

Accepted Manuscript

Some high-dimensional one-sample tests based on functions of interpoint distances

Enakshi Saha, Soham Sarkar, Anil K. Ghosh

PII: S0047-259X(17)30025-8
DOI: <http://dx.doi.org/10.1016/j.jmva.2017.07.006>
Reference: YJMVA 4271

To appear in: *Journal of Multivariate Analysis*

Received date : 11 January 2017

Please cite this article as: E. Saha, S. Sarkar, A.K. Ghosh, Some high-dimensional one-sample tests based on functions of interpoint distances, *Journal of Multivariate Analysis* (2017), <http://dx.doi.org/10.1016/j.jmva.2017.07.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Some high-dimensional one-sample tests based on functions of interpoint distances

Enakshi Saha^a, Soham Sarkar^{b,*}, Anil K. Ghosh^b

^aDepartment of Statistics, University of Chicago, 5747 South Ellis Avenue Chicago, IL 60637, USA

^bTheoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, India

Abstract

The multivariate one-sample location problem is well studied in the literature, and several tests are available for it. But most of the existing one-sample tests perform poorly for high-dimensional data, and many of them are not even applicable when the dimension of the data exceeds the sample size. In this article, we develop and investigate some nonparametric one-sample tests based on functions of interpoint distances. These proposed tests can be conveniently used in high dimension, low sample size (HDLSS) situations, and good power properties of these tests for HDLSS data have been established using both theoretical as well as numerical results.

Keywords: High-dimensional consistency, HDLSS data, Rotation invariance, Scale invariance.

1. Introduction

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be n independent realizations of a d -dimensional random vector \mathbf{X} having a continuous distribution $F_{\boldsymbol{\theta}}$, which is symmetric about $\boldsymbol{\theta} \in \mathbb{R}^d$ (i.e., $\mathbf{X} - \boldsymbol{\theta}$ and $\boldsymbol{\theta} - \mathbf{X}$ have the same distribution). In the one-sample problem, we test the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative hypothesis $\mathcal{H}_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a pre-specified point in \mathbb{R}^d (without loss of generality, we will assume $\boldsymbol{\theta}_0 = \mathbf{0}$ throughout this article). This problem is well investigated in the literature. If $F_{\boldsymbol{\theta}}$ is assumed to be normal, one uses the Hotelling's T^2 statistic to perform the test. There are several nonparametric tests as well. Sign and signed-rank tests for bivariate data include [7, 9–11, 14, 24, 26, 31]. Puri and Sen [34] used co-ordinate wise signs and ranks to develop some nonparametric tests for the multivariate one-sample location problem. Randles [35, 36] proposed one-sample tests based on interdirections, which can be viewed as multivariate generalizations of Blumen's test [7]. Chaudhuri and Sengupta [15] generalized Hodges' sign test [24] to higher dimension. Hallin and Paindaveine [21] proposed a test based on interdirections and pseudo-Mahalanobis distances. In [3, 22, 23, 29] also, the authors proposed some nonparametric tests for the multivariate one-sample problem. A brief overview of most of these tests can be found in [27, 30, 32].

However, most of these multivariate tests usually yield poor performance in high dimensions. Moreover, none of them can be used when the dimension exceeds the sample size. In the recent past, some Hotelling's T^2 -type one-sample tests have been proposed in the literature, which can be used in high dimension, low sample size (HDLSS) situations; see e.g., [16, 33, 37, 38]. These tests are based on the asymptotic null distribution of the test statistic, where the dimension is assumed to increase with the sample size. However, these tests are mainly concerned with the mean vector of a high-dimensional distribution. They are not robust against outliers and usually yield poor performance when the underlying distribution has heavy tails. Recently, Biswas et al. [6] took a graph-theoretic approach and proposed some distribution-free one-sample tests based on the shortest covering path. Among them the one-sample run test had the best overall performance in their experiments. However, finding the shortest covering path is computationally expensive. In fact, this is equivalent to solving the traveling salesman problem, which is NP complete; see [18]. Wang et al. [40] proposed a one-sample sign test for high-dimensional data, which is motivated by the

*Corresponding author

Email addresses: enu2.saha@gmail.com (Enakshi Saha), sohamsarkar1991@gmail.com (Soham Sarkar), akghosh@isical.ac.in (Anil K. Ghosh)

elliptic symmetry of the underlying distribution. But this test only uses the directions of the observations from θ_0 , and ignores important information contained in their distances. Another class of tests based on higher criticism (see, e.g., [1, 13, 17, 19, 42]) is available in the literature. But these tests are particularly useful for sparse alternatives. In practice, we often have situations, where most of the measurement variables contain negligible evidence against \mathcal{H}_0 , but their combined effect is not negligible. The higher criticism idea may not be helpful in such situations.

In this article, we develop two classes of tests based on functions of interpoint distances between the observations. These tests are fairly simple, computationally efficient and can be conveniently used for high-dimensional data even when the dimension is much larger than the sample size. In the next section, we introduce the first class of tests based on functions of the Euclidean distance. So, the resulting tests are invariant under location shift and rotation of the data. In Section 3, we introduce another class of tests based on functions of other distances, which are scale invariant. High-dimensional consistencies of these two classes of tests are established under appropriate regularity conditions. Several simulated and real data sets are analyzed in Sections 4 and 5, respectively, to compare the performance of the proposed tests with some existing one-sample tests. Some additional numerical results are presented in an Online Supplement. Finally, a brief summary of the work and some concluding remarks are given in Section 6. All proofs and mathematical details are postponed to the Appendix.

2. Tests based on Euclidean distances

Let \mathbf{X}_1 and \mathbf{X}_2 be two independent copies of $\mathbf{X} \sim F_{\theta}$. Note that if $\theta = \mathbf{0}$, then for $i \in \{1, 2\}$, \mathbf{X}_i and $-\mathbf{X}_i$ have the same distribution. Therefore, if $E(\|\mathbf{X}\|) < \infty$, we have $E(\|\mathbf{X}_1 + \mathbf{X}_2\|) = E(\|\mathbf{X}_1 - \mathbf{X}_2\|)$, where $\|\cdot\|$ denotes the Euclidean distance. But if $\theta \neq \mathbf{0}$, $E(\|\mathbf{X}_1 + \mathbf{X}_2\|)$ exceeds $E(\|\mathbf{X}_1 - \mathbf{X}_2\|)$ (follows from Lemma 1 stated below). Similarly, we have $E(\|\mathbf{X}_1 + \mathbf{X}_2\|^2) - E(\|\mathbf{X}_1 - \mathbf{X}_2\|^2) = 4E(\mathbf{X}_1^\top \mathbf{X}_2) = 4\|\theta\|^2 \geq 0$, where the equality holds if and only if $\theta = \mathbf{0}$. So, one can use the empirical analog

$$\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} (\|\mathbf{X}_i + \mathbf{X}_j\| - \|\mathbf{X}_i - \mathbf{X}_j\|) \quad \text{or} \quad \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} (\|\mathbf{X}_i + \mathbf{X}_j\|^2 - \|\mathbf{X}_i - \mathbf{X}_j\|^2)$$

as the test statistic, and reject \mathcal{H}_0 for large values of it. In fact, the test statistic

$$T_{CQ} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{X}_i^\top \mathbf{X}_j$$

considered by Chen and Qin [16] is equivalent to the latter one. But the performance of this test is affected by the presence of outliers, and it performs poorly when the underlying distribution has heavy tails. Recently, Wang et al. [40] used $\mathbf{X}_i/\|\mathbf{X}_i\|$, the spatial sign of \mathbf{X}_i , to come up with the one-sample test statistic

$$T_{WPL} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \frac{\mathbf{X}_i^\top \mathbf{X}_j}{\|\mathbf{X}_i\| \|\mathbf{X}_j\|}.$$

Compared to T_{CQ} , the test based on T_{WPL} is more robust against outliers generated from heavy-tailed distributions. But in order to achieve robustness, it completely ignores the valuable information on the magnitudes of the observations and only considers the angles between all pairs of data points. As a result, it often fails to achieve satisfactory power. The tests we consider in this section do not ignore this information on the magnitudes of the observations, but at the same time, they are robust against extreme values and contaminating observations. These tests are mainly motivated by the following result stated as Lemma 1.

Lemma 1. Let \mathbf{X}_1 and \mathbf{X}_2 be two independent random vectors having a common distribution F_{θ} , which is symmetric about $\theta \in \mathbb{R}^d$. Consider a function $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ (here \mathbb{R}_+ denotes the interval $[0, \infty)$) such that $\psi'(t)/t$ is a non-constant monotone function on $(0, \infty)$. If $E\{\psi(\|\mathbf{X}_1 + \mathbf{X}_2\|)\}$ and $E\{\psi(\|\mathbf{X}_1 - \mathbf{X}_2\|)\}$ are finite, then $E\{\psi(\|\mathbf{X}_1 + \mathbf{X}_2\|) - \psi(\|\mathbf{X}_1 - \mathbf{X}_2\|)\} \geq 0$, where the equality holds if and only if $\theta = \mathbf{0}$.

