basis such as a wavelet basis. Under such circumstances, under the assumed additive, isotropic noise model, we can easily obtain a consistent estimate of $\sigma^2$ by utilizing the inherent sparsity of the signal, especially in finer scales. This can be done, for example, by computing the variance of the wavelet coefficients in the finest scale [16].
Fig 4: The plots show the empirical estimation error for all six algorithms for the \( u \) and \( v \) described in (7.2). We include results from TPower, ITSPCA and the SVD for comparison.

One can obtain a more robust estimate by taking the median absolute deviation of the coefficients about their median and then by multiplying its square with a known scale factor (assuming normality). Alternatively, procedures such as those proposed in [23, 24, 29] could be employed.

8. A geometric view: which algorithm to use?

We have stated detectability results for each algorithm in Section 6 and provided a numerical verification and comparison in Section 7. In this section, we wish to analytically compare the algorithms. In particular, we have seen that the right singular vector \( v \) plays a critical role in the detectability and estimability of \( u \), and we will characterize this behavior carefully.

In this section, will use the following notational convenience: we absorb \((\theta u_i)\) into \( v \in \mathbb{R}^n \), and write the detectability of coordinates in terms of \( v \). That is, if \( v^T \) is a row of \( X \), we specify when that row is selected. Moreover, we take \( \sigma = 1 \) for simplicity.

There are two ‘classes’ of detectability: in terms of \( |v|_1 \) and in terms of \( |v|_2 \). The sum-, HC-sum, and FDR-SEPCA algorithms select a coordinate if \( \sum_k v_k = \|v\|_1 \) is large enough for a \( v \) in the orthant with all non-negative or all non-positive coordinates. Geometrically, the vector \( v \) is selected if it is ‘outside’ a hyperplane with a normal vector proportional to the vector of all 1s. The \( \ell_1 \)-SEPCA algorithm is similar, as it selects a coordinate when \( \|v\|_1 \) is large enough, or if \( v \) lies outside an \( \ell_1 \)-ball of some radius. The connection between
(a) The image shows the mean intensity of pixels from the top-left $72 \times 64$ pixels for 89 frames. White indicates the presence of a star.

(b) The plots show the singular values of the video data. The spacing suggests a low-rank-plus-noise structure.

Table 2
Video Example Figures

the previous three algorithms and $\ell_1$-SEPCA comes from noting that the faces of an $\ell_1$-ball are sections of hyperplanes with normal vectors proportional to a vector of $\pm 1$s. Finally, the $\ell_2$- and HC-$\ell_2$-SEPCA algorithms select a coordinate when $\|v\|_2$ is large enough. I.e., when $v$ lies outside some $\ell_2$-ball.

Our goal in this section is to derive comparisons between the six algorithms. Specifically, for a given vector $v$, which algorithm will have the greatest detection ability (we are, for the moment, only concerned with maximizing power)? Note that when $v$ has a large norm, it does not matter which algorithm is used. Questions only arise when $\|v\|_1$ or $\|v\|_2$ are relatively small and are close to the thresholds.

8.1. Intersection of a hyperplane and a hypersphere

We may think of the $\ell_1$ ball as a hyperplane when restricted to a single orthant. If a hypersphere of radius $r$ intersects a hyperplane with a normal vector proportional to the vector of all $\pm 1$s and minimum distance to the origin of $r - h$, a hyperspherical cap of height $h$ is formed; see Figure 7 for a simple illustration. Geometrically, a right triangle is formed, with hypotenuse $r$ and leg $r - h$. Hence, the angle between the center of the cap and the edge is:

$$\theta_{\text{lim}} = \cos^{-1} \left( \frac{r - h}{r} \right).$$  \hfill (8.1)
Fig 5: The left plot shows the True Positive Rate of the various algorithms as a function of the noise level $\sigma$. The right plot shows the False Discovery Rates.

It is sufficient to guarantee that

$$0 \leq \frac{r - h}{r} \leq 1$$

for the hyperspherical cap to exist. Moreover, a vector $v$ has a direction contained inside the cap when the angle between $v$ and the vector of $\pm 1$ in the orthant containing $v$ is smaller than $\theta_{\text{lim}}$. In other words, defining the angle for a vector $v$ as

$$\theta(v) = \cos^{-1} \frac{\|v\|_1}{\|v\|_2 \sqrt{n}}, \quad (8.2)$$

we need $\theta(v) \leq \theta_{\text{lim}}$. 


Fig 6: A zoomed-in view of the three top-right stars in the video example. White indicates a false negative (missed star), Red a false positive (a guessed pixel where there was nothing), and Blue a true positive (correctly identified pixel).

Fig 7: A spherical cap in $\mathbb{R}^2$
8.2. Comparison: $\ell_2$-based versus sum-based statistics

We begin with a summary of the performance of each individual algorithm in Table 3. We first compare $\ell_2$-SEPCA and then compare HC-$\ell_2$-SEPCA with sum-, HC-sum, FDR-SEPCA in Tables 4 and 5. In our comparisons, we consider when the hyperspherical cap exists and give the angle of the cap. These are routine calculations, so we omit the details. We also omit $\ell_1$-SEPCA from our comparisons, as we lack a closed-form expression for $t_{\ell_1}$.

Note that the existence of this cap is a proxy for the equivalent statement that there exist vectors for which the sum-based algorithms are more powerful than the $\ell_2$-based algorithm. While this existence is not the same as attributing uniformly greater power to the sum-based algorithms relative to the $\ell_2$-based algorithm, the cap not existing is equivalent to the $\ell_2$-based algorithm having uniformly greater power.

