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

Enakshi Saha, Soham Sarkar and Anil K. Ghosh

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

Email: enu2.saha@gmail.com, sohamsarkar1991@gmail.com, akghosh@isical.ac.in

June 2016

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 tests perform poorly for high dimensional data, and many of them are not applicable when the dimension 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, law of large numbers, level and power of a test, rotation invariance, scale invariance.

1 Introduction

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n independent realizations of a d-dimensional random vector $\mathbf{X}$ having a continuous distribution $F_{\boldsymbol{\theta}}$, which is symmetric about some $\boldsymbol{\theta} \in \mathbb{R}^d$ (i.e $\mathbf{X} - \boldsymbol{\theta} \stackrel{\mathcal{D}}{=} \boldsymbol{\theta} - \mathbf{X}$). In the one-sample problem, we test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative hypothesis $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, 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. For bivariate data, we have sign and signed-rank tests by Hodges (1955), Blumen (1958), Chatterjee (1966), Brown and Hettmansperger (1987, 1989), Oja and Nyblom (1989), Brown et al. (1992) and Larocque et al. (2000). Puri and Sen (1971) used co-ordinate wise signs and ranks to develop some nonparametric tests for the multivariate one-sample location problem. Randles (1989, 2000) proposed one-sample tests based on the idea of inter-directions. Chaudhuri and Sengupta (1993) generalized Hodges' sign test to higher dimension. Hallin and Paindaveine (2002) proposed a test based on interdirections and pseudo-Mahalanobis distances. Bickel (1965), Hettmansperger et al. (1994, 1997), Möttönen et al. (1997) and Chakraborty and Chaudhuri

(1999) also proposed some nonparametric tests for the multivariate one-sample problem. A brief overview of most of these tests can be found in Marden (1999), Oja and Randles (2004) and Oja (2010).

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 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., Srivastava and Du (2008); Srivastava (2009); Chen and Qin (2010); Park and Ayyala (2013)). These tests are based on the asymptotic null distribution of the test statistic, where the dimension also increases 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 have heavy tails. Recently, Biswas et al. (2015) took a graph-theoretic approach and proposed some one-sample tests based on the shortest covering path. These tests are distribution-free, but finding the shortest covering path is computationally expensive. In fact, this is equivalent to solving the traveling salesman problem, which is NP complete (see Garey and Johnson (1979)). Wang et al. (2015) proposed a one-sample sign test for high dimensional data, which is motivated by the elliptic symmetry of the underlying distribution. But this test only uses directions of the observations from $\boldsymbol{\theta}_0$, and ignores important information contained in their distances.

In this article, we develop two classes of one-sample 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, rotation and homogeneous scale transformation 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 high dimensional 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. 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 $\mathbf{X} \sim F_{\boldsymbol{\theta}}$. Note that if $\boldsymbol{\theta} = \mathbf{0}$, $\mathbf{X}_i$ and $-\mathbf{X}_i$ $(i = 1, 2)$ have the same distribution. Therefore, under the condition $E(\|\mathbf{X}\|) < \infty$, we have $E(\|\mathbf{X}_1 + \mathbf{X}_2\|) = E(\|\mathbf{X}_1 - \mathbf{X}_2\|)$, where $\|\cdot\|$ denotes the Euclidean distance. But, if $\boldsymbol{\theta} \neq \mathbf{0}$, $E(\|\mathbf{X}_1 + \mathbf{X}_2\|)$ exceeds $E(\|\mathbf{X}_1 - \mathbf{X}_2\|)$ (follows from Lemma 1 stated below). Similarly, we have $E(\|\mathbf{X}_1 + \mathbf{X}_2\|^2) - E(\|\mathbf{X}_1 - \mathbf{X}_2\|^2) = E(\mathbf{X}_1'\mathbf{X}_2) = \|\boldsymbol{\theta}\|^2 \geq 0$, where the equality holds if and only if $\boldsymbol{\theta} = \mathbf{0}$. So, one can use the empirical analog $\binom{n}{2}^{-1} \sum_{i<j} \left[\|\mathbf{X}_i + \mathbf{X}_j\| - \|\mathbf{X}_i - \mathbf{X}_j\|\right]$ or $\binom{n}{2}^{-1} \sum_{i<j} \left[\|\mathbf{X}_i + \mathbf{X}_j\|^2 - \|\mathbf{X}_i - \mathbf{X}_j\|^2\right]$ as the test statistic, and reject H_0 for large values of it. In fact, Chen and Qin (2010) considered an equivalent test statistic given by $T_{CQ} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{X}_i'\mathbf{X}_j$. But, the performance of this test is affected by the presence of outliers, and it yields poor performance when the underlying distribution has heavy tails. Recently, Wang et al. (2015) used $\mathbf{X}_i / \|\mathbf{X}_i\|$, the spatial sign of $\mathbf{X}_i$ $(i = 1, \ldots, n)$, to come up with the one-sample test statistic $T_{WPL} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \frac{\mathbf{X}_i'\mathbf{X}_j}{\|\mathbf{X}_i\|\|\mathbf{X}_j\|}$. Compared to T_{CQ}, the test based on T_{WPL} is more robust against outliers generating 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 below.

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

Here $\mathbf{R}_+ = [0, \infty)$. This lemma shows that for an appropriate choice of ψ, one can consider the test statistic

$$T_\psi = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \left[\psi(\|\mathbf{X}_i + \mathbf{X}_j\|) - \psi(\|\mathbf{X}_i - \mathbf{X}_j\|)\right] \tag{1}$$

and reject H_0 for large values of T_ψ. In fact, such tests were briefly considered in Székely and Rizzo (2013), where the authors also discussed about their large sample consistency. Clearly, the test 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 using $\psi(t) = t^2$, we get the test statistic proposed by Chen and Qin (2010), as a special case. Though $\psi(t) = t^2$ does not satisfy the condition stated in Lemma 1, but the same result holds for such ψ as we have seen before. However, the use of $\psi(t) = t^2$ makes the test sensitive against 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 the functions that diverge slowly (e.g., $\psi(t) = \log(1+t)$) 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, $T_\psi \xrightarrow{P} E(T_\psi) = E\left[\psi(\|\mathbf{X}_1 + \mathbf{X}_2\|) - \psi(\|\mathbf{X}_1 - \mathbf{X}_2\|)\right]$ as $n \to \infty$ (follows from the result on probability convergence of U-statistics). Now, if ψ satisfies the condition of Lemma 1, we have $E(T_\psi) = 0$ when $\boldsymbol{\theta} = \mathbf{0}$ and $E(T_\psi) > 0$ when $\boldsymbol{\theta} \neq \mathbf{0}$. Therefore, the power of the test based on T_ψ converges to 1 as $n \to \infty$. However, as mentioned in Wei et al. (2016), this type of consistency (in 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 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 HDLSS set up (see discussions in Biswas and Ghosh (2014); Wei et al. (2016)). Here, we investigate the power property of the test based on T_ψ in 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 $\phi : \mathbf{R}_+ \to \mathbf{R}_+$. There are several functions of this type that satisfy the conditions given in Lemma 1. For instance, one can use $\phi_1(t) = t$, $\phi_2(t) = \log(1+t)$, $\phi_3(t) = t^2/(1+t^2)$ or $\phi_4(t) = 1 - e^{-t}$. So, we consider the test statistic of the form

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

and reject H_0 for larger values of T_ϕ^d. The critical value is computed using a resampling method. Each time, we generate a new sample of the form $\{\mathbf{X}_1^* = a_1\mathbf{X}_1, \mathbf{X}_2^* = a_2\mathbf{X}_2, \ldots, \mathbf{X}_n^* = a_n\mathbf{X}_n\}$ for $a_1, a_2, \ldots, a_n$ being i.i.d. variables taking values 1 and -1 each with probability 1/2. The test 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 $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are n independent realizations of a d-dimensional random vector $\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})'$ following the distribution $F_{\boldsymbol{\theta}}$, which satisfy the conditions given below.

