

Research Article

T_3 -Plot for Testing Spherical Symmetry for High-Dimensional Data with a Small Sample Size

Jiajuan Liang

College of Business, University of New Haven, 300 Boston Post Road, West Haven, CT 06516, USA

Correspondence should be addressed to Jiajuan Liang, jliang@newhaven.edu

Received 3 August 2012; Accepted 16 October 2012

Academic Editor: Jose Sarabia

Copyright © 2012 Jiajuan Liang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High-dimensional data with a small sample size, such as microarray data and image data, are commonly encountered in some practical problems for which many variables have to be measured but it is too costly or time consuming to repeat the measurements for many times. Analysis of this kind of data poses a great challenge for statisticians. In this paper, we develop a new graphical method for testing spherical symmetry that is especially suitable for high-dimensional data with small sample size. The new graphical method associated with the local acceptance regions can provide a quick visual perception on the assumption of spherical symmetry. The performance of the new graphical method is demonstrated by a Monte Carlo study and illustrated by a real data set.

1. Introduction

Studies on highly complicated random systems or structures pose the problem of measuring a large number of variables simultaneously. Modern technology makes it possible to collect high-dimensional data. Because of the high cost or the great difficulty in measuring a large number of variables at the same time, it is quite common that high-dimensional data are usually associated with small sample size. For example, microarray data are usually obtained by measuring thousands of variables, but the sample size is possibly less than 100; image data could be obtained by measuring more than 10,000 variables at the same time but possibly with sample size of only several hundreds. Analysis of high-dimensional data with a small sample size has become an important research topic in statistics. Many authors have been making great efforts in developing various dimension-reduction techniques, among which sliced inverse regression (called SIR, see, Li [1]) and its extension (Li [2]; Cook [3]; Cook and Li [4]) are powerful. The critical assumption of SIR on population is spherical or elliptical symmetry. Thus, test of high-dimensional spherical or elliptical symmetry

will play an important role in practical implementation of those dimension-reduction techniques.

Data visualization techniques are welcome by data analyst in explaining statistical ideas and conclusions to nonstatisticians. Although many graphical methods such as the Q-Q (quantile-quantile) plots have been developed for testing high-dimensional normality (see, e.g., Healy [5]; Small [6]; Ahn [7]; Koziol [8]; Brown and Hettmansperger [9]; Liang and Bentler [10]), there is none for testing spherical symmetry except the Q-Q plots proposed by Li et al. [11]. These existing plotting methods require a sufficiently large sample size to be used effectively. The goal of this paper is to tackle the challenge of high dimension with a small sample size in testing spherical symmetry by a plotting method.

Let \mathbf{x} be a d -dimensional random vector. \mathbf{x} is said to have a spherical distribution if for any $d \times d$ orthogonal constant matrix Γ , $\Gamma\mathbf{x}$, and \mathbf{x} have the same distribution. It is well known that if a random vector \mathbf{x} has a spherical distribution with a density, then its density must be of the form $f(\mathbf{x}'\mathbf{x})$ for some nonnegative scalar function $f(\cdot)$ (see, e.g., Fang et al. [12]). Many well-known multivariate distributions, such as the multivariate standard normal distribution, the multivariate t -distribution with zero mean and an identity matrix as its covariance matrix, and any scale mixture of the multivariate standard normal distribution, are spherical distributions. See Chapter 3 of Fang et al. [12] for more examples of spherical distributions.

A goodness-of-fit test of spherical symmetry means testing if a set of i.i.d. (i.e., independently identically distributed) d -dimensional ($d \geq 2$) sample $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is from a population with a spherical distribution. Fang and Liang [13] gave an up-to-date overview of existing methods for tests of spherical symmetry and their extensions to testing elliptical symmetry. Several other statistics for testing spherical symmetry have been proposed in the last few years (e.g., Koltchinskii and Li [14]; Diks and Tong [15]). There is no empirical study on the performance of these existing methods when applied to testing spherical symmetry for high-dimensional data with a small sample size. It is for this purpose that we develop a new graphical method in this paper to tackle this problem. The new graphical method integrates the property of spherical distributions and the idea of the T_3 -plot, which was proposed by Ghosh [16] for detecting univariate nonnormality. Hence, we will call this version of the T_3 -plot as ST_3 -plot, where “ S ” stands for sphericity. The ST_3 -plot possesses a special property that it is still useful when the sample size n is smaller than the dimension d of \mathbf{x} . This property gives an superiority of the ST_3 -plot in the case of high dimension with small sample size. This will be demonstrated in Section 3.