This lemma shows that for an appropriate choice of ψ , one can consider the test statistic

$$T_\psi = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \{\psi(\|\mathbf{X}_i + \mathbf{X}_j\|) - \psi(\|\mathbf{X}_i - \mathbf{X}_j\|)\}$$

and reject \mathcal{H}_0 for large values of T_ψ . In fact, such tests were briefly considered in [39], where the authors also discussed about their large sample consistency. Clearly, tests based on T_ψ can be conveniently used for high-dimensional data even when the dimension is much larger than the sample size. Note that if we use $\psi(t) = t^2$, we get

$$T_\psi = 4T_{CQ} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} 4\|\mathbf{X}_i\| \|\mathbf{X}_j\| \cos(\Theta_{ij}),$$

which not only uses the information on the angles Θ_{ij} between the observations, but also the information contained in the magnitudes of the observations. Although $\psi(t) = t^2$ does not satisfy the condition stated in Lemma 1, the same result holds for this choice of ψ , as we have seen before. However, the use of $\psi(t) = t^2$ makes the test sensitive to outliers and extreme values. We can take care of this problem by using other appropriate functions. Note that if ψ is bounded (e.g., $\psi(t) = t^2/(1+t^2)$ or $\psi(t) = 1 - e^{-t}$), the moment condition in Lemma 1 holds trivially. Tests based on bounded ψ functions or tests based on functions that diverge slowly (e.g., $\psi(t) = \ln(1+t)$) are expected to be more robust than the test based on $\psi(t) = t$ or $\psi(t) = t^2$. These bounded or slowly increasing functions utilize the information on the magnitudes of the observations in a controlled manner to make the tests based on T_ψ robust.

Under the moment condition stated in Lemma 1,

$$T_\psi \xrightarrow{\text{Pr}} E(T_\psi) = E\{\psi(\|\mathbf{X}_1 + \mathbf{X}_2\|) - \psi(\|\mathbf{X}_1 - \mathbf{X}_2\|)\}$$

as $n \rightarrow \infty$ (follows from the result on convergence of U -statistics). Now, if ψ satisfies the condition of Lemma 1, we have $E(T_\psi) = 0$ when $\theta = \mathbf{0}$ and $E(T_\psi) > 0$ when $\theta \neq \mathbf{0}$. Therefore, the power of the test based on T_ψ converges to 1 as $n \rightarrow \infty$. However, as mentioned in [41], this type of consistency in the classical asymptotic regime is a rather trivial property of a test. The power of any reasonable test usually converges to unity as the sample size increases. But in the HDLSS asymptotic regime, where the sample size remains fixed and the dimension increases, consistency of a test is not a trivial property. Many well known tests fail to have consistency in the HDLSS setup; see discussions in [4, 41].

Here, we investigate the power property of the test based on T_ψ in the HDLSS asymptotic regime. However, first note that both $\|\mathbf{X}_1 - \mathbf{X}_2\|$ and $\|\mathbf{X}_1 + \mathbf{X}_2\|$ usually diverge with the dimension d . Therefore, to make T_ψ applicable to HDLSS data, we use an appropriate scaling and consider ψ of the form $\psi(t) = \phi(t/\sqrt{d})$ for some $\phi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$. There are several functions of this type that satisfy the conditions given in Lemma 1. For instance, one can use $\phi(t) = t$, $\phi(t) = \ln(1+t)$, $\phi(t) = t^2/(1+t^2)$ or $\phi(t) = 1 - e^{-t}$. So, we consider test statistics of the form

$$T_\phi^d = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left\{ \phi\left(\frac{\|\mathbf{X}_i + \mathbf{X}_j\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{d}}\right) \right\},$$

and reject \mathcal{H}_0 for large values of T_ϕ^d . The critical value is computed using a resampling method. Each time, we generate a new sample of the form $\{\mathbf{X}_1^* = a_1\mathbf{X}_1, \dots, \mathbf{X}_n^* = a_n\mathbf{X}_n\}$, where a_1, \dots, a_n are iid random variables taking values 1 and -1 each with probability $1/2$. The statistic T_ϕ^d is computed based on this new sample. This procedure is repeated several times to simulate the null distribution of T_ϕ^d and hence to determine the cut-off.

In order to investigate the high-dimensional behavior of this test, we assume that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are n independent realizations of a d -dimensional random vector $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})^\top$ following the distribution F_θ , which satisfy the conditions given below.

(A1) The fourth moments of the component variables $X^{(1)}, X^{(2)}, \dots$ are uniformly bounded.

(A2) For $\mathbf{U} = \mathbf{X}_1$ and $\mathbf{V} = \pm\mathbf{X}_2$, $\sum_{q_1 \neq q_2} \text{corr}\{(U^{(q_1)} - V^{(q_1)})^2, (U^{(q_2)} - V^{(q_2)})^2\}$ is of the order $\mathbf{o}(d^2)$.

Under (A1) and (A2), high-dimensional consistency of the test based on T_ϕ^d is given by the following theorem, where the sample size n is assumed to be fixed and the dimension d diverges to infinity.

Theorem 1. Suppose that F_{θ} satisfies (A1) and (A2), and $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a strictly increasing function. If $\liminf_{d \rightarrow \infty} \|\theta\|^2/d > 0$ and $2^{n-1} \geq 1/\alpha$, then the power of the test based on T_{ϕ}^d converges to 1 as $d \rightarrow \infty$.

The assumption (A2) indicates a form of weak dependence among the measurement variables. In case of a time series, (A2) holds if the series has the ρ -mixing property, stationarity of the time series is not required here. Recently, the authors in [5, 6] assumed similar conditions for investigating the high-dimensional behavior of their one-sample and two-sample tests. Similar assumptions were also considered by [20] for studying the high-dimensional behavior of some popular classifiers and by [25] to establish the high-dimensional consistency of their estimated principal component directions.

We need the moment assumption (A1) and the weak dependence assumption (A2) to have the weak law of large numbers (WLLN) for dependent and non-identically distributed random variables $\{(X_1^{(q)} \pm X_2^{(q)})^2 : q \geq 1\}$; the proof is straightforward and hence it is omitted. If the component variables $X^{(1)}, X^{(2)}, \dots$ are iid, (A2) holds automatically and instead of (A1), one needs the existence of second order moments only. In the classical asymptotic regime, we consider the dimension to be fixed and expect to get more information as the sample size increases. But in HDLSS asymptotic regime, the sample size is considered to be fixed, and under (A1)–(A2), we expect to get more information as the dimension increases. The condition $\liminf_{d \rightarrow \infty} \|\theta\|^2/d > 0$ ensures that the evidence against \mathcal{H}_0 remains significant as the dimension diverges. One can relax this condition on $\|\theta\|^2$ if further assumptions are made on the underlying distribution. For instance, if one assumes $\{X^{(q)} : q \geq 1\}$ to be a strictly stationary ρ -mixing sequence with mixing coefficient ρ satisfying $\sum_{k \geq 1} \rho(2^k) < \infty$, following Theorem 2.1 in [12], the high-dimensional consistency of the test based on T_{ϕ}^d can be achieved under a slightly weaker condition on $\|\theta\|^2$.

3. Scale invariant tests based on other distance functions

In Section 2, we considered some tests based on functions of the Euclidean distance and also discussed about some suitable functions to make the tests robust. The resulting tests turn out to be consistent in classical asymptotic regime under appropriate moment conditions. But in order to prove their consistency in the HDLSS asymptotic regime, we had to assume a fourth moment condition. Though this moment condition is only sufficient, and such an assumption is pretty common in the HDLSS literature [20, 25], for a nonparametric method, one would ideally like to relax this assumption. We can do that if we construct tests based on other distance functions. For instance, the distance between two observations $\mathbf{X}_1 = (X_1^{(1)}, \dots, X_1^{(d)})$ and $\mathbf{X}_2 = (X_2^{(1)}, \dots, X_2^{(d)})$ can be computed as $h\{\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|)\}$, where $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are suitable monotonically increasing functions with $h(0) = \varphi(0) = 0$. Clearly, this class of distance functions includes all ℓ_p distances with $p \geq 1$. However, $\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|)$ diverges with the dimension. Therefore, to use it meaningfully for HDLSS data, we scale it by a factor of $1/d$ and use the test statistic

$$T_{h,\varphi}^d = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left[h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi(|X_i^{(q)} + X_j^{(q)}|) \right\} - h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi(|X_i^{(q)} - X_j^{(q)}|) \right\} \right].$$

If second moments of the $\varphi(|X_1^{(q)} \pm X_2^{(q)}|)$'s are uniformly bounded and $\sum_{q_1 \neq q_2} \text{corr}\{\varphi(|X_1^{(q_1)} \pm X_2^{(q_1)}|), \varphi(|X_1^{(q_2)} \pm X_2^{(q_2)}|)\}$ is of the order $\mathbf{o}(d^2)$, the test based on $T_{h,\varphi}^d$ has the high-dimensional consistency when

$$\liminf_{d \rightarrow \infty} \mathbb{E} \left[d^{-1} \sum_{q=1}^d \{\varphi(|X_i^{(q)} + X_j^{(q)}|) - \varphi(|X_i^{(q)} - X_j^{(q)}|)\} \right] > 0.$$