Essentially, we observe that for $n$ and $p$ sufficiently large, the cap will exist. Moreover, for $v$ that is sufficiently dense ($\|v\|_1$ is sufficiently large), $\theta_{lim}$ will lie inside the cap. Hence, in these situations we would prefer a sum-based algorithm over an $\ell_2$-based algorithm.

Table 3: A summary of the six algorithms.

| Algorithm | Threshold | Geometric Quantity |
|-----------|-----------|--------------------|
| sum       | $\|v\|_1 \geq C_U \sqrt{\log p}$ | $r - h = C_U \sqrt{\frac{\log p}{n}}$ |
| HC-sum    | $\|v\|_1 \geq \sqrt{2\rho(\beta) \log p}$ | $r - h = \sqrt{2\rho(\beta) \frac{\log p}{n}}$ |
| FDR       | $\|v\|_1 \geq (1 - o(1)) \frac{\sqrt{\zeta}}{1 + \sqrt{\log (\nu p/\hat{k})}}$ | $r - h = (1 - o(1)) \frac{\sqrt{\zeta}}{1 + \sqrt{2 \log (\nu p/\hat{k})}}$ |
| $\ell_1$  | $\|v\|_1 \geq t_{\ell_1}$ | $r - h = t_{\ell_1}$ |
| $\ell_2$  | $\|v\|_2 \geq \sqrt{C_2} \sqrt{\frac{\log p}{\sqrt{n}}}$ | $r = \sqrt{C_2} \sqrt{\frac{\log p}{\sqrt{n}}}$ |
| HC-$\ell_2$ | $\|v\|_2 \geq 2\rho(\beta) \frac{\log p}{\sqrt{n}}$ | $r = 2\rho(\beta) \frac{\sqrt{\log p}}{\sqrt{n}}$ |

Table 4: The relative performance of $\ell_2$-SEPCA.

| Algorithm | $\cos \theta_{lim}$ | Cap exists if |
|-----------|----------------------|----------------|
| sum       | $C_U/\sqrt{C_2} \sqrt{\frac{\log p}{(1+\log p)\sqrt{n}}}$ | $n \geq C_U^2/C_2^2, p \geq 1$ |
| HC-sum    | $\sqrt{2\rho(\beta)/C_2} \sqrt{\frac{\log p}{(1+\log p)\sqrt{n}}}$ | $n \geq 1, p \geq 1$ |
| FDR-sum   | $\frac{1+\sqrt{2 \log \nu p/k}}{\sqrt{C_2 \sqrt{n \log \nu p}}}$ | $p \geq 11, n \geq 1$ |
Table 5: The relative performance of HC-$\ell_2$-SEPCA.

| Algorithm | $\cos \theta_{\text{lim}}$ | Cap exists if |
|-----------|-----------------------------|---------------|
| sum       | $\frac{C_U}{2\rho(\beta)\sqrt{\log p}}$ | $p \geq \exp \left( \frac{C_U}{4\rho(\beta)^2} \right)$ |
| HC-sum    | $\left[2\rho(\beta) \log p\right]^{-1}$ | $p \geq \exp \left( \frac{1}{2\rho(\beta)} \right)$ |
| FDR-sum   | $\frac{1+\sqrt{2\log np/k}}{2\rho(\beta)\log p}$ | $\log p \geq \frac{1}{4\rho(\beta)^2} \left(1 + 2\rho(\beta) + \sqrt{8\rho(\beta)^2 + 4\rho(\beta) + 1}\right)$ |

### 8.3. HC-$\ell_2$-SEPCA versus $\ell_2$-SEPCA

Now, we consider when HC-$\ell_2$-SEPCA is more powerful than $\ell_2$-SEPCA. The ratio of the radii is given by

$$\frac{2\rho(\beta)}{\sqrt{C_U}n^{1/4} \sqrt{1 + \log p}} \quad (8.3)$$

If this ratio is smaller than 1, HC-$\ell_2$-SEPCA is more powerful than $\ell_2$-SEPCA. Note that the quantity

$$\frac{2\sqrt{2}}{\sqrt{C_U}} \frac{\log p}{\sqrt{n}}$$

is an upper bound for (8.3), so that if

$$\frac{\log p}{\sqrt{n}} < \frac{e}{4\sqrt{2}},$$

the original ratio is smaller than 1 and HC-$\ell_2$-SEPCA is preferable to $\ell_2$-SEPCA.

### 8.4. Comparing the sum-based algorithms

Finally, we compare sum-, HC-sum-, and FDR-SEPCA. First, the ratio of the thresholds for HC-sum- and sum-SEPCA is

$$\frac{\sqrt{2\rho(\beta)}}{C_U}. \quad (8.4)$$

Noting that $\rho(\beta) \leq 1$ and that $C_U \geq \sqrt{2} + 1/3\sqrt{2}$, it is clear that this ratio is always smaller than 1 so that HC-sum-SEPCA is a strict improvement on sum-SEPCA.

Next, we compute the ratio of the thresholds for FDR- and sum-SEPCA:

$$\frac{1 + \sqrt{2\log np/k}}{C_U \sqrt{\log p}}. \quad (8.5)$$
Using the lower bound on $C_U$, we find that if $\hat{k} \geq 11$ (and $p \geq \hat{k}$, naturally), FDR-SEPCA is always more powerful than sum-SEPCA. For smaller values of $\hat{k}$, for sufficiently large values of $p$, the ratio will be smaller than 1.