The main idea to construct the ST_3 -plot is to project the high dimensional data to one dimensional space such that the T_3 statistic [16] with the projected data behaves like the one with an i.i.d. sample from the univariate standard normal distribution. Thus, a critical issue arising from the projection is how to select some “good” projection directions. We address this issue carefully in Section 2.

This paper is organized as follows. In Section 2 we will give a brief review on Ghosh’s [16] T_3 -plot for detecting univariate nonnormality. The theoretical principle for deriving the projection directions is discussed. In Section 3, we will provide some local acceptance regions for the ST_3 -plot and simulation results on its performance in testing high-dimensional spherical symmetry based on a Monte Carlo study. An application of the ST_3 -plot to a real high-dimensional data set with small sample size is illustrated. Some concluding remarks are given in Section 4.

2. The ST_3 -Plot for Testing High-Dimensional Spherical Symmetry

2.1. Background of the ST_3 -Plot

The ST_3 -plot is focused on looking for evidence of departure from a spherical distribution for an i.i.d. d -dimensional sample $\{x_1, \dots, x_n\}$. Spherical distributions possess very similar marginal distributions to those of the multivariate normal distribution. For example, all univariate (1-dimensional) marginals of a multivariate normal distribution are still normal. In comparison with this marginal property, all univariate marginals of a spherical distribution are scale-invariant. This scale invariance is an important characteristic of the family of spherical distributions. More discussions and induced results from the scale invariance for a spherical distribution can be found from Fang et al. [12]. Because most existing plotting methods for testing goodness of fit are only applicable for univariate distributions, the key idea of the ST_3 -plot is to extend the existing univariate T_3 -plot [16] to testing multivariate spherical symmetry from some "special principal component" directions by employing the same idea as in Fang et al. [17]. Section 2.2 gives a simple review on Ghosh's [16] T_3 -plot and Section 2.3 summarizes the theoretical details on the extension of Ghosh's [16] T_3 -plot to the ST_3 -plot.

2.2. A Review of the T_3 -Plot

Let x_1, \dots, x_n be an i.i.d. univariate sample and $\mathbf{x} = (x_1, \dots, x_n)'$. Testing goodness-of-fit for univariate normality is to test whether the underlying distribution of the sample is normal $N(\mu, \sigma^2)$ with unknown μ and σ^2 . The EMGF (empirical moment generating function) for the studentized data is defined as

$$M(\mathbf{x}, t) = \frac{1}{n} \sum_{j=1}^n \exp \left\{ \frac{t(x_j - \bar{x})}{s} \right\}, \quad t \in R^1, \quad (2.1)$$

where \bar{x} and s are the sample mean and the sample standard deviation, respectively, R^1 stands for the set of all real numbers. Denote the i th derivative of $M(\mathbf{x}, t)$ with respect to t by $M^{(i)}(\mathbf{x}, t)$ ($i \geq 0$). Ghosh [16] defined the $T_3^{(n)}(\mathbf{x}, t)$ (a function of t for fixed \mathbf{x}) as

$$\begin{aligned} T_3^{(n)}(\mathbf{x}, t) &= \sqrt{n} \frac{d^3}{dt^3} \{ \log M(\mathbf{x}, t) \} \\ &= \sqrt{n} M^{-2}(\mathbf{x}, t) \left\{ M^{(3)}(\mathbf{x}, t) M(\mathbf{x}, t) - 3M^{(2)}(\mathbf{x}, t) M^{(1)}(\mathbf{x}, t) + 2 \frac{M^{(1)}(\mathbf{x}, t)^3}{M(\mathbf{x}, t)} \right\}. \end{aligned} \quad (2.2)$$

The graphical method for detecting nonnormality of the underlying distribution of the sample $\{x_1, \dots, x_n\}$ is based on $T_3^{(n)}(\mathbf{x}, t)$.

