ASYMPTOTICALLY INDEPENDENT U-STATISTICS IN HIGH-DIMENSIONAL TESTING*

By Yinqiu He†, Gongjun Xu† Chong Wu‡ and Wei Pan§

University of Michigan†, Florida State University ‡, and University of Minnesota§

Many high-dimensional hypothesis tests aim to globally examine marginal or low-dimensional features of a high-dimensional joint distribution, such as testing of mean vectors, covariance matrices and regression coefficients. This paper constructs a family of U-statistics as unbiased estimators of the ℓ_p-norms of those features. We show that under the null hypothesis, the U-statistics of different finite orders are asymptotically independent and normally distributed. Moreover, they are also asymptotically independent with the maximum-type test statistic, whose limiting distribution is an extreme value distribution. Based on the asymptotic independence property, we propose an adaptive testing procedure which combines p-values computed from the U-statistics of different orders. We further establish power analysis results and show that the proposed adaptive procedure maintains high power against various alternatives.

1. Introduction.

Motivation. Analysis of high-dimensional data, whose dimension p could be much larger than the sample size n, has emerged as an important and active research area [e.g., 19, 62, 23, 21]. In many large-scale inference problems, one is often interested in globally testing some overall patterns of low-dimensional features of the high-dimensional random observations. One example is genome-wide association studies (GWAS), whose primary goal is to identify single nucleotide polymorphisms (SNPs) associated with certain complex diseases of interest. A popular approach in GWAS is to perform univariate tests which examine each SNP one by one. This however may lead to low statistical power due to the weak effect size of each SNP [46] and the small statistical significance threshold ($\sim 10^{-8}$) chosen to control the multiple-comparison type I error [39]. Researchers therefore have proposed to globally test a genetic marker set with many SNPs [63, 39] in order

*This research is supported by NSF grants DMS-1711226, DMS-1712717, SES-1659328, CAREER SES-1846747, and NIH grants R01GM113250, R01GM126002, R01HL105397 and R01HL116720

MSC 2010 subject classifications: Primary 62F03, 62F05

Keywords and phrases: High-dimensional hypothesis test, U-statistics, Adaptive testing

to achieve higher statistical power and to better understand the underlying genetic mechanisms.

In this paper, we focus on a family of global testing problems in the high-dimensional setting, including testing of mean vectors, covariance matrices and regression coefficients in generalized linear models. These problems can be formulated as $H_0 : \mathcal{E} = \mathbf{0}$, where $\mathbf{0}$ is an all zero vector, $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$ is a parameter vector with $\mathcal{L}$ being the index set, and e_l's being the corresponding parameters of interest, e.g., elements in mean vectors, covariance matrices or coefficients in generalized linear models. For the global testing problem $H_0 : \mathcal{E} = \mathbf{0}$ versus $H_A : \mathcal{E} \neq \mathbf{0}$, two different types of methods are often used in the literature. One is sum-of-squares-type statistics. They are usually powerful against "dense" alternatives, where $\mathcal{E}$ has a high proportion of nonzero elements with a large $\|\mathcal{E}\|_2 = \sum_{l\in\mathcal{L}} e_l^2$ or its weighted variants. See examples in mean testing [e.g., 4, 25, 59, 12, 11, 26, 61] and covariance testing [e.g., 3, 41, 13, 44]. The other is maximum-type statistics. They are usually powerful against "sparse" alternatives, where $\mathcal{E}$ has few nonzero elements with a large $\|\mathcal{E}\|_\infty$ [e.g., 35, 45, 27, 7, 8, 9, 57]. More recently, [20, 69] also proposed to combine these two kinds of test statistics. However, for denser or only moderately dense alternatives, neither of these two types of statistics may be powerful, as will be further illustrated in this paper both theoretically and numerically. Importantly, in real applications, the underlying truth is usually unknown, which could be either sparse, dense, or in-between. As global testing could be highly underpowered if an inappropriate testing method is used [e.g., 15], it is desired in practice to have a testing procedure with high statistical power against a variety of alternatives.

A Family of Asymptotically Independent U-Statistics. To address these issues, we propose a U-statistics framework and introduce its applications to adaptive high-dimensional testing. The U-statistics framework constructs *unbiased* and *asymptotically independent* estimators of $\|\mathcal{E}\|_a^a := \sum_{l\in\mathcal{L}} e_l^a$ for different (positive) integers a, where $a = 2$ corresponds to a sum-of-squares-type statistic, and an even integer $a \to \infty$ yields a maximum-type statistic. The adaptive testing then combines the information from different $\|\mathcal{E}\|_a^a$'s, and our power analysis shows that it is powerful against a wide range of alternatives, from highly sparse, moderately sparse to dense, to highly dense.

To illustrate our idea, suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are n independent and identically distributed (i.i.d.) copies of a random vector $\mathbf{z}$. We consider the setting where each parameter e_l has an unbiased kernel function estimator $K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})$, and γ_l is the smallest integer such that for any $1 \leq i_1 \neq \ldots \neq i_{\gamma_l} \leq n$, $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})] = e_l$. This includes many testing problems on moments

of low orders, such as entries in mean vectors, covariance matrices and score vectors of generalized linear models, which shall be discussed in details. The family of U-statistics can be constructed generally as follows. For integers $a \geq 1$, and $1 \leq i_1 \neq \ldots \neq i_{\gamma_l} \neq \ldots \neq i_{(a-1)\times\gamma_l+1} \ldots \neq i_{a\times\gamma_l} \leq n$, since the $\mathbf{z}$'s are i.i.d., we have $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}}) \cdots K_l(\mathbf{z}_{i_{(a-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{a\times\gamma_l}})] = e_l^a$. Therefore, we can construct an unbiased estimator of the parameters of augmented powers e_l^a with different a. Then $\|\mathcal{E}\|_a^a$ has an unbiased estimator

$$\mathcal{U}(a) = \sum_{l\in\mathcal{L}} (P_{a\times\gamma_l}^n)^{-1} \sum_{1\leq i_1\neq\ldots\neq i_{a\times\gamma_l}\leq n} \prod_{k=1}^{a} K_l(\mathbf{z}_{i_{(k-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{k\times\gamma_l}}), \tag{1.1}$$

where $P_k^n = n!/(n-k)!$ denotes the number of k-permutations of n. We call a the order of the U-statistic $\mathcal{U}(a)$. If $a > b$, we say $\mathcal{U}(a)$ is of higher order than $\mathcal{U}(b)$ and vice versa.

This construction procedure can be applied to many testing problems. We give three common examples below for illustration and more detailed case-studies will be discussed in Sections 2 and 4.

EXAMPLE 1. *Consider one-sample mean testing of $H_0 : \boldsymbol{\mu} = \mathbf{0}$, where $\mathcal{E} = \boldsymbol{\mu}$ is the mean vector of a p-dimensional random vector $\mathbf{x}$. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of $\mathbf{x}$. For each $i = 1, \ldots, n$, $j = 1, \ldots, p$, $x_{i,j}$ is a simple unbiased estimator of μ_j, then we can take the kernel function $K_j(\mathbf{x}_i) = x_{i,j}$. Following (1.1), we know the U-statistic*

$$\mathcal{U}(a) = (P_a^n)^{-1} \sum_{j=1}^{p} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^{a} x_{i_k,j}$$

is an unbiased estimator of $\|\mathcal{E}\|_a^a = \|\boldsymbol{\mu}\|_a^a = \sum_{j=1}^p \mu_j^a$. Please see Section 4.1 for the two-sample mean testing example and related theoretical properties.

EXAMPLE 2. *Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of a random vector $\mathbf{x}$ with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2}\}_{p\times p}$. For covariance testing $H_0 : \sigma_{j_1,j_2} = 0$ for any $1 \leq j_1 \neq j_2 \leq p$, we have $\mathcal{E} = \{\sigma_l : l \in \mathcal{L}\}$ with $\mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. Since $x_{i,j_1}x_{i,j_2}$ is a simple unbiased estimator of σ_{j_1,j_2}, then for each pair $l = (j_1, j_2) \in \mathcal{L}$, we can take the kernel function $K_l(\mathbf{x}_i) = x_{i,j_1}x_{i,j_2}$. Following (1.1), the U-statistic*

$$\mathcal{U}(a) = (P_a^n)^{-1} \sum_{1\leq j_1\neq j_2\leq p} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^{a} (x_{i_k,j_1}x_{i_k,j_2})$$

is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{1\leq j_1\neq j_2\leq p} \sigma_{j_1,j_2}^a$. Please see Section 2 for the general case with unknown $\boldsymbol{\mu}$.

EXAMPLE 3. *Consider a response variable y and its covariates $\mathbf{x} \in \mathbb{R}^p$ following a generalized linear model: $\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal\boldsymbol{\beta})$, where g is the canonical link function and $\boldsymbol{\beta} \in \mathbb{R}^p$ are the regression coefficients. Suppose that $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, are i.i.d. copies of $(\mathbf{x}, y)$. For testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, the score vectors $(S_{i,j} = (y_i - \mu_{0,i})x_{i,j} : j = 1, \ldots, p)^\intercal$ are often used in the literature, where $\mu_{0,i} = g^{-1}(\mathbf{x}_i^\intercal\boldsymbol{\beta}_0)$. Note that $\mathrm{E}(S_{i,j}) = 0$ under H_0. Thus to test H_0, we can take $\mathcal{E} = \{\mathrm{E}(S_{i,j}) : j = 1, \ldots, p\}$ and use the U-statistic*

$$\mathcal{U}(a) = (P_a^n)^{-1} \sum_{j=1}^{p} \sum_{1 \le i_1 \ne \ldots \ne i_a \le n} \prod_{k=1}^{a} S_{i_k,j},$$

which is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{j=1}^p \{\mathrm{E}(S_{i,j})\}^a$. Please see Section 4.3.

Related Literature. For high-dimensional testing, some other adaptive testing procedures have recently been proposed in [51, 66, 64]. These works combine the p-values of a family of sum-of-powered statistics that are powerful against different $\|\mathcal{E}\|_a^a$'s. However in these existing works, to evaluate the p-value of the adaptive test statistic, the joint asymptotic distribution of the statistics is difficult to obtain or calculate. Accordingly computationally expensive resampling methods are often used in practice [51, 39, 68]. For some special cases such as testing means and the coefficients of generalized linear models, [66] and [64] derived the limiting distributions of the test statistics under the framework of a family of von Mises V-statistics. However, the constructed V-statistics are usually *correlated* and *biased* estimators of the target $\|\mathcal{E}\|_a^a$. It follows that in [66] and [64], numerical approximations are still needed to calculate the tail probabilities of the adaptive test statistics; see Remark 4.1 and Section 4.3. In addition, these existing adaptive testing works mainly focus on the first-order moments, and their results do not directly apply to testing second-order moments, such as covariance matrices.

To overcome these issues, this paper considers the proposed family of unbiased U-statistics. There are some other recent works providing important results on high-dimensional U-statistics [e.g., 14, 42, 71]. For instance, [71] considered testing the regression coefficients in linear models using the fourth-order U-statistic; [42] studied the limiting distributions of rank-based U-statistics; and [14] studied bootstrap approximation of the second-order U-statistics. However, these results do not directly apply to the high-order U-statistics considered in this paper.

Our Contributions. We establish the theoretical properties of the U-statistics in various high dimensional testing problems, including testing mean vectors,

regression coefficients of generalized linear models, and covariance matrices. Our contributions are summarized as follows.

Under the null hypothesis, we show that the normalized U-statistics of different finite orders are jointly normally distributed. The result applies generally for any asymptotic regime with $n \to \infty$ and $p \to \infty$. In addition, we prove that all the finite-order U-statistics are asymptotically independent with each other under the null hypothesis. Moreover, we prove that U-statistics of finite orders are also asymptotically independent of the maximum-type test statistic with a limiting extreme value distribution.

Under the alternative hypothesis, we further analyze the asymptotic power for U-statistics of different orders. We show that when $\mathcal{E}$ has denser nonzero entries, $\mathcal{U}(a)$'s of lower orders tend to be more powerful; and when $\mathcal{E}$ has sparser nonzero entries, $\mathcal{U}(a)$'s of higher orders tend to be more powerful. More interestingly, we show that in the boundary case of "moderate" sparsity levels, $\mathcal{U}(a)$ with a finite $a > 2$ gives the highest power among the family of U-statistics, clearly indicating the inadequacy of both the sum-of-squares- and the maximum-type statistics.

An important application of the independence property among $\mathcal{U}(a)$'s is to construct adaptive testing procedures by combining the information of different $\mathcal{U}(a)$'s, whose univariate distributions or p-values can be easily combined to form a joint distribution to calculate the p-value of an adaptive test statistic. Compared with other existing works [e.g., 66, 64], numerical approximations of tail probabilities are no longer needed. As shown in the power analysis, an adaptive integration of information across different tests leads to a powerful testing procedure.

The rest of the paper is organized as follows. In Sections 2 and 3, we illustrate the framework by a covariance testing problem. Particularly, in Section 2.1, we study the U-statistics under null hypothesis; in Section 2.2, we analyze the power of the U-statistics; in Section 2.3, we develop an adaptive testing procedure. In Sections 3.1 and 3.2, we report simulations and a real dataset analysis. In Section 4, we study other high-dimensional testing problems, including testing means, regression coefficients and two-sample covariances. In Section 5, we discuss several extensions of the proposed framework. We give proofs and other stimulations in Supplementary Material.

2. Motivating Example: One-Sample Covariance Testing. The constructed family of U-statistics and adaptive testing procedure can be applied to various high-dimensional testing problems. In this section, we illustrate the framework with a motivating example of one-sample covariance testing. Analogous results for other high-dimensional testing problems in

Section 4 can be obtained following similar analyses. We showcase the study of one-sample covariance testing problem since this is more challenging than mean testing due to the two-way dependency structure and the one-sample problem can be used as the building block for more general cases.