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

(A2) For $\mathbf{U} = \mathbf{X}_1$ and $\mathbf{V} = \pm\mathbf{X}_2$, $\sum_{q_1 \neq q_2} corr\{(U^{(q_1)} - V^{(q_1)})^2, (U^{(q_2)} - V^{(q_2)})^2\}$ 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 $F_{\boldsymbol{\theta}}$ satisfies $(A1)$ and $(A2)$, and $\phi : \mathbf{R}_+ \to \mathbf{R}_+$ is a strictly increasing function. If $\liminf_{d\to\infty} \|\boldsymbol{\theta}\|^2/d > 0$ and $2^{n-1} \geq 1/\alpha$, the power of the test based on T_ϕ^d converges to 1 as $d \to \infty$.*

The assumption $(A2)$ indicates a form of weak dependence among the measurement variables. In the case of time series data, $(A2)$ holds if the series has the ρ-mixing property, stationarity of the time series is not required here. Recently, Biswas et al. (2014, 2015) assumed similar conditions for investigating high dimensional consistency of their one-sample and two-sample tests. Similar assumptions were also considered by Hall et al. (2005) for studying high dimensional behavior of some popular classifiers and by Jung and Marron (2009) to study 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 straight forward and hence it is omitted). If the component variables $X^{(1)}, X^{(2)}, \ldots$ are i.i.d., $(A2)$ holds automatically and instead of $(A1)$, one needs the existence of second order moments only. In 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 $\liminf_{d\to\infty} \|\boldsymbol{\theta}\|^2/d > 0$ ensures that the evidence against H_0 remains significant as the dimension diverges.

3 Scale invariant tests based on other distance functions

In Section 2, we considered some tests based on functions of Euclidean distances and also discussed about some suitable functions ϕ 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 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 HDLSS literature (see e.g., Hall et al. (2005); Jung and Marron (2009), in the case of 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 $\mathbf{X}_1 = (X_1^{(1)}, X_1^{(2)}, \ldots, X_1^{(d)})$ and $\mathbf{X}_2 = (X_2^{(1)}, X_2^{(2)}, \ldots, X_2^{(d)})$ can be computed as $h\Big\{\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|)\Big\}$, where $h : \mathbf{R}_+ \to \mathbf{R}_+$ and $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ 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

$$T_{h,\varphi}^d = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \left[h\left\{ \frac{1}{d} \sum_{q=1}^d \varphi\left(|X_i^{(q)} + X_j^{(q)}| \right) \right\} - h\left\{ \frac{1}{d} \sum_{q=1}^d \varphi\left(|X_i^{(q)} - X_j^{(q)}| \right) \right\} \right]. \tag{3}$$

If second moments of the $\varphi(X_1^{(q)} \pm X_2^{(q)})$'s are uniformly bounded and $\sum_{q_1 \neq q_2} corr\{\varphi(|X_1^{(q_1)} \pm X_2^{(q_1)}|), \varphi(|X_1^{(q_2)} \pm X_2^{(q_2)}|)\}$ is of the order $o(d^2)$, the test based on $T_{h,\varphi}^d$ has the high dimensional consistency when $\liminf_{d\to\infty} E\left[\frac{1}{d}\sum_{q=1}^d \left\{\varphi\left(|X_i^{(q)} + X_j^{(q)}|\right) - \varphi\left(|X_i^{(q)} + X_j^{(q)}|\right)\right\}\right] > 0$ (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 $E\left[\frac{1}{d}\sum_{q=1}^d \left\{\varphi\left(|X_i^{(q)} + X_j^{(q)}|\right) - \varphi\left(|X_i^{(q)} + X_j^{(q)}|\right)\right\}\right] \geq 0$ for any fixed d, where the equality holds if and only if $\boldsymbol{\theta} = \mathbf{0}$. So, it is reasonable to assume its limit inferior to be positive under the alternative. The distance function $h\Big\{\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|)\Big\}$ is location invariant. We can also make it scale invariant if we divide $|X_1^{(q)} - X_2^{(q)}|$ by $s^{(q)}$, an equivariant estimate of the scale corresponding to the q-th component variable ($q = 1, \ldots, d$) and use $h\Big\{\frac{1}{d}\sum_{q=1}^d \varphi(|X_1^{(q)} - X_2^{(q)}|/s^{(q)})\Big\}$ as the distance between $\mathbf{X}_1$ and $\mathbf{X}_2$. The test statistic based on this distance function is given by

$$\widetilde{T}_{h,\varphi}^d = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \left[h\left\{ \frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) \right\} - h\left\{ \frac{1}{d} \sum_{q=1}^d \varphi\left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right]. \tag{4}$$

Note that while we have $E(\widetilde{T}_{h,\varphi}^d) = 0$ under H_0, for suitable choices of h and φ, $E(\widetilde{T}_{h,\varphi})$ 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 random vectors having a common distribution $F_{\boldsymbol{\theta}}$ with marginal densities $f^{(q)}$ ($q = 1, 2, \ldots, d$), where $f^{(q)}(x)$ is decreasing in $|x - \theta^{(q)}|$. Also assume that $h : [0, \infty) \to [0, \infty)$ and $\varphi : [0, \infty) \to [0, \infty)$ are monotonically increasing functions. If h is differentiable and $E\left[h\left\{\frac{1}{d}\sum_{q=1}^d \varphi\left(\frac{|X_1^{(q)} \pm X_2^{(q)}|}{s^{(q)}}\right)\right\}\right] < \infty$, we have $E(\widetilde{T}_{h,\varphi}^d) \geq 0$, where the equality holds if and only if $\boldsymbol{\theta} = \mathbf{0}$.*

Note that Lemma 2 holds irrespective of the choice of the data based estimates $s^{(q)}$ for $q = 1, 2, \ldots, d$, and it suggests us to reject H_0 for large values of $\widetilde{T}^d_{h,\varphi}$. The cutoff is computed using the resampling technique discussed before. Since $\mathbf{X}$ and $-\mathbf{X}$ have the same distribution under H_0, along with the observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, we take their negatives $-\mathbf{X}_1, -\mathbf{X}_2, \ldots, -\mathbf{X}_n$ 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 makes our method computationally efficient. This type of scale invariant tests are particularly useful when the measurement variables are not of comparable units and scales. One should also notice that in order to construct a meaningful test, it is not necessary for $h\left\{\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right\}$ 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 $E(\widetilde{T}^d_{h,\varphi}) \geq 0$, where the equality will hold if and only if $\boldsymbol{\theta} = \mathbf{0}$. However, if $\varphi(|t_1 - t_2|)$ gives a distance between t_1 and t_2 in $\mathbb{R}$, $h(t) = t$ makes $h\left\{\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right\}$ a distance in $\mathbb{R}^d$. For $\varphi(|t_1 - t_2|) = |t_1 - t_2|^p$ and $h(t) = t$, $h\left\{\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right\}$ becomes a function of the ℓ_p-distance.