It is noted that $T_3^{(n)}(\mathbf{x}, t)$ is a stochastic process with an index $t \in [-a, a]$ (for some $a > 0$). Under the normal assumption, Ghosh [16] obtained the asymptotic distribution of

$T_3^{(n)}(\mathbf{x}, t)$, which is a zero-mean Gaussian process with a covariance function $K(t, s)$ ($t, s \in [-a, a]$). In particular,

$$K(t, t) = (t^6 + 9t^4 + 18t^2 + 6) \exp(t^2) - 2t^6. \quad (2.3)$$

The behavior of $T_3^{(n)}(\mathbf{x}, t)$ at $t = 0$ reflects evidence of departure from normality for the sample $\{x_1, \dots, x_n\}$. For example, $T_3^{(n)}(\mathbf{x}, 0)$ is proportional to the sample skewness and the slope of $T_3^{(n)}(\mathbf{x}, t)$ at $t = 0$ is proportional to the sample kurtosis. Fang et al. [17] proposed the following Kolmogorov-Smirnov (K-S) type statistic KS_n :

$$KS_n(\mathbf{x}) = \sup_{-1 \leq t \leq 1} \frac{|T_3^{(n)}(\mathbf{x}, t)|}{\sqrt{K(t, t)}} \quad (2.4)$$

as an analytical test associated with the T_3 -plot. Large values of $KS_n(\mathbf{x})$ indicate nonnormality of the univariate sample $\mathbf{x} = (x_1, \dots, x_n)'$. The exact finite-sample distribution of KS_n is not readily obtained under the normal assumption but its percentiles can be well fitted by a quadratic function of $1/\sqrt{n}$ using the least squares method. Fang et al. [17] provided these quadratic functions of $1/\sqrt{n}$ for significance levels $\alpha = 1\%$, 5% , and 10% based on a Monte Carlo study.

2.3. Extension of the T_3 -Plot to the ST_3 -Plot

Let x_1, \dots, x_n be an i.i.d. d -dimensional sample from a population characterized by a d -dimensional random vector \mathbf{x} . We want to test spherical symmetry of the sample. The theoretical principle for extending the T_3 -plot to the ST_3 -plot is based on Lemma 2.1 and Theorem 2.2 as follows.

Lemma 2.1 (Theorem 2.22 of Fang et al., Chapter 2 [12]). *Let $t(\mathbf{x})$ be a statistic based on a (not necessarily i.i.d.) sample $\mathbf{x} = (x_1, \dots, x_n)'$ such that $t(a\mathbf{x}) = t(\mathbf{x})$ (scale invariance) for any constant $a > 0$. If \mathbf{x} has a spherical distribution, then*

$$t(\mathbf{x}) \stackrel{d}{=} t(\mathbf{z}_0), \quad \mathbf{z}_0 \sim N_n(\mathbf{0}, \mathbf{I}_n), \quad (2.5)$$

where the sign " $\stackrel{d}{=}$ " means that the two sides of the equality have the same distribution.

Lemma 2.1 implies that for any scale-invariant statistic $t(\mathbf{x})$ associated with a spherically distributed random vector \mathbf{x} , its distribution is the same as that from taking the spherical random vector \mathbf{x} to be the standard normal $N_n(0, \mathbf{I}_n)$.

Theorem 2.2. *Let x_1, \dots, x_n be i.i.d. with a d -dimensional spherical distribution and $P(x_1 = 0) = 0$, and matrix $\tilde{\mathbf{X}} = (\mathbf{x}_1, \dots, \mathbf{x}_n) : d \times n$. Assume that*

$$\tilde{\mathbf{d}} = f(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}), \quad n \times 1 \quad (2.6)$$

is a vector function that is uniquely determined by $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$. Define the random vector

$$\mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}} : d \times 1. \quad (2.7)$$

Then \mathbf{z} has a d -dimensional spherical distribution.