Specifically, we focus on testing

$$H_0 : \sigma_{j_1,j_2} = 0 \quad \forall\ 1 \le j_1 \ne j_2 \le p, \tag{2.1}$$

where $\mathbf{\Sigma} = \{\sigma_{j_1,j_2} : 1 \le j_1, j_2 \le p\}$ is the covariance matrix of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$. The observed data include n i.i.d. copies of $\mathbf{x}$, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\intercal$. In factor analysis, testing H_0 in (2.1) can be used to examine whether $\mathbf{\Sigma}$ has any significant factor or not [2].

Global testing of covariance structure plays an important role in many statistical analysis and applications; see a review in [6]. Conventional tests include the likelihood ratio test, John's test, and Nagao's test, etc. [2, 49]. These methods, however, often fail in the high-dimensional setting when both $n, p \to \infty$. To address this issue, new procedures have been recently proposed [e.g., 3, 36, 37, 58, 56, 52, 41, 13, 35, 45, 7, 44, 57, 40]. However these methods might suffer from loss of power when the sparsity level of the alternative covariance matrix varies. In the following subsections, we introduce the general U-statistics framework, study their asymptotic properties, and develop a powerful adaptive testing procedure.

We introduce some notation. For two series of numbers $u_{n,p}$, $v_{n,p}$ that depend on n, p: $u_{n,p} = o(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| = 0$; $u_{n,p} = O(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} = \Theta(v_{n,p})$ denotes $0 < \liminf_{n,p\to\infty} |u_{n,p}/v_{n,p}| \le \limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} \simeq v_{n,p}$ denotes $\lim_{n,p\to\infty} u_{n,p}/v_{n,p} = 1$. Moreover, $\xrightarrow{P}$ and $\xrightarrow{D}$ represent the convergence in probability and distribution respectively. For p-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and $\forall j_1, \ldots, j_t \in \{1, \ldots, p\}$, we write the central moment as

$$\Pi_{j_1,\ldots,j_t} = \mathrm{E}[(x_{j_1} - \mu_{j_1}) \ldots (x_{j_t} - \mu_{j_t})]. \tag{2.2}$$

2.1. *Asymptotically Independent U-Statistics.* For testing (2.1), the set of parameters that we are interested in is $\mathcal{E} = \{\sigma_{j_1,j_2} : 1 \le j_1 \ne j_2 \le p\}$. Following the previous analysis of (1.1), since σ_{j_1,j_2} has a simple unbiased estimator $x_{i_1,j_1}x_{i_1,j_2} - x_{i_1,j_1}x_{i_2,j_2}$ with $1 \le i_1 \ne i_2 \le n$, then for integers $a \ge 1$, an unbiased U-statistic of $\|\mathcal{E}\|_a^a = \sum_{1\le j_1\ne j_2\le p} \sigma_{j_1,j_2}^a$ is

$$\mathcal{U}(a) = (P_{2a}^n)^{-1} \sum_{1\le j_1\ne j_2\le p} \sum_{1\le i_1\ne\ldots\ne i_{2a}\le n} \prod_{k=1}^{a} (x_{i_{2k-1},j_1}x_{i_{2k-1},j_2} - x_{i_{2k-1},j_1}x_{i_{2k},j_2}).$$

This is equivalent to

$$\mathcal{U}(a) = \sum_{1\le j_1\neq j_2\le p}\sum_{c=0}^{a}(-1)^c\binom{a}{c}\frac{1}{P_{a+c}^n}\sum_{1\le i_1\neq\ldots\neq i_{a+c}\le n}\prod_{k=1}^{a-c}(x_{i_k,j_1}x_{i_k,j_2})\prod_{s=a-c+1}^{a}x_{i_s,j_1}\prod_{t=a+1}^{a+c}x_{i_t,j_2}. \tag{2.3}$$

REMARK 2.1. *The U-statistics can be constructed by another method equivalently. Given $1 \le j_1 \ne j_2 \le p$, define $\varphi_{j_1,j_2} = \sigma_{j_1,j_2} + \mu_{j_1}\mu_{j_2}$. Then*

$$\sum_{1\le j_1\neq j_2\le p}\sigma_{j_1,j_2}^a = \sum_{1\le j_1\neq j_2\le p}\sum_{c=0}^{a}\binom{a}{c}\varphi_{j_1,j_2}^{a-c}\times(-\mu_{j_1}\mu_{j_2})^c, \tag{2.4}$$

which is a polynomial function of the moments μ_j and φ_{j_1,j_2}. Since μ_j and φ_{j_1,j_2} have unbiased estimators $x_{i,j}$ and $x_{i,j_1}x_{i,j_2}$ respectively, then for $1 \le i_1 \ne \ldots \ne i_{a+c} \le n$, $\mathrm{E}(\prod_{k=1}^{a-c}x_{i_k,j_1}x_{i_k,j_2}\prod_{s=a-c+1}^{a}x_{i_s,j_1}\prod_{t=a+1}^{a+c}x_{i_t,j_2}) = \varphi_{j_1,j_2}^{a-c}\mu_{j_1}^c\mu_{j_2}^c$. Given this and (2.4), the U-statistics (2.3) can be obtained.

REMARK 2.2. *The summed term with $c = 0$ in (2.3) is*

$$\tilde{\mathcal{U}}(a) := (P_a^n)^{-1}\sum_{1\le i_1\neq\cdots\neq i_a\le n}\sum_{1\le j_1\neq j_2\le p}\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2}), \tag{2.5}$$

which has the same form as the simplified U-statistic for mean zero observations in Example 2, and is shown to be the leading term of (2.3) in proof.

We next introduce some nice properties of the U-statistics (2.3). The first one is the following location invariant property.

PROPOSITION 2.1. *$\mathcal{U}(a)$ constructed as in (2.3) is location invariant; that is, for any vector $\boldsymbol{\Delta} \in \mathbb{R}^p$, the U-statistic constructed based on the transformed data $\{\mathbf{x}_i + \boldsymbol{\Delta} : i = 1, \ldots, n\}$ is still $\mathcal{U}(a)$.*

The following proposition verifies that the constructed U-statistics are unbiased estimators of $\|\mathcal{E}\|_a^a = \sum_{1\le j_1\neq j_2\le p}\sigma_{j_1,j_2}^a$.

PROPOSITION 2.2. *For any integer a, $\mathrm{E}[\mathcal{U}(a)] = \sum_{1\le j_1\neq j_2\le p}\sigma_{j_1,j_2}^a$. Under H_0 in (2.1), $\mathrm{E}[\mathcal{U}(a)] = 0$.*

We next study the limiting properties of the constructed U-statistics under H_0 given the following assumptions on the random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$.

CONDITION 2.1 (Moment assumption). $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(x_j-\mu_j)^8 < \infty$ *and* $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(x_j-\mu_j)^2 > 0$.

CONDITION 2.2 (Dependence assumption). *For a sequence of random variables* $\mathbf{z} = \{z_j : j \ge 1\}$ *and integers* $a < b$, *let* $\mathcal{Z}_a^b$ *be the* σ*-algebra generated by* $\{z_j : j \in \{a, \dots, b\}\}$. *For each* $s \ge 1$, *define the* α*-mixing coefficient* $\alpha_{\mathbf{z}}(s) = \sup_{t\ge 1}\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{Z}_1^t, B \in \mathcal{Z}_{t+s}^{\infty}\}$. *We assume that under* H_0, $\mathbf{x}$ *is* α*-mixing with* $\alpha_{\mathbf{x}}(s) \le M\delta^s$, *where* $\delta \in (0,1)$ *and* $M > 0$ *are some constants.*

CONDITION 2.2* (Alternative dependence assumption to Condition 2.2). *Following the notation in* (2.2), *we assume that under* H_0, *for any* $j_1, j_2, j_3 \in \{1, \dots, p\}$, $\Pi_{j_1,j_2,j_3} = 0$; *for any* $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$, $\Pi_{j_1,j_2,j_3,j_4} = \kappa_1(\sigma_{j_1,j_2}\sigma_{j_3,j_4} + \sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3})$ *for some constant* $\kappa_1 < \infty$; *and for* $t = 6, 8$, *and any* $j_1, \cdots, j_t \in \{1, \dots, p\}$, $\Pi_{j_1,\cdots,j_t} = 0$ *when at least one of these indexes appears odd times in* $\{j_1, \cdots, j_t\}$.

Condition 2.1 assumes that the eighth marginal moments of $\mathbf{x}$ are uniformly bounded from above and the second moments are uniformly bounded from below, which are true for most light-tailed distributions. Condition 2.2 assumes weak dependence among different x_j's under H_0, since the uncorrelatedness of x_j's under H_0 may not imply the independence of them, especially when x_j's are non-Gaussian. Under H_0, Condition 2.2 automatically holds when $\mathbf{x}$ is Gaussian or m-dependent. The mixing-type weak dependence is similarly considered in previous works such as [5, 11, 66] and also commonly assumed in time series and spatial statistics [24, 54]. Moreover, the variables in our motivating genome-wide association studies have a local dependence structure, with their associations often decreasing to zero as the corresponding physical distances on a chromosome increase. We note that it suffices to have Condition 2.2 hold up to a permutation of the variables.

Alternatively, we can substitute Condition 2.2 with Condition 2.2*. Condition 2.2* specifies some higher order moments of $\mathbf{x}$ and is satisfied when $\mathbf{x}$ follows an elliptical distribution with finite eighth moments and covariance $\mathbf{\Sigma}$ [see 2, 22, 49, 50]. Conditions 2.2* and 2.2 become equivalent when $\mathbf{x}$ follows a multivariate Gaussian distribution. The fourth moment condition is also assumed in other high-dimensional research [8]. In this work, the eighth moment condition is needed to establish the asymptotic joint distribution of different U-statistics.

The following theorem specifies the asymptotic variances of the finite order U-statistics and their joint limiting distribution. Since the U-statistics

are degenerate under H_0, an analysis different from the asymptotic theory on non-degenerate U-statistics [e.g., 31] is needed in the proof.

THEOREM 2.1. *Under H_0 in (2.1) and Conditions 2.1 and 2.2 (or 2.2*), for $\mathcal{U}(a)$'s defined in (2.3) and any distinct finite (and positive) integers $\{a_1, \ldots, a_m\}$, as $n, p \to \infty$,*

$$\Big[\frac{\mathcal{U}(a_1)}{\sigma(a_1)}, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)}\Big]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m), \tag{2.6}$$

where

$$\sigma^2(a) := \mathrm{var}[\mathcal{U}(a)] \simeq \frac{a!}{P_a^n} \sum_{1 \le j_1 \ne j_2 \le p;\, 1 \le j_3 \ne j_4 \le p} (\Pi_{j_1,j_2,j_3,j_4})^a, \tag{2.7}$$

with Π_{j_1,j_2,j_3,j_4} defined in (2.2). Note that $\sigma^2(a) = \Theta(p^2 n^{-a})$.

Theorem 2.1 shows that after normalization, the finite-order U-statistics have a joint normal limiting distribution with an identity covariance matrix, which implies that they are asymptotically independent as $n, p \to \infty$. The nice independence property makes it easy to combine these U-statistics and apply our proposed adaptive testing later. Moreover, the conclusion holds on general asymptotic regime for $n, p \to \infty$, without any constraint on the relationship between n and p. We will also see in Section 4 that similar results hold generally for some other testing problems.

REMARK 2.3. *Theorem 2.1 discusses the U-statistics of finite orders, i.e., the a values do not grow with n, p. When $\{x_1, \ldots, x_p\}$ are independent, Theorem 2.1 can be extended when $a = O(1) \min\{\log^\epsilon n, \log^\epsilon p\}$ for some $\epsilon > 0$. On the other hand, we will show in Section 2.2 that it is usually enough to include $\mathcal{U}(a)$'s of finite a. Therefore, we do not pursue the general case when a grows with n, p in this work.*

In the following, we further discuss the maximum-type test statistic $\mathcal{U}(\infty)$, which corresponds to the ℓ_∞-norm of the parameter vector $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$, that is, $\|\mathcal{E}\|_\infty = \max_{l \in \mathcal{L}} |e_l|$. In the existing literature, there is already some corresponding established work [35, 7] on the test statistic:

$$M_n^* := \max_{1 \le j_1 \ne j_2 \le p} |\hat{\sigma}_{j_1,j_2} / \sqrt{\hat{\sigma}_{j_1,j_1} \hat{\sigma}_{j_2,j_2}}|, \tag{2.8}$$

where $(\hat{\sigma}_{j_1,j_2})_{p \times p} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\intercal / n$ and $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i / n$. We will take $\mathcal{U}(\infty) = M_n^*$ below. The limiting distribution of $\mathcal{U}(\infty)$ was first studied in [35] and extended by [7, 45, 57]. Next we restate the result in [7], which gives the limiting distribution of (2.8) under the following condition.

CONDITION 2.3. *Consider the random vector* $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ *with mean vector* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$ *and covariance matrix* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{1,1}, \ldots, \sigma_{p,p})$. $(x_j - \mu_j)/\sqrt{\sigma_{j,j}}$ *are i.i.d. for* $j = 1, \ldots, p$. *Furthermore,* $\mathrm{E}e^{t_0(|x_1-\mu_1|/\sqrt{\sigma_{1,1}})^\varsigma} < \infty$ *for some* $0 < \varsigma \leq 2$ *and* $t_0 > 0$.

THEOREM 2.2 (Cai and Jiang [7, Theorem 2]). *Assume Condition 2.3 and* $\log p = o(n^\beta)$, *where* $\beta = \varsigma/(4+\varsigma)$. *Then* $P(n \times \mathcal{U}(\infty)^2 + \varpi_p \leq u) \to G(u) = e^{-(1/\sqrt{8\pi})e^{-u/2}}$, *where* $\varpi_p = -4\log p + \log\log p$ *and* $G(u)$ *is an extreme value distribution of type I.*

Theorems 2.1 and 2.2 give the limiting distributions of $\mathcal{U}(a)$ of finite orders and $\mathcal{U}(\infty)$ respectively; it is of interest to examine their joint distribution. The following theorem shows that although $\mathcal{U}(\infty)$ has limiting distribution different from $\mathcal{U}(a)$, $a < \infty$, they are still asymptotically independent.