For investigating the high dimensional behavior of the test based on $\widetilde{T}^d_{h,\varphi}$, here we consider assumptions similar to (A1) and (A2). For n independent observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \overset{i.i.d}{\sim} F_{\boldsymbol{\theta}}$, we define $\mathbf{Z}_i = (X_i^{(1)}/s^{(1)}, X_i^{(2)}/s^{(2)}, \ldots)$ for $i = 1, 2, \ldots, n$, and assume

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

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

Under (B1) and (B2), WLLN holds for the sequence $\{\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|); q \geq 1\}$, i.e.,

$$\left|\frac{1}{d}\sum_{q=1}^{d}\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|) - \frac{1}{d}\sum_{q=1}^{d}E\left[\varphi(|Z_1^{(q)} \pm Z_2^{(q)}|)\right]\right| \xrightarrow{P} 0 \quad \text{as } d \to \infty. \tag{5}$$

Now, let us define $\tau_{d,\varphi}(\boldsymbol{\theta}) = \frac{1}{d}\sum_{q=1}^{d}E\left\{\varphi\left(|Z_1^{(q)} + Z_2^{(q)}|\right) - \varphi\left(|Z_1^{(q)} - Z_2^{(q)}|\right)\right\}$. If φ satisfies the condition of Lemma 2, for any fixed d, we have $\tau_{d,\varphi}(\boldsymbol{\theta}) \geq 0$ where the equality holds if and only if $\boldsymbol{\theta} = \mathbf{0}$ (follows from the proof of Lemma 2). So, it is reasonable to assume that $\liminf_{d\to\infty} \tau_{d,\varphi}(\boldsymbol{\theta}) > 0$. The following theorem shows the high dimensional consistency of the scale invariant test under that condition.

Theorem 2. *Suppose that $F_{\boldsymbol{\theta}}$ satisfies (B1) and (B2), and $h : \mathbf{R}_+ \to \mathbf{R}_+$ is monotonically increasing. If $\tau_{\varphi} = \liminf_{d\to\infty} \tau_{d,\varphi}(\boldsymbol{\theta}) > 0$ and $2^{n-1} \geq 1/\alpha$, the power of the test based on $\widetilde{T}^d_{h,\varphi}$ converges to 1 as $d \to \infty$.*

4 Analysis of simulated data sets

We analyzed several 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 Hotelling's T^2 test, coordinate-wise sign and rank tests (see e.g., Puri and Sen (1971)), spatial sign and rank tests (see e.g. Möttönen and Oja (1995)), one-sample tests proposed by Srivastava (2009), Chen and Qin (2010), Park and Ayyala (2013), Wang et al. (2015) (referred to as SR, CQ, PA and WPL tests, respectively) and tests based on run and longest run statistics proposed in Biswas et al. (2015) (referred to as run and long run tests, respectively) for comparison. For nonparametric sign and rank tests, we used both, the test based on the large sample distribution of the test statistic and the conditional test based on the permutation principle. In each case, the best one (which happened to be the permutation test in almost all cases) has been reported. Recall that the first five tests cannot be used in HDLSS situations. Therefore, to make comparison with them, we first consider some data sets with dimension smaller than the sample size.

We began with examples involving d-dimensional normal, t_3 (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. In each case, we generated 50 observations from a distribution having diagonal scatter matrix and location parameter with the first $d/2$ elements equal to 0 and the rest equal to δ. We considered two sets of experiments. In the first (respectively, second) set, the first $d/2$ diagonal elements of the scatter matrix were 3 (respectively, 1/3) and the rest were 1. We used $d = 40$ and four different values of δ (0, 0.1, 0.2 and 0.3) to test $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$. Each experiment was repeated 500 times to estimate the powers (sizes in the case of $\delta = 0$) of different tests by the proportion of times they rejected H_0. These results are reported in Table 1 and Table 2, respectively. For our tests based on the Euclidean distance, here we report the results for $\phi(t) = t$, $\log(1+t)$, $t^2/(1+t^2)$ and $1 - e^{-t}$. The corresponding test statistics are denoted by T_{lin}, T_{log}, T_{frac} and T_{exp}, respectively. Analogous versions of scale invariant tests were constructed using $\varphi(t) = t$, $\log(1 + t)$, $t^2/(1 + t^2)$ and $1 - e^{-t}$, when $h(t) = t$ was used in all these cases. The corresponding statistics are denoted by $\widetilde{T}_{lin}$, $\widetilde{T}_{log}$,$\widetilde{T}_{frac}$,$\widetilde{T}_{exp}$, respectively. Note that $\varphi(t) = t$ and $\varphi(t) = 1 - e^{-t}$ lead to proper distance functions in $\mathbb{R}^d$. We also considered another test statistic with $\varphi_0(t) = t^2$ and $h_0(t) = \sqrt{t}$), which basically leads to a test based on the Euclidean distance computed based on the standardized versions of the component variables. The corresponding test statistic is denoted by $\widetilde{T}_0$.

Tables 1 and 2 show that in the examples involving normal distributions, all tests had sizes close to 0.05, but in cases of t_3 (with 3 df) and Cauchy distributions, the Hotelling's T^2 test and the SR test had sizes below the nominal level. SR test had size 0 in the case of Cauchy distributions, where the PA test also had size below 0.05. Observed levels of CQ and PA tests were slightly higher for t_3 distributions. In all other occasions, the true null hypothesis was rejected in nearly 5% of the cases.

As it is expected, all affine invariant (Hotelling's T^2, sign and rank tests) and scale invariant (PA, SR and all tests proposed in Section 3) tests had similar powers in these two sets of examples, but the rotation invariant tests (CQ, WPL, run, long run and all tests proposed in Section 2) had significantly different powers in these two sets. Note that in all these examples, only the last $d/2$ variables contain evidence against H_0, while the first $d/2$ variables can be viewed as noise. In the first (respectively, second) set of examples, the first $d/2$ variables had higher (respectively, lower) variance than the last $d/2$ variables. So, these unimportant variables had higher (respectively, lower) contribution to the Euclidean norm compared to other $d/2$ important variables. Thus, the test statistics based on Euclidean norm contained more noise (respectively, information). However, the scale invariant tests treated all the variables equally in both the cases. So, the overall performance of the Euclidean norm-based tests were much inferior (respectively, superior) compared to the scale invariant tests in the first (respectively, second) set of examples.

