Spherical Cap Packing Asymptotics and Rank-Extreme Detection

Kai Zhang

Abstract—We study the spherical cap packing problem with a probabilistic approach. Such probabilistic considerations result in an asymptotic sharp universal uniform bound on the maximal inner product between any set of unit vectors and a stochastically independent uniformly distributed unit vector. When the set of unit vectors are themselves independently uniformly distributed, we further develop the extreme value distribution limit of the maximal inner product, which characterizes its uncertainty around the bound.

As applications of the above asymptotic results, we derive (1) an asymptotic sharp universal uniform bound on the maximal spurious correlation, as well as its uniform convergence in distribution when the explanatory variables are independently Gaussian distributed; and (2) an asymptotic sharp universal bound on the maximum norm of a low-rank elliptically distributed vector, as well as related limiting distributions. With these results, we develop a fast detection method for a low-rank structure in high-dimensional Gaussian data without using the spectrum information.

Index Terms—Spherical cap packing, extreme value distribution, spurious correlation, low-rank detection and estimation, high-dimensional inference.

I. INTRODUCTION

In modern data analysis, datasets often contain a large number of variables with complicated dependence structures. This situation is especially common in important problems such as the relationship between genetics and cancer, the association between brain connectivity and cognitive states, the effect of social media on consumers' confidence, etc. Hundreds of research papers on analyzing such dependence have been published in top journals and conferences proceedings. For a comprehensive review of these challenges and past studies, see [1].

One of the most important measures on the dependence between variables is the correlation coefficient, which describes their linear dependence. In the new paradigm described above, understanding the correlation and the behavior of correlated variables is a crucial problem and prompts data scientists to develop new theories and methods. Among the important challenges of a large number of variables on the correlation, we focus particularly on the following two questions:

- **The maximal spurious sample correlation in high dimensions.** The Pearson's sample correlation coefficient between two random variables X and Y based on n

Kai Zhang is with the Department of Statistics and Operations Research, University of North Carolina, Chapel Hill (e-mail: zhangk@email.unc.edu)

observations can be written as

$$\widehat{C}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{I.1}$$

where X_i's and Y_i's are the n independent and identically distributed (i.i.d.) observations of X and Y respectively, and $\bar{X}$ and $\bar{Y}$ are the sample means of X and Y respectively. The sample correlation coefficient possesses important statistical properties and was carefully studied in the classical case when the number of variables is small compared to the number of observations. However, the situation has dramatically changed in the new high-dimensional paradigm [1, 2] as the large number of variables in the data leads to the failure of many conventional statistical methods. For sample correlations, one of the most important challenges is that when the number of explanatory variables, p, in the data is high, simply by chance, some explanatory variable will appear to be highly correlated with the response variable even if they are all scientifically irrelevant [3, 4]. Failure to recognize such spurious correlations can lead to false scientific discoveries and serious consequences. Thus, it is important to understand the magnitude and distribution of the maximal spurious correlation to help distinguish signals from noise in a large-p situation.

- **Detection of low-rank correlation structure.** Detecting a low-rank structure in a high-dimensional dataset is of great interest in many scientific areas such as signal processing, chemometrics, and econometrics. Current rank estimation methods are mostly developed under the factor model and are based on the principal component analysis (PCA) [5–21], where we look for the "cut-off" among singular values of the covariance matrix when they drop to nearly 0. These methods also usually assume a large sample size. However, in practice often a large number of variables are observed while the sample size is limited. In particular, PCA based methods will fail when the number of observations is less than the rank. Moreover, although we may get low-rank solutions to many problems, more detailed inference on the rank as a parameter is not very clear. Probabilistic statements on the rank, such as confidence intervals and tests, would provide useful information about the accuracy of these solutions. The computation complexity of the matrix calculations can be an additional issue in practice. In summary, it is desirable to have a fast detection and inference method of a low-rank structure in high dimensions from a small sample.

Our study of the above two problems starts with the following question: Suppose p points are placed on the unit sphere $\mathcal{S}^{n-1}$ in $\mathbb{R}^n$. If we now generate a new point on $\mathcal{S}^{n-1}$ according to the uniform distribution over the sphere, how far will it be away from these existing p points?

Intuitively, this minimal distance between the new point and the existing p points should depend on n and p, in a manner that it is decreasing in p and increasing in n. Yet, no matter how these existing p points are located, this new point cannot get arbitrarily close to the existing p points due to randomness. In other words, for any n and p, there is an intrinsic lower bound on this distance that the new point can get closer to the existing points only with very small probability.

Studies of this intrinsic lower bound in the above question have a long history under the notion of spherical cap packing, and this question has been one of the most fundamental questions in mathematics [22–24]. In fact, this question is closely related to the 18th question on the famous list from David Hilbert [25]. This question is also a very important problem in information theory and has been studied in coding, beamforming, quantization, and many other areas [26–33].

Besides the importance in mathematics and information theory, this question is closely connected to the two problems that we propose to investigate. For instance, the sample correlation between X and Y can be written as the inner product

$$\widehat{C}(X,Y) = \langle \frac{\boldsymbol{X} - \bar{X}\mathbf{1}_n}{\|\boldsymbol{X} - \bar{X}\mathbf{1}_n\|_2}, \frac{\boldsymbol{Y} - \bar{Y}\mathbf{1}_n}{\|\boldsymbol{Y} - \bar{Y}\mathbf{1}_n\|_2} \rangle, \quad \text{(I.2)}$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and $\mathbf{1}_n$ is the vector in $\mathbb{R}^n$ with all ones. In general, if we observe n i.i.d. samples from the joint distribution of $(X_1, \ldots, X_p, Y)$, the sample correlations between X_j's and Y can be regarded as inner products in $\mathbb{R}^n$ between the p unit vectors corresponding to X_j's and another unit vector corresponding to Y. Note that these unit vectors are all orthogonal to the vector $\mathbf{1}_n$ due to the centering process. Thus, they lie on an "equator" of the unit sphere $\mathcal{S}^{n-1}$ in $\mathbb{R}^n$, which is in turn equivalent to $\mathcal{S}^{n-2}$. Through this connection, the problem about the maximal spurious correlation is equivalent to the packing of the inner products, and existing methods and results from the packing literature can be borrowed to analyze this problem. In this paper, we particularly focus on probabilistic statements about such packing problems.

An important advantage of this packing perspective is a view of data that is free of an increasing p. Suppose we view the data as n points in a p-dimensional space, then if p exceeds n, all the n points will lie on a low-dimensional hyperplane in $\mathbb{R}^p$. This degeneracy forces us to change the methodology towards statistical problems, i.e., changing from the classical statistical methods to recent high-dimensional methods [34, 35]. However, if we view the data as p vectors in $\mathbb{R}^n$, then we will never have such a degeneracy problem. No matter how large p is, a packing problem is always a well-defined packing problem. Neither the theory nor the methodology needs to be changed due to an increase in p. Thus, with the packing perspective, theory and methodology can be set free from the restriction of an increasing p.

We summarize below our results on the asymptotic theories of the maximal inner products and spurious correlations. One major advantage of the packing approach is that instead of usual iterative asymptotic results which set $p = p(n)$ and let $n \to \infty$, our convergence results are uniform in n, which leads to double limits in both n and p.

- We characterize the largest magnitude of independent inner products (or spurious correlations) through an asymptotic bound. This bound is universal in the sense that it holds for arbitrary distributions of $\boldsymbol{L}_j$'s (or that of X_j's). This bound is uniform in the sense that it holds asymptotically in p but is uniform over n. This bound is sharp in the sense that it can be attained, especially when the unit vectors $\boldsymbol{L}_j$'s are i.i.d. uniform (or when X_j's are independently Gaussian). Thus, in an analogy, this bound is to the distribution of independent inner products (or to that of spurious correlations) as the fundamental bound $\sqrt{2\log p}$ is to the p-dimensional Gaussian distribution [36]. We refer this bound as the Sharp Asymptotic Bound for indEpendent inner pRoducts (or spuRious corrElations), abbreviated as the SABER (or SABRE).

- In the special important case when the set of unit vectors are i.i.d. uniformly distributed (or when X_j's are independently Gaussian distributed), we show the sharpness of the SABER (or SABRE) and describe a smooth phase transition phenomenon of them according to the limit of $\frac{\log p}{n}$. Furthermore, we develop the limiting distribution by combing the packing approach with extreme value theory in statistics [36, 37]. The extreme value theory results accurately characterize the deviation from the observed maximal magnitude of independent inner products (or that of spurious correlations) to the SABER (or SABRE). One important feature of these results is that they are not only finite sample results but also are uniform-n-large-p asymptotics that are widely applicable in the high-dimensional paradigm.

The spherical cap packing asymptotics can be also applied to the problem of the detection of a low-rank linear dependency. For this problem, we observe that the largest magnitude among p standard elliptical variables is closely related to the rank d of their correlation matrix. This is seen by decomposing elliptically distributed random vectors into the products of common Euclidean norms and inner products of unit vectors in $\mathbb{R}^d$, thus reducing the problem to one of spherical cap packing. As a consequence, the previous asymptotics can be applied here. We thus obtained a universal sharp asymptotic bound for the maximal magnitude of a degenerate elliptical distribution, as well as its limiting distribution when the unit vectors in the decomposition are i.i.d. uniform. Although many asymptotic bounds and limiting distributions on full rank maxima are well developed under different situations (see [36–38] for reviews of the extensive existing literature), we are not able to find similar theory in literature on low-rank maxima from elliptical distributions, not even in the special case of Gaussian distribution. We refer the connection we found between the maximal magnitude and the rank as the rank-extreme (ReX)

association.

Based on the asymptotic results on the degenerate elliptical distributions, we show that one can make statistical inference on a low-rank through the distributions of the extreme value as a statistic. One feature of this procedure is that it does not require the spectrum information from PCA. Thus, the new method works when $n < d$, when PCA based methods fail to work. It is also computationally fast since no matrix multiplication is needed in the algorithm. These advantages allow a fast detection of a low-dimensional correlation structure in high-dimensional data.

A. Related Work

We are not able to find similar probabilistic statements on uniform-n-large-p asymptotics. The following statistical papers are related to the study on the maximal spurious correlation.