Lastly, we compare FDR-SEPCA to HC-sum-SEPCA, wherein the ratio of the thresholds is (FDR to HC-sum):

$$\frac{1 + \sqrt{2 \log \nu p / \hat{k}}}{\sqrt{2 \rho(\beta) \log p}}.$$  \hfill (8.6)

Because of involvement of $\rho(\beta)$, this quantity is hard to analyze. If in an oracle manner, FDR-SEPCA obtained $\hat{k}$ correctly as $p^{1-\beta}$, we would find that this ratio is always larger than 1 for $p > 1$. That is if $\hat{k}$ assumes the correct value, HC-sum-SEPCA is more powerful than FDR-SEPCA. Alternatively, we can note that $\rho(\beta) \in (0, 1]$ and ask when the ratio is larger than 1. Based on the ratio above, we can see that in the following scenarios

\begin{align*}
\begin{cases}
\hat{k} = 1 \quad \text{and} \quad p > 1, \\
2 \leq \hat{k} \leq 18 \quad \text{and} \quad p \geq \hat{k} \ (\text{always}), \\
\hat{k} \geq 19 \quad \text{and} \quad \log p \geq \frac{1}{8} \left(4(\log \hat{k})^2 - 4 \log \hat{k} + 1\right),
\end{cases}
\end{align*}

HC-sum-SEPCA is more powerful than FDR-SEPCA.

To summarize, we prefer the FDR-controlling alternatives to sum-SEPCA, but depending on the output of FDR-SEPCA, HC-sum-SEPCA may be more powerful. However, as the simulations in Section 7 revealed (see Figure 2), the number of false positives with HC-sum-SEPCA may be higher than with FDR-SEPCA.

### 8.5. Overall Message

We have seen that for $n$ and $p$ sufficiently large and $\nu$ that is sufficiently dense (in the sense of $\|\nu\|_1$ being large), a sum-based statistic and algorithm leads to better performance. This is expected behavior, as by using a sum-based method, we are taking advantage of the equisigned nature of $\nu$. Moreover, within the class of sum-based algorithms, controlling the FDR leads to greater power, as expected. It is difficult to clearly identify which of HC-sum- and FDR-SEPCA will have the greatest power, and the end result may come down to a practitioner’s tolerance for false discoveries.

### 9. Conclusions

We have considered the setting where the left singular vector of the underlying rank one signal matrix plus noise data matrix is assumed to be sparse and the right singular vector is assumed to be equisigned. We have proposed six different SEPCA algorithms for estimating the sparse principal component
based on different decision statistics and provided sparsistency conditions for
the same. Our analysis reveals conditions where a coordinate selection scheme
based on a sum-based decision statistic outperforms schemes that utilize the $\ell_1$
and $\ell_2$ decision statistics. Thereby, the proposed algorithm outperforms known
schemes such as diagonal thresholded PCA [17] in terms of estimation of the
singular vectors associated with the rank-1 component. We have derived lower
bounds on the size of detectable coordinates of the principal left singular vector,
utilized these lower bounds to derive lower bounds on the worst-case risk and
verified our findings with numerical simulations. Finally, we have discussed the
results of our simulations analytically, by providing a geometric interpretation
of the differences in power among the algorithms. We note that while we have
stated our results for Gaussian noise with identity covariance, we can extend the
FWER-controlling results to any log-concave noise distribution, and the FDR-
controlling procedures to Gaussian noise with certain non-identity covariances.

A. Proof of Theorem 1

a. Note that

$$\mathbb{P}(T_i \geq \tau) \leq \mathbb{P}(\max_{j \in I^c} T_j \geq \tau)$$

for $i \in I^c$. Taking the maximum over the left-hand side and noting that the
right-hand side has limit zero yields the result. This follows from (3.1). □

b. We consider when true positives occur with probability approaching 1. We
want to find the smallest coordinate $(\theta u_i)$ such that the following probability
approaches 1:

$$\mathbb{P}(T_i > \tau_{n,p}) = \mathbb{P}\left(\frac{T_i - \mathbb{E}T_i}{\sqrt{\text{Var} T_i}} > \frac{\tau_{n,p} - \mathbb{E}T_i}{\sqrt{\text{Var} T_i}}\right). \quad (A.1)$$

Note that if $(\tau_{n,p} - \mathbb{E}T_i)$ is negative and not tending to zero as $n$ grows, and
if the variance of $T_i$ decays to zero as $n$ grows, the quantity

$$\frac{\tau_{n,p} - \mathbb{E}T_i}{\sqrt{\text{Var} T_i}} \quad (A.2)$$

tends toward negative infinity. Hence, we will specify conditions so that
Var $T_i$ decays to zero as $n$ grows and then compute when a coordinate is
detectable by considering when $\tau_{n,p}$ is strictly less than $\mathbb{E}T_i$. For brevity, we
omit the computations in solving $\tau_{n,p} < \mathbb{E}T_i$ for $|\theta u_i|$ and present verifications
that the variance of $T_i$ has limit 0. These results show that above the
decision boundary, we have uniform detection.

In sum-SEPCA, $T_i$ is a Gaussian random variable with mean $\frac{(\theta u_i)}{\sqrt{n}} \sum_k v_k$
and variance $\sigma^2$. Since $\sigma$ does not grow with $n$, Var $T_i$ always decays to
zero.