In almost all examples, the Hotelling's T^2 test and nonparametric sign and rank tests had lower powers compared to most of the tests considered here, especially for higher values of δ. Among the rotation invariant tests, performances of run and long run tests were somewhat inferior. The CQ test had the highest power when the underlying distribution was normal, while the proposed tests and the WPL test also had comparable performance. But in cases of t_3 and Cauchy distributions, tests based on T_{frac} and T_{exp} outperformed all other rotation invariant tests considered here. Performances of the WPL test and that based on T_{log} were almost similar. The test based on T_{lin} had relatively low power, but its performance was much better than the CQ test. Among the scale invariant tests, in the case of normal distribution, SR, PA and tests based on $\widetilde{T}_0$ and $\widetilde{T}_{lin}$ yielded higher powers, while other proposed tests also performed reasonably well. In the case of t_3 distribution, all proposed tests outperformed SR and PA tests. As expected, in the case of Cauchy distribution, while non-robust tests like SR and PA tests had powers close to the nominal level, the proposed tests, especially those based on $\widetilde{T}_{log}$, $\widetilde{T}_{frac}$ and $\widetilde{T}_{exp}$ had excellent performance.

Next, we considered some examples where the dimension was larger than the sample size. For these examples, we used $n = 20$ and $d = 30, 60, 120, 250$ and 500. Note that the Hotelling's T^2 test and nonparametric sign and rank tests could not be used in this setup. We again considered two sets of experiments with normal, t_3

Table 1: Powers (in %) of different tests on simulated data sets with $\sigma_1^2 = 3$ and $\sigma_2^2 = 1$ ($n = 50$, $d = 40$).

	Normal				t_3				Cauchy			
	δ=0.0	δ=0.1	δ=0.2	δ=0.3	δ=0.0	δ=0.1	δ=0.2	δ=0.3	δ=0.0	δ=0.1	δ=0.2	δ=0.3
Hot. T^2	5.0	11.6	32.2	71.2	3.0	9.8	30.4	62.0	3.4	6.6	21.4	49.4
Sp-sign	5.0	11.8	32.4	70.0	5.8	9.8	27.8	50.6	4.2	8.6	17.4	27.8
Sp-rank	5.2	11.8	31.8	71.4	5.4	10.6	31.6	62.0	4.6	8.8	24.6	47.8
coord-sign	5.6	8.4	20.4	47.4	6.0	6.6	18.0	35.0	6.2	7.8	13.8	22.6
coord-rank	6.0	12.2	33.8	71.2	4.8	9.6	28.8	62.2	5.2	10.2	24.2	50.0
PA	5.0	23.2	94.0	100.0	7.6	13.6	42.2	81.2	2.4	2.4	3.6	6.2
SR	4.0	21.8	93.6	100.0	1.8	5.4	30.2	76.4	0.0	0.0	0.0	0.0
CQ	5.2	11.8	52.8	98.4	6.6	9.2	19.6	51.0	5.4	5.6	6.6	7.4
WPL	3.8	8.4	45.6	96.8	5.0	8.0	35.8	90.8	4.8	7.4	25.2	70.0
Run	4.8	8.6	24.8	73.0	4.4	7.8	22.0	54.2	5.2	7.8	18.8	44.2
Long Run	4.2	7.8	14.8	41.2	4.6	7.6	14.2	39.6	4.6	6.4	14.0	38.6
T_{lin}	4.2	9.2	50.2	98.0	4.8	9.2	30.0	80.8	4.6	6.0	11.4	24.2
T_{log}	3.8	9.6	52.6	98.4	5.4	9.0	38.8	91.6	6.0	8.8	29.6	75.2
T_{frac}	4.4	9.4	53.4	98.2	5.0	11.2	51.6	96.6	4.2	9.8	46.6	93.4
T_{exp}	4.2	9.4	53.4	98.4	4.4	11.2	51.2	96.6	3.8	10.2	47.0	93.2
$\widetilde{T}_0$	4.6	22.8	93.0	100.0	5.0	15.6	70.8	98.2	3.8	6.0	19.4	49.4
$\widetilde{T}_{lin}$	4.8	24.6	91.2	100.0	5.2	16.0	69.6	98.4	3.6	6.6	21.0	48.6
$\widetilde{T}_{log}$	5.0	22.0	86.0	100.0	5.4	15.2	70.2	99.0	4.6	10.8	48.6	87.0
$\widetilde{T}_{frac}$	4.6	19.6	77.0	100.0	4.8	14.4	65.6	98.2	5.2	12.4	52.4	90.8
$\widetilde{T}_{exp}$	4.4	18.8	73.2	99.8	4.8	13.6	63.2	97.2	5.6	13.2	51.8	89.8

Table 2: Powers (in %) of different tests on simulated data sets with $\sigma_1^2 = 1/3$, $\sigma_2^2 = 1$ ($n = 50$, $d = 40$).

	Normal				t_3				Cauchy			
	δ=0.0	δ=0.1	δ=0.15	δ=0.2	δ=0.0	δ=0.1	δ=0.15	δ=0.2	δ=0.0	δ=0.1	δ=0.15	δ=0.2
Hot. T^2	5.0	11.6	18.2	32.2	3.0	9.8	17.0	30.4	3.4	6.6	11.8	21.4
Sp-sign	5.0	11.8	20.0	32.4	5.8	9.8	16.8	27.8	4.2	8.6	11.4	17.4
Sp-rank	5.2	11.8	18.8	31.8	5.4	10.6	19.4	31.6	4.6	8.8	16.8	24.6
coord-sign	5.6	8.4	13.8	20.4	6.0	6.6	11.4	18.0	6.2	7.8	9.2	13.8
coord-rank	6.0	12.2	21.6	33.8	4.8	9.6	17.0	28.8	5.2	10.2	16.8	24.2
PA	5.0	23.2	64.2	94.0	7.6	13.6	25.0	42.2	2.4	2.4	2.8	3.6
SR	4.0	21.8	61.6	93.6	1.8	5.4	12.8	30.2	0.0	0.0	0.0	0.0
CQ	5.0	39.0	81.8	99.2	7.2	20.0	40.6	63.6	3.6	4.6	5.4	7.4
WPL	4.4	30.6	75.2	97.8	4.0	25.8	64.4	94.0	4.0	16.8	46.0	79.0
Run	5.4	11.4	27.8	52.0	4.8	11.6	19.0	35.6	5.4	9.0	16.8	29.2
Long Run	4.6	10.4	18.6	32.2	4.4	10.2	14.0	23.8	4.6	8.2	17.0	28.0
T_{lin}	3.8	34.8	78.6	98.6	4.2	23.8	56.0	88.4	3.4	7.6	18.8	34.4
T_{log}	3.8	34.0	78.6	98.4	3.6	27.8	66.8	93.0	4.4	17.8	44.8	78.6
T_{frac}	3.8	33.8	77.8	98.4	4.0	30.6	75.6	96.2	4.8	25.2	66.0	91.8
T_{exp}	4.0	33.8	77.6	98.6	4.0	29.8	75.4	96.2	4.8	24.8	65.6	92.0
$\widetilde{T}_0$	4.6	22.8	61.2	93.0	5.0	15.6	37.2	70.8	3.8	6.0	11.2	19.4
$\widetilde{T}_{lin}$	4.8	24.6	59.4	91.2	5.2	16.0	37.8	69.6	3.6	6.6	12.0	21.0
$\widetilde{T}_{log}$	5.0	22.0	53.2	86.0	5.4	15.2	38.6	70.2	4.6	10.8	26.2	48.6
$\widetilde{T}_{frac}$	4.6	19.6	45.4	77.0	4.8	14.4	34.8	65.6	5.2	12.4	27.0	52.4
$\widetilde{T}_{exp}$	4.4	18.8	42.6	73.2	4.8	13.6	31.8	63.2	5.6	13.2	26.4	51.8

and Cauchy distributions having diagonal scatter matrices. In both cases, the location of the distribution was of the form $(\delta, \delta, \ldots, \delta)'$. In the fist set of examples, all component variables had the same scale parameter

Table 3: Powers (in %) of different tests in different dimensions ($n = 20$).