- In [4], the authors obtain a result on the order of the maximal spurious correlations in the regime that $\frac{\log p}{n} \to 0$. Through the packing approach, we derive the explicit limiting distribution of the extreme spurious correlations for entire scope of n and p.
- In [39], the authors develop a threshold for marginal correlation screening with large p and small n. The threshold appears in a similar form as the SABRE. We note two major differences between the results: (1) The results in [39] focus on the regime when $\frac{\log p}{n} \to \infty$ (i.e., when the threshold converges to 1), while our asymptotic results cover the entire scope of n and p, and the SABRE is shown to be valid from 0 to 1; (2) we derive the explicit limiting distribution of the maximal spurious correlation in the most important case when the variables are i.i.d. Gaussian.
- In [40–42], the minimal pairwise angles between i.i.d. uniformly random points on spheres are considered. A similar phase transition is described, and results on the limiting distribution are developed. We note two major differences between their results and ours: (1) Due to different motivations of the research, we focus on the marginal correlation between one response variable and p explanatory variables. We also develop a universal uniform bound for marginal correlations. (2) The extreme limiting distributions in their papers are stated separately according to if the limit of $\frac{\log p}{n}$ is 0, a proper constant, or ∞. From the packing perspective, we are able to state the convergence in a uniform manner with standardizing constants that are adaptive in n and p. Since in real data, the limit of $\frac{\log p}{n}$ is usually not known, this uniform convergence with adaptive standardizing constants makes the result easy to apply in practice.
- During the review process of this paper, we noticed the results in [43] which focus on the coupling and bootstrap approximations of the maximal spurious correlation when $\frac{\log^7 p}{n} \to 0$. Again our different focus is on explicit limiting distributions with adaptive standardizing constants from the packing perspective.

We are not able to find existing literature on the rank-extreme association. To evaluate the performance of our low-rank detection method, we compare our method with the algorithm in [14] which studies a similar problem. During the review process of the paper, we also noticed recent work by [21]. The most important difference from these papers is that they focus on the case when n and p are comparable and both large, while we consider the case when n is small and p is large.

B. Outline of the Paper

In Section II, we derive the asymptotic bound on the spherical packing problem, as well as that of the maximal spurious correlation and the related extreme value distributions. In Section III, we describe the rank-extreme association of elliptically distributed vectors. In Section IV we develop a fast detection method of a low-rank by using the rank-extreme association reversely. In Section V, we study the performance of the detection method through simulations. We conclude and discuss future work in Section VI.

II. ASYMPTOTIC THEORY OF THE SPHERICAL CAP PACKING PROBLEM

A. The Sharp Asymptotic Bound for independent inner products (SABER) and Spurious Correlations (SABRE)

We first observe that as described in [44], when U is uniformly distributed over $\mathcal{S}^{n-1}$, $|\langle L, U \rangle|^2 \sim Beta\left(\frac{1}{2}, \frac{n-1}{2}\right), \forall L \in \mathcal{S}^{n-1}$. By borrowing strength from the packing literature [22, 26] on the total area of non-overlap spherical caps on $\mathcal{S}^{n-2}$, we develop the following theorem on a sharp asymptotic bound for independent inner products.

Theorem II.1. Sharp Asymptotic Bound for Independent Inner Products (SABER).
For arbitrary deterministic unit vectors $L_1, \ldots, L_p$ and a uniformly distributed unit vector U over $\mathcal{S}^{n-1}$, the random variable $M_{p,n} = \max_{1 \leq j \leq p} |\langle L_j, U \rangle|$ satisfies that $\forall \delta > 0$,

$$\mathbf{P}\left(M_{p,n} > \sqrt{(1+\delta)(1 - p^{-2/(n-1)})} \right)$$
$$\leq \frac{\sqrt{2} p^{1/(n-1)} \exp\left(-\frac{1}{2}\delta(n-1)(p^{2/(n-1)} - 1) \right)}{\sqrt{\pi(1+\delta)(n-1)(p^{2/(n-1)} - 1)}}. \quad \text{(II.1)}$$

Therefore, $\forall \delta > 0$, as $p \to \infty$,

$$\sup_{n \geq 2} \mathbf{P}\left(M_{p,n} > \sqrt{(1+\delta)(1 - p^{-2/(n-1)})} \right) \to 0. \quad \text{(II.2)}$$

In particular, if $n \to \infty$, then we have the double limit

$$\lim_{p,n \to \infty} \mathbf{P}\left(M_{p,n} \leq \sqrt{1 - p^{-2/(n-1)}} \right) = 1. \quad \text{(II.3)}$$

Theorem II.1 provides an explicit answer to the question at the beginning of Section I with a probabilistic statement: No matter how L_j's are located on the unit sphere, the magnitude of the inner products (or cosines of the angle) between these p points and a uniformly random point cannot exceed $\sqrt{1 - p^{-2/(n-1)}}$ with high probability for large p. This upper bound on the inner products is equivalent to a lower bound on the minimal angle between the new random point to the existing p points.

The SABER possesses the following important properties:

1) This bound is universal in the sense that it holds for any configuration of L_j's.
2) This bound is uniform in the sense that it holds uniformity for $n \geq 2$.
3) This bound is sharp in the sense that it can be attained for some configuration of L_j's, especially when L_j's are i.i.d. uniformly distributed, as will be discussed in Section II-B.

Thus, in an analogy, the SABER $\sqrt{1 - p^{-2/(n-1)}}$ is to the distributions of the independent inner products as the fundamental bound $\sqrt{2 \log p}$ is to the p-dimensional Gaussian distribution.

A technical note here is that when n is finite, the fraction $\frac{2}{n-1}$ in the exponent of p can be replaced by $\frac{2}{n-n_1}$ with any fixed integer $0 < n_1 < n$. This change would not alter the asymptotic result in p due to a uniform convergence in the proof. The number n_1 only has an effect when the dimension n is finite. For example, see [39] for a similar but different bound when n is fixed. We focus on the bound $\sqrt{1 - p^{-2/(n-1)}}$ due to its connection to the $Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$ distribution. When $n \to \infty$, all these bounds are equivalent.

Another technical note is that although Theorem II.1 is for a deterministic set of L_j's, we note here that this set of unit vectors can be random as well. As long as L_j's are stochastically independent of U, Theorem II.1 can be applied to random L_j's by a conditioning argument on any realization of L_j's.

Figure 1 illustrates the SABER $\sqrt{1 - p^{-2/(n-1)}}$ in Theorem II.1 as a function of n and p. It can be seen that the SABER has a range of $(0, 1)$ as an increasing function in p and a decreasing function in n.
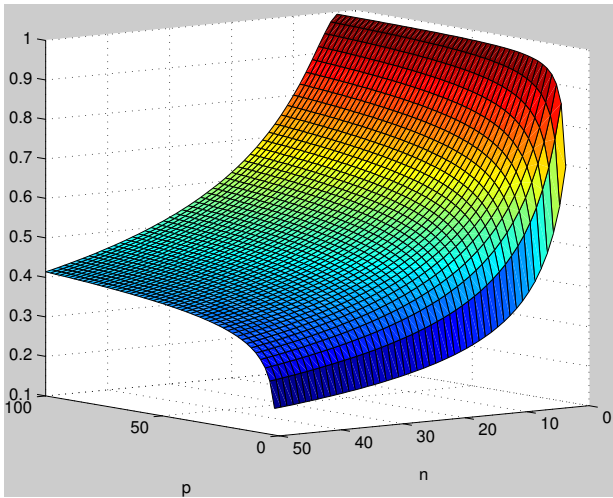

Fig. 1. The SABER $\sqrt{1 - p^{-2/(n-1)}}$ for $3 \leq n \leq 50$ and $2 \leq p \leq 100$. The SABER ranges from 0 to 1. The regions of the same color represent the smooth phase transition curves $\frac{\log p}{n} \approx \beta$ for $\beta > 0$ as described in Section II-B.

Due to the connection between sample correlations and the inner products (I.2), this bound is immediately applicable to spurious correlations. Suppose $Y = (Y_1, \ldots, Y_n)$ records n i.i.d. samples of a Gaussian variable Y, then it is well-known (see [44]) that $\frac{Y - \bar{Y} 1_n}{\|Y - \bar{Y} 1_n\|_2}$ is a uniformly distributed unit vector

over $\mathcal{S}^{n-2}$. Thus, we have the following bound on the maximal spurious correlation.

Corollary II.2. Sharp Asymptotic Bound for Spurious Correlations (SABRE).

Suppose we observe n i.i.d. samples of arbitrary random variables $X_1, \ldots, X_p$ and a Gaussian variable Y that is independent of X_j's. The maximal absolute sample correlation $M_{XY} = \max_{1 \leq j \leq p} |\hat{C}(X_j, Y)|$ satisfies that $\forall \delta > 0$, as $p \to \infty$,

$$\sup_{n \geq 3} \mathbf{P}\left(M_{XY} > \sqrt{(1 + \delta)(1 - p^{-2/(n-2)})} \right) \to 0. \quad \text{(II.4)}$$

In particular, if $n \to \infty$, then we have the double limit

$$\lim_{p, n \to \infty} \mathbf{P}\left(M_{XY} \leq \sqrt{1 - p^{-2/(n-2)}} \right) = 1. \quad \text{(II.5)}$$

Similarly as the interpretation for the SABER, the implication of the SABRE is as follows: Uniformly for $n \geq 3$, no matter how the p variables $X_1, \ldots, X_p$ are distributed, the magnitude of the sample correlations between X_j's and a Gaussian Y cannot exceed the SABRE with high probability for large p. Note here that in practice, the requirement of Gaussianity of Y can be easily relaxed through a transformation of distributions. Since the SABRE is universal, uniform, and sharp as the SABER, this bound provides a way to distinguish true signals from spurious correlations. We shall investigate this application in future work.

B. Limiting Distributions in the i.i.d. Case

In this section, we describe the asymptotics of the maximal inner product when L_j's are i.i.d. uniformly distributed and the asymptotics of spurious correlations when X_j's are independently Gaussian distributed. We first observe that when L_j's are i.i.d. uniformly unit vectors over $\mathcal{S}^{n-1}$, then for any random unit vector U that is independent of L_j's, we have the following two properties about the inner products $\langle L_j, U \rangle | U$:

1) Conditioning on U, the variables $\langle L_j, U \rangle | U$'s are independent since L_j's are independent;
2) For each j, the variable $|\langle L_j, U \rangle|^2 | U$ is distributed as $Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$. Since this conditional distribution does not depend on U, it implies that unconditionally $|\langle L_j, U \rangle|^2$ is stochastically independent of U.