In $\ell_2$-SEPCA, $T_i$ has

$$\mathbb{E}T_i = (\theta u_i)^2 + \sigma^2$$
and

\[ \text{Var } T_i = \frac{2\sigma^2}{n} \left( \sigma^2 + 2 (\theta u_i)^2 \right). \]

Since \( \sigma \) and \( \theta \) are fixed, the variance always decays to 0. Let \( x_{i,k} = \left( \sqrt{n} \frac{v_k (\theta u_i)}{\sigma} \right) \). In \( \ell_1 \)-SEPCA, \( T_i \) has

\[
\text{Var } T_i = \frac{\sigma^2}{n^2} \sum_k x_{i,k}^2 \left( 1 - \left( \text{Erf} \left( x_{i,k} \sqrt{\frac{2}{\sigma^2}} \right) \right)^2 \right) \\
+ \frac{\sigma^2}{n} \left( 1 - \frac{2}{n\pi} \sum_k \exp \left( -x_{i,k}^2 \right) \right) \\
- 2 \sqrt{\frac{\sigma^2}{\pi n^2}} \sum_k x_{i,k} \exp \left( -\frac{x_{i,k}^2}{2} \right) \text{Erf} \left( \frac{x_{i,k}}{\sqrt{2}} \right),
\]

which is less than or equal to

\[
\frac{(\theta u_i)^2}{n} \sum_k v_k^2 + \frac{\sigma^2}{n} + 2 \sqrt{\frac{2}{\pi n^2}} |\theta u_i| \sum_k |v_k|.
\]

Since \( \|v\|_2 = 1 \), the variance of \( T_i \) has limit 0. Because we cannot solve the inequality \( \tau_{n,p} < \mathbb{E} T_i \) analytically, we leave the bound in the form given previously. \( \square \)

In the proof above, note that if \( (\tau_{n,p} - \mathbb{E} T_i) \) is positive and not tending to zero as \( n \) grows, the quantity in (A.2) tends to positive infinity when the variance decays to zero. Hence, modifying the proof by solving \( \tau_{n,p} > \mathbb{E} T_i \) for \( |\theta u_i| \) yields when a coordinate is not detectable with probability approaching 1: i.e., when \( |\theta u_i| \) is smaller than the values given in (6.1). \( \square \)

B. Proof of Theorem 2

B.1. Size of Detectable Coordinates

B.1.1. Sum: HC-SEPCA

If \( \mathbf{v} \) is equisigned, summing across the rows of \( \mathbf{X} \) yields a normally distributed quantity with mean \( (\theta u_i) \| \mathbf{v} \|_1 \) and variance \( \sigma^2 \). Dividing by \( \sigma \) and adopting the notation of \( HC \), we have that under the alternative hypothesis, \( \mu_i = \sqrt{2r \log p} \), so that

\[ r = \left( \frac{|\theta u_i| \| \mathbf{v} \|_1}{\sqrt{2 \log p}} \right)^2. \]

Rearranging the inequality \( r > \rho(\beta) \) yields

\[ |\theta u_i| > \sigma \sqrt{\rho(\beta)} \frac{\sqrt{2 \log p}}{\| \mathbf{v} \|_1}. \]

(A.1)
Note that sum-SEPCA can detect coordinates of size
\[ |\theta_{ui}| > \sigma C_U \frac{\sqrt{\log p}}{\|v\|_1}. \]  
\text{(B.2)}

However, \( C_U \) is strictly larger than \( \sqrt{2} + 1/(3\sqrt{2}) \). Thus, using HC yields a threshold of the same order, but with a strictly smaller scaling.

\textbf{B.1.2. Sum of squares: HC-}\( \ell_2 \)-SEPCA

If we sum the squares of the entries of rows of \( X \), abusing notation slightly and using \( \mathcal{N}(\mu, \sigma^2) \) to indicate a Gaussian random variable with mean \( \mu \) and variance \( \sigma^2 \), the statistic for the \( i^{th} \) coordinate is of the form
\[ \sum_{k=1}^{n} \left( \frac{\sigma}{\sqrt{n}} \mathcal{N} \left( \frac{\theta_{ui}}{\sigma} v_k \sqrt{n}, 1 \right) \right)^2. \]

Assuming oracular knowledge of \( \sigma \), the statistic
\[ \frac{n}{\sigma^2} \sum_k \chi_{ik}^2 \]
places us in the setting of (4.5). The non-centrality parameter \( \delta \) is given by
\[ \delta = \sqrt{\sum_{k=1}^{n} \left( \frac{\theta_{ui}}{\sigma} v_k \sqrt{n} \right)^2} = \frac{|\theta_{ui}|}{\sigma} \sqrt{n}. \]

Setting \( \delta = 2r \log p \) and solving \( r > \rho(\beta) \) yields
\[ |\theta_{ui}| > \sigma \rho(\beta) \frac{2 \log p}{\sqrt{n}}. \]  
\text{(B.3)}

We have that \( \ell_2 \)-SEPCA can detect coordinates with
\[ |\theta_{ui}| > \sigma \sqrt{e} \sqrt{2} \sqrt{\frac{1 + \log p}{\sqrt{n}}}. \]  
\text{(B.4)}

Using HC offers a significant improvement over \( \ell_2 \)-SEPCA. However, we also expect HC with the \( \chi^2_n \) statistic to have a smaller detectable coordinate: \( \|v\|_1 \leq \sqrt{n} \), so that for fixed \( \beta \) and \( p \), the threshold in (B.2) is asymptotically larger than that in (B.4) (but potentially of the same order). This result is strange in context of the non-FDR results. In any case, HC improves on \( \ell_2 \)-SEPCA.
B.1.3. FDR-SEPCA