Normal	$\delta = 0.1$; $\sigma_i^2 = 1$ for all $i = 1, \ldots, d$					$\delta = 0.4$; $\sigma_i^2 = i \mod 30$ for $i = 1, \ldots, d$				
d	30	60	120	250	500	30	60	120	250	500
PA	15.0	24.2	33.2	58.2	77.2	35.6	52.6	78.2	96.6	99.6
SR	16.6	16.2	14.4	10.2	6.6	38.4	45.8	56.0	70.2	77.2
CQ	21.8	29.8	43.4	65.6	88.6	22.4	28.0	39.4	59.8	83.2
WPL	15.0	23.0	33.0	54.2	79.4	16.0	22.2	30.4	50.6	74.4
Run	11.6	14.0	14.6	20.0	37.0	9.8	14.6	21.6	32.0	52.0
Long run	11.2	10.2	11.8	16.2	24.2	10.2	12.0	17.6	24.2	37.2
T_{lin}	17.2	26.4	41.4	63.6	87.2	18.8	24.8	36.2	56.8	82.4
T_{log}	17.4	26.2	40.6	63.8	87.2	17.4	24.8	36.8	57.4	82.4
T_{frac}	17.6	27.4	40.4	63.8	86.8	16.8	25.8	36.2	57.8	81.6
T_{exp}	17.6	26.8	40.6	63.4	86.8	14.4	25.0	35.8	57.4	81.8
$\widetilde{T}_0$	16.6	21.0	34.0	53.0	77.6	33.8	50.2	75.8	95.8	100.0
$\widetilde{T}_{lin}$	17.4	24.4	36.0	56.0	79.2	35.0	55.4	79.2	96.8	100.0
$\widetilde{T}_{log}$	15.4	21.0	33.0	50.2	73.6	31.2	52.2	73.4	95.2	100.0
$\widetilde{T}_{frac}$	12.4	18.6	27.0	44.0	65.4	26.8	43.4	65.6	87.8	99.2
$\widetilde{T}_{exp}$	12.2	17.6	25.0	40.2	61.4	24.8	39.2	61.4	86.2	98.8

t (3 df)	$\delta = 0.12$; $\sigma_i^2 = 1$ for all $i = 1, \ldots, d$					$\delta = 0.45$; $\sigma_i^2 = i \mod 30$ for $i = 1, \ldots, d$				
d	30	60	120	250	500	30	60	120	250	500
PA	10.6	12.0	20.2	30.8	42.0	19.4	26.6	38.4	57.8	72.0
SR	4.6	3.0	1.6	0.8	0.2	13.0	9.4	9.2	7.4	4.6
CQ	13.6	21.4	29.4	44.8	62.2	13.0	14.8	22.2	32.8	51.4
WPL	17.0	29.2	44.8	67.0	89.0	13.8	20.4	29.8	52.6	76.6
Run	12.6	13.4	15.8	22.8	30.4	11.0	12.6	18.8	29.6	39.6
Long run	11.0	12.2	14.4	24.2	32.6	11.4	13.8	17.2	27.8	38.0
T_{lin}	16.8	27.8	42.0	61.4	82.6	12.4	19.0	30.2	47.4	72.0
T_{log}	18.8	32.6	47.8	70.6	90.4	15.6	23.4	36.4	59.8	83.4
T_{frac}	25.0	36.6	53.4	77.0	95.4	16.6	26.2	40.4	66.8	87.8
T_{exp}	23.8	35.6	52.4	76.0	94.6	16.2	25.6	36.4	61.6	83.2
$\widetilde{T}_0$	14.6	23.0	33.6	51.6	73.2	26.0	42.8	62.4	85.2	96.0
$\widetilde{T}_{lin}$	13.6	22.8	34.6	53.0	75.4	29.0	46.8	65.8	87.6	97.2
$\widetilde{T}_{log}$	14.2	23.6	33.2	54.8	78.8	31.8	47.2	68.2	89.4	97.8
$\widetilde{T}_{frac}$	11.6	19.6	28.8	49.0	72.2	28.4	43.4	62.0	83.8	96.8
$\widetilde{T}_{exp}$	10.8	18.2	26.4	46.0	69.2	27.4	41.4	60.0	80.8	95.0

Cauchy	$\delta = 0.15$; $\sigma_i^2 = 1$ for all $i = 1, \ldots, d$					$\delta = 0.5$; $\sigma_i^2 = i \mod 30$ for $i = 1, \ldots, d$				
d	30	60	120	250	500	30	60	120	250	500
PA	4.0	4.6	4.8	5.8	6.2	4.6	6.0	6.4	8.4	9.2
SR	0.6	0.2	0.0	0.0	0.0	0.6	0.4	0.0	0.0	0.0
CQ	8.2	9.2	10.0	13.0	16.2	7.2	8.2	9.0	8.8	9.6
WPL	22.8	33.2	48.6	71.2	88.6	13.6	20.6	29.0	43.0	69.0
Run	0.122	18.6	21.6	28.6	37.4	14.2	15.4	19.4	24.4	35.8
Long run	12.8	19.6	25.6	35.4	52.2	13.2	15.8	21.6	34.0	48.8
T_{lin}	13.0	17.8	25.6	35.4	53.2	8.4	13.4	16.8	22.8	34.0
T_{log}	25.6	38.0	51.6	74.2	88.6	16.8	24.8	36.0	53.0	77.0
T_{frac}	35.8	52.6	71.2	90.0	97.8	22.4	32.8	50.0	72.0	90.2
T_{exp}	35.6	51.8	71.0	89.4	97.6	21.2	30.8	45.2	68.4	87.8
$\widetilde{T}_0$	10.8	16.4	20.4	27.0	43.6	13.6	20.2	27.0	38.0	56.8
$\widetilde{T}_{lin}$	11.4	16.6	20.4	28.8	45.2	14.6	21.0	30.6	42.2	59.2
$\widetilde{T}_{log}$	17.2	25.0	35.8	55.0	75.0	25.6	37.6	52.0	73.8	89.2
$\widetilde{T}_{frac}$	17.2	26.8	37.4	57.0	78.4	25.2	38.2	55.8	75.0	90.6
$\widetilde{T}_{exp}$	18.0	26.8	36.6	56.6	78.2	25.4	37.6	55.6	75.2	91.2

1, but in the other set, the i-th variable ($i = 1, 2, \ldots, d$) had the scale parameter $i \mod 30$. As expected, our proposed rotation invariant tests had better overall performance in the first set of examples, while in the second set of examples, they were outperformed by their scale invariant analogs. For examples involving normal distributions, CQ and our proposed rotation invariant tests had comparable powers though CQ had a slight edge. But in cases of t_3 and Cauchy distributions, our tests based on T_{frac} and T_{exp} outperformed all other rotation invariant tests considered here. The performance of WPL test was somewhat similar to that of the test based on T_{log}. Just like the CQ test, scale invariant PA test also performed well for normal distributions, but it failed in the case of Cauchy distributions. Except for the second set of examples with normal distributions, the SR test had poor performance throughout. Our scale invariant tests, particularly those based on on $\widetilde{T}_0$, $\widetilde{T}_{lin}$ and $\widetilde{T}_{log}$ could compete with the PA test in the case of normal distributions, but they had much better performance than the PA test in cases of t_3 and Cauchy distributions. However, in these examples with heavy tailed distributions, the tests based on $\widetilde{T}_{frac}$ and $\widetilde{T}_{exp}$ performed even better.