The proof is similar to the proof of Theorem 1. Therefore, if φ is bounded, we can completely remove the moment condition. Note that if $\varphi(t)$ satisfies the condition in Lemma 1, we have

$$\mathbb{E} \left[d^{-1} \sum_{q=1}^d \{\varphi(|X_i^{(q)} + X_j^{(q)}|) - \varphi(|X_i^{(q)} - X_j^{(q)}|)\} \right] \geq 0$$

for any fixed d , where the equality holds if and only if $\theta = \mathbf{0}$. So, it is reasonable to assume its limit inferior to be positive under the alternative. The distance function $h\{\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|)\}$ is location invariant. We can also

make it scale invariant if for each $q \in \{1, \dots, d\}$, we divide $|X_1^{(q)} - X_2^{(q)}|$ by $s^{(q)}$, an equivariant estimate of the scale corresponding to the q th component variable, and use $h\{d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})\}$ as the distance between \mathbf{X}_1 and \mathbf{X}_2 . The test statistic based on this distance function is given by

$$\tilde{T}_{h,\varphi}^d = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left[h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} - h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right].$$

Note that while we have $E(\tilde{T}_{h,\varphi}^d) = 0$ under \mathcal{H}_0 , for suitable choices of h and φ , $E(\tilde{T}_{h,\varphi}^d)$ turns out to be positive under the alternative. This result is stated as Lemma 2 below.

Lemma 2. *Let \mathbf{X}_1 and \mathbf{X}_2 be two independent and identically distributed random vectors having a common distribution F_θ with marginal densities $f^{(1)}, \dots, f^{(d)}$, where $f^{(q)}(x)$ is decreasing in $|x - \theta^{(q)}|$. Also assume that $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are strictly increasing functions. If h is differentiable and $E[h\{d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} \pm X_2^{(q)}|/s^{(q)})\}] < \infty$, then $E(\tilde{T}_{h,\varphi}^d) \geq 0$, where the equality holds if and only if $\theta = \mathbf{0}$.*

Note that Lemma 2 holds irrespective of the choice of the data based estimates $s^{(1)}, \dots, s^{(d)}$, and it suggests us to reject \mathcal{H}_0 for large values of $\tilde{T}_{h,\varphi}^d$. The cutoff is computed using the resampling technique discussed before. Since \mathbf{X} and $-\mathbf{X}$ have the same distribution under \mathcal{H}_0 , along with $\mathbf{X}_1, \dots, \mathbf{X}_n$, we take their negatives $-\mathbf{X}_1, \dots, -\mathbf{X}_n$ to have a collection of $2n$ observations. Marginal semi inter-quartile ranges computed based on these $2n$ observations are used as the $s^{(q)}$'s. Because of this choice of the $s^{(q)}$'s, we do not need to recompute them during resampling, and this leads to substantial saving in computing time.

This type of scale invariant tests is particularly useful when the measurement variables are not of comparable units and scales. Note that, in order to construct a meaningful test, it is not necessary for $h\{d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})\}$ to be a distance between \mathbf{X}_1 and \mathbf{X}_2 . For instance, if φ satisfies the properties mentioned in Lemma 2, and we take $h(t) = t$, it may not always lead to a distance function, but it will ensure $E(\tilde{T}_{h,\varphi}^d) \geq 0$, where the equality will hold if and only if $\theta = \mathbf{0}$. However, if $\varphi(|t_1 - t_2|)$ gives a distance between t_1 and t_2 in \mathbb{R} , $h(t) = t$ makes $h\{d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})\}$ a distance in \mathbb{R}^d . For $\varphi(|t_1 - t_2|) = |t_1 - t_2|^p$ and $h(t) = t$, it becomes a function of the ℓ_p -distance.

For investigating the high-dimensional behavior of the test based on $\tilde{T}_{h,\varphi}^d$, here we consider assumptions similar to (A1) and (A2). For n independent observations $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} F_\theta$, we define, for each for $i \in \{1, \dots, n\}$, $\mathbf{Z}_i = (X_i^{(1)}/s^{(1)}, X_i^{(2)}/s^{(2)}, \dots)$, and we assume

(B1) The second moments of the $\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|)$'s are uniformly bounded.

(B2) $\sum_{q_1 \neq q_2} \text{corr}\{\varphi(|Z_1^{(q_1)} \pm Z_2^{(q_1)}|), \varphi(|Z_1^{(q_2)} \pm Z_2^{(q_2)}|)\}$ is of the order $\mathbf{o}(d^2)$.

Note that (B1) holds automatically if φ is bounded. The condition (B2) holds when the measurement variables $\{X^{(q)} : q \geq 1\}$ are either independent or they have weak dependence among themselves. For instance, (B2) holds when the measurement variables satisfy the ρ -mixing property; see Theorem 5.2 in [8]. Under (B1) and (B2), WLLN holds for the sequence $\{\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|) : q \geq 1\}$, i.e.,

$$\left| \frac{1}{d} \sum_{q=1}^d \varphi(|Z_1^{(q)} \pm Z_2^{(q)}|) - \frac{1}{d} \sum_{q=1}^d E\{\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|)\} \right| \xrightarrow{\text{Pr}} 0 \text{ as } d \rightarrow \infty.$$

Now, let us define

$$\tau_{d,\varphi}(\theta) = d^{-1} \sum_{q=1}^d E\{\varphi(|Z_1^{(q)} + Z_2^{(q)}|) - \varphi(|Z_1^{(q)} - Z_2^{(q)}|)\}.$$

If φ satisfies the condition of Lemma 2, for any fixed d , we have $\tau_{d,\varphi}(\theta) \geq 0$ where the equality holds if and only if $\theta = \mathbf{0}$ (follows from the proof of Lemma 2). So, it is reasonable to assume that $\liminf_{d \rightarrow \infty} \tau_{d,\varphi}(\theta) > 0$. The following theorem asserts the high-dimensional consistency of our scale invariant tests under that condition.

Theorem 2. *Suppose that F_θ satisfies (B1) and (B2), and $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a strictly increasing function. If $\tau_\varphi = \liminf_{d \rightarrow \infty} \tau_{d,\varphi}(\theta) > 0$ and $2^{n-1} \geq 1/\alpha$, then the power of the test based on $\tilde{T}_{h,\varphi}^d$ converges to 1 as $d \rightarrow \infty$.*

4. Analysis of simulated data sets

We analyzed several high-dimensional simulated data sets to compare the performance of our proposed methods with some popular one-sample tests available in the literature. In particular, we used the one-sample tests proposed by Srivastava [37], Chen and Qin [16], Park and Ayyala [33], Wang et al. [40] (referred to as SR, CQ, PA and WPL tests, respectively) and the run test developed in [6] for comparison. While CQ, WPL and run tests are rotation invariant, the other two tests have the scale invariance property. The run test is distribution-free. In all other cases, we used the test based on asymptotic null distribution of the test statistic, where d is assumed to grow with n . Note that among other regularity conditions, CQ, PA and SR tests require the existence of fourth order moments for the asymptotic distribution, and they are not robust. The run test and the WPL test are known to have better robustness properties.

For our tests based on the Euclidean distance, we used $\phi(t) = t$, $\ln(1 + t)$, $1 - e^{-t}$ and $t^2/(1 + t^2)$, but for the last two choices, the results were almost the same. So, here we report the results for the first three choices only, and the corresponding test statistics are denoted by T_{lin} , T_{log} and T_{exp} , respectively. Analogous versions of scale invariant tests were constructed using $\varphi(t) = t$, $\ln(1 + t)$ and $1 - e^{-t}$ while $h(t) = t$ was used in all these cases. The corresponding test statistics are denoted by \tilde{T}_{lin} , \tilde{T}_{log} and \tilde{T}_{exp} , respectively. Note that $\varphi(t) = t$ and $\varphi(t) = 1 - e^{-t}$ lead to proper distance functions in \mathbb{R}^d . As discussed in Sections 2 and 3, in all these cases, we used permutation tests, where the critical value was computed based 10,000 random generations of the sign vector $(a_1, \dots, a_n)^\top$. We also considered another test with $\varphi_0(t) = t^2$ and $h_0(t) = \sqrt{t}$, which basically leads to a test based on the Euclidean distance computed using standardized versions of the component variables. But the results were similar to those obtained for \tilde{T}_{lin} . Therefore, we do not report them in this article. Throughout this article, all tests are considered to have 5% nominal level.

We began with examples involving normal, t_3 (Student's t with 3 degrees of freedom) and Cauchy distributions. These distributions were chosen for varying degrees of heaviness of their tails. Note that normal distributions have finite moments of all order, t_3 distributions have finite first and second order moments only, and Cauchy distributions do not have finite moments of any order. For studying the level properties of different tests, first we generated 20 observations from a d -variate normal distribution with the location parameter $\mathbf{0} = (0, \dots, 0)^\top$ and the scatter matrix $\mathbf{S}_d = (s_{ij})$, where $s_{ij} = 0.5^{|i-j|}$ for all $i, j \in \{1, \dots, d\}$. These observations were used to test $\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ against $\mathcal{H}_A : \boldsymbol{\theta} \neq \mathbf{0}$. This experiment was repeated 500 times to estimate the sizes of different tests by the proportion of times they rejected \mathcal{H}_0 . Similar experiment was carried out with t_3 and Cauchy distributions. We considered five different choices of d , viz., 30, 60, 120, 250 and 500, to study the behavior of these tests as the dimension increases.