Recall that taking sums across the rows of $X$, we obtain a vector $y$ where $y_i = \mu_i + \sigma z_i$, with $\mu_i = (\theta u_i) \|v\|_1$. Moreover, we have noted that $t_k \approx \sqrt{\zeta}(1 + \sqrt{2 \log(\nu p/k)})$, where $t_k$ is the level at which $y$ is thresholded. It follows that, entries of $y$ that are of size at least $y_i > (1 - o(1)) \sqrt{\zeta} \sigma \left(1 + \sqrt{2 \log(\nu p/k)}\right)$ are selected, or, since $\mu_i = (\theta u_i) \|v\|_1$ (when $v$ is equisigned), if we select $\hat{k}$ coordinates, we expect to detect

$$|\theta u_i| > (1 - o(1)) \sqrt{\zeta} \sigma \left(1 + \sqrt{2 \log(\nu p/k)}\right) \|v\|_1$$

are of size at least

$$y_i > (1 - o(1)) \sqrt{\zeta} \sigma \left(1 + \sqrt{2 \log(\nu p/k)}\right),$$

where $t_k$ is the level at which $y$ is thresholded. It follows that, entries of $y$ that are of size at least

$$y_i > (1 - o(1)) \sqrt{\zeta} \sigma \left(1 + \sqrt{2 \log(\nu p/k)}\right) \|v\|_1$$

are selected, or, since $\mu_i = (\theta u_i) \|v\|_1$ (when $v$ is equisigned), if we select $\hat{k}$ coordinates, we expect to detect

$$|\theta u_i| > (1 - o(1)) \sqrt{\zeta} \sigma \left(1 + \sqrt{2 \log(\nu p/k)}\right) \|v\|_1 \quad \text{(B.5)}$$

Relative to HC and sum-SEPCA, the gain here is found when there are many smaller coordinates of $u$ and $\hat{k}$ is large.

B.2. Proofs for the Higher Criticism-Based Methods

a. From (2.8) in [13],

$$P(T_i \geq \tau) \leq P\left(\max_{j \in I^c} T_j \geq \tau\right)$$

has limit zero.

b. Let $I_1 \subseteq I$ be the set of coordinates with signal larger than the detection limit ($i \in I$ such that $|\theta u_i| > \beta_{\text{crit}}(1 + \epsilon)$), and let $I_2 \subseteq I$ contain the rest of the coordinates ($i \in I$ such that $|\theta u_i| < \beta_{\text{crit}}(1 - \epsilon)$). By Theorem 1 in [1], the asymptotic power for detecting signals below the detection limit is one, and that for signals below the limit is zero. Hence, for $i \in I_1$,

$$\min_{i \in I : |\theta u_i| > \beta_{\text{crit}}(1 + \epsilon)} P(i \text{ selected}) \rightarrow 1,$$

and for $i \in I_2$,

$$\max_{i \in I : |\theta u_i| < \beta_{\text{crit}}(1 - \epsilon)} P(i \text{ selected}) \rightarrow 0.$$

As with Theorem 1, we omit the computation of $\beta_{\text{crit}}$, as it follows from the discussion in Section 4.1.
B.3. FDR-SEPCA

The details of these computations are in Appendix D.1, so we will summarize the properties here.

a. The choice of $\nu = 2^{1/\omega}$ controls the FDR at level $\omega$ [18]. Choosing $\omega(p) \to 0$ as $p \to \infty$ leads to an asymptotic FDR of zero. I.e., for $i \in I^c$,
$$\max_{i \in I^c} \mathbb{P}(i \text{ selected}) \to 0. \quad \square$$

b. Noting that the consistency of estimating the mean vector $\mu = (\theta \|v\|_1)u$ encompasses the estimation of the support of $u$, risk bounds for the estimation of $\mu$ yield the result. To be precise, if the expected risk $\mathbb{E}\|\mu - \hat{\mu}\|_2^2 \leq B$ for some bound $B$, we expect to detect coordinates of size larger than $B$ and to not detect those smaller than $B$. \quad \square

C. Risk bounds under $\ell_q$ sparsity

In this section, we simultaneously generalize our setting to approximate sparsity, and specify the risk lower-bounds. We omit the $\ell_1$-SEPCA algorithm from consideration.

Let $u \in \mathbb{R}^p$ have unit $\ell_2$-norm and belong to an $\ell_q$ ball with radius $C$ for $q \in (0, 2]$. I.e.,
$$\sum_{i=1}^{p} |u_i|^q \leq C^q. \quad (C.1)$$

When $q = 0$, we replace $C^q$ with $s$, the level of ‘hard’ sparsity. We have the following result:

**Theorem 3.** Let
$$L(\hat{u}, u) = \|u - \text{sign}(u, \hat{u})\hat{u}\|_2^2 \quad (C.2)$$
be the risk of the estimator $\hat{u}$ of $u$, where $u$ is as specified in (2.1) and the estimators are the six algorithms that we have previously described. Then,

a. sum-, HC-sum, and FDR-SEPCA have expected risks lower-bounded by
$$\mathbb{E}L(\hat{u}, u) \geq O\left(\left|C^q - 1\right|\|v\|_1^{(2-q)}\right). \quad (C.3)$$

b. $\ell_2$-SEPCA has a risk lower-bounded by
$$\mathbb{E}L(\hat{u}, u) \geq O\left(\left|C^q - 1\right|n^{-\frac{1}{2}(1-q/2)}\right). \quad (C.4)$$

c. HC-$\ell_2$-SEPCA has a risk lower-bounded by
$$\mathbb{E}L(\hat{u}, u) \geq O\left(\left|C^q - 1\right|n^{-\frac{1}{2}(1-q/2)}\right). \quad (C.5)$$

The rest of this section contains the proof of Theorem 3.
C.1. Proof of Theorem 3

We construct a 'worst-case' sparse \( u \). Note that \( C^q \geq 1 \) necessarily, and that if \( C \geq p^{1-q/2} \), every unit norm vector is in the \( \ell_q \) ball. Hence, we take \( C \in [1, p^{1-q/2}) \).