Proof. First, we point out that the random matrix $\tilde{\mathbf{X}} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ has a left spherical matrix distribution [18]. That is, it satisfies

$$\Gamma\tilde{\mathbf{X}} \stackrel{d}{=} \tilde{\mathbf{X}} \quad (2.8)$$

for any $d \times d$ orthogonal matrix Γ that is independent of $\tilde{\mathbf{X}}$. We can write the random vector \mathbf{z} in (2.7) as

$$\mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}} = \tilde{\mathbf{X}}f(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}) \stackrel{d}{=} \Gamma\tilde{\mathbf{X}}f\left[\left(\Gamma\tilde{\mathbf{X}}\right)'(\Gamma\tilde{\mathbf{X}})\right] = \Gamma\mathbf{z}, \quad (2.9)$$

that is, we obtain

$$\Gamma\mathbf{z} \stackrel{d}{=} \mathbf{z} \quad (2.10)$$

for any $d \times d$ orthogonal matrix Γ that is independent of \mathbf{z} . This shows that the random vector \mathbf{z} given by (2.7) has a spherical distribution by definition. This completes the proof. \square

The random vector $\tilde{\mathbf{d}}$ in (2.7) acts as a direction for projecting a left spherically distributed random matrix $\tilde{\mathbf{X}}$ into a spherically distributed random vector \mathbf{z} by (2.7). This idea is due to Lauter [19] in constructing tests for the multivariate normal mean $\boldsymbol{\mu}$ in $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with an unknown $\boldsymbol{\Sigma}$.

Based on Lemma 2.1 and Theorem 2.2, we can extend the T_3 -plot to the ST_3 -plot for testing high-dimensional spherical symmetry. At first, we point out that the T_3 function $T_3^{(n)}(\mathbf{x}, t)$ given by (2.2) is scale-invariant, that is,

$$T_3^{(n)}(a\mathbf{x}, t) = T_3^{(n)}(\mathbf{x}, t), \quad (2.11)$$

for any constant $a > 0$. Therefore, by Lemma 2.1, if \mathbf{x} is spherically distributed, then

$$T_3^{(n)}(\mathbf{x}, t) \stackrel{d}{=} T_3^{(n)}(\mathbf{z}_0, t), \quad \mathbf{z}_0 \sim N_n(\mathbf{0}, \mathbf{I}_n). \quad (2.12)$$

Similarly, for the K-S type statistic given by (2.4), it is also true that

$$\text{KS}_n(\mathbf{x}) \stackrel{d}{=} \text{KS}_n(\mathbf{z}_0), \quad \mathbf{z}_0 \sim N_n(\mathbf{0}, \mathbf{I}_n), \quad (2.13)$$

if \mathbf{x} is spherically distributed.

Now we propose a series of necessary tests for high-dimensional (the dimension d of data is very large) spherical symmetry by Theorem 2.2. The meaning for necessary test is the same as in Fang et al. [20]. That is, when the null hypothesis is not rejected, it implies insufficient information to draw a statistical conclusion from the sampled data. Instead of testing the hypothesis of spherical symmetry for an i.i.d. sample $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ directly, we turn to test a series of hypotheses defined by

$$H_{0d} : \mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}} \text{ given by (2.7) has a spherical distribution} \quad (2.14)$$

versus $H_{1d} : \mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}}$ given by (2.7) is nonspherical. Hypothesis (2.14) is for any possible choices of $\tilde{\mathbf{d}}$ in Theorem 2.2. So hypothesis (2.14) comprises a family of tests for spherical symmetry of \mathbf{z} . If any of the possible choices of $\tilde{\mathbf{d}}$ leads to rejection of H_{0d} in (2.14), the hypothesis of spherical symmetry will be also rejected. So any test for (2.14) is a necessary test for the hypothesis of spherical symmetry for the original sample. The T_3 -function in (2.2) based on $\mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}} = (z_1, \dots, z_d)'$ (computed from (2.7)) becomes