In all these examples, the SR test had sizes below the nominal level, especially in higher dimensions. In the case of Cauchy distributions, this test had observed size close to 0 for all values of d . The PA test also had size somewhat lower than 0.05 for Cauchy distributions. However, all other tests rejected the true null hypothesis in nearly 5% of the cases in all the examples considered here. These results are reported in detail in the Online Supplement.

Next, we investigate the power properties of different tests. For this investigation, we generated 20 observations from the normal, t_3 or Cauchy distribution, where the location parameter had the first $d/2$ elements equal to δ and the rest equal to 0. The scatter matrix of the distribution was of the form $\boldsymbol{\Sigma}_d = \boldsymbol{\Lambda}_d^{1/2} \mathbf{S}_d \boldsymbol{\Lambda}_d^{1/2}$, where $\mathbf{S}_d = (0.5^{|i-j|})$ and $\boldsymbol{\Lambda}_d$ was a diagonal matrix with the first $d/2$ diagonal elements equal to λ and the rest equal to 1. We considered two sets of experiments. For the first set we used $\delta = 0.1$ and $\lambda = 1/3$, while for the second set we used $\delta = 0.25$ and $\lambda = 3$. Each experiment was carried out for values of d ranging between 30 and 2000. We repeated the experiments 500 times as before to compute the powers of different tests, and they are reported in Figure 1.

Note that in all these examples, only the first $d/2$ variables contained evidence against \mathcal{H}_0 , while the last $d/2$ variables can be viewed as noise. In the first (respectively, second) set of examples, the first $d/2$ variables had lower (respectively, higher) variance than the last $d/2$ variables. So, the important variables had lower (respectively, higher) contribution to the Euclidean norm compared to the unimportant variables. Thus, the test statistics based on the Euclidean norm contained more noise (respectively, information). In contrast, the scale invariant tests treated all the variables equally in both the cases. So, the overall performance of our rotation invariant tests was much inferior (respectively, superior) compared to our scale invariant tests in the first (respectively, second) set of examples. Recall that the CQ test is rotation invariant, and the PA test is scale invariant. So, in the first type (respectively, second type) of examples with normal distributions, the PA test had superior (respectively, inferior) performance compared to the CQ test. The SR test, however, had poor performance in all these examples, especially in higher dimensions.

In the case of normal distributions, CQ, WPL and our rotation invariant tests had similar performance. Among them, the CQ test had the best performance followed by T_{lin} . But we observed a completely different picture in cases

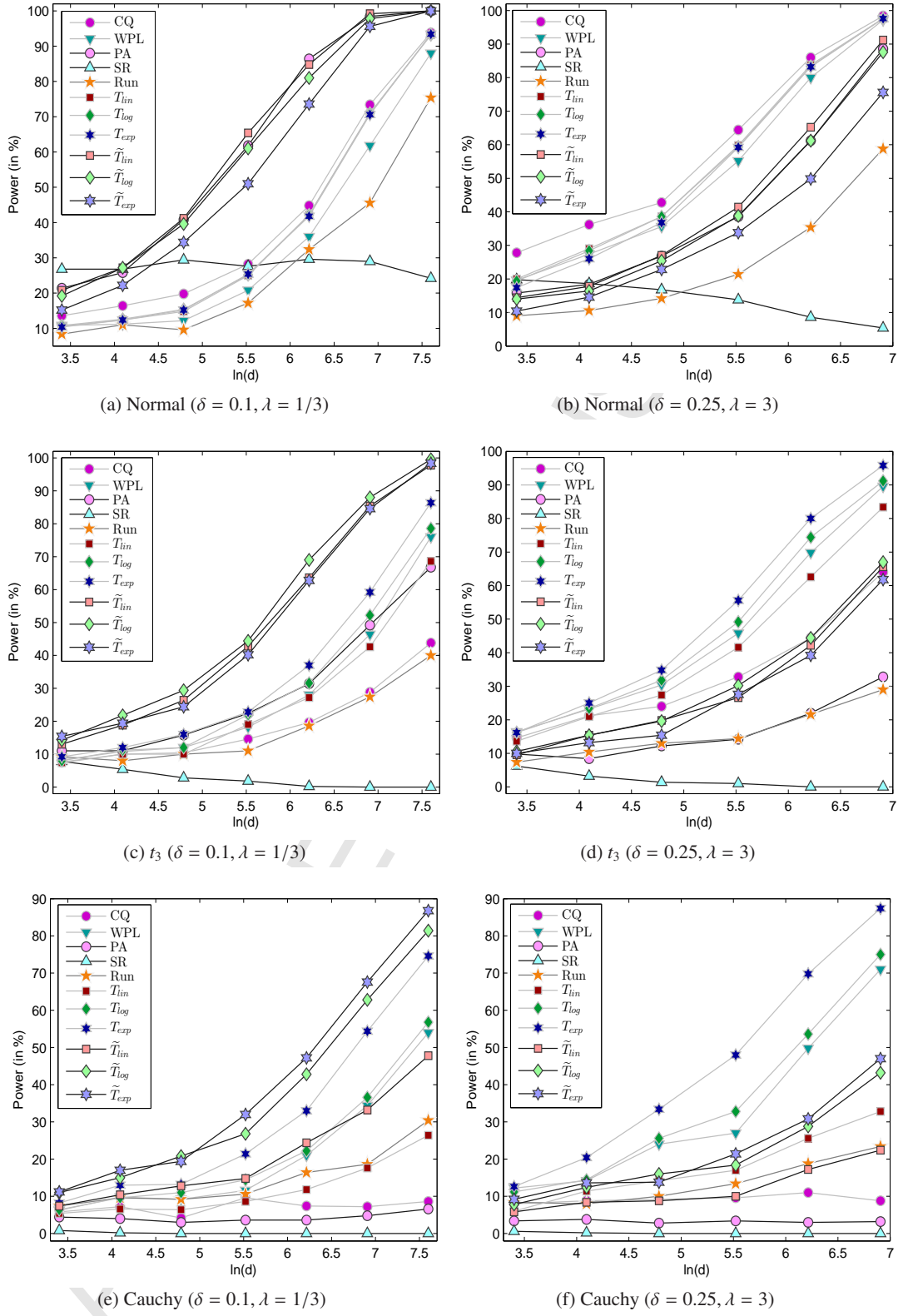


Figure 1: Powers of different tests for different choices of d .

of t_3 and Cauchy distributions, especially in the latter case. In such cases, the CQ test had poor performance, and T_{lin} also had lower power. But the tests based on T_{exp} and T_{log} , especially the previous one, performed much better than all other rotation invariant tests considered here. We observed similar results for scale invariant tests as well. In the case of normal distributions, PA and \tilde{T}_{lin} tests performed better than the tests based on \tilde{T}_{log} and \tilde{T}_{exp} . But in cases of t_3 and Cauchy distributions, while the PA and \tilde{T}_{lin} tests failed to yield satisfactory powers, tests based on \tilde{T}_{log} and \tilde{T}_{exp} had much better performance.

Next, we carried out these two sets of experiments for varying choices of δ when d was kept fixed at 250. Here also, our findings remained the same. The results reported in Figure 2 basically shows the same phenomena that were observed in Figure 1. We carried out similar experiments with $d = 40$ and $n = 50$ as well. In those cases, since the sample size was larger than the dimension, Hotelling's T^2 test and other multivariate nonparametric tests like the coordinate-wise sign and rank tests [34] and spatial sign and rank tests [28] could also be used for comparison. But in almost all cases, these five tests had lower powers compared to most of the high-dimensional tests considered in this section. Relative performances of different high-dimensional tests were almost the same as observed in Figures 1 and 2. So, instead of reporting those results in this article, we report them in the Online Supplement.

We also considered some examples with non-elliptic distributions. Table 1 presents the results for uniform, Laplace and mixture normal distributions, where we used $d = 250$ and three different sample sizes ($n = 20, 30$ and 40). In the first two examples, component variables were independent and identically distributed. In one case, we generated them from $\mathcal{U}(-2, 2)$ distribution, and in the other case, they were generated from the standard Laplace distribution. After that, we shifted their locations from the origin by adding $(0.1, \dots, 0.1)^\top$ to the measurement vector. Our proposed tests performed well in these two examples. In the case of uniform distribution, while CQ, PA and our rotation invariant tests had higher powers than their competitors, in the Laplace distribution, our scale invariant tests outperformed all other tests considered here. The SR test and the run test had poor performance in these examples. As expected, powers of different tests increased with the sample size, but their relative ordering remained almost the same.