THEOREM 2.3. *Assume that Condition 2.1 is satisfied, Condition 2.3 holds for* $\varsigma = 2$, *and* $\log p = o(n^{1/7})$. *For finite integers* $\{a_1, \ldots, a_m\}$, *under* H_0, $\mathcal{U}(a_1), \ldots, \mathcal{U}(a_m)$ *and* $\mathcal{U}(\infty)$ *are mutually asymptotically independent. In specific, for any* $z_1, \ldots, z_m, y \in \mathbb{R}$, *as* $n, p \to \infty$,

$$\Big| P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y\Big) \times \prod_{r=1}^{m} P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \leq z_r\Big)\Big| \to 0.$$

Theorem 2.1 suggests that all the finite-order U-statistics are asymptotically independent with each other. Given this, Theorem 2.3 further shows that the maximum-type test statistic $\mathcal{U}(\infty)$ is also asymptotically mutually independent with those finite-order U-statistics. The conclusion shares similarity with some classical results on the asymptotic independence between the sum-of-squares-type and maximum-type statistics. Specifically, for random variables $w_1, \ldots, w_n$, [32, 29] proved the asymptotic independence between $\sum_{i=1}^n w_i^2$ and $\max_{i=1,\ldots,n} |w_i|$ for weakly dependent observations. The similar independence properties were extensively studied in literature [e.g. 47, 30, 53, 33, 66, 43]. However, there are several differences between existing literature and the results in this paper. First, we discuss a family of U-statistics $\mathcal{U}(a)$'s, which takes different a values, and $\mathcal{U}(2)$ here corresponding to the sum-of-squares-type statistic is only a special case of general $\mathcal{U}(a)$. Furthermore, we have shown not only the asymptotic independence between $\mathcal{U}(a)$ and $\mathcal{U}(\infty)$, but also the asymptotic independence among $\mathcal{U}(a)$'s of finite a values. Second, the constructed $\mathcal{U}(a)$'s are unbiased estimators, which

are different from the sum-of-squares statistics usually examined in the literature. Moreover, the x's are allowed to be dependent and the theoretical development in the covariance testing involves a two-way dependence structure, which requires different proof techniques from the existing studies.

REMARK 2.4. *An alternative way to construct $\mathcal{U}(\infty)$ is to standardize $\hat{\sigma}_{j_1,j_2}$ by its variance $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})$. Specifically, following Cai et al. [8], we take $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2}) = n^{-1}\sum_{i=1}^{n}\{(x_{i,j_1}-\bar{x}_{j_1})(x_{i,j_2}-\bar{x}_{j_2})-\hat{\sigma}_{j_1,j_2}\}^2$. Define $M_n^{\dagger} = \max_{1\le j_1\neq j_2\le p}|\hat{\sigma}_{j_1,j_2}|/\{\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})\}^{1/2}$ and we take $\mathcal{U}(\infty) = M_n^{\dagger}$. Theoretically, we prove that Theorem 2.3 still holds with $\mathcal{U}(\infty) = M_n^{\dagger}$ in Supplementary Material Section **??**. Numerically, we provide the simulations in Supplementary Material Section **??**, which shows that M_n^* in (2.8) generally has higher power than $M_n^{\dagger}$.*

To apply hypothesis testing using the asymptotic results in Theorems 2.1 and 2.3, we need to estimate $\mathrm{var}\{\mathcal{U}(a)\}$. In particular, we propose the following moment estimator of (2.7):

$$\mathbb{V}_u(a) = \frac{2a!}{(P_a^n)^2}\sum_{1\le j_1\neq j_2\le p}\ \sum_{1\le i_1\neq\ldots\neq i_a\le n}\ \prod_{t=1}^{a}(x_{i_t,j_1}-\bar{x}_{j_1})^2(x_{i_t,j_2}-\bar{x}_{j_2})^2. \tag{2.9}$$

The next result establishes the statistical consistency of $\mathbb{V}_u(a)$.

CONDITION 2.4. *For integer a, $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^{8a}<\infty$.*

THEOREM 2.4. *Under H_0 in (2.1), assume Conditions 2.1, 2.2 and 2.4 hold. Then $\mathbb{V}_u(a)/\mathrm{var}\{\mathcal{U}(a)\}\xrightarrow{P}1$.*

Theorem 2.4 implies that the asymptotic results in Theorems 2.1 and 2.3 still hold by replacing $\mathrm{var}\{\mathcal{U}(a)\}$ with its estimator $\mathbb{V}_u(a)$. Specifically, under H_0, $[\mathcal{U}(a_1)/\sqrt{\mathbb{V}_u(a_1)},\ldots,\mathcal{U}(a_m)/\sqrt{\mathbb{V}_u(a_m)}]^{\intercal}\xrightarrow{D}\mathcal{N}(0,I_m)$ under Conditions 2.1, 2.2 and 2.4. Moreover, Theorem 2.3 implies that $\{\mathcal{U}(a)/\sqrt{\mathbb{V}_u(a)}\}$'s are asymptotically independent with $\mathcal{U}(\infty)$.

2.2. *Power Analysis.* In this section, we analyze the asymptotic power of the U-statistics. The power of $\mathcal{U}(2)$ has been studied in the literature. In particular, [10] studied the hypothesis testing of a high-dimensional covariance matrix with $H_0:\mathbf{\Sigma}=I_p$. The authors characterized the boundary that distinguishes the testable region from the non-testable region in terms of the Frobenius norm $\|\mathbf{\Sigma}-I_p\|_F$, and showed that the test statistic proposed by [13, 10], which corresponds to $\mathcal{U}(2)$ in this paper, is rate optimal over their

considered regime. However in practice, $\mathcal{U}(2)$ may be not powerful if the alternative covariance matrix is sparse with a small $\|\mathbf{\Sigma} - I_p\|_F$. When the alternative covariance has different sparsity levels, it is of interest to further examine which $\mathcal{U}(a)$ achieves the best power performance among the constructed family of U-statistics.

To study the test power, we establish the limiting distributions of $\mathcal{U}(a)$'s under the alternative hypothesis $H_A : \mathbf{\Sigma} = \mathbf{\Sigma}_A$, where the alternative covariance matrix $\mathbf{\Sigma}_A = (\sigma_{j_1,j_2})_{p\times p}$ is specified in the following Condition 2.5. Define $J_A = \{(j_1, j_2) : \sigma_{j_1,j_2} \neq 0, 1 \leq j_1 \neq j_2 \leq p\}$, which indicates the nonzero off-diagonal entries in $\mathbf{\Sigma}_A$. The cardinality of J_A, denoted by $|J_A|$, then represents the sparsity level of $\mathbf{\Sigma}_A$.

CONDITION 2.5. *Assume* $|J_A| = o(p^2)$ *and for* $(j_1, j_2) \in J_A$, $|\sigma_{j_1,j_2}| = \Theta(\rho)$, *where* $\rho = \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1,j_2}|/|J_A|$.

Here ρ represents the average signal strength of $\mathbf{\Sigma}_A$. In our following power comparison of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$, we say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$, if, under the same test power, $\mathcal{U}(a)$ can detect a smaller average signal strength ρ (please see the specific definition in Criterion 1 on Page 13). Condition 2.5 specifies a general family of "local" alternatives, which include banded covariance matrices, block covariance matrices, and sparse covariance matrices whose nonzero entries are randomly located.

THEOREM 2.5. *Suppose Conditions 2.1, 2.5, and* **??** *(an analogous condition to Condition 2.2* under* H_A*) in the Supplementary Material hold. For* $\mathcal{U}(a)$ *in* (2.3) *and finite integers* $\{a_1, \ldots, a_m\}$, *if* $\rho = O(|J_A|^{-1/a_t} p^{1/a_t} n^{-1/2})$ *for* $t = 1, \ldots, m$, *then as* $n, p \to \infty$,

$$\left[\frac{\mathcal{U}(a_1) - \mathrm{E}[\mathcal{U}(a_1)]}{\sigma(a_1)}, \ldots, \frac{\mathcal{U}(a_m) - \mathrm{E}[\mathcal{U}(a_m)]}{\sigma(a_m)}\right]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m),$$

where for $a \in \{a_1, \ldots, a_m\}$, $\mathrm{E}[\mathcal{U}(a)] = \sum_{(j_1,j_2)\in J_A} \sigma^a_{j_1,j_2}$ *and* $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] \simeq 2a!\kappa_1^a n^{-a} \sum_{1\leq j_1\neq j_2\leq p} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}$, *which is of order* $\Theta(p^2 n^{-a})$.

Theorem 2.5 shows that for a single U-statistic $\mathcal{U}(a)$ of finite order a,

$$P\left(\frac{\mathcal{U}(a)}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} > z_{1-\alpha}\right) \to 1 - \Phi\left(z_{1-\alpha} - \frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}}\right), \tag{2.10}$$

where $z_{1-\alpha}$ is the upper α quantile of $\mathcal{N}(0,1)$ and $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0,1)$. By Theorem 2.5, the asymptotic power of

$\mathcal{U}(a)$ of the one-sided test depends on

$$\frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} \simeq \frac{\sum_{(j_1,j_2)\in J_A} \sigma_{j_1,j_2}^a}{\{2a!\kappa_1^a n^{-a} \sum_{1\le j_1\ne j_2\le p} (\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a\}^{1/2}}. \tag{2.11}$$

By Theorem 2.5, (2.11) $= \Theta(|J_A|\rho^a p^{-1} n^{a/2})$. It follows that when $\mathrm{E}[\mathcal{U}(a)]$ is of the same order of $\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., $\mathrm{E}[\mathcal{U}(a)] = O(1)\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, the constraint of ρ in Theorem 2.5 is satisfied.

In the following power analysis, we will first compare $\mathcal{U}(a)$'s of finite a and then compare them with $\mathcal{U}(\infty)$. As we focus on studying the relationship between the sparsity level and power, we consider an ideal case where $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$ and $\sigma_{j,j} = \nu^2 > 0$ for $j = 1, \ldots, p$. Then

$$(2.11) \simeq |J_A|\rho^a/(\sqrt{2a!\kappa_1^a}\nu^{2a} p n^{-a/2}). \tag{2.12}$$

We next show how the order of the "best" U-statistics changes when the sparsity level $|J_A|$ varies. To be specific of the meaning of "best", we compare the ρ values needed by different U-statistics to achieve the same asymptotic power. Particularly, we fix $\mathrm{E}[\mathcal{U}(a)]/\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., (2.12) to be some constant $M/\sqrt{2}$ for different a's and the asymptotic power of each $\mathcal{U}(a)$ is (2.10) $= 1 - \Phi(z_{1-\alpha} - M/\sqrt{2})$. Then by (2.12), the ρ value such that $\mathcal{U}(a)$ attains the power above is

$$\rho_a = \sqrt{\kappa_1}(a!)^{\frac{1}{2a}} \nu^2 (Mp/|J_A|)^{\frac{1}{a}} n^{-\frac{1}{2}}. \tag{2.13}$$

By the definition in (2.13), we compare the power of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$ with $a \ne b$ following the Criterion 1 below.

CRITERION 1. *We say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$ if $\rho_a < \rho_b$.*

Given values of $n, p, |J_A|$ and M, (2.13) is a function of a. Therefore, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the smallest ρ_a value in (2.13). We then have the following proposition discussing the optimality among the U-statistics of finite orders in (2.3).

PROPOSITION 2.3. *Given $n, p, |J_A|$ and any constant $M \in (0, +\infty)$, we consider ρ_a in (2.13) as a function of integer a, then*

(i) when $|J_A| \ge Mp$, the minimum of ρ_a is achieved at $a_0 = 1$;
(ii) when $|J_A| < Mp$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_A|$ increases.

By Proposition 2.3, the order a_0 that attains the smallest value of ρ_a depends on the value of $Mp/|J_A|$ and does not have a closed form solution. We use numerical plots to demonstrate the relationship between a_0 and the sparsity level. Particularly, let $|J_A| = p^{2(1-\beta)}$, where $\beta \in (0,1)$ denotes the sparsity level. To have a better visualization, we use $g(a) = \log(\rho_a n^{1/2} \kappa_1^{-1/2} \nu^{-2}) = (1/2a) \log a! + a^{-1} \log(Mp^{2\beta-1})$ instead of ρ_a. We plot $g(a)$ curves in Figure 1 for each $\beta \in \{0.1, \ldots, 0.9\}$ with $M = 4$ and $p \in \{100, 10000\}$. Other values of M and p are also taken, which give similar patterns to Figure 1 and are not presented.