From these two properties, we conclude that unconditionally, $|\langle L_j, U \rangle|^2$'s are i.i.d. $Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$ distributed. We thus show the sharpness of the SABER and SABRE by studying the maximum of i.i.d. $Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$ variables.

Theorem II.3.

1) (Sharpness of SABER)
   Suppose L_j's are i.i.d. uniformly distributed over the $(n-1)$-sphere $\mathcal{S}^{n-1}$, then for arbitrary random unit vector U that is independent of L_j's, uniformly for all $n \geq 2$, as $p \to \infty$, the random variable $M_{p,n} = \max_j |\langle L_j, U \rangle|$ has the following convergence:

$$M_{p,n} / \sqrt{1 - p^{-2/(n-1)}} \xrightarrow{prob.} 1, \quad \text{(II.6)}$$

i.e., $\forall \delta > 0$, *as* $p \to \infty$,

$$\sup_{n \geq 2} \mathbf{P}(|M_{p,n}^2/(1 - p^{-2/(n-1)}) - 1| > \delta) \longrightarrow 0. \quad \text{(II.7)}$$

2) (Sharpness of SABRE)

Similarly, suppose we observe n i.i.d. samples of independent Gaussian variables $X_1, \ldots, X_p$ and an arbitrarily distributed random variable Y that is independent of X_j's. Consider the maximal absolute sample correlation $M_{XY} = \max_{1 \leq j \leq p} |\widehat{C}(X_j, Y)|$. Uniformly for all $n \geq 3$, as $p \to \infty$, we have

$$M_{XY}/\sqrt{1 - p^{-2/(n-2)}} \xrightarrow{prob.} 1. \quad \text{(II.8)}$$

Theorem II.3 shows the sharpness of the SABER and the SABRE. It further describes a smooth phase transition of $M_{p,n}$ (also M_{XY}) depending on the limit of $\frac{\log p}{n}$:

(i) If $\lim_{p \to \infty} \log p/n = \infty$, then $M_{p,n} \xrightarrow{prob.} 1$ and $M_{p,n}/\sqrt{1 - p^{-2/(n-1)}} \xrightarrow{prob.} 1$.

(ii) If $\lim_{p \to \infty} \log p/n = \beta$ for fixed $0 < \beta < \infty$, then $M_{p,n} \xrightarrow{prob.} \sqrt{1 - e^{-2\beta}}$.

(iii) If $\lim_{p \to \infty} \log p/n = 0$, then $M_{p,n} \xrightarrow{prob.} 0$ and $M_{p,n}/\sqrt{2 \log p/n} \xrightarrow{prob.} 1$.

Note in particular that when $\lim_{p \to \infty} \log p/n = 0$, the SABRE satisfies

$$\begin{aligned}
&\sqrt{1 - p^{-2/(n-2)}} \\
=&\sqrt{1 - e^{-2\log p/(n-2)}} \\
\sim&\sqrt{1 - (1 - 2\log p/n)} \\
=&\sqrt{2 \log p/n}.
\end{aligned} \quad \text{(II.9)}$$

The rate $\sqrt{2 \log p/n}$ has appeared in hundreds of books and papers and is very-well known in high-dimensional statistics literature [35]. However, it is just a special case of the general rate $\sqrt{1 - p^{-2/(n-2)}}$, which is obtained through the packing perspective. This fact demonstrates the power of this packing approach. In Figure 1, the smooth phase transition curves $\frac{\log p}{n} \approx \beta$ are represented as regions of the same color.

Below are some geometric intuitions on why the phase transition depends on the limit of $\frac{\log p}{n}$: Note that the number of orthants in $\mathbb{R}^n$ is 2^n and is growing exponentially in n. Therefore, if the growth of p is faster than the exponential rate in n, then the p unit vectors on $\mathcal{S}^{n-1}$ would be so "dense" that they would cover the sphere, making the magnitude of the maximal inner product converging to 1; if the growth of p is exponential in n, then there would be a constant number (depending on the limit of $\frac{\log p}{n}$) of points in each orthant, so that the new random point would stay around some proper angle to the existing points; if the growth of p is slower than the exponential rate, then many orthants would be empty of points asymptotically, thus the new random point can be almost orthogonal to the existing points.

When $\boldsymbol{L}_j$'s are i.i.d. uniformly distributed or when X_j's are independently Gaussian, by combining the results in packing literature [22, 26] and classical extreme value theory [36, 37], we further develop the following uniform convergence in distribution of the corresponding maxima.

Theorem II.4.

1) (Limiting Distribution of the Maximal Independent Inner Product)

Suppose $\boldsymbol{L}_j$'s are i.i.d. uniformly unit vectors over $\mathcal{S}^{n-1}$. For arbitrary random unit vector $\boldsymbol{U}$ that is independent of $\boldsymbol{L}_j$'s, consider $M_{p,n} = \max_{1 \leq j \leq p} |\langle \boldsymbol{L}_j, \boldsymbol{U} \rangle|$. Let

$$a_{p,n} = 1 - p^{-2/(n-1)}c_{p,n}, \qquad b_{p,n} = \frac{2}{n-1}p^{-2/(n-1)}c_{p,n},$$

where $c_{p,n} = \left(\frac{n-1}{2}B\left(\frac{1}{2}, \frac{n-1}{2}\right)\sqrt{1 - p^{-2/(n-1)}}\right)^{2/(n-1)}$ is a correction factor with $B(s,t)$ being the Beta function. Then for any fixed x, as $p \to \infty$,

$$\begin{aligned}
\sup_{n \geq 2} &\left| \mathbf{P}\left(\frac{M_{p,n}^2 - a_{p,n}}{b_{p,n}} < x \right) - I\left(x > \frac{n-1}{2} \right) \right. \\
&\left. - \exp\left(- \left(1 - \frac{2}{n-1}x\right)^{(n-1)/2} \right) I\left(x \leq \frac{n-1}{2} \right) \right| \to 0.
\end{aligned} \quad \text{(II.10)}$$

In particular, if $n \to \infty$ and $p \to \infty$, then for any fixed x, we have the double limit

$$\mathbf{P}\left(\frac{M_{p,n}^2 - a_{p,n}}{b_{p,n}} < x \right) \to \exp\left(- e^{-x} \right). \quad \text{(II.11)}$$

2) (Limiting Distribution of the Maximal Spurious Correlation)

Similarly, suppose we observe n i.i.d. samples of independent Gaussian variables $X_1, \ldots, X_p$ and an arbitrarily distributed random variable Y that is independent of X_j's. Consider the maximal absolute sample correlation $M_{XY} = \max_{1 \leq j \leq p} |\widehat{C}(X_j, Y)|$. Then for any fixed x, as $p \to \infty$,

$$\begin{aligned}
\sup_{n \geq 3} &\left| \mathbf{P}\left(\frac{M_{XY}^2 - a_{p,n-1}}{b_{p,n-1}} < x \right) - I\left(x > \frac{n-2}{2} \right) \right. \\
&\left. - \exp\left(- \left(1 - \frac{2}{n-2}x\right)^{(n-2)/2} \right) I\left(x \leq \frac{n-2}{2} \right) \right| \to 0.
\end{aligned} \quad \text{(II.12)}$$

In particular, if $n \to \infty$ and $p \to \infty$, then for any fixed x, we have the double limit

$$\mathbf{P}\left(\frac{M_{XY}^2 - a_{p,n-1}}{b_{p,n-1}} < x \right) \to \exp\left(- e^{-x} \right). \quad \text{(II.13)}$$

Theorem II.4 characterizes the uncertainty of the maximal independent inner product and the maximal spurious correlation from the SABER and SABRE respectively. This result possesses the following desirable properties for practice: (1) The convergence of $M_{p,n}$ (M_{XY}) is uniform for $n \geq 2$ ($n \geq 3$) and is applicable provided the dataset contains two (three) observations. This uniformity over n is due to the packing perspective. (2) The convergence is arbitrary for any distribution of Y. This arbitrariness results from the invariance property of the uniform distribution over the sphere. (3) The convergence is adaptive to the number of variables p: Despite the phase transition phenomenon, the normalizing constants $a_{p,n}$ and $b_{p,n}$ adaptively adjust themselves for different n and p to guarantee a good approximation to a proper limiting distribution. (4) Instead of the "curse of dimensionality," the convergence is a "blessing of dimensionality": The larger p is, the better the approximation is. These properties make

the result widely applicable in the high-dimension-and-low-sample size situations.

We also remark here that for statistical applications, although in principle the empirical distribution of M_{XY} can be simulated based on the Gaussian assumptions, in a large-p situation, for example $p = 10^{10}$, such simulation can incur extremely high time and computation cost. On the other hand, these quantiles can be easily obtained through the formulas of $a_{p,n}$ and $b_{p,n}$ for an arbitrary large p. Indeed, in modern data analysis, it is more and more often to encounter datasets with a number of variables in millions, billions, or even larger scales [45]. The uniform-n-large-p type asymptotics presented in this paper can be especially useful in these situations.

III. RANK-EXTREME ASSOCIATION OF DEGENERATE ELLIPTICAL VECTORS

A. Rank-Extreme Bound of Degenerate Elliptical Vectors

In this section we consider the maximal magnitude of an elliptically distributed vector. A p-dimensional random vector V is said to be elliptically distributed and is denoted as $V \sim \mathcal{EC}_p(\boldsymbol{\xi}, \boldsymbol{\Theta})$ if its density $f(\boldsymbol{v})$ satisfies that

$$f(\boldsymbol{v}) \propto g((\boldsymbol{v} - \boldsymbol{\xi})^T \boldsymbol{\Theta}^{-1} (\boldsymbol{v} - \boldsymbol{\xi})) \quad \text{(III.1)}$$

for some continuous integrable function $g(\cdot)$ so that its isodensity contours are ellipses. The family of elliptical distributions is a generalization of multivariate Gaussian distributions and is an important and general class of distributions in practice [46].