Let \( \theta \) and \( \sigma \) be fixed. We want a sparse vector with several coordinates guaranteed to be missed (the probability of not detecting them is asymptotically 1). For this vector \( u \) to be sparse and for the loss to not be 1, set \( u_1 \) to be \( \sqrt{1 - r_n^2} \), where \( r_n = o(1) \), and take \( u_2, \ldots, u_{m_n + 1} \) to be \( r_n / \sqrt{m_n} \). The other coordinates of \( u \) are 0, so that \( u \) has unit \( \ell_2 \)-norm.

We assume that \( u_1 \) is detected with probability 1 as \( n \to \infty \), and want to set \( u_2, \ldots, u_{m_n + 1} \) so that the expected loss is lower bounded by:

\[
E_L(u, \hat{u}) \geq \sum_{k=1}^p |u_k|^2 \mathbb{P}(\text{Not Selecting Coordinate } k) \geq \sum_{k=2}^{m_n+1} |u_k|^2 \mathbb{P}(\text{Not Selecting Coordinate } k). 
\]  

If coordinates of size \( r_n / \sqrt{m_n} \) are not detected, the expected loss is lower bounded by \( r_n^2 \).

Let \( m_n = \lfloor m \rfloor \) where

\[
m = \delta n^\phi r_n^\psi \|v\|^\eta. 
\]

Note that we have not specified the norm used in \( \|v\| \): we will choose the norm at the very end of the calculation. Let

\[
r_n = [C^q - 1]^{\alpha_n \beta + \gamma q} \|v\|^\kappa, 
\]

so that,

\[
r_n / \sqrt{m_n} \approx \frac{r_n}{\sqrt{m}} = \frac{1}{\delta} n^{-\phi/2} \|v\|^{\kappa - n/2} r_n^{1-\psi/2}. 
\]

We will choose \( \delta, \phi, \eta, \alpha, \beta, \gamma, \kappa, \psi \) so that the \( \ell_q \) sparsity constraint is met and the lower bound \( r_n^2 \) is maximized. The sparsity constraint requires:

\[
\sum_{i=1}^p |u_i|^q = (1 - r_n)^{q/2} + m_n^{1-q/2} r_n^q \leq 1 + m_n^{1-q/2} r_n^q \leq C^q. 
\]  

First, we will assume (for now) that \( r_n = o(1) \) and that via other parameters we may control the scaling of the coordinate sizes; hence, we set \( \psi = 2 \). Then,

\[
r_n m_n^{1-q/2} = \delta^{1-q/2} n^{2(\beta + \gamma q) + (1-q/2)\phi} [C^q - 1]^{2\alpha} \|v\|^{2\kappa + q(1-q/2)}. 
\]  

We need this quantity to be smaller than \( C^q - 1 \). To eliminate the \( n \) dependence, we set \( \beta = \frac{\delta}{2} \) and \( \gamma = \frac{\delta}{2} \). We choose \( \alpha = \frac{1}{2} \) to match powers of \( [C^q - 1] \) on both sides of the inequality. Defining another parameter \( \rho \), let \( \delta = \rho \|v\|^{\eta} \). Then, the inequality is

\[
\rho^{1-q/2} \|v\|^{2\kappa [C^q - 1]} \leq [C^q - 1]. 
\]
Choosing \( \rho \leq \|v\|^{-2\kappa/(1-q/2)} \) is enough.

With these choices of parameters,

\[
\begin{align*}
    r_n &= \sqrt{[C^q - 1]} n^{1/2} \|v\|^{\kappa}, \\
    m &= \rho n^{\phi/r_n^2},
\end{align*}
\]

so that

\[
\frac{r_n}{\sqrt{m}} = \frac{1}{\sqrt{\rho}} n^{-\phi/2}.
\]

Noting that

\[
\frac{1}{\sqrt{\rho}} \geq \|v\|^{\kappa/(1-q/2)},
\]

choosing \( \rho = \|v\|^{-2\kappa/(1-q/2)} \) leads to the smallest possible choice of coordinate.

In summary:

\[
\begin{align*}
    r_n &= \sqrt{[C^q - 1]} n^{1/2} \|v\|^{\kappa}, \\
    r_n^2 &= [C^q - 1] n^{-\phi/(1-q/2)} \|v\|^{2\kappa}, \\
    m &= \|v\|^{-2\kappa/(1-q/2)} n^{\phi/r_n^2},
\end{align*}
\]

and

\[
\frac{r_n}{\sqrt{m}} = \|v\|^{\kappa/(1-q/2)} n^{-\phi/2}. 
\]

So, for a given algorithm, it remains to choose \( \phi \) and \( \kappa \) so that the worst-case risk is lower-bounded by \( r_n^2 \). Sum-SEPCA misses coordinates of size \( O\left(\frac{\sqrt{\log p}}{\|v\|_1}\right) \) and \( \ell_2\)-SEPCA misses coordinates of size \( O\left(\frac{n^{1/4}}{\|v\|_2}\right) \). For sum-SEPCA, \( \kappa = q - \frac{1}{2} \), and for \( \ell_2\)-SEPCA, \( \kappa \) is irrelevant, as \( \|v\|_2 = 1 \). Sum-SEPCA uses \( \phi = 0 \) and \( \ell_2\)-SEPCA uses \( \phi = \frac{1}{2} \). Hence, sum-SEPCA has a risk lower-bounded by

\[
O\left(\frac{[C^q - 1]}{\|v\|_1^{-(2-q)}}\right).
\]

Noting that \( \|v\|_2 = 1 \), \( \ell_2\)-SEPCA has

\[
\begin{align*}
    O\left(\frac{[C^q - 1]}{n^{-1/2} (1-q/2)} \|v\|_2^{-(1-q/2)}\right) = \\
    O\left(\frac{[C^q - 1]}{n^{-1/2} (1-q/2)}\right).
\end{align*}
\]