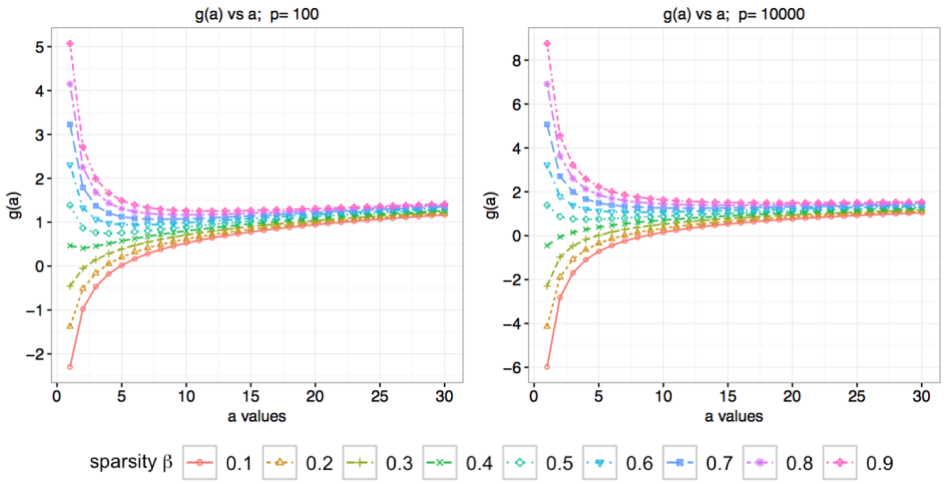

Fig 1: $g(a)$ versus a with different sparsity level β for $p = 100, 10000$

Figure 1 shows that the a_0 such that $g(a)$ attains the smallest value increases when the sparsity level β increases. In particular, when the sparsity level $\beta \leq 0.3$, that is, when $|J_A|$ is "very" large and then $\mathbf{\Sigma}_A$ is "very" dense, $g(a)$ has the smallest value at $a_0 = 1$. This is consistent with the conclusion in Proposition 2.3 (i). When the sparsity level β is between 0.4 and 0.5, we note that $a_0 = 2$ achieves the minimum of $g(a)$. This shows that when $|J_A|$ is "moderately" large and $\mathbf{\Sigma}_A$ is "moderately" dense, $\mathcal{U}(2)$ is more powerful than $\mathcal{U}(1)$. When the sparsity level $\beta > 0.5$, we find that $a_0 > 2$. This implies that when $|J_A|$ becomes smaller and $\mathbf{\Sigma}_A$ becomes sparser, U-statistics of higher orders are more powerful. Additionally, we note that a_0 increases slowly as β increases, which verifies Proposition 2.3 (ii). Moreover, the curves converge as a increases and the differences of $g(a)$ for large a values ($a \geq 6$) are small. This implies that when selecting the range of considered orders of U-statistics, it suffices to select an upper bound with

$a = 6$ or 8, which gives better or similar ρ_a values to those larger a's.

In summary, when $|J_A|$ is large, i.e., $\mathbf{\Sigma}_A$ is dense, a small a tends to obtain a smaller lower bound in terms of ρ. But when $|J_A|$ decreases, i.e., $\mathbf{\Sigma}_A$ becomes sparse, a U-statistic of large finite order (or the maximum-type U-statistic as shown next) tends to obtain a smaller lower bound in ρ. This observation is consistent with the existing literature [13, 7, 10, 6].

Next, we proceed to examine the power of the maximum-type test statistic $\mathcal{U}(\infty)$, and compare it with the U-statistics $\mathcal{U}(a)$ of finite a defined in (2.3). By [7], the rejection region for $\mathcal{U}(\infty)$ with significance level α is

$$|\mathcal{U}(\infty)| \geq t_p := n^{-1/2}\sqrt{4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}}.$$

Note $t_p \simeq 2\sqrt{\log p/n}$ and under alternative, the power for $\mathcal{U}(\infty)$ is

$$P(|\mathcal{U}(\infty)| \geq t_p). \tag{2.14}$$

As discussed, we consider the alternatives satisfying Conditions 2.2* and 2.5, $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$, and $\sigma_{j,j} = \nu^2$ for $j = 1, \ldots, p$. For simplicity, we assume $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ and ν^2 are given, and focus on the simplified

$$\mathcal{U}(\infty) = \max_{1\leq j_1<j_2\leq p} \left|\nu^{-2}n^{-1}\sum\nolimits_{i=1}^{n}(x_{i,j_1} - \mu_{j_1})(x_{i,j_2} - \mu_{j_2})\right|. \tag{2.15}$$

We show in the following proposition when the power of $\mathcal{U}(\infty)$ asymptotically converges to 1 or is strictly smaller than 1 under alternative.

PROPOSITION 2.4. *Under the considered alternative $\mathbf{\Sigma}_A$ above, suppose $\max_{j=1,\ldots,p}\mathrm{E}e^{t_0|x_j-\mu_j|^{\varsigma}} < \infty$ for some $0 < \varsigma \leq 2$ and $t_0 > 0$, and $\log p = o(n^{\beta})$ with $\beta = \varsigma/(4+\varsigma)$. Then for (2.15), when $n, p \to \infty$,*

(i) there exists a constant $c_1 > 2$ such that if $\rho \geq c_1\sqrt{\log p/n}$, (2.14) $\to 1$;

(ii) there exists another constant $0 < c_2 < 2$ such that when $\rho \leq c_2\sqrt{\log p/n}$, Condition 2.2 holds for $\kappa_1 \leq 1$ and $|J_A| = o(1)p^{\frac{2(1-c_2/2)^2}{\kappa_1+m}}(\log p)^{\frac{1}{2}-\frac{1}{2(\kappa_1+m)}}$ for some $m > 0$, we have (2.14) $\leq \log(1-\alpha)^{-1}$.*

Recall that Proposition 2.3 shows that there exists a finite integer a_0, such that ρ_{a_0} is the minimum of (2.13), and ρ_{a_0} is a lower bound of ρ value for the finite-order U-statistics to achieve the given asymptotic power. With Propositions 2.3 and 2.4, we next compare the finite-order U-statistics defined in (2.3) with the maximum-type test statistic $\mathcal{U}(\infty)$.

PROPOSITION 2.5. *Under the conditions of Theorem 2.5 and Proposition 2.4, for any finite integer a, there exist constants c_1 and c_2 such that when p is sufficiently large,*

(i) For any M, when $|J_A| < c_1^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$, $\mathcal{U}(\infty)$ has higher asymptotic power than $\mathcal{U}(a)$.

(ii) When M is big enough and $|J_A| > c_2^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$, $\mathcal{U}(a)$ has higher asymptotic power than $\mathcal{U}(\infty)$.

From Proposition 2.3, we know when $Mp/|J_A| = O(1)$, there exists a finite a_0 such that $\mathcal{U}(a_0)$ is the "best" among all the finite-order U-statistics; in this case, Proposition 2.5 (ii) further indicates that $\mathcal{U}(a_0)$ has higher asymptotic power than $\mathcal{U}(\infty)$. Specifically, if $Mp/|J_A| < 1$, $a_0 = 1$, then $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(p|J_A|^{-1}n^{-1/2})$. More interestingly, when $\mathbf{\Sigma}_A$ is moderately dense or moderately sparse with $Mp/|J_A| > 1$ and bounded, some U-statistic of finite order $a_0 > 1$ would become the "best". By Figure 1, the value of a_0 increases as $\mathbf{\Sigma}_A$ becomes denser. On the other hand, when $\mathbf{\Sigma}_A$ is "very" sparse with $|J_A| < c_1^{-a_0}(a_0!)^{\frac{1}{2}}\kappa_1^{\frac{a_0}{2}}(\log p)^{-\frac{a_0}{2}}Mp$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$.

REMARK 2.5. *The above power comparison results are under the constructed family of U-statistics. We note that additional formulation may further enhance the test power. For instance, [11, 72] showed that an adaptive thresholding in certain ℓ_p-type test statistics can achieve high power under the alternatives with sparse and faint signals. It is of interest to incorporate the adaptive thresholding into the constructed family of U-statistics, which is left for future study.*

REMARK 2.6. *The analysis above focuses on the ideal case where the nonzero off-diagonal entries of $\mathbf{\Sigma}_A$ are the same for illustration. When these entries of $\mathbf{\Sigma}_A$ are different, similar analysis still applies by Theorem 2.5 for general covariance matrices. In specific, the asymptotic power of $\mathcal{U}(a)$ depends on the mean variance ratio (2.11) and $\rho_a = \sqrt{\kappa_1}n^{-1/2}(a!)^{1/2a} \times (M\sum_{j=1}^p \sigma_{j,j}^a/\sum_{1\le j_1,j_2\le p}\sigma_{j_1,j_2}^a)^{1/a}$. We can then obtain conclusions similar to Propositions 2.3–2.5. One interesting case is when $\mathbf{\Sigma}_A$ contains both positive and negative entries; the same analysis applies for even-order U-statistics, since σ_{j_1,j_2}^a's are all non-negative for even a. On the other hand, the odd-order U-statistics would have low power, since $\sum_{1\le j_1\ne j_2\le p}\sigma_{j_1,j_2}^a$ could be small due to the cancellation of positive and negative σ_{j_1,j_2}^a's. We have conducted simulations when the nonzero σ_{j_1,j_2}'s are different in Section 3.1, and the results exhibit consistent patterns as expected.*

2.3. *Application to Adaptive Testing & Computation.*

Adaptive Testing. Power analysis in Section 2.2 shows that when the sparsity level of the alternative changes, the test statistic that achieves the highest power could vary. However, since the truth is often unknown in practice, it is unclear which test statistic should be chosen. Therefore, we develop an adaptive testing procedure by combining the information from U-statistics of different orders, which would yield high power against various alternatives.

In particular, we propose to combine the U-statistics through their p-values, which is widely used in literature [48, 51, 70]. One popular method is the minimum combination, whose idea is to take the minimum p-value to approximate the maximum power [51, 70, 66]. Specifically, let Γ be a candidate set of the orders of U-statistics, which contains both finite values and ∞. We compute p-values p_a's of the U-statistics $\mathcal{U}(a)$'s satisfying $a \in \Gamma$. The minimum combination takes the statistic $T_{\text{adpUmin}} = \min\{p_a : a \in \Gamma\}$ and has the asymptotic p-value $p_{\text{adpUmin}} = 1 - (1 - T_{\text{adpUmin}})^{|\Gamma|}$, where $|\Gamma|$ denotes the size of the candidate set Γ. We reject H_0 if $p_{\text{adpUmin}} < \alpha$. Under H_0, p_a's are asymptotically independent and uniformly distributed by the theoretical results in Section 2.1. The type I error is asymptotically controlled as $P(p_{\text{adpUmin}} < \alpha) = P(\min_{a\in\Gamma} p_a < p_\alpha^*) \to \alpha$, where $p_\alpha^* = 1 - (1 - \alpha)^{1/|\Gamma|}$. Since $P(\min_{a\in\Gamma} p_a < p_\alpha^*) \geq P(p_a < p_\alpha^*)$, the power of the adaptive test goes to 1 if there exists $a \in \Gamma$ such that the power of $\mathcal{U}(a)$ goes to 1. We note that the power of the adaptive test is not necessarily higher than that of all the U-statistics. This is because the power of $\mathcal{U}(a)$ is $P(p_a < \alpha)$, and is different from $P(p_a < p_\alpha^*)$ since $p_\alpha^* < \alpha$ when $|\Gamma| > 1$. Based on our extensive simulations, we find that the adaptive test is usually close to or even higher than the maximum power of the U-statistics.

REMARK 2.7. *Fisher's method [48] is another popular method for combining independent p-values. It has the test statistic* $T_{\text{adpUf}} = -2\sum_{k=1}^{|\Gamma|} \log p_k$, *which converges to* $\chi^2_{2|\Gamma|}$ *under* H_0. *By our simulations, the minimum combination and Fisher's method are generally comparable, while Fisher's method has higher power under several cases. Moreover, we can also use other methods to combine the p-values, such as higher criticism [16, 17]. We leave the study of how to efficiently combine the p-values for future research.*

We select the candidate set Γ by the power analysis in Section 2.2. We would recommend including $\{1, 2, \ldots, 6, \infty\}$, which can be powerful against a wide spectrum of alternatives. In particular, by Propositions 2.3 and 2.5, we include $a = 1, 2$ that are powerful against dense signals; $a = \infty$ that is powerful against sparse signals; and also $a = \{3, \ldots, 6\}$ for the moderately dense and moderately sparse signals. By Figure 1, it generally suffices to choose finite a up to 6–8, which often give similar/better performance

to/than larger a values. The simulations in Section 3.1 confirm the good performance of this choice of Γ; and the proposed adaptive test appears to well approximate the "best" performance even when Γ may not always contain the unknown "optimal" U-statistics.

We would like to mention that the adaptive procedure can be generalized to other testing problems, as long as similar theoretical properties are given, such as the examples in Section 4.

Computation. Next we discuss the computation in the adaptive testing. A direct calculation following the form of $\mathcal{U}(a)$ in (2.3) and $\mathbb{V}(a)$ in (2.9) would be computationally expensive for large a with a cost of $O(p^2n^{2a})$. To address this issue, we introduce a method that can reduce the cost.

We first consider a simplified setting when $\mathrm{E}(x_{i,j}) = 0$ to illustrate the idea. As discussed in Remark 2.2, we examine $\tilde{\mathcal{U}}(a)$ defined in (2.5). Let $\mathcal{L} = \{(j_1, j_2) : 1 \le j_1 \ne j_2 \le p\}$ denote the set of index tuples, and for each index tuple $l = (j_1, j_2) \in \mathcal{L}$, define $s_{i,l} = x_{i,j_1}x_{i,j_2}$. Note that $\tilde{\mathcal{U}}(a) = (P_a^n)^{-1}\sum_{l\in\mathcal{L}}\mathcal{U}_l(a)$, where $\mathcal{U}_l(a) = \sum_{1\le i_1\ne\cdots\ne i_a\le n}\prod_{k=1}^{a} s_{i_k,l}$. Calculating $\mathcal{U}_l(a)$ directly is of order $O(n^a)$. We then focus on reducing the computational cost of $\mathcal{U}_l(a)$. For $l \in \mathcal{L}$ and finite integers $t_1, \ldots, t_k$, define

$$V_l^{(t_1,\ldots,t_k)} = \prod_{r=1}^{k}\Big(\sum_{i=1}^{n} s_{i,l}^{t_r}\Big), \quad U_l^{(t_1,\cdots,t_k)} = \sum_{1\le i_1\ne\ldots\ne i_k\le n}\prod_{r=1}^{k} s_{i_1,l}^{t_r}. \tag{2.16}$$

We can see that $\mathcal{U}_l(a) = U_l^{\mathbf{1}_a}$ with $\mathbf{1}_a$ being an a-dimensional vector of all ones, and $U_l^{(a)} = V_l^{(a)}$ for any finite integer a. To reduce the computational cost of $\mathcal{U}_l(a)$, the main idea is to obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\ldots,t_k)}$, whose computational cost is $O(n)$. In particular, $\mathcal{U}_l(a)$ can be attained iteratively from $V_l^{(t_1,\ldots,t_k)}$ based on the following equation

$$U_l^{(k,\mathbf{1}_{r-k})} = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k)\times U_l^{(k+1,\mathbf{1}_{r-k-1})}, \tag{2.17}$$

which follows from the definitions. Algorithm 1 below summarizes the steps.