In this paper, we focus on an elliptical distributed vector $X \sim \mathcal{EC}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ with a covariance matrix $\boldsymbol{\Sigma}$ that has unit diagonals. Through a packing argument, we find a functional link between the distribution of $\max_{1 \le j \le p} |X_j|$ and the rank of $\boldsymbol{\Sigma}$. we thus refer this link as the rank-extreme (ReX) association.

Below are the observations that connect these results to the packing problem: Consider any $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ that is positive semi-definite, has ones on the diagonal, and has rank d. Through its eigen-decomposition, we can write $\boldsymbol{\Sigma} = \mathbf{L}^T \mathbf{L}$, where $\mathbf{L} = [\boldsymbol{L}_1, \dots, \boldsymbol{L}_p]$ is a $d \times p$ matrix with columns $\boldsymbol{L}_j$'s such that $\|\boldsymbol{L}_j\|_2 = 1$. Thus, we can write $X = \mathbf{L}^T Z$ where $Z \sim \mathcal{EC}_d(\boldsymbol{0}, \mathbf{I})$. Moreover, for any $Z \sim \mathcal{EC}_d(\boldsymbol{0}, \mathbf{I})$, if we consider the spherical coordinates, then we have $Z = \|Z\|_2 U$ where $U \sim Unif(\mathcal{S}^{d-1})$. Note that $\|Z\|_2$ is a random variable which depends only on d. We thus assume $\|Z\|_2$ is a random variable such that $\frac{\|Z\|_2^2 - u_d}{v_d} \xrightarrow{dist.} F_\infty$ and $\frac{\|Z\|_2^2}{u_d} \xrightarrow{prob.} 1$ where u_d and v_d are sequences of constants that depends only on d, and F_∞ is a proper random variable. Note also that $\|Z\|_2$ and U are independent. Based on the above consideration, we obtain the following decomposition

$$\|X\|_\infty = \max_j |X_j| = \max_j |\langle \boldsymbol{L}_j, Z \rangle| = \|Z\|_2 \max_j |\langle \boldsymbol{L}_j, U \rangle|. \quad \text{(III.2)}$$

Since the distribution of the maximal absolute inner products $\max_j |\boldsymbol{L}_j, U|$ is studied in Section II, we can apply these asymptotic results to study the distribution of $\|X\|_\infty = \max_j |X_j|$. In particular, we develop the following universal bound on a degenerate elliptically distributed vector X with

a particular case of a degenerate Gaussian vector, where $\|Z\|_2^2 \sim \chi_d^2$ with $u_d = d$.

Theorem III.1.

1) (**ReX Bound for Degenerate Elliptical Vectors**)
   For any vector of p standard elliptical variables $X \sim \mathcal{EC}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $rank(\boldsymbol{\Sigma}) = d$, the random variable $\|X\|_\infty = \max_j |X_j|$ satisfies that for any fixed $\delta > 0$,

$$\lim_{p,d \to \infty} \mathbf{P}\left(\|X\|_\infty / \sqrt{u_d(1 - p^{-2/(d-1)})} > 1 + \delta \right) = 0. \quad \text{(III.3)}$$

2) (**ReX Bound for Degenerate Gaussian Vectors**)
   In particular, for any vector of p standard Gaussian variables $X \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $rank(\boldsymbol{\Sigma}) = d$, the random variable $\|X\|_\infty = \max_j |X_j|$ satisfies that for any fixed $\delta > 0$,

$$\lim_{p,d \to \infty} \mathbf{P}\left(\|X\|_\infty / \sqrt{d(1 - p^{-2/(d-1)})} > 1 + \delta \right) = 0. \quad \text{(III.4)}$$

   If further $d = d(p)$ with $\lim_{p \to \infty} (\log \log p)^2 d / (\log p)^2 \to \infty$, then

$$\lim_{p \to \infty} \mathbf{P}\left(\|X\|_\infty / \sqrt{d(1 - p^{-2/(d-1)})} \le 1 \right) = 1. \quad \text{(III.5)}$$

Similar to the SABER $\sqrt{1 - p^{-2/(n-1)}}$, this bound is universal over any correlation structures of rank d. We also show that this bound is sharp, as described in Section III-B.

B. Attainment of the ReX Bound and Related Limiting Distributions

The sharpness of the bound in Theorem III.1 was shown by considering the case when $\boldsymbol{L}_j$'s in the decomposition (III.2) are i.i.d. uniformly distributed over $\mathcal{S}^{d-1}$.

Theorem III.2. (**Sharpness of ReX Bounds**) *If $\boldsymbol{L}_j$'s are i.i.d. uniformly distributed over the $(d-1)$-sphere $\mathcal{S}^{d-1}, \forall j$ and are independent of $Z \sim \mathcal{EC}_d(\boldsymbol{0}, \mathbf{I})$, then as $d \to \infty$ and $p \to \infty$,*

$$\max_j |\langle \boldsymbol{L}_j, Z \rangle| / \sqrt{u_d(1 - p^{-2/(d-1)})} \xrightarrow{prob.} 1, \quad \text{(III.6)}$$

i.e., $\forall \delta > 0$,

$$\lim_{p,d \to \infty} \mathbf{P}\left(\left| \max_j |\langle \boldsymbol{L}_j, Z \rangle| / \sqrt{u_d(1 - p^{-2/(d-1)})} - 1 \right| > \delta \right) = 0. \quad \text{(III.7)}$$

In particular, if $Z \sim \mathcal{N}_d(\boldsymbol{0}, \mathbf{I})$, then as $d \to \infty$ and $p \to \infty$,

$$\max_j |\langle \boldsymbol{L}_j, Z \rangle| / \sqrt{d(1 - p^{-2/(d-1)})} \xrightarrow{prob.} 1. \quad \text{(III.8)}$$

One remark here is that though each realization of $\boldsymbol{L}_j$'s results in a degenerate elliptically distributed X, unconditionally the joint distribution of X is not elliptically distributed. Nevertheless, Theorem III.2 shows the existence of configurations of $\boldsymbol{L}_j$ that attains the bound in Theorem III.1.

The limit in Theorem III.2 indicates the following phase transition for the extreme value in degenerate Gaussian vectors, again depending on the limit of $\frac{\log p}{d}$:

(i) If $d \rightarrow \infty$ and $\lim_{p \rightarrow \infty} \log p / d = \infty$, then $\max_j |\langle \boldsymbol{L}_j, \boldsymbol{Z} \rangle| / \sqrt{d} \xrightarrow{prob.} 1$.

(ii) If $\lim_{p \rightarrow \infty} \log p / d = \beta$ for fixed $0 < \beta < \infty$, then $\max_j |\langle \boldsymbol{L}_j, \boldsymbol{Z} \rangle| / \sqrt{\log p} \xrightarrow{prob.} \sqrt{(1 - e^{-2\beta})/\beta}$.

(iii) If $\lim_{p \rightarrow \infty} \log p / d = 0$, then $\max_j |\langle \boldsymbol{L}_j, \boldsymbol{Z} \rangle| / \sqrt{2 \log p} \xrightarrow{prob.} 1$.

Note that the function $f(\beta) = (1 - e^{-2\beta})/\beta$ is a smooth function for $\beta > 0$ and its range is $(0, 2)$. Thus, as the phase transition in Section II-B, the above phase transition is smooth. Moreover, the regime (iii) in the phase transition implies that when the rank d is high compared to $\log p$, the maximum magnitude of a degenerate Gaussian vector can behave as that of i.i.d. Gaussian vectors.

Note that by (III.2), we have the decomposition of the squared maximum norm

$$\|\boldsymbol{X}\|_\infty^2 = \max_{1 \le j \le p} |\langle \boldsymbol{L}_j, \boldsymbol{Z} \rangle|^2 = \|\boldsymbol{Z}\|_2^2 M_{p,d}^2. \qquad (III.9)$$

Thus, by the results in Section II-B, we also develop the following result on the limiting distribution of a degenerate elliptical vector when $\boldsymbol{L}_j$'s are i.i.d. uniform.

Theorem III.3.

1) *(**Limiting Distribution of the Maximum of Degenerate Elliptical Vectors**)*
   Suppose $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_p \overset{iid}{\sim} Unif(\mathcal{S}^{d-1})$ and $\boldsymbol{Z} \sim \mathcal{EC}_d(\boldsymbol{0}, \mathbf{I})$ with $\frac{\|\boldsymbol{Z}\|_2^2 - u_d}{v_d} \xrightarrow{dist.} F_\infty$ for some sequences u_d, v_d and a proper random variable F_∞. Then with the constants $a_{p,d}$ and $b_{p,d}$ as in Theorem II.4, the random variable $K_{p,d} = \max_{1 \le j \le p} |\langle \boldsymbol{L}_j, \boldsymbol{Z} \rangle|^2 = \|\boldsymbol{Z}\|_2^2 M_{p,d}^2$ has the following limiting distribution:

   a) *If d is fixed and $p \rightarrow \infty$, then $K_{p,d} \xrightarrow{dist.} \|\boldsymbol{Z}\|_2^2$.*
   b) *Suppose $d \rightarrow \infty$ and $p \rightarrow \infty$.*
      i) *If $d \rightarrow \infty, p \rightarrow \infty$, and $\frac{v_d a_{p,d}}{u_d b_{p,d}} \rightarrow \infty$, then $\frac{K_{p,d} - u_d a_{p,d}}{v_d a_{p,d}} \xrightarrow{dist.} F_\infty$.*
      ii) *If $d \rightarrow \infty, p \rightarrow \infty$, and $\frac{v_d a_{p,d}}{u_d b_{p,d}} \rightarrow c$ with $0 < c < \infty$, then $\frac{K_{p,d} - u_d a_{p,d}}{v_d a_{p,d}} \xrightarrow{dist.} F_\infty + \frac{1}{c} H$. where $H \sim Gumbel(0,1)$, and F_∞ and H are independent.*
      iii) *If $d \rightarrow \infty, p \rightarrow \infty$, and $\frac{v_d a_{p,d}}{u_d b_{p,d}} \rightarrow 0$, then $\frac{K_{p,d} - u_d a_{p,d}}{u_d b_{p,d}} \xrightarrow{dist.} H$ where $H \sim Gumbel(0,1)$.*