In the \( \ell_0 \) case, i.e., when \( u \) has no more than \( s \) non-zero entries, the preceding analysis goes through with \( C^q \) replaced by \( s \) and \( q \) set to zero.
C.1.1. FDR Algorithms

For HC-sum-SEPCA, the $\beta_{\text{crit}}$ is of the same order as that for sum-SEPCA. Similarly, for FDR-SEPCA, if $\hat{k}$ is much smaller than $p$, $\beta_{\text{crit}}$ is of roughly the same order. Hence, these two algorithms have the same risk bound as sum-SEPCA. For HC-$\ell_2$-SEPCA, $\kappa = 0$ and $\phi = 1$. The risk is therefore lower-bounded by

$$O\left(\left[C^q - 1\right]n^{-(1-q/2)}\right).$$  \hfill (C.15)

D. FDR-SEPCA: Further Details

Let $y_i = \mu_i + \sigma z_i$, where $i \in \{1, \cdots, p\}$, the vector $z$ of the $z_i$ is normally distributed with mean 0 and covariance $\Sigma$, and $\Sigma$ satisfies

$$\xi_0 I_p \leq \Sigma \leq \xi_1 I_p.$$  

Here, $\xi_0$ is the smallest eigenvalue of $\Sigma$ and $\xi_1$ is the largest. The mean vector $\mu$ of the $\mu_i$ is assumed to be sparse; the goal is to estimate $\mu$. The following penalized least squares formulation yields an estimator for $\mu$:

$$\hat{\mu} = \arg \min_\mu \|y - \mu\|_2^2 + \sigma^2 \text{pen}(\|\mu\|_0),$$  \hfill (D.1)

where $\text{pen}(k)$ is defined as

$$\text{pen}(k) = \xi_1 \zeta k \left(1 + \sqrt{2L_{p,k}}\right)^2,$$  \hfill (D.2)

with $\zeta > 1$ and

$$L_{p,k} = (1 + 2\beta) \log(\nu p / k).$$  \hfill (D.3)

The parameter $\beta$ may be set to 0 here, and $\nu$ is chosen to be no smaller than $e^{1/(1+2\beta)}$. We define $\|\mu\|_0$ to be number of non-zero coordinates of $\mu$.

The solution to (D.1) is given by hard-thresholding. Let $|y|_{(i)}$ be the $i^{th}$ order statistic of $|y_i|$, namely $|y|_{(1)} \geq \cdots \geq |y|_{(p)}$. Then if

$$\hat{k} = \arg \min_{k \geq 0} \sum_{i > k} |y|_{(i)}^2 + \sigma^2 \text{pen}(k),$$  \hfill (D.4)

defining

$$t_k^2 = \text{pen}(k) - \text{pen}(k-1),$$  \hfill (D.5)

the solution is to hard threshold at $t_{\hat{k}}$.

In this set-up, we have that

$$t_k \approx \lambda_{p,k} = \sqrt{\xi_1 \zeta (1 + \sqrt{2L_{p,k}})},$$
with $|t_k - \lambda_{p,k}| < c/\lambda_k$. More precisely, Lemma 11.7 of [16] says that

$$\lambda_{p,k} - \frac{4\zeta b}{\lambda_{p,k}} \leq t_k \leq \lambda_{p,k}.$$ 

When $\nu \geq e^2$, we may take $b = (1 + 2\beta)$. In any case, if $k = o(n)$, $\lambda_{p,k} \approx \sqrt{\log p}$.

Hence, entries of $y$ that are of size at least

$$y_i > (1 - o(1))\sqrt{\xi_1\sigma} \left(1 + \sqrt{2\log(\nu p/k)}\right)$$

are selected, or, since $\mu_i = (\theta u_i)\|v\|_1$ (when $v$ is equisigned), if we select $\hat{k}$ coordinates, we expect to detect

$$|\theta u_i| > (1 - o(1))\sqrt{\xi_1\sigma} \frac{(1 + \sqrt{2\log(\nu p/k)})}{\|v\|_1}$$

$$= O \left(\sigma \sqrt{\xi_1} \frac{\sqrt{2\log(\nu p/k)}}{\|v\|_1}\right). \quad (D.6)$$

### D.1. Risk Behavior

Recalling that (D.1) solves a penalized least squares problem for $\hat{y}$ close to $y$, we may discuss the statistical behavior of this estimator. The following discussion follows and reproduces that in [18].

First, note that for $\beta = 0$, the parameter $\nu$ directly controls the FDR (where a false positive corresponds to selecting a zero coordinate in $y$): a choice of $\nu = 2^{1/\omega}$ for $\omega \in (0, 1)$ bounds the FDR at a level $\omega$.

Second, the expected risk, $E\|y - \hat{y}\|^2_2$, is bounded as follows. By Proposition 4.1 in [18],

$$E\|y - \hat{y}\|^2_2 \leq D \left[2M'_{\nu} \xi_1 \sigma^2 + R(y, \sigma)\right], \quad (D.7)$$

where $D$ is a constant $2\zeta(\zeta + 1)^3(\zeta - 1)^{-3} = \Theta(1)$, we assume that $\xi_1 = 1$, and $0 \leq M'_\nu \leq C_\beta p^{-2\beta} \nu^{-1}$, for some $C_\beta > 0$. Since $\beta = 0$, $M'_\nu = O(1/\nu) = O(\omega)$, if we control the FDR at level $\omega$.