We illustrate the idea of the algorithm by some examples. By definition, $U_l^{(1)} = V_l^{(1)}$, which can be computed with cost $O(n)$. Next consider in (2.17), if $r = 2$ and $k = 1$, then $U_l^{(1,1)} = V_l^{(1)} \times U_l^{(1)} - (2-1)\times U_l^{(2)} = V_l^{(1)} \times V_l^{(1)} - V_l^{(2)}$, which yields $U_l^{\mathbf{1}_2}$ with cost $O(n)$. For $U_l^{\mathbf{1}_3}$, we first take $r = 3$ and $k = 2$ in (2.17), then with cost $O(n)$, we have $U_l^{(2,1)} = V_l^{(2)} \times U_l^{(1)} - U_l^{(3)} = V_l^{(2)} \times V_l^{(1)} - V_l^{(3)}$, as $V_l^{(k)} = U_l^{(k)}$ by the definition. Given $U_l^{\mathbf{1}_2}$ and $U_l^{(2,1)}$, we obtain $U_l^{(1,\mathbf{1}_2)} = V_l^{(1)} \times U_l^{\mathbf{1}_2} - 2\times U_l^{(2,\mathbf{1}_1)}$. Thus $U_l^{\mathbf{1}_3}$ is also computed with cost

Data: $s_{i,l}$ $(1 \le i \le n,\, l \in \mathcal{L})$.
Result: $\tilde{\mathcal{U}}(a)$.
for $l \in \mathcal{L}$ **do**
  Compute and store $V_l^{(k)} = U_l^{(k)} = \sum_{i=1}^n s_{i,l}^k$, $(k = 1, \cdots, a)$ during the algorithm;
  $U_l^{\mathbf{1}_1} = V_l^{(1)}$, $U_l^{\mathbf{1}_2} = U_l^{\mathbf{1}_1} V_l^{(1)} - U_l^{(2)}$;
  while $3 \le r \le a$ **do**
    $T_l = U_l^{(r)}$
    for $k \leftarrow r-1$ **to** 1 **do**
      $T_l = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k) \times T_l$
    end
    $U_l^{\mathbf{1}_r} = T_l$
  end
end
$\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{l \in \mathcal{L}} U_l^{\mathbf{1}_a}$

Algorithm 1: Iterative Computation Implementation

$O(n)$. Iteratively, for any finite integer a, we can obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\ldots,t_k)}$ whose computational cost is $O(n)$. More closed form formulae representing $U_l^{\mathbf{1}_a}$ by $V_l^{(t_1,\ldots,t_k)}$ are given in Section **??** of Supplementary Material.

Algorithm 1 reduces the computational cost of $\tilde{\mathcal{U}}(a)$ from $O(p^2 n^a)$ to $O(p^2 n)$. Its idea is general and can be extended to compute other different U-statistics by changing the input $s_{i,l}$. In particular, the variance estimator $\mathbb{V}(a)$ can be computed with cost $O(p^2 n)$ by specifying $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})^2 (x_{i,j_2} - \bar{x}_{j_2})^2$, for each $l \in \mathcal{L} = \{(j_1, j_2) : 1 \le j_1 \ne j_2 \le p\}$. Then $\mathbb{V}(a) = 2a!(P_a^n)^{-2} \sum_{l \in \mathcal{L}} \sum_{1 \le i_1 \ne \ldots \ne i_a \le n} \prod_{k=1}^a s_{i_k,l}$ and the Algorithm 1 can be applied. Moreover, when $\mathrm{E}(x_{i,j})$ is unknown, $\mathcal{U}(a)$ can still be computed with cost $O(p^2 n)$ using the iterative method similar to Algorithm 1. The details are provided in Section **??** of Supplementary Material.

3. Simulations and Real Data Analysis.

3.1. *Simulations.* We conduct simulation studies to evaluate the performance of the proposed adaptive testing procedures, and investigate the relationship between the power and sparsity levels. For one-sample covariance testing discussed in Section 2, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and consider the following five simulation settings.

Setting 1: $\mathbf{x}_i$ has p i.i.d. entries of $\mathcal{N}(0,1)$ and Gamma$(2, 0.5)$ respectively. Under each case, we take $n = 100$ and $p \in \{50, 100, 200, 400, 600, 800, 1000\}$ to verify the theoretical results under H_0 and the validity of the adaptive test across different n and p combinations.

For the following settings 2–5, we generate $\mathbf{x}_i$ from multivariate Gaussian distributions with mean zero and different covariance matrices $\mathbf{\Sigma}_A$'s.

Setting 2: $\mathbf{\Sigma}_A = (1-\rho)I_p + \rho \mathbf{1}_{p,k_0}\mathbf{1}_{p,k_0}^\intercal$, where $\mathbf{1}_{p,k_0}$ is a p-dimensional vector with the first k_0 elements one and the rest zero. We take $(n,p) \in \{(100,300),(100,600),(100,1000)\}$, and study the power with respect to different signal sizes ρ and sparsity levels k_0.

Setting 3: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are ρ with random positions. We take $(n,p) \in \{(100,600),(100,1000)\}$ and let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 4: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are uniformly generated from $(0, 2\rho)$ with random positions. We take $(n,p) = (100,1000)$ and similarly let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 5: We consider the multivariate models in [13]. Specifically, for each $i = 1,\ldots,n$, $\mathbf{x}_i = \Xi \mathbf{z}_i + \boldsymbol{\mu}$, where Ξ is a matrix of dimension $p \times m$, and $\mathbf{z}_i$'s are i.i.d. Gaussian or Gamma random vectors. Under null hypothesis, $m = p$, $\Xi = I_p$ $\boldsymbol{\mu} = 2\mathbf{1}_p$; under alternative hypothesis, $m = p+1$, $\Xi = (\sqrt{1-\rho}I_p, \sqrt{2\rho}\mathbf{1}_p)$, $\boldsymbol{\mu} = 2(\sqrt{1-\rho}+\sqrt{2\rho})\mathbf{1}_p$. We also take the n and p combination in [13] with $(n,p) \in \{(40,159),(40,331),(80,159),(80,331),(80,642)\}$.

We compare several methods in the literature, including both maximum-type and sum-of-squares-type tests. In particular, the maximum-type test statistic in Jiang [35] is taken as $\mathcal{U}(\infty)$ in this framework. Since the convergence in [35] is known to be slow, we use permutation to approximate the distribution in the simulations. In addition, we consider some sum-of-squares-type methods. Specifically, we examine the identity and sphericity tests in Chen et al. [13], which are denoted as "Equal" and "Spher", respectively. We also compare the methods in Ledoit and Wolf [41] and Schott [56], which are referred to as "LW" and "Schott", respectively.

To illustrate, Figure 2 summarizes the numerical results for the setting 3 when $n = 100$ and $p = 1000$. All the results are based on 1000 simulations at the 5% nominal significance level. In Figure 2, we present the power of single U-statistics with orders in $\{1,\ldots,6,\infty\}$. "adpUmin" and "adpUf" represent the results of the adaptive testing procedure using the minimum combination and Fisher's method in Section 2.2 respectively. The simulation results show that the type I error rates of the U-statistics and adaptive test are well controlled under H_0. In addition, Figure 2 exhibits several patterns that are consistent with the power analysis in Section 2.2. First, it shows that among the U-statistics, when $|J_A|$ is very small, $\mathcal{U}(\infty)$ performs best; and when $|J_A|$ increases, the performances of some U-statistics of finite orders catch

up. For instance, when $|J_A| = 100$, $\mathcal{U}(6)$ and $\mathcal{U}(\infty)$ are similar and are better than the other U-statistics; when $|J_A| = 400$, $\mathcal{U}(4)$ and $\mathcal{U}(5)$ are similar and better than the other U-statistics. When $\mathbf{\Sigma}_A$ is relatively dense, $\mathcal{U}(2)$ and $\mathcal{U}(1)$ become more powerful. Particularly, when $|J_A| = 1600$, $\mathcal{U}(2)$ is powerful; when $|J_A|$ becomes larger, such as when $|J_A| = 3200$, $\mathcal{U}(1)$ is overall the most powerful. Second, Figure 2 shows that "LW", "Schott", "Equal", "Spher" and $\mathcal{U}(2)$ perform similarly under various cases. In particular, these methods are not powerful when the alternative is sparse but becomes more powerful when the alternative gets denser. This is because they are all sum-of-squares-type statistics that target at dense alternatives. Third and importantly, the two adaptive tests "adpUmin" and "adpUf" maintain high power across different settings. Specifically, they perform better than most single U-statistics: their powers are usually close to or even higher than the best single U-statistic. Moreover, "adpUmin" and "adpUf" generally have higher power than the compared existing methods. We also note that "adpUf" overall performs better than "adpUmin" in this simulation setting. In summary, Figure 2 demonstrates the relationship between the sparsity levels of alternatives and the power of the tests, confirming the theoretical conclusions in Section 2.2. Notably, the proposed adaptive testing procedure is powerful against a wide range of alternatives, and thus advantageous in practice when the true alternative is unknown.

Due to the space limitation, we provide other extensive numerical studies in Supplementary Material Section **??**. The conclusions are similar to those of Figure 2, and consistent with the theoretical results in Section 2.2. In particular, the results show that the empirical sizes of the tests are close to the nominal level, suggesting the good finite-sample performance of the asymptotic approximations. Moreover, under highly dense alternatives with only non-negative entries in the covariance matrix, $\mathcal{U}(1)$ is the most powerful one among the $\mathcal{U}(a)$'s and the other tests in [41, 56, 13], in agreement with the results in Propositions 2.3 and 2.5. Furthermore, the proposed adaptive testing procedures often have higher power than most single U-statistics.

3.2. *Real Data Analysis.* Alzheimer's disease (AD) is the most prevalent neurodegenerative disease [55] and is ranked as the sixth leading cause of death in the US [67]. Every 65 seconds, someone in the US develops AD [1]. To advance our understanding of AD, the Alzheimer's Disease Neuroimaging Initiative (ADNI) was started in 2004, collecting extensive genetic data for both healthy individuals and AD patients. To gain insight into the genetic mechanisms of AD, one can test a single SNP a time. However, due to a relatively small sample size of the ADNI data, scanning across all SNPs failed to

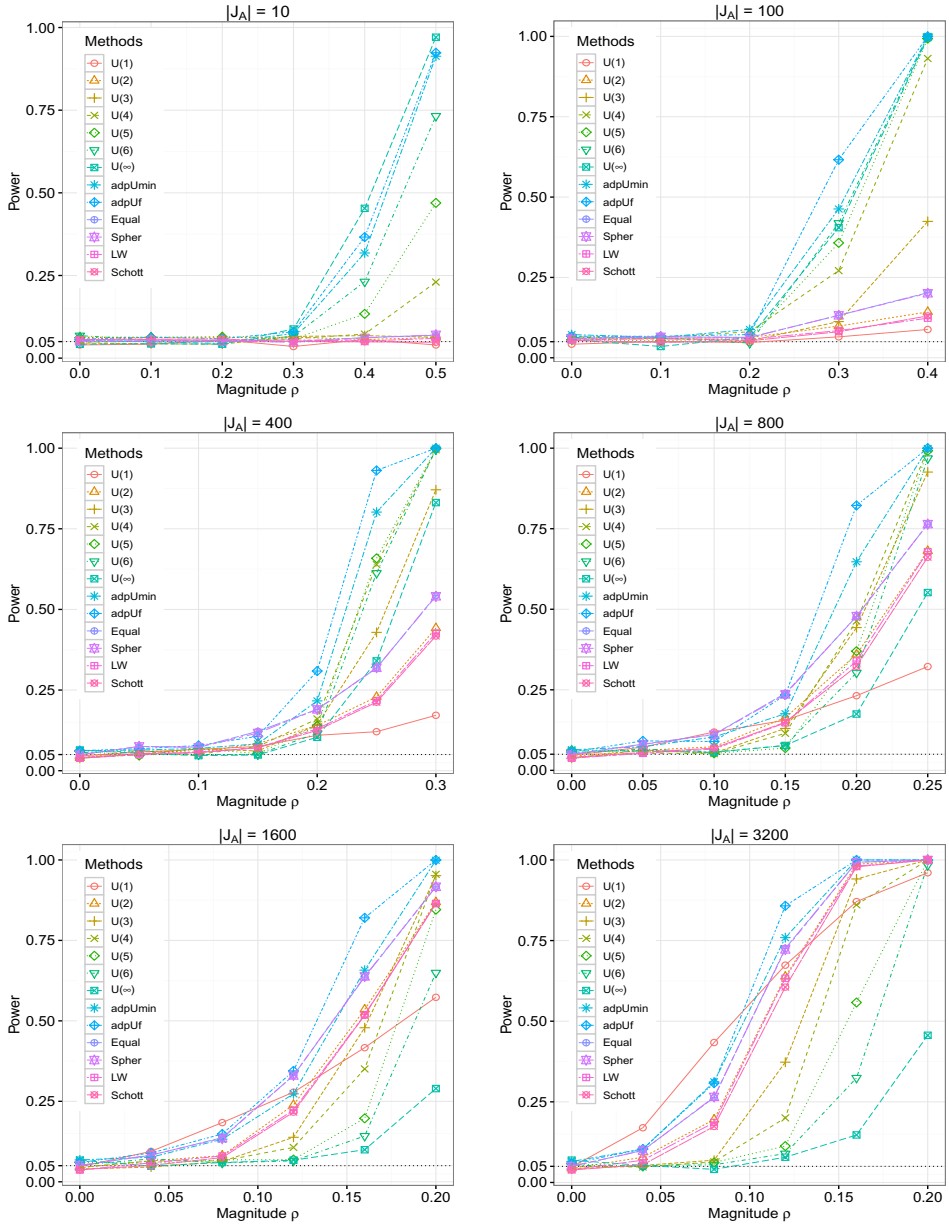

Fig 2: Power comparison.

identify any genome-wide significant SNP (with p-value $< 5 \times 10^{-8}$)[39]. To date, the largest meta-analysis of more than 600,000 individuals identified 29 significant risk loci [34] and can only explain a small proportion of AD vari-

ance. On the other hand, a group of functionally related genes as annotated in a biological pathway are often involved in the same disease susceptibility and progression [28]. Thus, pathway-based analyses, which jointly analyze a group of SNPs in a biological pathway, have become increasingly popular. We retrieve a total of 214 pathways from the KEGG database [38] for the subsequent analysis.