$$\begin{aligned} T_3^{(d)}(\mathbf{z}, t) &= \sqrt{d} \frac{d^3}{dt^3} \{ \log M(\mathbf{z}, t) \} \\ &= \sqrt{d} M^{-2}(\mathbf{z}, t) \left\{ M^{(3)}(\mathbf{z}, t) M(\mathbf{z}, t) - 3M^{(2)}(\mathbf{z}, t) M^{(1)}(\mathbf{z}, t) + 2 \frac{M^{(1)}(\mathbf{z}, t)^3}{M(\mathbf{z}, t)} \right\}, \end{aligned} \quad (2.15)$$

where

$$M(\mathbf{z}, t) = \frac{1}{d} \sum_{j=1}^d \exp \left\{ \frac{t(z_j - \bar{z})}{s_z} \right\}, \quad t \in R^1, \quad (2.16)$$

where \bar{z} and s_z are the sample mean and the sample standard deviation calculated from $\mathbf{z} = \tilde{\mathbf{X}}\tilde{\mathbf{d}} = (z_1, \dots, z_d)'$ by Theorem 2.2. The K-S type statistic in (2.4) becomes

$$KS_d(\mathbf{z}) = \sup_{-1 \leq t \leq 1} \frac{|T_3^{(d)}(\mathbf{z}, t)|}{\sqrt{K(t, t)}}, \quad (2.17)$$

where $T_3^{(d)}(\mathbf{z}, t)$ is given by (2.15) and $K(t, t)$ by (2.3).

By (2.12) and (2.13), the principle for using the ST_3 -plot to detect high-dimensional nonspherical symmetry can be summarized as: plot the T_3 -function $T_3^{(d)}(\mathbf{z}, t)$ given by (2.15) versus $t \in [-1, 1]$. If the plot shows a significant departure from the horizontal axis $y = 0$ in R^2 , hypothesis (2.14) is rejected, and as a result, the i.i.d. sample $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ can be considered from a population of nonspherical distribution. The K-S type statistic $KS_d(\mathbf{z})$ in (2.17) can be employed to evaluate the significance of departure from spherical symmetry.

It is obvious that there are numerous choices of the function $f(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ in Theorem 2.2 in constructing the projection direction $\tilde{\mathbf{d}}$. We will study the empirical performance of the choices recommended by Läuter [19], and Läuter et al. [21] in next section.

3. A Monte Carlo Study

3.1. An Overview

Graphical methods can only serve as descriptive statistical inference without associated acceptance regions with the plots. One of the impressive characteristics of Ghosh's [16] T_3 -plot for detecting departure of univariate normality is its associated acceptance regions, which make it possible the plotting method as analytical statistical inference. The purpose of the Monte Carlo study in this section is to provide simulated acceptance regions for the ST_3 -plot developed in Section 2 by using a similar Monte Carlo method to that in Ghosh [16] and in Fang et al. [17]. Based on the acceptance regions, the empirical performance of the ST_3 -plot can be partially evaluated through counting the rejection rate (type I error) from a selected set of spherical distributions (which serve as the null hypothesis) and counting the rejection rate (empirical power) from a selected set of nonspherical alternative distributions (which serve as the alternative hypothesis) when applying the ST_3 -plot.

3.2. The Local Acceptance Regions

A local acceptance region for the plot of $T_3^{(d)}(\mathbf{z}, t)$ given by (2.15) is a critical band for the plot on $t \in [-1, 1]$. If the plot goes outside the critical band, it is an indication that \mathbf{z} is not spherically distributed, and as a result, the hypothesis of spherical symmetry is rejected. A critical band can be constructed by simulating the percentiles of the finite-sample distribution of the K-S type statistic in (2.17). By Lemma 2.1, we have

$$KS_d(\mathbf{z}) \stackrel{d}{=} KS_d(\mathbf{z}_0), \quad \mathbf{z}_0 \sim N_d(\mathbf{0}, \mathbf{I}_d), \quad (3.1)$$