Table 1: Powers of different tests for non-elliptic distributions

n	Uniform			Laplace			Mixture Normal		
	20	30	40	20	30	40	20	30	40
PA	39.4	70.0	89.0	20.8	39.8	57.8	36.0	54.2	67.6
SR	4.0	18.6	52.6	1.2	8.6	23.0	40.2	57.0	68.2
CQ	47.8	76.0	91.6	31.2	50.2	66.2	40.2	57.2	68.2
WPL	36.8	65.2	86.0	24.4	37.8	55.0	13.2	17.6	25.0
Run	15.2	22.8	31.6	19.4	21.8	27.0	100	100	100
T_{lin}	45.6	74.0	91.0	28.6	48.2	64.6	29.2	50.8	69.4
T_{log}	45.6	73.6	90.6	28.6	48.4	65.0	46.2	90.4	99.2
T_{exp}	45.8	73.4	90.2	28.6	48.6	65.0	99.6	100	100
\tilde{T}_{lin}	31.4	54.8	75.6	44.6	68.8	89.4	29.0	50.4	67.8
\tilde{T}_{log}	28.4	51.0	72.2	50.2	75.4	93.0	34.8	67.6	93.4
\tilde{T}_{exp}	26.4	47.0	67.6	47.8	69.6	89.0	51.8	95.0	100

Next, we considered a mixture of three normal distributions all having the same scatter matrix $0.25\mathbf{I}_{250}$, and location parameters $-2\mathbf{1}_{250}$, $\mathbf{1}_{250}$ and $4\mathbf{1}_{250}$, respectively. Here \mathbf{I}_{250} is the 250×250 identity matrix, and $\mathbf{1}_{250}$ is a 250-dimensional vector with all elements equal to 1. The mixing proportions for these three distributions were taken as 0.4, 0.2 and 0.4, respectively, to make the distribution symmetric about $\mathbf{1}_{250}$. Our test based on T_{exp} had excellent performance in this example. The tests based on \tilde{T}_{exp} and T_{log} also yielded competitive powers for $n = 30$ and $n = 40$. For $n = 40$, \tilde{T}_{log} also rejected \mathcal{H}_0 in more than 93% cases. CQ, PA, SR tests and the tests based on T_{lin} and \tilde{T}_{lin} had similar performance, particularly for $n = 30$ and $n = 40$. But the WPL test had powers much lower than its competitors.

Finally, we considered some examples involving perturbed normal distributions to study the robustness of different tests. We generated 20 observations from a multivariate normal distribution with the location parameter $0.15\mathbf{1}_{250}$ and the dispersion matrix of the form $0.25\{(1 - \rho)\mathbf{I}_{250} + \rho\mathbf{1}_{250}\mathbf{1}_{250}^\top\}$. Each observation underwent a random operation,

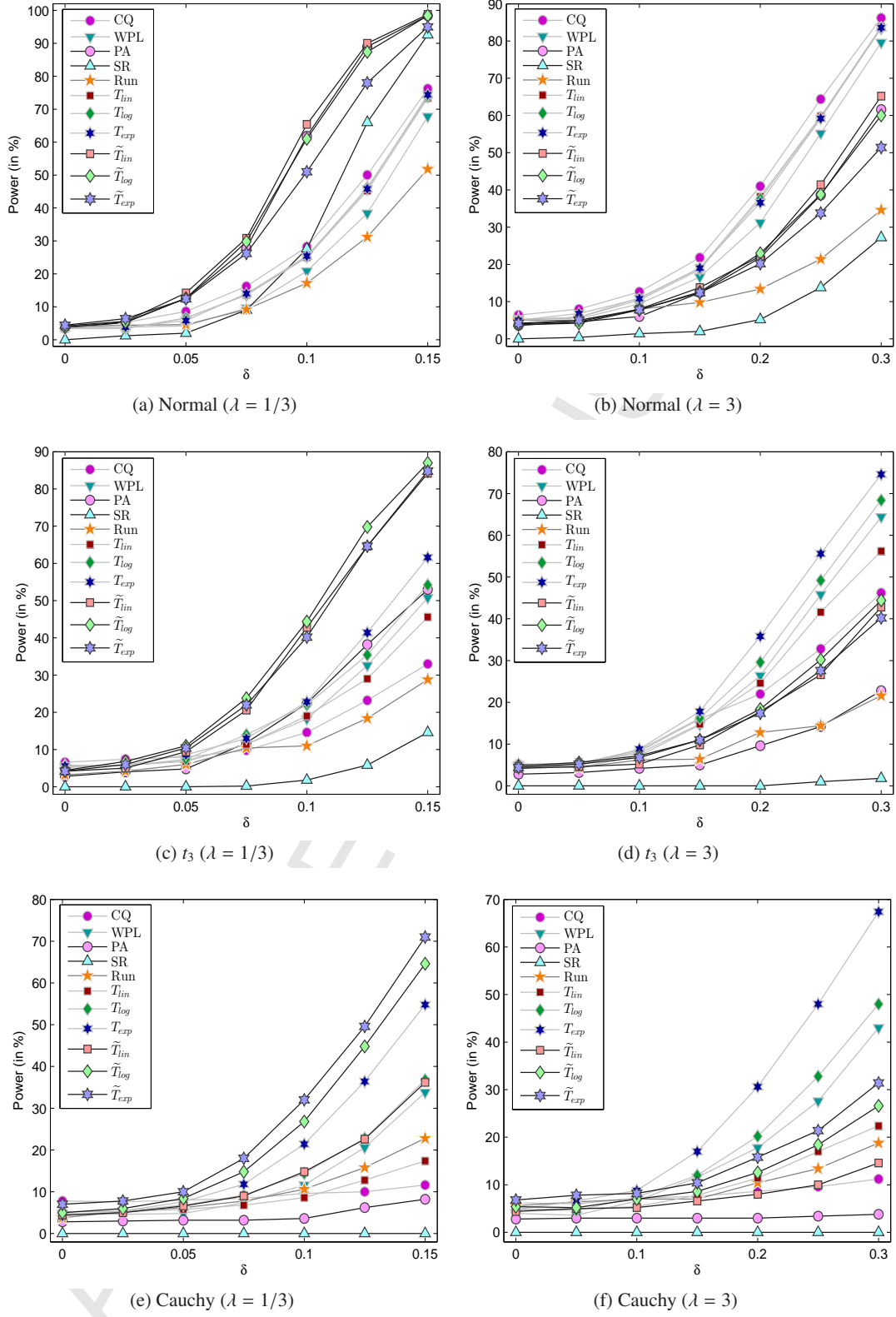


Figure 2: Powers of different tests for different choices of δ .

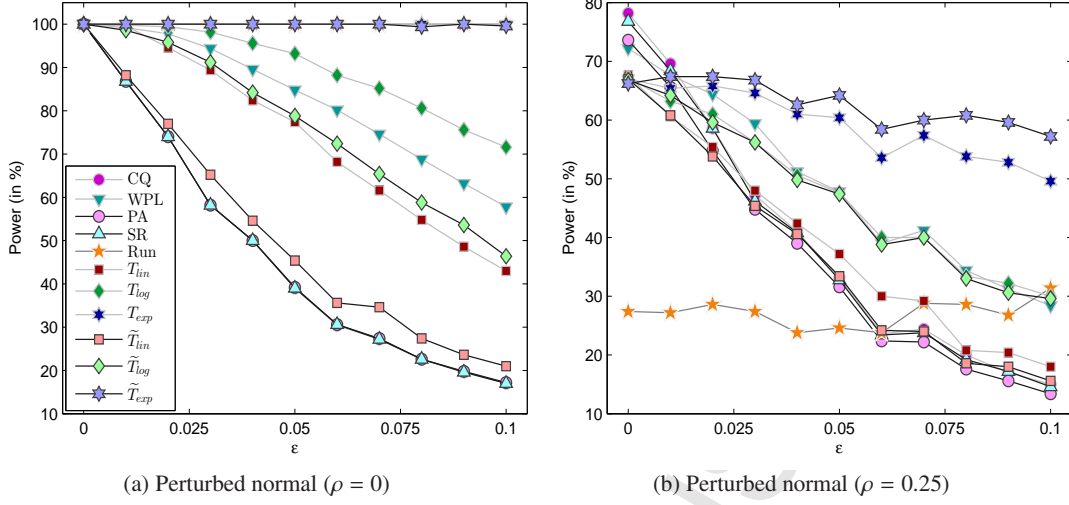


Figure 3: Powers of different tests in perturbed normal example.

where with probability ϵ , it was contaminated by subtracting $2\mathbf{1}_{250}$ from it. We used 2 values of ρ (0 and 0.25) and 11 equidistant values of ϵ in the range $[0, 0.1]$. Each experiment was repeated 500 times as before, and the observed powers of different tests are presented in Figure 3. This figure clearly shows better robustness properties of the tests based on T_{exp} and \tilde{T}_{exp} . In the case of $\rho = 0$, even for $\epsilon = 0.1$, these two tests and the run test rejected \mathcal{H}_0 on all occasions, when the powers of the other tests dropped down substantially. In the case of $\rho = 0.25$, the run test could not perform well, but the superiority of T_{exp} and \tilde{T}_{exp} was evident even in this correlated setup.