Next, we considered two examples involving normal distributions with mean $(0.2, 0.2, \ldots, 0.2)'$, where the component variables were correlated. In the first example (referred to as Correlated-1 in Table 4), the (i, j)-th element of the dispersion matrix was $(-0.99)^{|i-j|}$. In the other example (referred to as Correlated-2 in Table 4), all diagonal elements of the dispersion matrix were 1 and all off-diagonal elements were 0.25. In the first example, our rotation invariant tests, particularly those based on T_{frac} and T_{exp} had excellent performance. Performances of SR, CQ, WPL and run test were also comparable. Among the scale invariant tests, the test based on $\widetilde{T}_0$ had competitive performance for $n = 40$. But, all other tests yielded somewhat lower powers in this example. In the second example, SR, CQ, PA tests had best overall performance. Barring the tests based on run and longest run statistics, all tests yielded competitive powers.

We also considered two examples with non-elliptic distributions. In both of these cases, the component variables were independent. In one case, we generated them from $U(-3, 3)$ 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 $(0.1, 0.1, \ldots, 0.1)'$ to the measurement vector. In the example with Laplace distribution, our scale invariant tests based on $\widetilde{T}_{lin}, \widetilde{T}_{log}, \widetilde{T}_{frac}$ and $\widetilde{T}_{exp}$ outperformed all other tests considered here. Our rotation invariant tests and the CQ test also had satisfactory performance. CQ, PA and our rotation invariant tests performed better than others in the case of uniform distribution. The SR test, run and long run tests did not have good powers in these two examples.

Next, we carried out our experiment with a mixture of three normal distributions all having the same scatter matrix 0.25I. The locations of these normal distributions were $-2\mathbf{1}_{500}, \mathbf{1}_{500}$ and $4\mathbf{1}_{500}$, where $\mathbf{1}_{500}$ is a 500-

Table 4: Powers (in %) of different tests on simulated data sets ($d = 500$).

	Correlated-1		Correlated-2		Laplace		Uniform		Mixture normal		Perturbed normal	
	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$
PA	24.2	83.6	44.2	73.4	29.6	83.4	22.6	61.8	38.0	69.6	0.0	0.0
SR	36.2	92.0	47.2	74.6	0.0	19.8	0.0	2.4	39.8	70.8	0.0	0.0
CQ	33.4	90.2	48.4	75.0	43.2	90.2	27.2	66.0	39.6	70.2	0.0	0.0
WPL	40.0	99.0	43.6	70.4	34.8	84.4	19.8	54.2	16.2	24.8	68.4	100.0
Run	45.8	92.4	13.8	22.4	26.2	47.8	11.4	20.0	100.0	100.0	100.0	100.0
Long Run	31.8	60.0	16.6	26.8	19.0	23.2	9.8	11.6	92.8	100.0	100.0	100.0
T_{lin}	32.6	94.8	37.2	69.0	42.0	89.2	25.6	64.8	32.2	73.2	0.0	1.6
T_{log}	37.6	98.6	37.2	68.4	41.6	89.0	25.8	65.0	51.0	99.6	99.4	100.0
T_{frac}	44.6	99.8	36.4	67.8	41.8	89.4	26.0	64.4	100.0	100.0	100.0	100.0
T_{exp}	43.4	99.8	36.6	68.0	41.6	89.2	26.0	64.6	100.0	100.0	100.0	100.0
$\widetilde{T}_0$	28.6	95.8	38.4	68.8	34.4	85.0	20.8	56.4	33.6	74.0	0.0	0.0
$\widetilde{T}_{lin}$	21.8	73.0	38.8	69.6	64.2	98.4	19.6	45.2	32.2	71.4	0.0	0.6
$\widetilde{T}_{log}$	22.0	67.2	37.2	69.4	71.2	99.0	16.8	41.8	38.0	93.2	8.8	99.8
$\widetilde{T}_{frac}$	21.8	62.0	36.8	69.4	69.2	99.6	16.6	42.0	39.2	96.4	100.0	100.0
$\widetilde{T}_{exp}$	21.6	60.2	36.4	69.0	68.8	99.6	15.8	39.0	61.0	100.0	100.0	100.0

dimensional vector with all elements equal to 1. The mixing proportions for these three distributions were 0.4, 0.2 and 0.4, respectively. So, the distribution was symmetric about $\mathbf{1}_{500}$. For $n = 20$, the run test and our tests based on T_{frac} and T_{exp} rejected H_0 in all occasions, while the long run test also performed well. For $n = 40$, the test based on T_{exp} also had unit power. Tests based T_{log}, $\widetilde{T}_{log}$ and $\widetilde{T}_{log}$ had powers close to 1 as well. The WPL test had poor performance in this example.

Finally, we considered an experiment to study the robustness of different tests. In this experiment, we generated observations from a multivariate normal distribution with location $0.15\mathbf{1}_{500}$ and dispersion matrix $0.25\mathbf{I}$, but 10% of them were contaminated by subtracting 2 from each of the coordinates. In this example, run and long run tests had power 1 even for $n = 20$. But on the other hand, CQ, PA and SR tests had zero power even for $n = 40$. Our tests based on ℓ_1 and ℓ_2 norms also performed poorly, but the ones based on bounded ϕ and φ functions yielded excellent performance.

5 Analysis of average temperature data

The `Average Daily Temperature Archive` (http://academic.udayton.edu/kissock/http/Weather/) of the University of Dayton contains daily average temperature 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 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 observations. We did not consider any city that had more than 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. When we used the tests considered in Section 4 on these African and Asian datasets, all tests rejected the null hypothesis giving a strong indication of a change in the temperature patterns over the 10 years. 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 (reported in Table 5) 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 H_0.

Table 5: 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
Long Run	34.5	59.5	77.5	95.5	47.5	64.0	76.5	88.0
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_{frac}	55.0	95.0	100	100	69.5	99.0	100	100
T_{exp}	67.5	96.0	100	100	64.5	92.5	100	100
$\widetilde{T}_0$	33.0	81.0	93.0	100	51.5	90.5	97.5	100
$\widetilde{T}_{lin}$	57.0	92.5	99.5	100	59.5	93.0	99.0	100
$\widetilde{T}_{log}$	66.5	96.5	100	100	66.5	95.5	99.5	100
$\widetilde{T}_{frac}$	71.0	96.5	100	100	72.0	97.0	100	100
$\widetilde{T}_{exp}$	70.0	97.5	100	100	72.5	98.0	100	100

Table 5 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 methods are not applicable when dimension exceeds the sample size, these proposed tests based on pairwise 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 Biswas et al. (2015), here we do not need to solve any NP-complete problem to construct the test statistics. Unlike CQ, PA and SR tests, our tests have better robustness properties 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 for being robust. That information is used in a controlled manner by using appropriate transformations. As a result, these tests can 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 by Székely and Rizzo (2013). In this article, we have studied the high dimensional behavior of such tests. We have also developed the scale invariant analogs of these tests and investigated their behavior in HDLSS set up. These scale invariant tests are particularly useful when the measurement variables are not of comparable units and scales. In this article, we used several functions of Euclidean and other distances to construct different tests and showed that different functions are preferred in different situations. So, it will be great if one can come up with a data driven method for selecting the function to be used for the particular problem in hand.