if the null hypothesis H_{0d} in (2.14) is true. Therefore, in simulating the percentiles of the finite-sample null distribution of $KS_d(\mathbf{z})$ in (2.17), we can simply generate the data for \mathbf{z} from the standard normal $N_d(\mathbf{0}, \mathbf{I}_d)$. By generating the normal $N_d(\mathbf{0}, \mathbf{I}_d)$ data for \mathbf{z} with 2,000 replications, we record the $100(1 - \alpha)\%$ -percentiles (e.g., $\alpha = 5\%$) of $KS_d(\mathbf{z})$, where the $KS_d(\mathbf{z})$ in (3.1) is approximately calculated by taking the supremum on the discrete values of $t = 0.01, 0.02, \dots, 0.99$ (i.e., $t = 0.01, 0.02, \dots, 0.99$). A quadratic curve of $1/\sqrt{d} : a(\alpha) + b(\alpha)/\sqrt{d} + c(\alpha)/d$, is fitted for the $100(1 - \alpha)\%$ -percentiles $c_d(\alpha)$ of $KS_d(\mathbf{z})$ by the least squares method to find the approximate relation of $c_d(\alpha)$ to the sample dimension d . The following quadratic curves based on values $d = 10$ (2) 50 (i.e., $d = 10, 12, 14, \dots, 50$) were obtained from the least squares fitting:

$$\begin{aligned} c_d(\alpha) &= 4.7353 - 9.7698/\sqrt{d} + 2.1227/d, & \alpha &= 1\%, \\ c_d(\alpha) &= 2.6152 - 3.5985/\sqrt{d} - 1.4409/d, & \alpha &= 5\%, \\ c_d(\alpha) &= 2.2321 - 4.2514/\sqrt{d} + 1.8314/d, & \alpha &= 10\%. \end{aligned} \quad (3.2)$$

The quadratic curves given by (3.2) are suitable for estimating the percentiles of the K-S type statistic $KS_d(\mathbf{z})$ for d in the range $10 \leq d \leq 50$. Figure 1 shows the plots of the simulated percentiles of $KS_d(\mathbf{z})$ and the estimated percentiles $c_d(\alpha)$ given by (3.2) by the least squares method. The fit in Figure 1 seems to be acceptable.

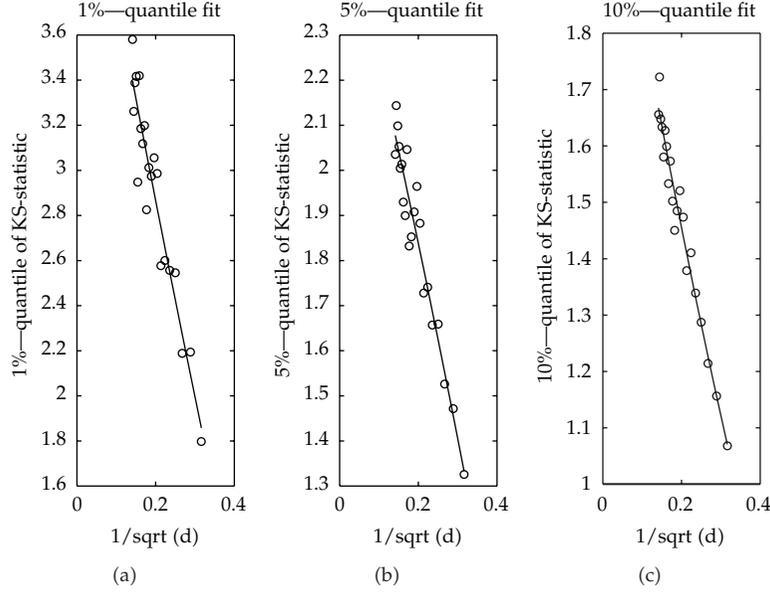


Figure 1: Quadratic curve fit for the percentiles of the K-S type statistic $KS_d(\mathbf{z})$ given by (2.17).