We carried out this experiment under heteroscedastic setup as well, where we used the dispersion matrix $\Sigma_0 = \Lambda_0^{1/2}\{(1 - \rho)\mathbf{I}_{250} + \rho\mathbf{1}_{250}\mathbf{1}_{250}^T\}\Lambda_0^{1/2}$ for Λ_0 being the diagonal matrix with its i th diagonal element equal to $0.3\{(i \bmod 10) + 1\}$ for each $i \in \{1, \dots, d\}$. The observations were randomly contaminated as before. Table 2 presents the results for two choices of ρ (0 and 0.25) and two choices of ϵ (0 and 0.1). In this heteroscedastic setup, our scale invariant tests performed better than their rotation invariant analogs. For $\rho = 0$, T_{exp} , \tilde{T}_{exp} and run tests were less affected by contamination. These three tests outperformed all other tests for $\epsilon = 0.1$. For $\rho = 0.25$, though the performance of the run test was least affected by contamination, \tilde{T}_{exp} had the highest power for $\epsilon = 0.1$.

Table 2: Powers of different tests in perturbed normal distributions under heteroscedastic setup

ρ	ϵ	PA	SR	CQ	WPL	Run	T_{lin}	T_{log}	T_{exp}	\tilde{T}_{lin}	\tilde{T}_{log}	\tilde{T}_{exp}
0	0.0	100	100	100	100	99.8	100	100	100	100	100	100
	0.1	17.0	16.8	16.8	43.2	99.4	31.0	46.0	93.8	29.0	50.4	98.4
0.25	0.0	77.6	81.4	60.6	56.2	34.0	52.4	52.8	51.8	72.2	71.6	71.6
	0.1	12.6	13.8	11.4	18.0	33.0	11.2	16.0	27.0	15.4	24.6	51.0

5. Analysis of average temperature data

The Average Daily Temperature Archive (<http://academic.udayton.edu/kissock/http/Weather/>) of the University of Dayton (Ohio, USA) contains daily average temperatures for several cities from January 1, 1995 till date. For our analysis, we considered the average temperatures from 1996 to 2015 and divided it into two parts of 10 consecutive years each. We considered the difference between the daily temperatures in these two parts as the random vector \mathbf{X} and used different methods to test whether the average daily temperature has changed over the period of 10 years. From the list of cities we considered African and Asian cities separately. Many of these cities had missing

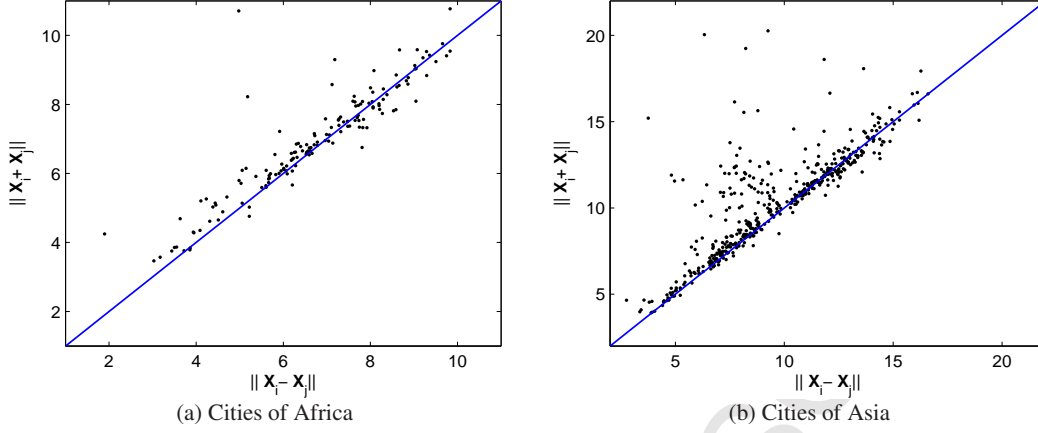


Figure 4: Scatterplot of $\|\mathbf{X}_i - \mathbf{X}_j\|$ vs. $\|\mathbf{X}_i + \mathbf{X}_j\|$.

observations. We did not consider any city that had more than a thousand missing observations, and after this filtration we were left with 18 African cities and 31 Asian cities. Missing observations for these cities were imputed by linear interpolation. Figure 4 shows the scatter plots of $\|\mathbf{X}_i - \mathbf{X}_j\|$ vs. $\|\mathbf{X}_i + \mathbf{X}_j\|$ for these African and Asian data sets. In each of these scatter plots, $\|\mathbf{X}_i + \mathbf{X}_j\|$ was higher than $\|\mathbf{X}_i - \mathbf{X}_j\|$ in majority of the cases, which gives an evidence against \mathcal{H}_0 . In this figure, one can observe that the evidence was stronger for the Asian data set.

When we used the tests considered in Section 4 on these African and Asian data sets, all tests rejected \mathcal{H}_0 giving a strong indication of a change in the temperature patterns over the 10 year periods considered here. However, based on that single experiment, it was not possible to compare among different test procedures. Therefore, to facilitate comparison, we used subsamples of different sizes (see Table 3) from the data and used different tests on those subsamples. For each subsample size, the experiment was repeated 200 times, and empirical powers of different tests were calculated as the proportions of times they rejected \mathcal{H}_0 .

Table 3: Powers of different tests for daily temperature data

	African Temperature				Asian Temperature			
	$n = 7$	$n = 10$	$n = 12$	$n = 15$	$n = 7$	$n = 10$	$n = 12$	$n = 15$
PA	6.0	28.0	70.0	100	8.0	54.0	80.0	100
SR	0.0	8.0	20.0	64.0	18.0	48.0	72.0	94.0
CQ	58.0	84.0	92.0	100	60.0	82.0	100	100
WPL	60.0	90.0	100	100	70.0	94.0	100	100
Run	24.5	41.5	50.5	66.0	41.5	67.0	83.5	96.0
T_{lin}	44.5	86.0	96.0	100	62.0	93.5	99.5	100
T_{log}	47.5	88.0	99.0	100	67.0	97.5	99.5	100
T_{exp}	67.5	96.0	100	100	64.5	92.5	100	100
\tilde{T}_{lin}	57.0	92.5	99.5	100	59.5	93.0	99.0	100
\tilde{T}_{log}	66.5	96.5	100	100	66.5	95.5	99.5	100
\tilde{T}_{exp}	70.0	97.5	100	100	72.5	98.0	100	100

Table 3 clearly shows that our proposed tests had excellent performance in both data sets. Among the existing tests, CQ and WPL tests, particularly the latter one performed well. But our tests based on bounded ϕ and φ functions outperformed them.

6. Concluding remarks

In this article, we have proposed and investigated two classes of tests for the multivariate one-sample problem. While many popular one-sample tests are not applicable when the dimension exceeds the sample size, these proposed tests based on interpoint distances can be conveniently used for HDLSS data or even for functional data taking values in an infinite dimensional Banach space. Our methods are conceptually and computationally simple. Unlike [6], here we do not need to solve any NP-complete problem to construct the test statistics. Compared to CQ, PA and SR tests, our tests become more robust for suitable choices of ϕ or φ . While the WPL test only considers the direction of the observations and ignores the information on their magnitudes, our tests do not need to sacrifice that valuable information to gain in robustness. That information is used in a controlled manner by using appropriate transformations. As a result, these tests outperform the WPL test in a wide variety of examples. Using several simulated and real data sets, we have amply demonstrated these important features of our tests in this article.

The one-sample tests we proposed in Section 2 can be viewed as one-sample versions of so-called two-sample energy statistics considered in [39]. In this article, we have studied the high-dimensional behavior of such tests. We have also developed a class of scale invariant tests and investigated their behavior in HDLSS setup. These scale invariant tests are particularly useful when the measurement variables are not of comparable units and scales. When the measurement variables containing stronger signals against \mathcal{H}_0 have higher (respectively, lower) variances than the variables containing weaker signals, rotation invariant (respectively, scale invariant) tests have better performance. In this article, we have used several functions of Euclidean and other distances to construct different tests. When the underlying distribution has exponential tails (e.g., normal distribution), the use of the linear function $\phi(t) = t$ or $\varphi(t) = t$ usually leads to better performance. But in cases of distributions with heavy polynomial tails (e.g., Student's t and Cauchy distributions), bounded ϕ and φ functions are preferred. The tests based on these bounded functions are more robust. In the presence of outliers or contaminating observations, they usually outperform the tests based on linear functions.

Acknowledgments. We would like to thank the Editor-in-Chief and two anonymous reviewers for their careful reading of an earlier version of the manuscript and for providing us with several insightful comments, which helped us to improve the manuscript substantially.