2) *(**Limiting Distribution of the Maximum of Degenerate Gaussian Vectors**)*
   In particular, if $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \mathbf{I})$, then the random variable $K_{p,d}$ has the following limiting distribution:

   a) *If d is fixed and $p \rightarrow \infty$, then $K_{p,d} \xrightarrow{dist.} \chi_d^2$.*
   b) *Suppose $d \rightarrow \infty$ and $p \rightarrow \infty$.*
      i) *If $d \rightarrow \infty, p \rightarrow \infty$, and $(\log p)^2 / d \rightarrow \infty$, then $\frac{K_{p,d} - d a_{p,d}}{\sqrt{2d} a_{p,d}} \xrightarrow{dist.} G$ where $G \sim \mathcal{N}(0,1)$.*
      ii) *If $d \rightarrow \infty, p \rightarrow \infty$, and $(\log p)^2 / d \rightarrow c$ with $0 < c < \infty$, then $\frac{K_{p,d} - d a_{p,d}}{\sqrt{2d} a_{p,d}} \xrightarrow{dist.} G + \frac{1}{\sqrt{2c}} H$ where $G \sim \mathcal{N}(0,1)$, $H \sim Gumbel(0,1)$, and G and H are independent.*
      iii) *If $d \rightarrow \infty, p \rightarrow \infty$, and $(\log p)^2 / d \rightarrow 0$, then $\frac{K_{p,d} - d a_{p,d}}{d b_{p,d}} \xrightarrow{dist.} H$ where $H \sim Gumbel(0,1)$.*

Theorem III.3 characterizes the limiting distribution of the squared maximum norm of degenerate elliptical vectors for the entire scope of the rank. The limiting distribution takes on a phase transition phenomenon according to the cross ratio between standardizing constants in the convergence of the norm and the convergence of the maximal squared inner product. This phenomenon is similar as the phase transitions in the classical extreme value theory for correlated random variables [36–38]. When $\boldsymbol{Z}$ is standard Gaussian distributed, the limiting distribution can be either χ_d^2, standard Gaussian, a mixture of the standard Gaussian and Gumbel, or Gumbel depending on the relationship between d and p.

IV. ReX Detection of Low-Dimensional Linear Dependency

In this section we consider the problem of detection of low-rank dependency in high-dimensional Gaussian data. Suppose we have n observations of a Gaussian vector $\boldsymbol{W} \in \mathbb{R}^p$ whose covariance matrix $\boldsymbol{\Sigma}$ has rank is $rank(\boldsymbol{\Sigma}) = d \ll p$. One common technique in estimating d is eigenvalue thresholding based on the principal component analysis (PCA). However, such methods become inaccurate when n is small. Moreover, statistical inference, such as tests and confidence intervals, about d as a parameter is not completely clear.

We propose to apply the rank-extreme association to obtain the information about d. We consider the following generating process of the data matrix $\mathbf{W}_{n \times p}$ from a factor model:

$$\mathbf{W}_{n \times p} = \mathbf{1}_n \boldsymbol{\mu}^T + \mathbf{Z}_{n \times d} \mathbf{L}_{d \times p} \mathbf{T}_{p \times p} + \sigma \mathbf{G}_{n \times p}, \qquad (IV.1)$$

where $\boldsymbol{\mu}$ is a fixed p-dimensional vector, $\mathbf{Z}_{n \times d}$ has i.i.d. $\mathcal{N}(0,1)$ entries, $\mathbf{L}_{d \times p}$ has columns of unit vectors, $\mathbf{T}_{p \times p}$ is a diagonal matrix with positive diagonal elements $\tau_1, \ldots, \tau_p$, $\mathbf{G}_{n \times p}$ has i.i.d. $\mathcal{N}(0,1)$ entries as the observation noises, and $\sigma \ge 0$ is the standard deviation of the noise. $\mathbf{Z}$ and $\mathbf{G}$ are mutually independent so that each entry W_{ij} is marginally distributed as $\mathcal{N}(\mu_j, \tau_j^2 + \sigma^2)$. All of the above variables are not observed except for the data matrix $\mathbf{W}$, and our goal is to estimate the rank d with these observations.

Conventional estimate of d is through a proper threshold over the eigenvalues of the sample covariance matrix of $\mathbf{W}$. Such an approach requires the eigenvalues to be at least $O(\sigma^2 \sqrt{\frac{p}{n}})$ for possible detection, as shown in equation (7) and Theorem 1 in [14]. In [14], the authors consider the case when $p = O(n)$ so that this required magnitude is $O(1)$. In general, to set this required magnitude to be $O(1)$ is equivalent to set $\sigma^2 = O(\sqrt{n/p})$.

In what follows, we introduce our ReX method for the inference of d based on the observed extreme values. We consider both the case when the columns are i.i.d. uniform unit vectors and the general case.

A. The Case When the Columns of $\mathbf{L}$ are i.i.d. Uniform Unit Vectors

We first consider the case when the columns of $\mathbf{L}$ are realizations of i.i.d. uniform unit vectors over $\mathcal{S}^{d-1}$. To explain our ReX method, we start with the elementary noiseless case when it is known that $\boldsymbol{\mu} = \boldsymbol{0}$, $\sigma = 0$, and $\tau_j = 1$. In this

case, we propose to approximate the asymptotic distribution of the maximal squared entry in each row of $\mathbf{W}$ by that of $K_{p,d}$. This approximation is particularly useful when $n \ll p$, where obtaining the spectrum information is difficult from PCA based methods. The accuracy of the approximation is due to the following two reasons: (1) the theorems in Section III are for each row of $\mathbf{W}$ and have no requirement on n; (2) for each row, the condition $\sigma^2 = O(\sqrt{n/p})$ in turn shows that the largest magnitude of noise in each row of $\mathbf{W}$ is in the order of $O_p(\sqrt{2\log p}(n/p)^{1/4})$. Thus, when $n/p \to 0$, this magnitude is $o_p(1)$ and will not affect the limiting distributions.

Note that for a large p, Theorem II.4, the χ_d^2 distribution, and the generalized extreme value distribution [37] imply that

$$
\begin{aligned}
&\mathbf{E}[M_{p,d}^2] \\
=&\mathbf{E}[\max_j |\langle \mathbf{L}_j, \mathbf{U}\rangle|^2] \\
\sim& m_{p,d} := a_{p,d} + \frac{d-1}{2}(1 - \Gamma(1 + 2/(d-1)))b_{p,d} \\
&\mathrm{Var}[M_{p,d}^2] \\
\sim& v_{p,d} := \frac{(d-1)^2 b_{p,d}^2}{4}\left(\Gamma(1 + 4/(d-1)) - \Gamma(1 + 2/(d-1))^2\right)
\end{aligned}
$$
(IV.2)

where $a_{p,d}$ and $b_{p,d}$ are as in Theorem II.4. Thus, through (III.9) and Theorem III.2:

$$
\begin{aligned}
\mathbf{E}[K_{p,d}] &\sim E_{p,d} := dm_{p,d}, \\
\mathrm{Var}[K_{p,d}] &\sim V_{p,d} := 2d(v_{p,d} + m_{p,d}^2) + d^2 v_{p,d}.
\end{aligned}
$$
(IV.3)

Suppose we observe n i.i.d. samples of $K_{p,d}$ which are denoted as $K_{1,p,d}, \ldots, K_{n,p,d}$. By the central limit theorem we have

$$
\sqrt{n}\frac{\bar{K}_{p,d} - E_{p,d}}{\sqrt{V_{p,d}}} \xrightarrow{dist.} G
$$
(IV.4)

where $\bar{K}_{p,d} = \frac{1}{n}\sum_{i=1}^n K_{i,p,d}$ and $G \sim \mathcal{N}(0,1)$. An easy estimate of d is thus the solution of the equation

$$
\bar{K}_{p,d} = E_{p,d}.
$$
(IV.5)

The estimators from this approach usually have a right-skewed distribution, as the distribution of χ_d^2 and $M_{p,n}^2$ are both right-skewed. To reduce the right-skewness in the distribution of $K_{p,d}$, we take the square-root transformation and use the delta method as in [47] to obtain the following approximate probabilistic statement

$$
\mathbf{P}\left(\bar{K}_{p,d} \geq \left(z_\alpha\sqrt{\frac{V_{p,d}}{4nE_{p,d}}} + \sqrt{E_{p,d}}\right)^2\right) \approx 1 - \alpha
$$
(IV.6)

where $0 < \alpha < 1$ and z_α is the α-quantile of the standard Gaussian distribution. One then solves the inequality

$$
\bar{K}_{p,d} \geq \left(z_\alpha\sqrt{\frac{V_{p,d}}{4nE_{p,d}}} + \sqrt{E_{p,d}}\right)^2
$$
(IV.7)

in d to obtain the $(1 - \alpha)$-left-sided confidence interval from 0 to this solution. Thus, probability statements about an unknown d can be made. Note that n needs not to be larger than d throughout this approach.

Another advantage of the proposed inference method is the speed. Note that through the rank-extreme approach, there is

no need of matrix multiplications. By quickly checking the maximal entry in each row, we may get a good sense of the rank as a parameter. Thus, much computation cost can be saved from the rank-extreme approach, and the proposed inference method for d can be used for a fast detection of a low-rank.

When the parameters $\boldsymbol{\mu}$, σ, and τ_j's are unknown, we would need to estimate them. Since we are considering the case when n is small while p is large, the estimation of each component variance $\tau_j^2 + \sigma^2$ is difficult. However, when it is known that τ_j's are equal to some unknown τ, we can estimate the variance $\tau^2 + \sigma^2$ by borrowing strength from all variables. Specifically, we propose the following procedure for the inference of d:

1) Center each column of $\mathbf{W}$ by subtracting the column averages. Denote the resulting data matrix by $\mathbf{W}_0$.
2) Stack the columns of $\mathbf{W}_0$ into an $(np) \times 1$ vector and estimate the component standard deviation $\sqrt{\tau^2 + \sigma^2}$ with this the sample standard deviation of this vector. Denote the estimate by s.
3) Standardize $\mathbf{W}_0$ by dividing s. Denote the resulting data matrix by $\mathbf{W}_s$.
4) Apply the approach in the noiseless case above to $\mathbf{W}_s$ for inference about d.