The second term in (D.7) is the ideal risk, or, the infimum of the penalized least squares objective. If $y$ belongs to an $\ell_q$ ball with radius $C$ and $0 < q < 2$, and we define

$$r_{p,q}(C) = \begin{cases} 
C^2 & \text{if } C \leq \sqrt{1 + \log p}, \\
C^q[1 + \log(p/C^q)]^{1-q/2} & \text{if } \sqrt{1 + \log p} \leq C \leq p^{1/q}, \\
p & \text{if } C \geq p^{1/q},
\end{cases} \quad (D.8)$$
the ideal risk is bounded as
\[
\sup_{y \in \mathbb{R}^p : \sum_i |y_i| \leq C} R(y, \sigma) \leq c (\log \nu) \sigma^2 r_{p,q} (C/\sigma),
\] (D.9)
for some \(c > 0\). The supplementary results in [18] yield that \(R(y, \sigma)\) is bounded by \(C^2 \log \nu\), and by \(C^2\) when \(C \leq \sqrt{1 + \log p}\).

As in Appendix C, we may replace \(q\) with 0 and \(C^q\) with \(s\) in the case of hard sparsity with \(s\) non-zero coordinates. Doing so leads to the bound:
\[
E \| y - \hat{y} \|_2^2 \leq s \sigma \log \nu \log \frac{\sigma p \nu}{s} + \sigma_1 \sigma^2 \frac{2}{\nu},
\] (D.10)
Note that we have recovered the factor of \(\log \nu p/s\) in \(\beta_{\text{crit}}\).

References

[1] ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Annals of Statistics* 2533–2556.
[2] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis* 111 120–135.
[3] BERTHET, Q., RIGOLLET, P. et al. (2013). Optimal detection of sparse principal components in high dimension. *Annals of Statistics* 41 1780–1815.
[4] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Annals of Statistics* 41 1055.
[5] BOBKOV, S. G. and NAZAROV, F. L. (2003). On convex bodies and log-concave probability measures with unconditional basis. In *Geometric aspects of functional analysis* 53–69. Springer.
[6] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press.
[7] Boucheron, S. and Thomas, M. (2012). Concentration inequalities for order statistics. *Electronic Communications in Probability* 17 no. 51, 1-12.
[8] d’ASPREMONT, A., EL GHAOUI, L., JORDAN, M. I. and LANCKR IET, G. R. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review* 49 434–448.
[9] Ding, C. H., Li, T. and JORDAN, M. I. (2010). Convex and semi-nonnegative matrix factorizations. *IEEE Transactions on Pattern analysis and Machine Intelligence* 32 45–55.
[10] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* 962–994.
[11] DONOHO, D., JIN, J. et al. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science* 30 1–25.
[12] Gao, C., Moore, B. E. and Nadakuditi, R. R. (2017). Augmented robust PCA for foreground-background separation on noisy, moving camera video. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) 1240–1244. IEEE.

[13] Hall, P., Jin, J. et al. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Annals of Statistics* 38 1686–1732.

[14] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research* 5 1457–1469.

[15] Huang, K., Sidiropoulos, N. D. and Swami, A. (2014). Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Transactions on Signal Processing* 62 211–224.

[16] Johnstone, I. M. (2013). Function estimation and gaussian sequence models. *Unpublished manuscript*.

[17] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* 104 682.

[18] Johnstone, I. M. and Paul, D. (2014). Adaptation in some linear inverse problems. *Stat* 3 187–199.

[19] Latala, R. (2011). Order statistics and concentration of norms for log-concave vectors. *Journal of Functional Analysis* 261 681 - 696.

[20] Liu, H., Wu, Z., Li, X., Cai, D. and Huang, T. S. (2012). Constrained nonnegative matrix factorization for image representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 1299–1311.

[21] Ma, Z. et al. (2013). Sparse principal component analysis and iterative thresholding. *Annals of Statistics* 41 772–801.

[22] Moore, B., Gao, C. and Nadakuditi, R. R. (2019). Panoramic robust PCA for foreground-background separation on noisy, free-motion camera video. *IEEE Transactions on Computational Imaging*.

[23] Passemier, D., Li, Z. and Yao, J. (2017). On estimation of the noise variance in high dimensional probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79 51–67.

[24] Pastor, D. and Socheleau, F.-X. (2012). Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences. *IEEE transactions on Signal Processing* 60 1545–1555.

[25] Piccardi, M. (2004). Background subtraction techniques: A review. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)* 4 3099-3104. IEEE.

[26] Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). High-dimensional Ising model selection using $\ell_1$-regularized logistic regression. *Annals of Statistics* 38 1287–1319.

[27] Ren, B., Pueyo, L., Zhu, G. B., Debes, J. and Duchêne, G. (2018). Non-negative matrix factorization; robust extraction of extended structures. *The Astrophysical Journal* 852 104.

[28] Ross, P. *Stars*. https://archive.org/details/Stars_2D. Accessed: 2016 November 16.
[29] Socheleau, F.-X. and Pastor, D. (2014). Testing the energy of random signals in a known subspace: An optimal invariant approach. *IEEE Signal Processing Letters* 21 1182–1186.

[30] Taslaman, L. and Nilsson, B. (2012). A framework for regularized non-negative matrix factorization, with application to the analysis of gene expression data. *PloS one* 7 e46331.

[31] Vaswani, N., Bouwmans, T., Javed, S. and Narayanamurthy, P. (2018). Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery. *IEEE signal processing magazine* 35 32–55.

[32] Wang, Y.-X. and Zhang, Y.-J. (2013). Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering* 25 1336–1353.

[33] Yuan, X.-T. and Zhang, T. (2013). Truncated power method for sparse eigenvalue problems. *Journal of Machine Learning Research* 14 899–925.