Although pathway-based analyses with KEGG pathways are common in real studies, formally testing the correlations of the genes in a KEGG pathway has been largely untouched. Here, we apply our method and other competing methods in [13] to test if all the genes in a pathway have correlated gene expression levels. Perhaps as expected, all methods reject the null hypothesis for all pathways with highly significant p-values, since the KEGG pathways are constructed to include only the genes with similar function into the same pathway [38], while similar function often implies co-expression (and vice versa). To compare the performance of the different tests, for each pathway we randomly select 50 subjects and restrict our analysis to pathways of at least 50 genes, leading to 103 pathways for the following analysis. Then we perturb the data by shuffling the gene expression levels of randomly selected $100(1-\alpha)\%$ genes in a pathway before applying each test. Figure 3 shows the performance of the tests with two significance cutoffs, where "$\mathcal{U}(2)$" represents the single $\mathcal{U}(2)$ statistic, "adpU" represents our proposed adaptive testing procedure using the minimum combination with candidate U-statistics of orders in $\{1, \ldots, 6, \infty\}$, and "Equal" and "Spher" represent the identity and sphericity tests in [13] respectively. Because all pathways are highly significant with all samples, we can treat all pathways as the true positives. Due to the adaptiveness of our proposed testing procedure, "adpU" identifies more significant pathways than the competing methods across all the levels of data perturbation (mimicking the varying sparsity levels of the alternatives).

4. Other High-Dimensional Examples. In this section, we apply the proposed U-statistics framework to other high-dimensional testing problems. Similar theoretical results to Section 2 are developed, with detailed proofs and related simulation studies provided in Supplementary Material.

4.1. *Mean Testing.* Testing mean vectors is widely used in many statistical analysis and applications [2, 49]. Under high-dimensional scenarios, e.g., in genome-wide studies, dimension of the data is often much larger than the sample size, so traditional multivariate tests such as Hotelling's T^2-test either cannot be directly applied or have low power [18]. To address this issue, several new procedures for testing high-dimensional mean vectors have been

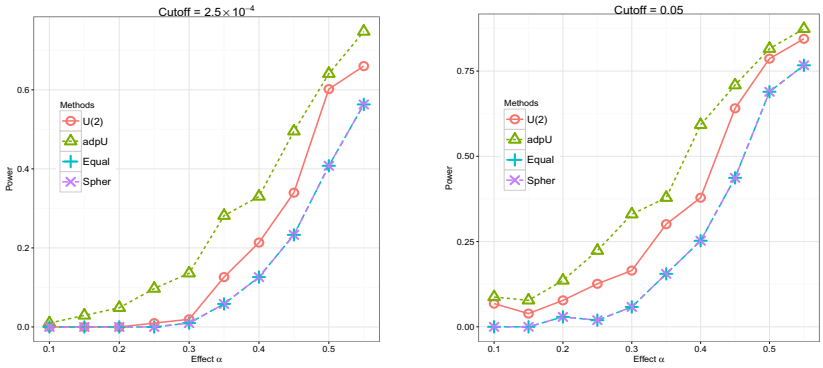

Fig 3: Power comparison of different methods with ADNI data.

proposed [4, 16, 25, 59, 12, 27, 9, 11, 26, 17, 61, 66]. However, many of the statistics only target at either sparse or dense alternatives, and suffer from loss of power for other types of alternatives. We next apply the U-statistics framework to one-sample and two-sample mean testing problems.

One-sample mean testing. We first discuss the one-sample mean vector testing. Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$, covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \le j_1, j_2 \le p\}$. We want to conduct the global test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0 = (\mu_{1,0}, \ldots, \mu_{p,0})^\intercal$ is given.

Similar to previous discussion, the parameter set that we are interested in is $\mathcal{E} = \{\mu_1 - \mu_{1,0}, \ldots, \mu_p - \mu_{p,0}\}$. For each $j = 1, \ldots, p$, $\mathrm{E}(x_{i,j}) = \mu_j$, so $K_j(\mathbf{x}_i) = x_{i,j} - \mu_{j,0}$ is a kernel function, which is a simple unbiased estimator of the target. Following our construction, the U-statistic for finite a is

$$\mathcal{U}(a) = \sum_{j=1}^{p} \frac{1}{P_a^n} \sum_{1 \le i_1 \ne \cdots \ne i_a \le n} \prod_{k=1}^{a} (x_{i_k,j} - \mu_{j,0}), \tag{4.1}$$

which targets at $\|\mathcal{E}\|_a^a = \sum_{j=1}^{p} (\mu_j - \mu_{j,0})^a$, and the U-statistic corresponding to $\|\mathcal{E}\|_\infty$ is $\mathcal{U}(\infty) = \max_{1 \le j \le p} \sigma_{j,j}^{-1} (\bar{x}_j - \mu_{0,j})^2$ with $\bar{x}_j = \sum_{i=1}^{n} x_{i,j}/n$.

Given the statistics, we have the theoretical results similar to Theorems 2.1–2.3. The following Theorems 4.1–4.2 are established under similar conditions to that of Theorems 2.1–2.3. Due to the limited space, we provide the conditions and corresponding discussions in Supplementary Material.

THEOREM 4.1. *Under H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, assume Condition ?? in Supplementary Material. Then for any finite integers $\{a_1, \ldots, a_m\}$, as $n, p \to \infty$,*

$[\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where* $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] = \sum_{i=1}^p \sum_{j=1}^p a!\sigma_{i,j}^a / P_a^n$ *with the order of* $\Theta(a!pn^{-a})$.

THEOREM 4.2. *Under* H_0*:* $\boldsymbol{\mu} = \boldsymbol{\mu}_0$*, assume Condition* **??** *in Supplementary Material. Then* $\forall u \in \mathbb{R}$, $P(n\mathcal{U}(\infty) - \tau_p \le u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, *as* $n, p \to \infty$*, where* $\tau_p = 2\log p - \log\log p$*. In addition, for any finite integer* a, $\{\mathcal{U}(a)/\sigma(a)\}$ *and* $\{n\mathcal{U}(\infty) - \tau_p\}$ *are asymptotically independent.*

By Theorems 4.1 and 4.2, we obtain the asymptotic independence among the U-statistics and the corresponding limiting distributions of the U-statistics under H_0. Under the alternative hypothesis, since the power analysis of the one-sample mean testing is similar to that of the two-sample case, we delay the power analysis after presenting the asymptotic independence property of the proposed U-statistics in the two-sample mean testing problem.

Two-sample mean testing. Next we discuss the two-sample mean testing problem. Suppose we have two groups of p-dimensional observations $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$, which are i.i.d. copies of two independent random vectors $\mathbf{x} = (x_1,\ldots,x_p)^\intercal$ and $\mathbf{y} = (y_1,\ldots,y_p)^\intercal$ respectively. Suppose $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1,\ldots,\mu_p)^\intercal$, $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = (\nu_1,\ldots,\nu_p)^\intercal$, $\mathrm{cov}(\mathbf{x}) = \boldsymbol{\Sigma}_x$ and $\mathrm{cov}(\mathbf{y}) = \boldsymbol{\Sigma}_y$. We write $n = n_x + n_y$ and assume $n_x = \Theta(n_y)$. For easy illustration, we first consider $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \le j_1, j_2 \le p\}$. We will then discuss the case when $\boldsymbol{\Sigma}_x \ne \boldsymbol{\Sigma}_y$, where similar analysis applies.

The two-sample mean testing examines $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$ versus $H_A : \boldsymbol{\mu} \ne \boldsymbol{\nu}$, then $\mathcal{E} = (\mu_1 - \nu_1,\ldots,\mu_p - \nu_p)^\intercal$. For $1 \le j \le p$, $1 \le k \le n_x$, $1 \le s \le n_y$, $K_j(\mathbf{x}_k, \mathbf{y}_s) = x_{k,j} - y_{s,j}$ is a simple unbiased estimator of $\mu_j - \nu_j$, and thus we construct $\mathcal{U}(a) = \sum_{j=1}^p (P_a^{n_x} P_a^{n_y})^{-1} \sum_{\substack{1\le k_1\ne\ldots\ne k_a\le n_x \\ 1\le s_1\ne\ldots\ne s_a\le n_y}} \prod_{t=1}^a (x_{k_t,j} - y_{s_t,j})$, which is also equivalent to

$$\mathcal{U}(a) = \sum_{j=1}^p \sum_{c=0}^a \binom{a}{c} \frac{(-1)^{a-c}}{P_c^{n_x} P_{a-c}^{n_y}} \sum_{\substack{1\le k_1\ne\cdots\ne k_c\le n_x \\ 1\le s_1\ne\cdots\ne s_{a-c}\le n_y}} \prod_{t=1}^c x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}. \tag{4.2}$$

We can check that (4.2) satisfies $\mathrm{E}\{\mathcal{U}(a)\} = \sum_{j=1}^p (\mu_j - \nu_j)^a$, so $\mathcal{U}(a)$ is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{j=1}^p (\mu_j - \nu_j)^a$. On the other hand, for $\|\mathcal{E}\|_\infty$, following the maximum-type test statistic in Cai et al. [9], we have

$$\mathcal{U}(\infty) = \max_{1\le j\le p} \sigma_{j,j}^{-1} (\bar{x}_j - \bar{y}_j)^2, \tag{4.3}$$

where $\bar{x}_j = \sum_{i=1}^{n_x} x_{i,j}/n_x$, $\bar{y}_j = \sum_{i=1}^{n_y} y_{i,j}/n_y$. We then obtain results similar to Theorems 2.1, 2.3 and 2.5. As the conditions are similar to those in

Section 2, we only keep the key conclusions, and the details of conditions and discussions are given in Supplementary Material Section **??**.

THEOREM 4.3. *Under Condition* **??** *in Supplementary Material,* $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ *and* $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$*, for any finite integers* $(a_1, \ldots, a_m)$*, as* $n, p \to \infty$*,* $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$*, where* $\sigma^2(a) \simeq a! \sum_{j_1,j_2=1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a / (n_x n_y)^a$ *is of the order* $\Theta(a! p n^{-a})$*.*

THEOREM 4.4. *Under Condition* **??** *in Supplementary Material,* $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ *and* $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$*,* $\forall u \in \mathbb{R}$*,* $P(\frac{n_x n_y}{n_x + n_y} \mathcal{U}(\infty) - \tau_p \le u) \to \exp\{-\pi^{-1/2} \exp(-u/2)\}$*, as* $n, p \to \infty$*, where* $\tau_p = 2 \log p - \log \log p$*. Moreover,* $\{\mathcal{U}(a)/\sigma(a)\}$ *of finite integer* a *and* $\{n_x n_y \mathcal{U}(\infty)/(n_x + n_y) - \tau_p\}$ *are asymptotically independent.*

Theorems 4.3 and 4.4 provide the asymptotic properties of finite-order U-statistics and $\mathcal{U}(\infty)$ under H_0. To analyze the power of $\mathcal{U}(a)$'s, we derive the asymptotic results of $\mathcal{U}(a)$'s under the alternative hypotheses. We focus on the two-sample mean testing problem, while one-sample mean testing can be obtained similarly. Specifically, we consider the alternative $\mathcal{E}_A = \{\mu_j - \nu_j = \rho > 0 \text{ for } j = 1, \ldots, k_0; \mu_j - \nu_j = 0 \text{ for } j = k_0 + 1, \cdots, p\}$. We then obtain similar conclusions to Theorem 2.5.

THEOREM 4.5. *Assume Condition* **??** *in Supplementary Material and* $k_0 = o(p)$*. For any finite integers* $\{a_1, \ldots, a_m\}$*, if* ρ *in* $\mathcal{E}_A$ *satisfies* $\rho = O(k_0^{-1/a_t} p^{1/(2a_t)} n^{-1/2})$ *for* $t = 1, \ldots, m$*, then* $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$*, as* $n, p \to \infty$*. Here* $\mathrm{E}[\mathcal{U}(a)] = \|\mathcal{E}_A\|_a^a = k_0 \rho^a$ *and* $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq V_a$*, with* $V_a = a! \sum_{j_1,j_2=k_0+1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a / (n_x n_y)^a$ *of the order* $\Theta(a! p n^{-a})$*.*

Next we compare the power of different U-statistics under alternatives with different sparsity levels. Theorem 4.5 shows that under the local alternatives, the asymptotic power of $\mathcal{U}(a)$ mainly depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Therefore by Theorem 4.5, given constant $M > 0$, for each $\mathcal{U}(a)$, if $\rho = M^{1/a} k_0^{-1/a} V_a^{1/(2a)}$, then $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$; that is, different $\mathcal{U}(a)$'s have the same power asymptotically. For easy illustration, we consider $\sigma_{j_1,j_2} = 1$ when $j_1 = j_2 \in \{k_0 + 1, \ldots, p\}$, and $\sigma_{j_1,j_2} = 0$ when $j_1 \ne j_2 \in \{k_0 + 1, \ldots, p\}$, then $M^{1/a} k_0^{-1/a} V_a^{1/(2a)} \simeq \rho_a$ with

$$\rho_a := a!^{\frac{1}{2a}} (M\sqrt{p}/k_0)^{\frac{1}{a}} \{(n_x + n_y)/(n_x n_y)\}^{\frac{1}{2}}. \tag{4.4}$$

Therefore, similarly to the analysis in Section 2.2, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the minimum ρ_a in (4.4). We have the following result similar to Proposition 2.3.