The above consideration is also applicable to the situation when the variables can be grouped into several blocks and the component variances within each block are close. Tests of equality variances such as [48] are widely available. We will study the case with unequal variances in future work.

B. General Case

In this section we discuss the much more challenging situation when the columns of $\mathbf{L}$ are general unit vectors. For simplicity we restrict ourselves in the case when it is known that $\boldsymbol{\mu} = \mathbf{0}$, $\sigma = 0$, and $\tau_j = 1$. We observe that by the decomposition (III.2), we have the following proposition:

Proposition IV.1. *Suppose $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ has unit diagonals and $\mathrm{rank}(\boldsymbol{\Sigma}) = d$. If there exists a collection of deterministic unit vectors $\mathbf{L}_j$'s in $\mathbb{R}^d$ such that $\boldsymbol{\Sigma} = \mathbf{L}^T\mathbf{L}$ where $\mathbf{L} = [\mathbf{L}_1, \ldots, \mathbf{L}_p]$ and that for an independent uniformly distributed unit vector $\mathbf{U} \in \mathbb{R}^d$, $\max_j |\langle \mathbf{L}_j, \mathbf{U}\rangle| \xrightarrow{prob.} 1$ as $p \to \infty$, then as $p \to \infty$,*

$$
\max_j |X_j|^2 \xrightarrow{dist.} \chi_d^2.
$$
(IV.8)

With this proposition, we convert the inference about d as a parameter to a simple inference problem on the degrees of freedom of a χ^2 distribution. The condition $\max_j |\langle \mathbf{L}_j, \mathbf{U}\rangle| \xrightarrow{prob.} 1$ is a condition on $\boldsymbol{\Sigma}$ as $p \to \infty$. It requires that the p vectors $\mathbf{L}_j$'s be "densely" distributed over the unit sphere in $\mathbb{R}^d$ as p increases, so that the minimal angle between the collection of $\mathbf{L}_j$'s and the vector $\mathbf{U}$ converges to 0 as the number of points on the unit sphere increases. The existence of such a $\boldsymbol{\Sigma}$ is shown by the sharpness of the SABER. We aren't able to find a more precise condition on $\boldsymbol{\Sigma}$ to guarantee the convergence as it

relates to the challenging question of the optimal configuration of spherical cap packing and spherical code, on which some recent development includes [49]. However, as long as $\lim_{p \to \infty} \mathbf{P}(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U} \rangle| \geq 1 - \delta) \geq 1 - \varepsilon$ for some δ and ε, by conditioning on this event, inference such as confidence intervals can be made about d as a parameter. Unfortunately, as many conditions in statistical literature, neither of these above conditions can be checked in practice. We will consider further analysis on this approach in future work.

V. SIMULATION STUDIES

In this section we study the performance of the ReX detection of a low-rank from the model in Section IV-A. We consider two cases: (1) the case when it is known that $\boldsymbol{\mu} = \boldsymbol{0}$, $\sigma = 0$, and $\tau_j = 1$ and (2) the case when the unknown component variances $\tau_j^2 = \tau^2$ for some unknown τ.

A. Noiseless Case

In this subsection, we study the performance of the ReX detection when it is known that $\boldsymbol{\mu} = \boldsymbol{0}$, $\sigma = 0$, and $\tau_j = 1$. We set $p = 8000$, n to be from $\{10, 20, 30\}$, and d to be from $\{11, 16, 21\}$. In this case, the estimation of d can be obtained by solving (IV.5), and the confidence interval can be obtained by solving (IV.7). We evaluate the performance of the ReX inference for d with two criteria: (1) the sample mean squared error (MSE) of the point estimate of d which is defined by

$$MSE_{\widehat{d}} = \frac{1}{N} \sum_{k=1}^{N} (d - \widehat{d}_k)^2 \qquad \text{(V.1)}$$

where N is the number of simulations, and $\widehat{d}_k$ is the estimate of d from the k-th simulated data, $k = 1, \ldots, N$; and (2) the coverage and 95% upper bounds for d. As a comparison, we also study the MSE of an important PCA-based method, the KN method, proposed in [14] by applying $\mathbf{W}$ to the algorithm posted on the authors' website.

Table I represents simulation results on the performance of the ReX inference for different n's and d's. The results are based on 1000 simulated datasets. The first block in the table summarizes the MSE of the ReX estimation and the KN estimation. The second block shows the average coverage probability and the mean and median length of 95% left-sided confidence intervals for d. When (IV.7) does not have a solution, we record the confidence interval as not covering d.

In terms of estimation, although the MSE of the ReX estimation seems larger than that of the KN method in some cases, we noticed that in seven out of nine scenarios the KN method actually returns $n - 2$ as an estimate of d. Indeed, the consistency of the KN method is shown when n and p are large and comparable, whereas its consistency is not guaranteed in these difficult situations when p is much larger than n. In the scenarios in our simulations, the estimations of the KN method are not consistent and can lead to serious problems in practice, particularly when $n < d$. On the other hand, we see from Table I that the MSE of the ReX estimation of d gets better as n grows. When the KN method returns better estimates, such as the cases when $n = 30$ and $d = 11$ or $d = 16$, the ReX method has a much smaller MSE.

On the performance of ReX confidence intervals, note that the standard deviation of sample proportion of 1000 Bernoulli trials with success probability 0.95 is about 0.007. Thus, a scenario with an average coverage between 0.936 and 0.964 shows a satisfactory confidence interval without being too liberal or too conservative. With this criterion, all ReX confidence intervals are satisfactory except when $n = 10$ and $d = 21$. In this case, not being able to solve (IV.7) is the main reason of not covering d in this difficult situation, see discussions at the end of this section. The length of the ReX confidence intervals is decreasing as n increases. The median lengths are less than the mean lengths, showing the distribution of the upper bound of confidence intervals is indeed right-skewed, as expected in Section IV-A.

B. Equal Variance Case

In this case, we set $p = 8000$, n to be from $\{10, 20, 30\}$, and d to be from $\{11, 16, 21\}$ as in Section V-A. We set σ to be $(n/p)^{1/4}$ as discussed in Section IV, set $\boldsymbol{\mu}$ to be a regular sequence of length p from -5 to 5, and set τ to be 2. Table II shows the results based on 1000 simulated datasets.

On the estimation, Table II shows again the problem of PCA based methods when p is much larger than n: the KN method returns $n - 2$ for seven out of nine scenarios. When $n = 30$ and $d = 11$ or $d = 16$, the KN method returns better estimates, but its MSE is larger than that of the ReX estimation. Note that in these two scenarios for the KN method as well as in all nine scenarios for the ReX method, the MSEs are much smaller than those in Table I. One possible reason here is the standardization process. For the ReX method, recall from Section IV-A that the distribution of the estimators can be right-skewed. Since the variance estimation from the sample usually underestimates $\tau^2 + \sigma^2$, the row maximum $K_{p,d}$ from standardized data can often be larger than that in the noiseless case, leading to a larger estimate of d which offsets the right-skewness in the distribution.

On the ReX confidence intervals, Table II shows that the coverage probability of them is 1 for all nine scenarios. Although the coverage probability is conservative, the lengths of intervals are reasonably tight. Also, the median upper bounds are usually less than the mean ones, showing again the right-skewness. The problem of right-skewness is much more benign though.

In summary, in our simulation studies when p is much larger than n, the traditional PCA based methods such as the KN method (1) may have a large MSE in estimating d, (2) may not be able to provide confidence intervals for d, and (3) requires matrix-wise calculation. On the other hand, the ReX inference (1) has a small MSE in estimation, (2) provides confidence interval statements for d, and (3) only needs to scan through the row maxima in the matrix and is thus fast. These results demonstrate the advantages of using the ReX method for the detection of a low-rank structure in high dimensions with a small sample size.

The simulation results also reflect some issues of the ReX method that need further improvements. For example, for some cases in Table II, the MSE of the ReX method increases as n

TABLE I
PERFORMANCE OF REX INFERENCE FOR DIFFERENT n'S AND d'S WHEN $p = 8000$ IN THE UNIT VARIANCE AND NOISELESS CASE.

	$d = 11$			$d = 16$			$d = 21$		
	$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$
MSE of Estimation									
ReX	12.40	3.97	2.91	38.07	12.69	8.73	73.52	47.21	22.03
KN	9.00	49.00	148.16	64.00	4.00	143.98	169.00	9.00	49.00
95% left-sided ReX Confidence Interval									
Coverage	0.958	0.946	0.944	0.951	0.949	0.950	0.926	0.937	0.952
Mean Upper Bound	18.21	15.15	14.20	31.41	23.66	22.13	49.32	35.74	31.01
Median Upper Bound	16.55	14.65	13.87	26.27	22.34	21.48	38.11	31.38	29.30

TABLE II
PERFORMANCE OF REX INFERENCE FOR DIFFERENT n'S AND d'S WHEN $p = 8000$ IN THE EQUAL VARIANCE WITH NOISE CASE.

	$d = 11$			$d = 16$			$d = 21$		
	$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$
MSE of Estimation									
ReX	0.49	0.65	0.82	1.93	1.52	1.51	6.89	4.24	3.42
KN	9	49	3.66	64	4	124.02	169	9	49.00
95% left-sided ReX Confidence Interval									
Coverage	1	1	1	1	1	1	1	1	1
Mean Upper Bound	17.74	15.87	15.18	28.10	24.27	22.84	41.43	34.03	31.44
Median Upper Bound	17.70	15.83	15.18	27.77	24.21	22.76	40.35	33.56	31.32

increases. This problem could be related to the approximation error in (IV.3). Also, the ReX inference are based on solutions of (IV.5) and (IV.7). Such equations may not have a solution in difficult practical situations (This happens about 1% of the time when $n = 10$ and $d = 21$). Although this problem seems to disappear when n is above 10, a more stable algorithm is needed. We shall improve our method in these directions in future work.

VI. DISCUSSIONS

We develop a probabilistic upper bound for the maximal inner product between any set of unit vectors and a stochastically independent uniformly distributed unit vector, as well as the limiting distributions of the maximal inner product when the set of unit vectors are i.i.d. uniformly distributed. We demonstrate the applications of these results the problems of spurious correlations and low-rank detections.