Appendix

Proof of Lemma 1: The result directly follows from Baringhaus and Franz (2010) (p. 1335-1336) and Biswas et al. (2015) (p. 1427).

Proof of Theorem 1: Note that 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., $\left|\frac{\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2}{d} - E\frac{\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2}{d}\right| \xrightarrow{P} 0$ as $d \to \infty$. So, we have $\left|\frac{\|\mathbf{X}_1 \pm \mathbf{X}_2\|}{\sqrt{d}} - \sqrt{\frac{E\|\mathbf{X}_1 \pm \mathbf{X}_2\|^2}{d}}\right| \xrightarrow{P} 0$ and hence $\left|\frac{\|\mathbf{X}_1 + \mathbf{X}_2\|}{\sqrt{d}} - \frac{\|\mathbf{X}_1 - \mathbf{X}_2\|}{\sqrt{d}} - \tau_d(\boldsymbol{\theta})\right| \xrightarrow{P} 0$ as $d \to \infty$, where $\tau_d(\boldsymbol{\theta}) = \sqrt{E\frac{\|\mathbf{X}_1 + \mathbf{X}_2\|^2}{d}} - \sqrt{E\frac{\|\mathbf{X}_1 - \mathbf{X}_2\|^2}{d}}$.

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 $\liminf_{d\to\infty} \beta_d > 0$. This implies that there exists some $\delta > 0$ and $d_0 \geq 1$ such that $\beta_d > \delta$ for every $d \geq d_0$. Then, for every $d \geq d_0$, we get $\tau_d(\boldsymbol{\theta}) = \sqrt{\alpha_d + \beta_d} - \sqrt{\alpha_d} \geq \sqrt{\alpha_d + \delta} - \sqrt{\alpha_d} = \frac{\delta}{\sqrt{\alpha_d + \delta} + \sqrt{\alpha_d}} > \frac{\delta}{2\sqrt{M+\delta}}$. This implies $\liminf_{d\to\infty} \tau_d(\boldsymbol{\theta}) > 0$.

Since $\left|\frac{\|\mathbf{X}_1+\mathbf{X}_2\|}{\sqrt{d}} - \frac{\|\mathbf{X}_1-\mathbf{X}_2\|}{\sqrt{d}} - \tau_d(\boldsymbol{\theta})\right| \xrightarrow{P} 0$ as $d \to \infty$, and ϕ is strictly increasing, we have

$$P\left(\frac{\|\mathbf{X}_i+\mathbf{X}_j\|}{\sqrt{d}} > \frac{\|\mathbf{X}_i-\mathbf{X}_j\|}{\sqrt{d}}\right) \to 1 \Rightarrow P\left(\phi\left(\frac{\|\mathbf{X}_i+\mathbf{X}_j\|}{\sqrt{d}}\right) > \phi\left(\frac{\|\mathbf{X}_i-\mathbf{X}_j\|}{\sqrt{d}}\right)\right) \to 1 \text{ as } d \to \infty. \tag{6}$$

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

$$\phi\left(\frac{\|\mathbf{X}_i^*+\mathbf{X}_j^*\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i^*-\mathbf{X}_j^*\|}{\sqrt{d}}\right) = \begin{cases} \phi\left(\frac{\|\mathbf{X}_i+\mathbf{X}_j\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i-\mathbf{X}_j\|}{\sqrt{d}}\right) & \text{if } a_i = a_j \\ \phi\left(\frac{\|\mathbf{X}_i-\mathbf{X}_j\|}{\sqrt{d}}\right) - \phi\left(\frac{\|\mathbf{X}_i+\mathbf{X}_j\|}{\sqrt{d}}\right) & \text{if } a_i \neq a_j. \end{cases} \tag{7}$$

Let T_ϕ^{d*} be the test statistic computed from $\{\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*\}$. 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 equation (6) it follws that $P(T_\phi^d > T_\phi^{d*}) \to 1$ as $d \to \infty$. So, $P(T_\phi^{d*} \geq T_\phi^d) \to 2/2^n = 1/2^{n-1}$ as $d \to \infty$. Therefore, if $1/2^{n-1} < \alpha$, the power of a level α test based on T_ϕ^d converges to 1 as $d \to \infty$. $\square$

Proof of Lemma 2: Note that for any $q = 1, 2, \ldots, d$, $X^{(q)} \stackrel{\mathcal{D}}{=} 2\theta^{(q)} - X^{(q)}$ and hence $|X_1^{(q)} + X_2^{(q)}| \stackrel{\mathcal{D}}{=} |X_1^{(q)} - X_2^{(q)} + 2\theta^{(q)}|$. 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, $P(|W^{(q)}| \leq w) = P(-w \leq W^{(q)} \leq w)$ and $P(|W^{(q)} + 2\theta^{(q)}| \leq w) = P(-w - 2\theta^{(q)} \leq W^{(q)} \leq w - 2\theta^{(q)})$. Therefore, considering the fact that $g^{(q)}(w)$ is a decreasing function of $|w|$, we have $P(|W^{(q)}| \leq w) \geq P(|W^{(q)} + 2\theta^{(q)}| \leq w)$ 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 $E\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right) \geq E\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)$, where equality holds if and only if $\theta^{(q)} = 0$. Now, combining the results for all $q = 1, 2, \ldots, d$, we get

$$E\left[\frac{1}{d}\sum_{q=1}^{d} \varphi\left(\frac{|X_1^{(q)} + X_2^{(q)}|}{s^{(q)}}\right)\right] \geq E\left[\frac{1}{d}\sum_{q=1}^{d} \varphi\left(\frac{|X_1^{(q)} - X_2^{(q)}|}{s^{(q)}}\right)\right], \tag{8}$$

where the equality holds if and only if $\boldsymbol{\theta} = \mathbf{0}$.

Now, using the mean value theorem, one gets

$$
\begin{aligned}
h\left\{\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right)\right\} &- h\left\{\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right\} \\
&= \left[\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right) - \frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right] h^{'}(\xi) \\
&= \frac{1}{d}\sum_{q=1}^{d}\left[\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right) - \varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)\right] h^{'}(\xi),
\end{aligned} \tag{9}
$$

where ξ lies between $\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right)$ and $\frac{1}{d}\sum_{q=1}^{d}\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right)$. Since $h^{'}(\xi)$ is positive (note that h is a strictly increasing function) with probability 1, for all $q = 1, 2, \ldots, d$, we have stochastic ordering between $\varphi\left(\frac{|X_1^{(q)}+X_2^{(q)}|}{s^{(q)}}\right) h^{'}(\xi)$ and $\varphi\left(\frac{|X_1^{(q)}-X_2^{(q)}|}{s^{(q)}}\right) h^{'}(\xi)$. 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 equation (9). $\square$.