PROPOSITION 4.1. *Given any constant $M \in (0, +\infty)$ and n, p, k_0, we consider ρ_a in (4.4) as a function of positive integers a, then*

(i) when $k_0 \geq M\sqrt{p}$, the minimum of ρ_a is achieved at $a_0 = 1$;
(ii) when $k_0 < M\sqrt{p}$, the minimum of ρ_a is achieved at some a_0, which increases as $M\sqrt{p}/|J_D|$ increases.

Proposition 4.1 shows that when the sparsity level k_0 is large, i.e., $\mathcal{E}_a$ is dense, a small a tends to obtain a smaller lower bound in ρ, and vice versa. As (4.4) and (2.13) are similar, we have similar patterns to that in Figure 1 when examining the corresponding numerical plots of ρ_a. In addition, [9] shows that when $\rho = \rho_\infty := C_1\sqrt{\log p/n}$ for a large C_1, the power of $\mathcal{U}(\infty)$ converges to 1, and $\sqrt{\log p/n}$ is minimax rate optimal for sparse alternatives; see also [17]. Thus, if $\rho_\infty < \rho_{a_0}$, i.e., $k_0 < MC_1^{-a_0}\sqrt{pa_0!}/\log^{a_0/2} p$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$. On the other hand, Proposition 4.1 shows that when $\mathcal{E}_A$ is dense with $k_0 > \sqrt{Mp}$, $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{p}k_0^{-1}n^{-1/2})$. Moreover, for some large M and C_2, when $\mathcal{E}_A$ is "moderately dense" or "moderately sparse" with $C_2\sqrt{pa_0!}/\log^{a_0/2} p < k_0 < \sqrt{Mp}$, $\mathcal{U}(a_0)$ is the "best" and its lowest detectable order of ρ is $\Theta\{(\sqrt{p}/k_0)^{\frac{1}{a_0}}n^{-1/2}\}$, which is of a smaller order than the optimal detection boundary of the sparse case $\Theta(\sqrt{\log p/n})$.

More generally, when $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, similar results to Theorems 4.3 and 4.5 can be obtained. In particular, we have the following corollary.

COROLLARY 4.1. *When $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, under Condition ?? in Supplementary Material, Theorem 4.3 holds with $\sigma^2(a) \simeq a!\sum_{j_1,j_2=1}^{p}(\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$ and Theorem 4.5 holds with $V_a = a!\sum_{j_1,j_2=k_0+1}^{p}(\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$.*

Corollary 4.1 shows that the asymptotic power of finite-order U-statistics depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. By the construction of finite-order U-statistics and the proof, we obtain that $\mathrm{E}\{\mathcal{U}(a)\} = k_0\rho^a$ and $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(a!pn^{-a})$. We then know that for finite-order U-statistics, similar results to Proposition 4.1 still hold by examining $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$.

The above power analysis shows that the optimal U-statistic varies when the alternative hypothesis changes. To achieve high power across various alternatives, we can develop an adaptive test similar to that in Section 2.3. Specifically, we calculate the p-values of the U-statistics (4.1) and (4.2) following the theoretical results above and the algorithm in Section 2.3. By combining the p-values as discussed in Section 2.3, the asymptotic power of the adaptive test goes to 1 if there exists one $\mathcal{U}(a)$ whose power goes to 1.

REMARK 4.1. *Xu et al. [66] has also discussed the adaptive testing of two-sample mean that is powerful against various ℓ_p-norm-like sums of $\boldsymbol{\mu}-\boldsymbol{\nu}$. But [66] is under the framework of a family of von Mises V-statistics where $\mathcal{V}(a)=\sum_{j=1}^{p}(\bar{x}_j-\bar{y}_j)^a$. We note that $\mathcal{V}(a)$ is equivalent to*

$$\mathcal{V}(a)=\sum_{j=1}^{p}\sum_{c=0}^{a}(-1)^{a-c}\binom{a}{c}({n_x}^{c}{n_y}^{a-c})^{-1}\sum_{\substack{1\le k_1,\cdots,k_c\le n_x\\ 1\le s_1,\cdots,s_{a-c}\le n_y}}\prod_{t=1}^{c}x_{k_t,j}\prod_{m=1}^{a-c}y_{s_m,j},$$

which allows the indexes k's and s's to be the same and thus is different from the U-statistics in (4.2). [66] shows that the constructed V-statistics are biased estimators of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and $\mathcal{V}(a)$ and $\mathcal{V}(b)$ are asymptotically independent if $a+b$ is odd, but are asymptotically correlated if $a+b$ is even. The constructed U-statistics in this work extend the properties of those V-statistics such that $\mathcal{U}(a)$ in (4.2) is an unbiased estimator of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and all $\mathcal{U}(a)$'s are asymptotically independent with each other. Given these nice statistical properties, it becomes easier to obtain the joint asymptotic distribution of the U-statistics, and then apply the adaptive test.

4.2. *Two-Sample Covariance Testing.* The U-statistics framework can be applied similarly to testing the equality of two covariance matrices. Suppose $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$ are i.i.d. copies of two independent random vectors $\mathbf{x}=(x_1,\ldots,x_p)^\intercal$ and $\mathbf{y}=(y_1,\ldots,y_p)^\intercal$ respectively. Denote $\mathrm{E}(\mathbf{x})=\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^\intercal$, $\mathrm{E}(\mathbf{y})=\boldsymbol{\nu}=(\nu_1,\ldots,\nu_p)^\intercal$; $\mathrm{cov}(\mathbf{x})=\boldsymbol{\Sigma}_x=\{\sigma_{x,j_1,j_2}:1\le j_1,j_2\le p\}$ and $\mathrm{cov}(\mathbf{y})=\boldsymbol{\Sigma}_y=\{\sigma_{y,j_1,j_2}:1\le j_1,j_2\le p\}$. Consider $H_0:\boldsymbol{\Sigma}_x=\boldsymbol{\Sigma}_y=\boldsymbol{\Sigma}=(\sigma_{j_1,j_2})_{p\times p}$. Given $1\le j_1,j_2\le p$, $1\le k_1\ne k_2\le n_x$, and $1\le s_1\ne s_2\le n_y$, $K_{j_1,j_2}(\mathbf{x}_{k_1},\mathbf{x}_{k_2},\mathbf{y}_{s_1},\mathbf{y}_{s_2})=(x_{k_1,j_1}x_{k_1,j_2}-x_{k_1,j_1}x_{k_2,j_2})-(y_{s_1,j_1}y_{s_1,j_2}-y_{s_1,j_1}y_{s_2,j_2})$ is a simple unbiased estimator of $\sigma_{x,j_1,j_2}-\sigma_{y,j_1,j_2}$. Therefore, for a finite positive integer a, we have the U-statistic

$$\mathcal{U}(a)=\sum_{1\le j_1,j_2\le p}\frac{1}{P_{2a}^{n_x}P_{2a}^{n_y}}\sum_{\substack{1\le k_{1,1}\ne k_{1,2}\ne\ldots\\ \ne k_{a,1}\ne k_{a,2}\le n_x}}\sum_{\substack{1\le s_{1,1}\ne s_{1,2}\ne\ldots\\ \ne s_{a,1}\ne s_{a,2}\le n_y}}\prod_{t=1}^{a}K_{j_1,j_2}(\mathbf{x}_{k_{t,1}},\mathbf{x}_{k_{t,2}},\mathbf{y}_{s_{t,1}},\mathbf{y}_{s_{t,2}}).\tag{4.5}$$

As in Remark 2.1, another formulation of $\mathcal{U}(a)$ equivalent to (4.5) is

$$\mathcal{U}(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c}(-1)^{c-b_1+b_2}\sum_{1\le j_1,j_2\le p}\sum_{\substack{1\le i_1\ne\ldots\ne\\ i_{2c-b_1}\le n_x}}\sum_{\substack{1\le w_1\ne\ldots\ne\\ w_{2(a-c)-b_2}\le n_y}} C_{n_x,n_y,a,c,b_1,b_2}\times\prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2})\prod_{s=b_1+1}^{c}x_{i_s,j_1}\prod_{t=c+1}^{2c-b_1}x_{i_t,j_2}$$
$$\times\prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2})\prod_{l=b_2+1}^{a-c}y_{w_l,j_1}\prod_{q=a-c+1}^{2(a-c)-b_2}y_{w_q,j_2}, \tag{4.6}$$

where $C_{n_x,n_y,c,b_1,b_2} = (P_{2c-b_1}^{n_x}P_{2(a-c)-b_2}^{n_y})^{-1}a!/\{b_1!(c-b_1)!b_2!(a-c-b_2)!\}$, and (4.6) shall be used in the theoretical developments.

We next present the asymptotic results of the constructed U-statistics under the null hypothesis. Here we assume the regularity Condition **??** or **??**, whose details and discussions are provided in Section **??** of Supplementary Material due to the space limitation. We mention that Condition **??** is a mixing-type dependence assumption similar to Condition 2.2, and Condition **??** is a moment-type dependence assumption similar to Condition 2.2*. Particularly, Condition **??** extends the moment assumption for second-order U-statistics in Li and Chen [44] to U-statistics of general orders; please see the detailed discussions in Section **??**.

THEOREM 4.6. *Under H_0 and Condition* **??** *or* **??** *in Supplementary Material, for finite integers $\{a_1,\ldots,a_m\}$, $[\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0,I_m)$, where for $a\in\{a_1,\ldots,a_m\}$,*

$$\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\}$$
$$\simeq \sum_{1\le j_1,j_2,j_3,j_4\le p} a!\Big\{\frac{1}{n_x}(\Pi^x_{j_1,j_2,j_3,j_4}-\sigma_{j_1,j_2}\sigma_{j_3,j_4})+\frac{1}{n_y}(\Pi^y_{j_1,j_2,j_3,j_4}-\sigma_{j_1,j_2}\sigma_{j_3,j_4})\Big\}^a$$

with $\Pi^x_{j_1,j_2,j_3,j_4} = \mathrm{E}\{\prod_{t=1}^4(x_{1,j_t}-\mu_{j_t})\}$ and $\Pi^y_{j_1,j_2,j_3,j_4} = \mathrm{E}\{\prod_{t=1}^4(y_{1,j_t}-\nu_{j_t})\}$.

Theorem 4.6 provides the asymptotic independence and joint normality of the finite-order U-statistics, which are similar to Theorems 2.1, 4.1 and 4.3. To further study the power of these finite-order U-statistics, we next consider the alternative hypotheses where $\mathbf{\Sigma}_x \ne \mathbf{\Sigma}_y$. Let $\mathbb{J}_0$ be the largest subset of $\{1,\ldots,p\}$ such that $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2} = \sigma_{j_1,j_2}$ for any $j_1,j_2\in\mathbb{J}_0$. We then obtain the following theorem under the regularity conditions given in Section **??** of Supplementary Material.

THEOREM 4.7. *Under Conditions* **??** *and* **??** *in the Supplementary Material, for finite integers* $\{a_1, \ldots, a_m\}$, $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where*

$$\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq a! C_{\kappa,a} \sum_{j_1,j_2,j_3,j_4 \in \mathbb{J}_0} \sigma^a_{j_1,j_2} \sigma^a_{j_3,j_4},$$

and $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$ *with* κ_x *and* κ_y *given in Condition* **??**.

Given the asymptotic results under the alternatives, we next analyze the power of the finite-order U-statistics. By Theorem 4.7, the asymptotic power of $\mathcal{U}(a)$ depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Let $J_D = \{(j_1, j_2) : \sigma_{x,j_1,j_2} \neq \sigma_{y,j_1,j_2}, 1 \leq j_1, j_2 \leq p\}$, then $\mathrm{E}\{\mathcal{U}(a)\} = \sum_{(j_1,j_2)\in J_D} (\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2})^a$. Similarly to Section 2.2, to study the relationship between the sparsity level of $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ and the power of U-statistics, we consider the case where the nonzero differences between $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. Specifically, let $\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2} = \rho$ for $(j_1, j_2) \in J_D$, and then $\mathrm{E}\{\mathcal{U}(a)\} = |J_D|\rho^a$. Following the analysis in Section 2.2, we compare the ρ values needed by different $\mathcal{U}(a)$'s to achieve $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ for a given constant M. In particular, for given integer a, suppose $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ is achieved when $\rho = \rho_a$. For any $a \neq b$, we compare $\mathcal{U}(a)$ and $\mathcal{U}(b)$ following Criterion 1.

We use the following example as an illustration, where $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ satisfy the conditions of Theorem 4.7. Specifically, we assume that $\mathbf{\Sigma}_x = (\sigma_{x,j_1,j_2})_{p\times p}$ has the diagonal elements $\sigma_{x,j,j} = \nu^2$; and the off-diagonal elements $\sigma_{x,j_1,j_2} = h_{|j_1-j_2|} \in (0, \nu^2)$ with $h_{|j_1-j_2|} = \Theta(\nu^2)$ when $|j_1 - j_2| \leq s$, while $\sigma_{x,j_1,j_2} = 0$ when $|j_1 - j_2| > s$. This covers the moving average covariance structure of order s, and $\mathbf{\Sigma}_x$ is a banded matrix with bandwidth s. In addition, we assume the bandwidth $s = o(p)$ and $p - |\mathbb{J}_0| = o(p)$. By the definition of $\mathbb{J}_0$, the assumption $p - |\mathbb{J}_0| = o(p)$ implies that a large square sub-matrix of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. For simplicity, we let $n_x = n_y$ with $n = n_x + n_y$, and a similar analysis can be applied when $n_x \neq n_y$. By Theorem 4.7, $\mathrm{var}\{\mathcal{U}(a)\} \simeq (n/2)^{-a} a!\{2\kappa_1^a + \kappa_2^a\}\{p\nu^{2a} + 2\sum_{t=1}^s h_t^a (p-t)\}^2$, where $\kappa_1 = \kappa_x + \kappa_y$ and $\kappa_2 = \kappa_x + \kappa_y - 2$. Therefore we know for given finite integer a, $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ holds when $\rho = \rho_a$ defined as

$$\rho_a = \frac{(a!)^{\frac{1}{2a}} \sqrt{\kappa_1}\nu}{(n/2)^{1/2}} \Big(\frac{Mp}{|J_D|}\Big)^{1/a} \Big\{2 + \Big(\frac{\kappa_2}{\kappa_1}\Big)^a\Big\}^{\frac{1}{2a}} \Big\{1 + 2\sum_{t=1}^s \Big(\frac{h_t}{\nu^2}\Big)^a \Big(1 - \frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

We next compare the ρ_a's and obtain the following proposition.