We emphasize that we focus our asymptotic theory in the uniform-n-large-p paradigm. This type of asymptotics is motivated by the high-dimensional-low-sample-size framework [45] which is emerging in many areas of science. The proposed packing approach can be especially useful in this framework because (1) finite-sample properties can be studied, and (2) existing packing literature can be applied. In the future, we will continue to explore this type of asymptotics in more general situations. For the theory, we plan to investigate the distribution of the maximal inner products with more generally correlated $\boldsymbol{L}_j$'s. One of the applications of the new theory could be a more accurate detection method of a low rank. We

also plan to improve and generalize the ReX detection method in the case when τ_j's are different, as well as in the case when the data are not Gaussian distributed.

APPENDIX A
TECHNICAL LEMMAS

We provide some key proofs in the appendix. Proofs of other results are immediate corollaries of these results. We start with the key observation that the distribution of each $|\langle \boldsymbol{L}_j, \boldsymbol{U} \rangle|^2$ is $Beta(1/2, (n-1)/2)$, as discussed at the beginning in Section II-A and also in [44]. Based on this fact, we first derive a lemma on the tail bounds of the $Beta(1/2, (n-1)/2)$ distribution. This lemma is proved by integration by parts, and the details are omitted.

Lemma A.1. *For $0 < w \leq 1$, we have the following bounds for an incomplete beta integral:*

$$
\frac{2((n+2)w - 1)}{(n^2 - 1)} w^{-3/2} (1-w)^{\frac{n-1}{2}}
$$
$$
\leq \int_w^1 s^{-\frac{1}{2}} (1-s)^{\frac{n-3}{2}} ds \tag{A.1}
$$
$$
\leq \frac{2}{(n-1)} w^{-1/2} (1-w)^{\frac{n-1}{2}}
$$

We also find a lemma on the uniform convergence of the function $(n-1)(p^{2/(n-1)} - 1)$. This lemma is important for the uniform convergence in the paper. The proof is easy and is omitted.

Lemma A.2. *Uniformly for any $n \geq 2$, as $p \to \infty$, $(n-1)(p^{2/(n-1)} - 1) \to \infty$.*

We derive below a lemma summarizing the uniform convergence of standardizing constants in the theorems. Their proofs are routine analysis and are omitted.

Lemma A.3. *Consider the sequences $a_{p,n} = 1 - p^{-2/(n-1)}c_{p,n}$, $b_{p,n} = \frac{2}{n-1}p^{-2/(n-1)}c_{p,n}$ in Theorem II.4 where $c_{p,n} = \left(\frac{n-1}{2}B\left(\frac{1}{2}, \frac{n-1}{2}\right)\sqrt{1 - p^{-2/(n-1)}}\right)^{2/(n-1)}$ is a correction factor. For any fixed x, let $w_{p,n} = a_{p,n} + b_{p,n}x$. We have the following asymptotic results:*

1) *Uniformly for any $n \geq 2$, as $p \to \infty$, $c_{p,n}/\left(\frac{n-1}{2}B\left(\frac{1}{2}, \frac{n-1}{2}\right)\right)^{2/(n-1)} \to 1$, $b_{p,n} \to 0$, $\frac{a_{p,n}}{1 - p^{-2/(n-1)}} \to 1$, and $\frac{b_{p,n}}{a_{p,n}} \to 0$.*

2) *Uniformly for any $n \geq 2$, as $p \to \infty$, $\frac{(n+1)w_{p,n}}{(n+2)w_{p,n} - 1}I\left(x \leq \frac{n-1}{2}\right) + I\left(x > \frac{n-1}{2}\right) \to 1$.*

APPENDIX B
PROOFS IN SECTION II

Proof of Theorem II.1. To show (II.1), note that for $\delta \geq 1/(p^{2/(n-1)} - 1)$, $(1+\delta)(1 - p^{-2/(n-1)}) \geq 1$, thus the bound is trivial. Therefore, it is enough to show the convergence for any δ that $0 < \delta < 1/(p^{2/(n-1)} - 1)$. Similarly as the proof of Theorem 6.3 in [50], by Lemma A.1 and the inequalities that $\Gamma(x + 1/2)/\Gamma(x) < \sqrt{x}$ as in [51], we have

$$
\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1+\delta)(1 - p^{-2/(n-1)})}\right)
$$
$$
\leq p\mathbf{P}\left(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1+\delta)(1 - p^{-2/(n-1)})}\right)
$$
$$
\leq p\sqrt{\frac{2}{(n-1)\pi}}\frac{(1 - (1+\delta)(1 - p^{-2/(n-1)}))^{(n-1)/2}}{\sqrt{(1+\delta)(1 - p^{-2/(n-1)})}}
$$
$$
= \sqrt{\frac{2}{\pi(1+\delta)}}\frac{p^{1/(n-1)}\left(1 - \delta(p^{2/(n-1)} - 1)\right)^{(n-1)/2}}{\sqrt{(n-1)(p^{2/(n-1)} - 1)}}
$$
$$
\leq \sqrt{\frac{2}{\pi(1+\delta)}}\frac{p^{1/(n-1)}\exp\left(-\frac{1}{2}\delta(n-1)(p^{2/(n-1)} - 1)\right)}{\sqrt{(n-1)(p^{2/(n-1)} - 1)}}
$$
(B.1)

Thus, by Lemma A.2,

$$
\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1+\delta)(1 - p^{-2/(n-1)})}\right) \to 0
$$
(B.2)

as $p \to \infty$ regardless of n.

To see (II.3), note that if $\lim_{p\to\infty} n/\log p = \beta > 0$, then $p^{1/(n-1)} \to e^{1/\beta} < \infty$. Thus we may set $\delta = 0$ to get (II.3). Also, if $n \to \infty$ but $n/\log p \to 0$, then (B.1) is further bounded by $\sqrt{\frac{2}{\pi(n-1)}}(1 + o(1))$. Thus we have (II.3). $\qquad\square$

Proof of Theorem II.3. Since we already have the upper bound, it is enough to show that for any fixed δ such that $0 < \delta < 1/2$,

$$
\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| < \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right) \to 0.
$$
(B.3)

By the independence discussed at the beginning of Section II-B, we have that for $p \to \infty$,

$$
\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| < \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right)
$$
$$
= \left(\mathbf{P}\left(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| < \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right)\right)^p
$$
$$
\leq \exp\left(-p\mathbf{P}\left(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right)\right).
$$
(B.4)

We will lower-bound the absolute value of the exponent in (B.4). By the lower bound in Lemma A.1 and the inequality that $\Gamma(x+1)/\Gamma(x + 1/2) > \sqrt{x + 1/4}$ as in [51], we have

$$
p\mathbf{P}\left(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right)
$$
$$
= \frac{p}{B(1/2, (n-1)/2)}\int_{(1-\delta)(1-p^{-2/(n-1)})}^1 x^{-1/2}(1 - x)^{(n-3)/2}dx
$$
$$
\geq \sqrt{\frac{1}{\pi(1-\delta)^3}}\frac{\sqrt{2n-3}}{n^2 - 1}\frac{p^{1/(n-1)}}{\sqrt{(p^{2/(n-1)} - 1)^3}}\cdot
$$
$$
\left(1 + \delta(p^{2/(n-1)} - 1)\right)^{(n-1)/2}\left((1 - \delta)n(p^{2/(n-1)} - 1) - 1\right)
$$
$$
\geq \sqrt{\frac{2}{\pi(1-\delta)}}\frac{p^{1/(n-1)}\left(1 + \delta(p^{2/(n-1)} - 1)\right)^{(n-1)/2}}{\sqrt{n(p^{2/(n-1)} - 1)}}(1 + o(1)).
$$
(B.5)

In the last step of (B.5), we use Lemma A.2 again. It is now easy to see that

$$
p\mathbf{P}\left(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1 - \delta)(1 - p^{-2/(n-1)})}\right) \to \infty \quad \text{(B.6)}
$$

as $p \to \infty$ regardless of the rate of $n = n(p)$, which completes the proof of Theorem II.3. $\qquad\square$

Proof of Theorem II.4. If $x \geq (n-1)/2$, then $a_{p,n} + b_{p,n}x \geq 1$ and the result is trivial. For $x < (n-1)/2$, by Lemma A.1 and Lemma A.3, uniformly for any $n \geq 2$, as $p \to \infty$,

$$
-\log\mathbf{P}\left(\frac{M_{p,n}^2 - a_{p,n}}{b_{p,n}} < x\right)
$$
$$
= -\log\{\mathbf{P}(|\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle|^2 < b_{p,n}x + a_{p,n})^p\}
$$
$$
= \frac{2p(1 - a_{p,n} - b_{p,n}x)^{(n-1)/2}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)(n-1)\sqrt{a_{p,n} + b_{p,n}x}}(1 + o(1))
$$
$$
= \frac{2p\left(1 - 1 + c_{p,n}p^{\frac{-2}{(n-1)}} - \frac{2}{n-1}c_{p,n}p^{\frac{-2}{(n-1)}}x\right)^{\frac{(n-1)}{2}}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)(n-1)\sqrt{a_{p,n}(1 + o(1))}}(1 + o(1))
$$
$$
= \left(1 - \frac{2}{n-1}x\right)^{(n-1)/2}(1 + o(1)).
$$
(B.7)

When $n \to \infty$, $\left(1 - \frac{2}{n-1}x\right)^{(n-1)/2} \to e^{-x}$, which concludes the proof. $\qquad\square$

APPENDIX C
PROOFS IN SECTION III

Proof of Theorem III.1. It is easy to show (III.3) and (III.4). To show (III.5), note that for any $0 < \varepsilon < 1$,

$$
\begin{aligned}
&\mathbf{P}\left(\|\boldsymbol{X}\|_\infty / \sqrt{d(1 - p^{-2/(d-1)})} > 1 \right) \\
=&\mathbf{P}\left(\|\boldsymbol{Z}\|_2 \max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| / \sqrt{d(1 - p^{-2/(d-1)})} > 1 \right) \\
\leq&\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1-\varepsilon)(1 - p^{-2/(d-1)})} \right) \\
&+ \mathbf{P}\left(\|\boldsymbol{Z}\|_2 > \sqrt{(1+\varepsilon)d} \right)
\end{aligned}
\tag{C.1}
$$

We will show each of the two summands in the last line can be made small with a proper choice of $\varepsilon = \varepsilon(p)$.