Lemma 3. *Let X and Y be two 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 $X \pm Y$ 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

$$
\int_{S^+} f_X(y^*)[f_Y(x_1 - y^*) - f_Y(x_2 - y^*)]dy^* > \int_{S^-} f_X(y)[f_Y(x_2 - y) - f_Y(x_1 - y)]dy, \tag{10}
$$

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.,

$$
\left|\frac{1}{d}\sum_{q=1}^{d}\varphi\big(|Z_i^{(q)} \pm Z_j^{(q)}|\big) - E\Big[\frac{1}{d}\sum_{q=1}^{d}\varphi\big(|Z_i^{(q)} \pm Z_j^{(q)}|\big)\Big]\right| \xrightarrow{P} 0 \text{ as } d \to \infty, \tag{11}
$$

where $Z_i^{(q)} = X_i^{(q)}/s^q$ for $i = 1, 2, \ldots, n$ and $q = 1, 2, \ldots, d$. This implies that

$$\left| \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) - \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) - \tau_{d,\varphi}(\boldsymbol{\theta}) \right| \xrightarrow{P} 0 \text{ as } d \to \infty \tag{12}$$

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

$$P\left(\frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) > \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right) \to 1 \text{ as } d \to \infty \tag{13}$$

$$\Rightarrow P\left(h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) \right\} > h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right) \to 1 \text{ as } d \to \infty \tag{14}$$

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

$$\begin{aligned} & h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{*(q)} + X_j^{*(q)}|}{s^{*(q)}} \right) \right\} - h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{*(q)} - X_j^{*(q)}|}{s^{*(q)}} \right) \right\} \\ & \qquad = a_i a_j \left[h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} + X_j^{(q)}|}{s^{(q)}} \right) \right\} - h\left\{ \frac{1}{d} \sum_{q=1}^{d} \varphi\left(\frac{|X_i^{(q)} - X_j^{(q)}|}{s^{(q)}} \right) \right\} \right]. \end{aligned} \tag{15}$$

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

References

Baringhaus, L. and Franz, C. (2010) Rigid motion invariant two-sample tests. *Statistica Sinica*, **20**, 1333–1361.

Bickel, P. J. (1965) On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann. Math. Statist.*, **36**, 160–173.

Biswas, M. and Ghosh, A. K. (2014) A nonparametric two-sample test applicable to high dimensional data. *J. Multivariate Anal.*, **123**, 160–171.

Biswas, M., Mukhopadhyay, M. and Ghosh, A. K. (2014) A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika*, **101**, 913–926.

— (2015) On some exact distribution-free one-sample tests for high dimension low sample size data. *Statistica Sinica*, **25**, 1421–1435.

Blumen, I. (1958) A new bivariate sign test. *J. Amer. Statist. Assoc.*, **53**, 448–456.

Brown, B. and Hettmansperger, T. (1987) Affine invariant rank methods in the bivariate location model. *J. Royal Statist. Soc., Ser. B*, **49**, 301–310.

Brown, B., Hettmansperger, T., Nyblom, J. and Oja, H. (1992) On certain bivariate sign tests and medians. *J. Amer. Statist. Assoc.*, **87**, 127–135.

Brown, B. and Hettmansperger, T. P. (1989) An affine invariant bivariate version of the sign test. *J. Royal Statist. Soc., Ser. B*, **51**, 117–125.

Chakraborty, B. and Chaudhuri, P. (1999) On affine invariant sign and rank tests in one- and two-sample multivariate problems. *Multivariate Analysis, Design of Experiments and Survey Sampling, (Ed. S. Ghosh)*, 499–522.

Chatterjee, S. K. (1966) A bivariate sign test for location. *Ann. Math. Statist.*, **37**, 1771–1782.

Chaudhuri, P. and Sengupta, D. (1993) Sign tests in multidimension: inference based on the geometry of the data cloud. *J. Amer. Statist. Assoc.*, **88**, 1363–1370.

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

Garey, M. R. and Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. WH Freeman and Co., San Francisco.

Hall, P., Marron, J. and Neeman, A. (2005) Geometric representation of high dimension, low sample size data. *J. Royal Statist. Soc., Ser. B*, **67**, 427–444.

Hallin, M. and Paindaveine (2002) Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, **30**, 1103–1133.

Hettmansperger, T. P., Möttönen, J. and Oja, H. (1997) Affine-invariant multivariate one-sample signed-rank tests. *J. Amer. Statist. Assoc.*, **92**, 1591–1600.

Hettmansperger, T. P., Nyblom, J. and Oja, H. (1994) Affine invariant multivariate one-sample sign tests. *J. Royal Statist. Soc., Ser. B*, **56**, 221–234.

Hodges, J. L. (1955) A bivariate sign test. *Ann. Math. Statist.*, **26**, 523–527.

Jung, S. and Marron, J. (2009) PCA consistency in high dimension, low sample size context. *Ann. Statist.*, **37**, 4104–4130.

Larocque, D., Tardif, S. and van Eeden, C. (2000) Bivariate sign tests based on the Sup, L_1 and L_2 norms. *Ann. Inst. Statist. Math.*, **52**, 488–506.

Marden, J. I. (1999) Multivariate rank tests. *Multivariate Analysis, Design of Experiments and Survey Sampling, (Ed. S. Ghosh)*, 401–432.

Möttönen, J. and Oja, H. (1995) Multivariate spatial sign and rank methods. *J. Nonparametric Statist.*, **5**, 201–213.

Möttönen, J., Oja, H. and Tienari, J. (1997) On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.*, **25**, 542–552.

Oja, H. (2010) *Multivariate Nonparametric Methods With R: an approach based on spatial signs and ranks*. Springer, New York.

Oja, H. and Nyblom, J. (1989) Bivariate sign tests. *J. Amer. Statist. Assoc.*, **84**, 249–259.

Oja, H. and Randles, R. H. (2004) Multivariate nonparametric tests. *Statist. Science*, **19**, 598–605.

Park, J. and Ayyala, D. N. (2013) A test for the mean vector in large dimension and small samples. *J. Statist. Plann. Inf.*, **143**, 929–943.

Puri, M. L. and Sen, P. K. (1971) *Nonparametric methods in multivariate analysis*. Wiley, New York.

Randles, R. H. (1989) A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.*, **84**, 1045–1050.

— (2000) A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.*, **95**, 1263–1268.

Srivastava, M. S. (2009) A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.*, **100**, 518–532.

Srivastava, M. S. and Du, M. (2008) A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.*, **99**, 386–402.

Székely, G. J. and Rizzo, M. L. (2013) Energy statistics: A class of statistics based on distances. *J. Statist. Plann. Inf.*, **143**, 1249–1272.

Wang, L., Peng, B. and Li, R. (2015) A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.*, **110**, 1658–1669.

Wei, S., Lee, C., Wichers, L., Li, G. and Marron, J. (2016) Direction-projection-permutation for high dimensional hypothesis tests. *Journal of Computational and Graphical Statistics*, **25**, 549–569.