Appendix

Proof of Lemma 1. The result directly follows from [2] (pp. 1335–1336) and [6] (p. 1427). \square

Proof of Theorem 1. Under (A1) and (A2), the weak law of large numbers (WLLN) holds for the sequence $\{(X_1^{(q)} \pm X_2^{(q)})^2 : q \geq 1\}$, i.e., as $d \rightarrow \infty$,

$$|d^{-1}\{\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2 - E(\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2)\}| \xrightarrow{\text{Pr}} 0.$$

So, we have

$$|d^{-1/2}\{\|\mathbf{X}_1 \pm \mathbf{X}_2\| - \sqrt{E(\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2)}\}| \xrightarrow{\text{Pr}} 0$$

and hence

$$|d^{-1/2}\{\|\mathbf{X}_1 + \mathbf{X}_2\| - \|\mathbf{X}_1 - \mathbf{X}_2\| - \tau_d(\boldsymbol{\theta})\}| \xrightarrow{\text{Pr}} 0$$

as $d \rightarrow \infty$, where $\tau_d(\boldsymbol{\theta}) = d^{-1/2}\{\sqrt{E(\|\mathbf{X}_1 + \mathbf{X}_2\|^2)} - \sqrt{E(\|\mathbf{X}_1 - \mathbf{X}_2\|^2)}\}$.

Note that $E(\|\mathbf{X}_1 + \mathbf{X}_2\|^2) = E(\|\mathbf{X}_1 - \mathbf{X}_2\|^2) + 4\|\boldsymbol{\theta}\|^2$. Let us write, $\alpha_d = d^{-1}E(\|\mathbf{X}_1 - \mathbf{X}_2\|^2)$ and $\beta_d = 4d^{-1}\|\boldsymbol{\theta}\|^2$. Under (A1), since the second moments of the $(X_1^{(q)} \pm X_2^{(q)})$'s are uniformly bounded, we can find $M > 0$ such that $\alpha_d \leq M$ for every $d \geq 1$. Also, we have $\liminf_{d \rightarrow \infty} \beta_d > 0$. This implies that there exists some $\delta_0 > 0$ and $d_0 \geq 1$ such that $\beta_d > \delta_0$ for every $d \geq d_0$. Then, for every $d \geq d_0$, we get

$$\tau_d(\boldsymbol{\theta}) = \sqrt{\alpha_d + \beta_d} - \sqrt{\alpha_d} \geq \sqrt{\alpha_d + \delta_0} - \sqrt{\alpha_d} = (\sqrt{\alpha_d + \delta_0} + \sqrt{\alpha_d})^{-1} \delta_0 > (2\sqrt{M + \delta_0})^{-1} \delta_0.$$

This implies $\liminf_{d \rightarrow \infty} \tau_d(\boldsymbol{\theta}) > 0$. Since

$$|d^{-1/2}\{\|\mathbf{X}_1 + \mathbf{X}_2\| - \|\mathbf{X}_1 - \mathbf{X}_2\| - \tau_d(\boldsymbol{\theta})\}| \xrightarrow{\text{Pr}} 0$$

as $d \rightarrow \infty$, and ϕ is strictly increasing, we have, as $d \rightarrow \infty$,

$$\Pr\left(\frac{\|\mathbf{X}_i + \mathbf{X}_j\|}{\sqrt{d}} > \frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{d}}\right) \rightarrow 1 \Rightarrow \Pr\left\{\phi\left(\frac{\|\mathbf{X}_i + \mathbf{X}_j\|}{\sqrt{d}}\right) > \phi\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{d}}\right)\right\} \rightarrow 1. \quad (1)$$

Now, consider a resample $\{\mathbf{X}_1^* = a_1 \mathbf{X}_1, \dots, \mathbf{X}_n^* = a_n \mathbf{X}_n\}$. Note that for any $i \neq j$, we have

$$\phi\left(\frac{\|\mathbf{X}_i^* + \mathbf{X}_j^*\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i^* - \mathbf{X}_j^*\|}{\sqrt{d}}\right) = \begin{cases} \phi\left(\frac{\|\mathbf{X}_i + \mathbf{X}_j\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{d}}\right) & \text{if } a_i = a_j, \\ \phi\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i + \mathbf{X}_j\|}{\sqrt{d}}\right) & \text{if } a_i \neq a_j. \end{cases}$$

Let T_ϕ^{d*} be the test statistic computed from $\{\mathbf{X}_1^*, \dots, \mathbf{X}_n^*\}$. Clearly, if all a_i 's are of the same sign (which happens with probability $2/2^n = 1/2^{n-1}$), we have $T_\phi^d = T_\phi^{d*}$. Otherwise, from Eq. (1) it follows that $\Pr(T_\phi^d > T_\phi^{d*}) \rightarrow 1$ as $d \rightarrow \infty$. So, $\Pr(T_\phi^{d*} \geq T_\phi^d) \rightarrow 2/2^n = 1/2^{n-1}$ as $d \rightarrow \infty$. Therefore, if $1/2^{n-1} < \alpha$, the power of a level α test based on T_ϕ^d converges to 1 as $d \rightarrow \infty$. \square

Proof of Lemma 2. Note that for any $q \in \{1, \dots, d\}$, $X^{(q)} \stackrel{D}{=} 2\theta^{(q)} - X^{(q)}$ and hence $|X_1^{(q)} + X_2^{(q)}| \stackrel{D}{=} |X_1^{(q)} - X_2^{(q)} + 2\theta^{(q)}|$. Under the conditions assumed in the lemma, $X^{(q)} - \theta^{(q)}$ has a density $f^{(q)}(x)$ which is decreasing in $|x|$. This implies that $W^{(q)} = X_1^{(q)} - X_2^{(q)} = (X_1^{(q)} - \theta^{(q)}) - (X_2^{(q)} - \theta^{(q)})$ has a density, say $g^{(q)}(w)$, which is also decreasing in $|w|$ (follows from Lemma 3 below).

Now, $\Pr(|W^{(q)}| \leq w) = \Pr(-w \leq W^{(q)} \leq w)$ and $\Pr(|W^{(q)} + 2\theta^{(q)}| \leq w) = \Pr(-w - 2\theta^{(q)} \leq W^{(q)} \leq w - 2\theta^{(q)})$. Therefore, considering the fact that $g^{(q)}(w)$ is a decreasing function of $|w|$, we have $\Pr(|W^{(q)}| \leq w) \geq \Pr(|W^{(q)} + 2\theta^{(q)}| \leq w)$ for every $w \in \mathbb{R}$, where the equality holds if and only if $\theta^{(q)} = 0$. This in turn implies that $|W^{(q)} + 2\theta^{(q)}|$ is stochastically larger than $|W^{(q)}|$. Since $s^{(q)}$ is positive with probability 1, $|W^{(q)} + 2\theta^{(q)}|/s^{(q)}$ is also stochastically larger than $|W^{(q)}|/s^{(q)}$. Again, since φ is strictly increasing, stochastic ordering is retained under that transformation. So, we have $E\{\varphi(|X_1^{(q)} + X_2^{(q)}|/s^{(q)})\} \geq E\{\varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})\}$, where equality holds if and only if $\theta^{(q)} = 0$. Combining the results for all $q \in \{1, \dots, d\}$, we get

$$E\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right)\right\} \geq E\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right)\right\},$$

where the equality holds if and only if $\theta = \mathbf{0}$. Now, using the mean value theorem, one gets

$$\begin{aligned} h\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right)\right\} - h\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right)\right\} &= \left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right) - \frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right)\right\} h'(\xi) \\ &= \frac{1}{d} \sum_{q=1}^d \left\{\varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right) - \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right)\right\} h'(\xi), \end{aligned} \quad (2)$$

where ξ lies between $d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} + X_2^{(q)}|/s^{(q)})$ and $d^{-1} \sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})$. Since $h'(\xi)$ is positive (note that h is a strictly increasing function) with probability 1, for all $q \in \{1, \dots, d\}$, we have stochastic ordering between $\varphi(|X_1^{(q)} + X_2^{(q)}|/s^{(q)})h'(\xi)$ and $\varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})h'(\xi)$. This implies

$$E\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right) h'(\xi)\right\} \geq E\left\{\frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right) h'(\xi)\right\}.$$

Now the result follows from Eq. (2). \square

Lemma 3. Let X and Y be two independent random variables with probability density functions f_X and f_Y , respectively. If both $f_X(x)$ and $f_Y(x)$ are decreasing in $|x|$, then the density of $X \pm Y$ is also decreasing in $|x|$.