By the proof of Theorem II.1, we see that

$$
\begin{aligned}
&\mathbf{P}\left(\max_j |\langle \boldsymbol{L}_j, \boldsymbol{U}\rangle| > \sqrt{(1-\varepsilon)(1 - p^{-2/(d-1)})} \right) \\
\leq&\sqrt{\frac{2}{\pi(1-\varepsilon)}} \frac{p^{1/(d-1)} \exp\left(\frac{1}{2}\varepsilon(d-1)(p^{2/(d-1)} - 1) \right)}{\sqrt{(d-1)(p^{2/(d-1)} - 1)}}
\end{aligned}
\tag{C.2}
$$

Note also that $\|\boldsymbol{Z}\|_2^2 \sim \chi_d^2$. Thus by the Chernoff bound for χ_d^2 distribution,

$$
\begin{aligned}
&\mathbf{P}\left(\|\boldsymbol{Z}\|_2 > \sqrt{(1+\varepsilon)d} \right) \\
=&\mathbf{P}\left(\|\boldsymbol{Z}\|_2^2 > (1+\varepsilon)d \right) \\
\leq&((1+\varepsilon)e^{-\varepsilon})^{d/2} \\
\leq&e^{-d\varepsilon^2/6}
\end{aligned}
\tag{C.3}
$$

Due to (C.2) and (C.3), we let $\varepsilon = \varepsilon(p) = \log\log p / (4\log p)$. In the case when $\lim_{p\to\infty} \frac{(\log\log p)^2 d}{(\log p)^2} \to \infty$, both (C.2) and (C.3) converge to 0. $\qquad\square$

Proof of Theorem III.3. Note that,

$$
K_{p,d} - u_d a_{p,d} = M_{p,d}^2 (\|\boldsymbol{Z}\|_2^2 - u_d) + u_d(M_{p,d}^2 - a_{p,d})
\tag{C.4}
$$

Now note also that $a_{p,d}$ is bounded and that $M_{p,d}/a_{p,d} \xrightarrow{prob.} 1$. Therefore, the theorem follows from Slutsky's theorem by checking the limit of the ratio $v_d a_{p,d}$ and $u_d b_{p,d}$ and picking the one with a larger magnitude as the scaling factor. $\qquad\square$

REFERENCES

[1] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *National Science Review*, vol. 1, pp. 293–314, 2014.

[2] I. M. Johnstone and D. M. Titterington, "Statistical challenges of high-dimensional data," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4237–4253, 2009. [Online]. Available: http://rsta.royalsocietypublishing.org/content/367/1906/4237.abstract

[3] J. Fan, J. Lv, and L. Qi, "Sparse high dimensional models in economics," *Annual review of economics*, vol. 3, p. 291, 2011.

[4] J. Fan, S. Guo, and N. Hao, "Variance estimation using refitted cross-validation in ultrahigh dimensional regression," *J. R. Statist. Soc. B*, vol. 74, pp. 37–65, 2012.

[5] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, Applications*. New York: Springer, 2012.

[6] B. Yang, "Projection approximation subspace tracking," *Signal Processing, IEEE Transactions on*, vol. 43, no. 1, pp. 95–107, Jan 1995.

[7] D. J. Rabideau, "Fast, rank adaptive subspace tracking and applications," *Signal Processing, IEEE Transactions on*, vol. 44, no. 9, pp. 2229–2244, 1996.

[8] A. Kavčić and B. Yang, "Adaptive rank estimation for spherical subspace trackers," *Signal Processing, IEEE Transactions on*, vol. 44, no. 6, pp. 1573–1579, Jun 1996.

[9] E. C. Real, D. W. Tufts, and J. W. Cooley, "Two algorithms for fast approximate subspace tracking," *Signal Processing, IEEE Transactions on*, vol. 47, no. 7, pp. 1936–1945, 1999.

[10] M. Shi, Y. Bar-Ness, and W. Su, "Adaptive estimation of the number of transmit antennas," in *Military Communications Conference, 2007. MILCOM 2007. IEEE*. IEEE, 2007, pp. 1–5.

[11] R. Badeau, G. Richard, and B. David, "Fast and stable yast algorithm for principal and minor subspace tracking," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3437–3446, 2008.

[12] S. Bartelmaos and K. Abed-Meraim, "Fast principal component extraction using givens rotations," *Signal Processing Letters, IEEE*, vol. 15, pp. 369–372, 2008.

[13] X. G. Doukopoulos and G. V. Moustakides, "Fast and stable subspace tracking," *Signal Processing, IEEE Transactions on*, vol. 56, no. 4, pp. 1452–1465, 2008.

[14] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, no. 1, pp. 19 – 32, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743908001111

[15] ——, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *Signal Processing, IEEE Transactions on*, vol. 57, no. 10, pp. 3930–3941, Oct 2009.

[16] I. M. Johnstone and A. Y. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, vol. 104, no. 486, 2009.

[17] P. Perry and P. Wolfe, "Minimax rank estimation for subspace tracking," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 504–513, June 2010.

[18] Q. Berthet and P. Rigollet, "Optimal detection of sparse principal components in high dimension," *The Annals of Statistics*, vol. 41, no. 4, pp. 1780–1815, 08 2013. [Online]. Available: http://dx.doi.org/10.1214/13-AOS1127

[19] D. L. Donoho and M. Gavish, "The optimal hard threshold for singular values is 4/sqrt (3)," *arXiv preprint arXiv:1305.5870*, 2013.

[20] T. Cai, Z. Ma, and Y. Wu, "Optimal estimation and rank detection for sparse spiked covariance matrices," *Probability Theory and Related Fields*, pp. 1–35, 2014. [Online]. Available: http://dx.doi.org/10.1007/s00440-014-0562-z

[21] Y. Choi, J. Taylor, and R. Tibshirani, "Selecting the number of principal components: estimation of the true rank of a noisy matrix," *arXiv preprint arXiv:1410.8260*, 2014.

[22] R. A. Rankin, "The closest packing of spherical caps in n dimensions," *Proceedings of the Glasgow Mathematical Association*, vol. 2, pp. 139–144, 7 1955. [Online]. Available: http://journals.cambridge.org/article_S2040618500033219

[23] T. M. Thompson, *From Error-correcting Codes through Sphere Packings to Simple Groups*. Mathematical Association of America, 1983.

[24] J. H. Conway, N. J. A. Sloane, and E. Bannai, *Sphere-packings, Lattices, and Groups*. New York, NY, USA: Springer-Verlag New York, Inc., 1987.

[25] D. Hilbert, "Mathematical problems," *Bulletin of the American Mathematical Society*, vol. 8, no. 10, pp. 437–479, 1902.

[26] A. D. Wyner, "Random packings and coverings of the unit n-sphere," *Bell System Technical Journal*, vol. 46, pp. 2111–2118, 1967.

[27] A. Barg and D. Nogin, "Bounds on packings of spheres in the grassmann manifold," *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2450–2454, Sep 2002.

[28] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2562–2579, 2003.

[29] O. Henkel, "Sphere-packing bounds in the grassmann and stiefel manifolds," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3445–3456, 2005.

[30] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3579–3591, 2008.

[31] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on grassmann manifolds and applications to mimo communications," *Information Theory, IEEE Transactions on*, vol. 54, no. 3, pp. 1108–1123, 2008.

[32] R.-A. Pitaval, H.-L. Maattanen, K. Schober, O. Tirkkonen, and R. Wichman, "Beamforming codebooks for two transmit antenna systems based on optimum grassmannian packings," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6591–6602, 2011.

[33] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *Information Theory, IEEE Transactions on*, vol. 59, no. 12, pp. 8027–8056, Dec 2013.

[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*, 2nd ed. Springer Verlag., 2009.

[35] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*. Springer, 2011.

[36] M. R. Leadbetter, *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 1983.

[37] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*. Springer, 2006.

[38] R. J. Adler, "An introduction to continuity, extrema, and related topics for general gaussian processes," *Lecture Notes-Monograph Series*, pp. i–155, 1990.

[39] A. Hero and B. Rajaratnam, "Large-scale correlation screening," *Journal of the American Statistical Association*, vol. 106, pp. 1540–1552, 2012.

[40] T. T. Cai and T. Jiang, "Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices," *Ann. Statist*, vol. 39, pp. 1496–1525, 2011.

[41] ——, "Phase transition in limiting distributions of coherence of high-dimensional random matrices," *J. Multivariate Analysis*, vol. 107, pp. 24–39, 2012.

[42] T. T. Cai, J. Fan, and T. Jiang, "Distribution of angles in random packing on spheres," *J. Machine Learning Research*, vol. 14, pp. 1801–1828, 2013.

[43] J. Fan, Q.-M. Shao, and W.-X. Zhou, "Are discoveries spurious? distributions of maximum spurious correlations and their applications," *arXiv preprint arXiv:1502.04237*, 2015.

[44] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[45] P. Hall, J. S. Marron, and A. Neeman, "Geometric representation of high dimension, low sample size data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 3, pp. 427–444, 2005. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9868.2005.00510.x

[46] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric multivariate and related distributions*. Chapman and Hall, 1990.

[47] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

[48] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proceedings of the Royal Society of London. Series A*, vol. 160, no. 901, pp. 268–282, 1937.

[49] H. Cohn and Y. Zhao, "Sphere packing bounds via spherical codes," *Duke Mathematical Journal*, vol. 163, no. 10, pp. 1965–2002, 07 2014. [Online]. Available: http://dx.doi.org/10.1215/00127094-2738857

[50] R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao, "Valid post-selection inference," *The Annals of Statistics*, vol. 41, no. 2, pp. 802–837, 04 2013. [Online]. Available: http://dx.doi.org/10.1214/12-AOS1077

[51] G. J. O. Jameson, "Inequalities for gamma function ratios," *The American Mathematical Monthly*, vol. 120, pp. 936–940, 2013.

Kai Zhang received the Ph.D. degree in Mathematics from Temple University in 2007 and the Ph.D. degree in Statistics from the Wharton School, University of Pennsylvania, in 2012. He is now with the Department of Statistics and Operations Research at the University of North Carolina, Chapel Hill. His research interests include high-dimensional regression and inference, causal inference and observational studies, and quantum computing.