PROPOSITION 4.2. *There exists $\mathbb{D}_0$ that only depends on the given $\kappa_x, \kappa_y, \nu^2, s$, and $h_t, t = 1, \ldots, s$, and satisfies $\mathbb{D}_0 = \Theta(1/s^2)$ such that*

(i) When $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at $a_0 = 1$.
(ii) When $|J_D| < Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_D|$ increases.

Proposition 4.2 is similar to Propositions 2.3 and 4.1. Following the analysis in Section 2.2, Proposition 4.2 shows that when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "very" dense with $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, $\mathcal{U}(1)$ is the most powerful U-statistic; when $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ becomes sparser as $Mp/|J_D|$ decreases, a higher order U-statistic is more powerful; when the $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "moderately" dense or sparse, a U-statistic of finite order $a_0 > 1$ would be the most powerful one.

The power analysis above shows that the power of the U-statistics varies when the alternative changes. To maintain high power across different alternatives, we can develop an adaptive testing procedure similar to that in Section 2.3. Given the asymptotic independence in Theorem 4.6, an adaptive testing procedure using the constructed $\mathcal{U}(a)$'s is valid with the type I error asymptotically controlled. Also, the adaptive test achieves high power by combining the U-statistics as discussed in Section 2.3.

We provide simulation studies on two-sample covariance testing in Supplementary Material Section **??**. By the simulations, we first find that the type I errors of the U statistics and the adaptive test are well controlled under H_0. This verifies the theoretical results in Theorem 4.7. Second, similarly to the one-sample covariance testing, we find that generally when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is sparser, a U-statistic of higher order is more powerful, and vice versa. Moreover, under moderately sparse/dense alternatives, $\mathcal{U}(a_0)$ with $a_0 > 1$ could achieve the highest power. The results are consistent with Proposition 4.2. Third, we compare the proposed adaptive test with existing methods in literature including [56, 60, 44, 8], and find that the proposed adaptive testing procedure maintains high power across various alternatives.

REMARK 4.2. *Similarly to Section 2, we can let $\mathcal{U}(\infty)$ be the maximum-type test statistic in [8], and expect that the result similar to Theorem 2.3 holds under certain regularity conditions. However, as the dependence structure of two-sample covariance matrices is more complicated than the one-sample case, it is more challenging to establish the asymptotic joint distribution of $\mathcal{U}(\infty)$ and finite-order U-statistics. We leave this interesting problem for future study, while find in simulations that the performance of $\mathcal{U}(\infty)$ is similar to high-order U-statistics $\mathcal{U}(a)$'s.*

4.3. *Generalized Linear Model.* In this section, we consider the Example 3 of generalized linear models (on Page 4) to show that the proposed framework can be extended to other testing problems. Similarly to the results in Section 4.1, we show that the constructed U-statistics are asymptotically independent and normally distributed, and also establish the power analysis results of the U-statistics. We provide the details in Section **??** of Supplementary Material. Recently, Wu et al. [64] also discussed the adaptive testing of generalized linear model. But similarly to [66], [64] is under the framework of a family of von Mises V-statistics, and thus is different from the current paper as discussed in Remark 4.1. Moreover, the current work provides the theoretical power analysis while [64] did not.

5. Discussion. This paper introduces a general U-statistics framework for applications to high-dimensional adaptive testing. Particularly, we focus on the examples including testing of means, covariances and regression coefficients in generalized linear models. Under the null hypothesis, we prove that the U-statistics of finite orders have asymptotic joint normality, and establish the asymptotic mutual independence among the finite-order U-statistics and $\mathcal{U}(\infty)$. Moreover, under alternative hypotheses, we analyze the power of different U-statistics and demonstrate how the most powerful U-statistic changes with the sparsity level of the alternative parameters. Based on the theoretical results, we propose an adaptive testing procedure, which is powerful against different alternatives. The superior performance of this adaptive testing is confirmed in the simulations and real data analysis.

There are several possible extensions of the U-statistics framework in this paper. First, by our current proof, the convergence rate in Theorem 2.3 is bounded by $O(\log^{-1/2} p)$, which is an upper bound and not sharp. From our extensive simulations, we find that the type I error rate of the adaptive testing is well-controlled with a relatively small p, e.g., $p = 50$. We might obtain a shaper bound of the convergence rate, but more refined concentration property of the high-dimensional and high-order U-statistics is needed. Second, the proposed framework requires that the elements in the parameter set $\mathcal{E}$ have unbiased estimates. When we can not obtain unbiased estimates easily, e.g., for the precision matrix, the proposed construction may not follow directly. Nevertheless we may use "nearly" unbiased estimators to construct "U-statistics" for hypothesis testing, such as the "nearly" unbiased estimator of the precision matrix proposed in [65]; the main challenge is then to control the accumulative bias over the parameters under high-dimensions. Third, this paper discusses the examples where the elements in $\mathcal{E}$ are comparable. When the parameters in $\mathcal{E}$ are not comparable, such as $\mathcal{E}$ containing

both means and covariances parameters, the construction of U-statistics still follows but the theoretical derivation may require a careful case-by-case examination. Fourth, the construction of the U-statistics treats the parameters in $\mathcal{E}$ with equal weight. More generally, we could assign different weights to different parameter estimators. For instance, standardizing the data is one example of assigning different weights. As inappropriate weight assignments could lead to power loss, when the truth is unknown, how to effectively assign weights to maximize the test power is an interesting research question. We shall discuss these extensions in the future as a significant amount of additional work is still needed.

In addition to the examples in this paper, the proposed U-statistics framework can be applied to other high-dimensional hypothesis testing problems. For example, it can be applied to testing the block-diagonality of a covariance matrix, whose theoretical analysis would be similar to the considered one sample and two sample covariance testing problems. It can also be used to test high-dimensional regression coefficients in complex regression models other than the generalized linear models, following a similar construction based on the score functions. A key step is then to characterize the impact of nuisance parameters that are estimated under the null hypothesis, and challenges arise especially when the nuisance parameters are high-dimensional. Such interesting extensions will be further explored in our follow-up studies.

Acknowledgements. The authors thank Co-Editors Prof. Edward I. George and Prof. Richard J. Samworth, an Associate Editor, and three anonymous referees for their constructive comments. The authors also thank Prof. Ping-Shou Zhong for sharing the code of the paper [13] and Prof. Xuming He and Prof. Peter Song for helpful discussions.

SUPPLEMENTARY MATERIAL

Supplementary Material:
(doi: XXX; Supplementary.pdf). This supplementary material contains the technical proofs of the main paper and additional simulations.

REFERENCES

[1] Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 14*(3), 367–429.

[2] Anderson, T. W. (2009). *An introduction to multivariate statistical analysis.* John Wiley & Sons, New York.

[3] Bai, Z., D. Jiang, J.-F. Yao, and S. Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist. 37*(6B), 3822–3840.

[4] Bai, Z. and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 311–329.

[5] Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

[6] Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application 4*(1), 423–446.

[7] Cai, T. T. and T. Jiang (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist. 39*(3), 1496–1525.

[8] Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

[9] Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.

[10] Cai, T. T. and Z. Ma (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli 19*(5B), 2359–2388.

[11] Chen, S. X., J. Li, and P.-S. Zhong (2019). Two-sample and ANOVA tests for high dimensional means. *The Annals of Statistics 47*(3), 1443–1474.

[12] Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist. 38*(2), 808–835.

[13] Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

[14] Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *The Annals of Statistics 46*(2), 642–678.

[15] Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger, and J. E. Kleinman (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature 478*(7370), 519.

[16] Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist. 32*(3), 962–994.

[17] Donoho, D. and J. Jin (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science 30*(1), 1–25.

[18] Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association 91*(434), 674–688.

[19] Fan, J., F. Han, and H. Liu (2014). Challenges of big data analysis. *National science review 1*(2), 293–314.

[20] Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

[21] Fan, J., J. Lv, and L. Qi (2011). Sparse high-dimensional models in economics.

[22] Frahm, G. (2004). *Generalized elliptical distributions: theory and applications*. Ph. D. thesis, Universität zu Köln.

[23] Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science 326*(5951), 399–403.

[24] Gaetan, C. and X. Guyon (2010). *Spatial statistics and modeling*, Volume 90. Springer.

[25] Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493.

[26] Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association 110*(510), 837–849.

[27] Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist. 38*(3), 1686–1732.

[28] Heinig, M., E. Petretto, C. Wallace, L. Bottolo, M. Rotival, H. Lu, Y. Li, R. Sarwar, S. R. Langley, A. Bauerfeind, et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature 467*(7314), 460.

[29] Ho, H.-C. and T. Hsing (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of Applied Probability 33*(1), 138–145.

[30] Ho, H.-C. and W. P. McCormick (1999). Asymptotic distribution of sum and maximum for Gaussian processes. *Journal of Applied Probability 36*(4), 1031–1044.

[31] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist. 19*(3), 293–325.

[32] Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab. 23*(2), 938–947.

[33] James, B., K. James, and Y. Qi (2007). Limit distribution of the sum and maximum from multivariate Gaussian sequences. *Journal of multivariate analysis 98*(3), 517–532.

[34] Jansen, I. E., J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics 51*, 404–413.

[35] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab. 14*(2), 865–880.

[36] Jiang, T. and F. Yang (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist. 41*(4), 2029–2074.

[37] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*(2), 295–327.

[38] Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research 38*(suppl 1), D355–D360.

[39] Kim, J., Y. Zhang, and W. Pan (2016). Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics 203*(2), 715–731.

[40] Lan, W., R. Luo, C.-L. Tsai, H. Wang, and Y. Yang (2015). Testing the diagonality of a large covariance matrix in a regression setting. *Journal of Business & Economic Statistics 33*(1), 76–86.

[41] Ledoit, O. and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist. 30*(4), 1081–1102.

[42] Leung, D. and M. Drton (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics 46*(1), 280–307.

[43] Li, D. and L. Xue (2015). Joint limiting laws for high-dimensional independence tests. *ArXiv e-prints*.

[44] Li, J. and S. X. Chen (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist. 40*(2), 908–940.

[45] Liu, W.-D., Z. Lin, and Q.-M. Shao (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab. 18*(6), 2337–2366.

[46] Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, et al. (2009). Finding the missing heritability of complex diseases. *Nature 461*(7265), 747.

[47] McCormick, W. and Y. Qi (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *Journal of Applied Probability 37*(4), 958–971.

[48] Mosteller, F. and R. A. Fisher (1948). Questions and answers. *The American Statis-*

tician 2(5), 30–31.
[49] Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*. John Wiley & Sons, New York.
[50] Paindaveine, D. and G. Van Bever (2014). Inference on the shape of elliptical distributions based on the MCD. *Journal of Multivariate Analysis 129*, 125–144.
[51] Pan, W., J. Kim, Y. Zhang, X. Shen, and P. Wei (2014). A powerful and adaptive association test for rare variants. *Genetics 197*(4), 1081–1095.
[52] Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields 143*(3-4), 481–516.
[53] Peng, Z. and S. Nadarajah (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory of Probability & Its Applications 47*(4), 706–709.
[54] Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications 19*(2), 297–303.
[55] Prince, M., R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia 9*(1), 63–75.
[56] Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis 51*(12), 6535–6542.
[57] Shao, Q.-M. and W.-X. Zhou (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab. 42*(2), 623–648.
[58] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics 108*(5-6), 1033–1056.
[59] Srivastava, M. S. and M. Du (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 99*(3), 386–402.
[60] Srivastava, M. S. and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis 101*(6), 1319–1329.
[61] Srivastava, R., P. Li, and D. Ruppert (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics 25*(3), 954–970.
[62] Storey, J. D. and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences 100*(16), 9440–9445.
[63] Wang, L., P. Jia, R. D. Wolfinger, X. Chen, and Z. Zhao (2011). Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics 98*(1), 1–8.
[64] Wu, C., G. Xu, and W. Pan (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica*.
[65] Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika 102*(2), 247–266.
[66] Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.
[67] Xu, J., S. L. Murphy, K. D. Kochanek, B. Bastian, and E. Arias (2018). Deaths: Final data for 2016. *National Vital Statistics Reports 67*(5).
[68] Xu, Z., G. Xu, and W. Pan (2017). Adaptive testing for association between two random vectors in moderate to high dimensions. *Genetic epidemiology 41*(7), 599–609.
[69] Yang, Q. and G. Pan (2017). Weighted statistic in detecting faint and sparse alter-

natives for high-dimensional covariance matrices. *Journal of the American Statistical Association 112*(517), 188–200.

[70] Yu, K., Q. Li, A. W. Bergen, R. M. Pfeiffer, P. S. Rosenberg, N. Caporaso, P. Kraft, and N. Chatterjee (2009). Pathway analysis by adaptive combination of p-values. *Genetic epidemiology 33*(8), 700–709.

[71] Zhong, P.-S. and S. X. Chen (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association 106*(493), 260–274.

[72] Zhong, P.-S., S. X. Chen, and M. Xu (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist. 41*(6), 2820–2851.

Yinqiu He and Gongjun Xu
Department of Statistics
University of Michigan
E-mail: yqhe@umich.edu
gongjun@umich.edu

Chong Wu
Department of Statistics
Florida State University
E-mail: cwu3@fsu.edu

Wei Pan
Division of Biostatistics
School of Public Health
University of Minnesota
E-mail: panxx014@umn.edu