Proof. Let $|x_1| < |x_2|$ be two arbitrary points. Since both f_X and f_Y are decreasing in $|x|$, we have $f_X(x_1) > f_X(x_2)$ and $f_Y(x_1) > f_Y(x_2)$. Define $x = (x_1 + x_2)/2$. Clearly, both 0 and x_1 are either less than or greater than x . Let S^+ be the side of x where they belong and S^- be the other side. Now, for every $y \in S^-$, take $y^* \in S^+$ such that $|y - x| = |y^* - x|$. Then, $f_X(y^*) > f_X(y)$ and $f_Y(x_1 - y^*) - f_Y(x_2 - y^*) = f_Y(x_2 - y) - f_Y(x_1 - y) > 0$. Thus, we have

$$\int_{S^+} f_X(y^*) \{f_Y(x_1 - y^*) - f_Y(x_2 - y^*)\} dy^* > \int_{S^-} f_X(y) \{f_Y(x_2 - y) - f_Y(x_1 - y)\} dy,$$

which after simplification implies that $f_Z(x_1) > f_Z(x_2)$, where f_Z denotes the density of $Z = X + Y$. Note that since Y has a symmetric distribution, Y and $-Y$ have the same density function. Therefore, the result holds for the density of $X - Y$ as well. \square

Proof of Theorem 2. Note that under (B1) and (B2), for every $i \neq j$, WLLN holds for the sequence $\{\varphi(Z_i^{(q)} \pm Z_j^{(q)}) : q \geq 1\}$, i.e.,

$$\left| \frac{1}{d} \sum_{q=1}^d \varphi(|Z_i^{(q)} \pm Z_j^{(q)}|) - \mathbb{E} \left\{ \frac{1}{d} \sum_{q=1}^d \varphi(|Z_i^{(q)} \pm Z_j^{(q)}|) \right\} \right| \xrightarrow{\text{Pr}} 0 \text{ as } d \rightarrow \infty,$$

where $Z_i^{(q)} = X_i^{(q)}/s^{(q)}$ for all $i \in \{1, \dots, n\}$ and $q \in \{1, \dots, d\}$. This implies that, as $d \rightarrow \infty$,

$$\left| \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) - \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) - \tau_{d,\varphi}(\theta) \right| \xrightarrow{\text{Pr}} 0.$$

Since $\tau_\varphi = \liminf_{d \rightarrow \infty} \tau_{d,\varphi}(\theta) > 0$, and h is strictly increasing, we have

$$\Pr \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) > \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \rightarrow 1$$

as $d \rightarrow \infty$, and hence

$$\Pr \left[h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) \right\} > h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right] \rightarrow 1$$

as $d \rightarrow \infty$. Now consider a resample $\{\mathbf{X}_1^* = a_1 \mathbf{X}_1, \dots, \mathbf{X}_n^* = a_n \mathbf{X}_n\}$. Note that $s^{*(q)}$, the value of $s^{(q)}$ computed based on $\{\mathbf{X}_1^*, \dots, \mathbf{X}_n^*\}$ remains unchanged over resamples (see our discussion after Lemma 2). So, we have

$$\begin{aligned} h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{*(q)} + X_j^{*(q)}|}{s^{*(q)}} \right) \right\} - h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{*(q)} - X_j^{*(q)}|}{s^{*(q)}} \right) \right\} \\ = a_i a_j \left[h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) \right\} - h \left\{ \frac{1}{d} \sum_{q=1}^d \varphi \left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right]. \end{aligned}$$

Now, the proof follows using the same argument as used in the proof of Theorem 1. \square

References

- [1] E. Arias-Castro, E.J. Candès, Y. Plan, Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism, *Ann. Statist.* 39 (2011) 2533–2556.
- [2] L. Baringhaus, C. Franz, Rigid motion invariant two-sample tests, *Statist. Sinica* 20 (2010) 1333–1361.
- [3] P.J. Bickel, On some asymptotically nonparametric competitors of Hotelling's T^2 , *Ann. Math. Statist.* 36 (1965) 160–173.
- [4] M. Biswas, A.K. Ghosh, A nonparametric two-sample test applicable to high dimensional data, *J. Multivariate Anal.* 123 (2014) 160–171.

- [5] M. Biswas, M. Mukhopadhyay, A.K. Ghosh, A distribution-free two-sample run test applicable to high-dimensional data, *Biometrika* 101 (2014) 913–926.
- [6] M. Biswas, M. Mukhopadhyay, A.K. Ghosh, On some exact distribution-free one-sample tests for high dimension low sample size data, *Statist. Sinica* 25 (2015) 1421–1435.
- [7] I. Blumen, A new bivariate sign test, *J. Amer. Statist. Assoc.* 53 (1958) 448–456.
- [8] R.C. Bradley, Basic properties of strong mixing conditions: A survey and some open questions, *Prob. Surveys* 2 (2005) 107–144.
- [9] B.M. Brown, T.P. Hettmansperger, Affine invariant rank methods in the bivariate location model, *J. Royal Statist. Soc. Ser. B* 49 (1987) 301–310.
- [10] B.M. Brown, T.P. Hettmansperger, An affine invariant bivariate version of the sign test, *J. Royal Statist. Soc. Ser. B* 51 (1989) 117–125.
- [11] B.M. Brown, T.P. Hettmansperger, J. Nyblom, H. Oja, On certain bivariate sign tests and medians, *J. Amer. Statist. Assoc.* 87 (1992) 127–135.
- [12] A. Chakraborty, P. Chaudhuri, Tests for high-dimensional data based on means, spatial signs and spatial ranks, *Ann. Statist.* 45 (2017) 771–799.
- [13] J. Chang, C. Zheng, W.X. Zhou, W. Zhou, Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity, *Biometrics* (2017) doi:10.1111/biom.12695.
- [14] S.K. Chatterjee, A bivariate sign test for location, *Ann. Math. Statist.* 37 (1966) 1771–1782.
- [15] P. Chaudhuri, D. Sengupta, Sign tests in multidimension: Inference based on the geometry of the data cloud, *J. Amer. Statist. Assoc.* 88 (1993) 1363–1370.
- [16] S.X. Chen, Y.L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, *Ann. Statist.* 38 (2010) 808–835.
- [17] D. Donoho, J. Jin, Higher criticism for detecting sparse heterogeneous mixtures, *Ann. Statist.* 32 (2004) 962–994.
- [18] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, WH Freeman and Co., San Francisco, 1979.
- [19] P. Hall, J. Jin, Innovated higher criticism for detecting sparse signals in correlated noise, *Ann. Statist.* 38 (2010) 1686–1732.
- [20] P. Hall, J.S. Marron, A. Neeman, Geometric representation of high dimension, low sample size data, *J. Royal Statist. Soc. Ser. B* 67 (2005) 427–444.
- [21] M. Hallin, D. Paindaveine, Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, *Ann. Statist.* 30 (2002) 1103–1133.
- [22] T.P. Hettmansperger, J. Möttönen, H. Oja, Affine-invariant multivariate one-sample signed-rank tests, *J. Amer. Statist. Assoc.* 92 (1997) 1591–1600.
- [23] T.P. Hettmansperger, J. Nyblom, H. Oja, Affine invariant multivariate one-sample sign tests, *J. Royal Statist. Soc. Ser. B* 56 (1994) 221–234.
- [24] J.L. Hodges, A bivariate sign test, *Ann. Math. Statist.* 26 (1955) 523–527.
- [25] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, *Ann. Statist.* 37 (2009) 4104–4130.
- [26] D. Larocque, S. Tardif, C. van Eeden, Bivariate sign tests based on the Sup, L_1 and L_2 norms, *Ann. Inst. Statist. Math.* 52 (2000) 488–506.
- [27] J. Marden, *Multivariate rank tests, Multivariate Analysis, Design of Experiments and Survey Sampling*, (Ed. S. Ghosh), (1999) 401–432.
- [28] J. Möttönen, H. Oja, Multivariate spatial sign and rank methods, *J. Nonparametric Statist.* 5 (1995) 201–213.
- [29] J. Möttönen, H. Oja, J. Tienari, On the efficiency of multivariate spatial sign and rank tests, *Ann. Statist.* 25 (1997) 542–552.
- [30] H. Oja, *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*, Springer, New York, 2010.
- [31] H. Oja, J. Nyblom, Bivariate sign tests, *J. Amer. Statist. Assoc.* 84 (1989) 249–259.
- [32] H. Oja, R.H. Randles, Multivariate nonparametric tests, *Statist. Science* 19 (2004) 598–605.
- [33] J. Park, D.N. Ayyala, A test for the mean vector in large dimension and small samples, *J. Statist. Plann. Inference* 143 (2013) 929–943.
- [34] M.L. Puri, P.K. Sen, *Nonparametric Methods in Multivariate Analysis*, Wiley, New York, 1971.
- [35] R.H. Randles, A distribution-free multivariate sign test based on interdirections, *J. Amer. Statist. Assoc.* 84 (1989) 1045–1050.
- [36] R.H. Randles, A simpler, affine-invariant, multivariate, distribution-free sign test, *J. Amer. Statist. Assoc.* 95 (2000) 1263–1268.
- [37] M.S. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, *J. Multivariate Anal.* 100 (2009) 518–532.
- [38] M.S. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, *J. Multivariate Anal.* 99 (2008) 386–402.
- [39] G.J. Székely, M.L. Rizzo, Energy statistics: A class of statistics based on distances, *J. Statist. Plann. Inference* 143 (2013) 1249–1272.
- [40] L. Wang, B. Peng, R. Li, A high-dimensional nonparametric multivariate test for mean vector, *J. Amer. Statist. Assoc.* 110 (2015) 1658–1669.
- [41] S. Wei, C. Lee, L. Wichers, G. Li, J.S. Marron, Direction-projection-permutation for high dimensional hypothesis tests, *J. Comput. Graph. Statist.* 25 (2016) 549–569.
- [42] P.S. Zhong, S.X. Chen, M. Xu, Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence, *Ann. Statist.* 41(2013) 2820–2851.