By using the quadratic curves given by (3.2) for estimating the finite-sample percentiles of the statistic $KS_d(\mathbf{z})$ in (2.17) for dimension $10 \leq d \leq 50$, the acceptance region for the T_3 -function $T_3^{(d)}(\mathbf{z}, t)$ in (2.15) for testing spherical symmetry is given by

$$Q_d(\alpha) = \pm c_d(\alpha) \sqrt{K(t, t)}, \quad (3.3)$$

where $c_d(\alpha)$ is given by (3.2) and $K(t, t)$ is given by (2.3). We will call the acceptance region determined by (3.3) the local acceptance region for the ST_3 -plot in testing high-dimensional spherical symmetry. When the plot of $T_3^{(d)}(\mathbf{z}, t)$ ($t \in [-1, 1]$) goes outside the acceptance region determined by (3.3), hypothesis (2.14) is rejected, and as a result, the underlying distribution of the sample $\{x_1, \dots, x_n\}$ shows evidence of nonspherical symmetry.

3.3. Type I Error Rates

We already pointed out that there are numerous choices for the projection direction $\tilde{\mathbf{d}}$ in (2.7) to plot the T_3 -function (2.15). We will perform a Monte Carlo study on the following choices of $\tilde{\mathbf{d}}$ that were suggested by Lauter [19], and Lauter et al. [21].

- (1) Solution to the eigenvalue problem:

$$\tilde{\mathbf{X}} \tilde{\mathbf{X}} \tilde{\mathbf{D}} = \tilde{\mathbf{D}} \tilde{\Lambda}, \quad (3.4)$$

where $\tilde{\mathbf{D}}_{n \times n} = [\gamma_1, \dots, \gamma_n]$, $\tilde{\Lambda} = \text{Diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_n)$ with $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_n \geq 0$. To ensure the unique solution, the random matrix $\tilde{\mathbf{D}}$ is assumed to have positive diagonal elements. The following directions are chosen:

$$\tilde{\mathbf{d}}_1 = \gamma_1, \quad \tilde{\mathbf{d}}_2 = \gamma_k \quad \text{with } k = \begin{cases} \left\lfloor \frac{2}{n} \right\rfloor, & \text{if } n < d, \\ \left\lfloor \frac{2}{d} \right\rfloor, & \text{if } n \geq d, \end{cases} \quad \tilde{\mathbf{d}}_3 = \begin{cases} \gamma_{n'} & \text{if } n < d, \\ \gamma_{d'} & \text{if } n \geq d, \end{cases} \quad (3.5)$$

where the sign $[x]$ stands for the integer part ($\leq x$) of a real number x (e.g., $[2.9] = 2$ and $[3.2] = 3$).

- (2) The direction based on the SS-test discussed by Lauter et al. [19]. Choose

$$\tilde{\mathbf{d}} = \tilde{\mathbf{d}}_4 = \mathbf{d}_{ss} = \left[\text{Diag}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}) \right]^{-1/2} \mathbf{1}_n, \quad (3.6)$$

where $\text{Diag}(\tilde{\mathbf{X}}\tilde{\mathbf{X}})$ denotes the diagonal matrix with the same diagonal elements as those of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}$, and $\mathbf{1}_n$ is an $n \times 1$ vector of ones.

- (3) The direction based on the PC-test discussed by Lauter et al. [21]. Let $\tilde{\mathbf{D}}$ be the solution matrix to the eigenvalue problem

$$\tilde{\mathbf{X}}\tilde{\mathbf{X}}\tilde{\mathbf{D}} = \tilde{\mathbf{D}}\tilde{\Lambda}, \quad \tilde{\mathbf{D}} \text{Diag}(\tilde{\mathbf{X}}\tilde{\mathbf{X}})\tilde{\mathbf{D}} = \mathbf{I}_n, \quad (3.7)$$

where similar conditions to those on the matrices $\tilde{\mathbf{D}}$ and $\tilde{\Lambda}$ in (3.4) are imposed on the matrices $\tilde{\mathbf{D}}$ and $\tilde{\Lambda}$ in (3.7) to ensure the unique solution. Let $\tilde{\mathbf{D}} = [\tilde{\gamma}_1, \dots, \tilde{\gamma}_n]$. The three directions are chosen:

$$\tilde{\mathbf{d}}_5 = \tilde{\gamma}_1, \quad \tilde{\mathbf{d}}_6 = \tilde{\gamma}_{k'}, \quad \tilde{\mathbf{d}}_7 = \begin{cases} \tilde{\gamma}_{n'} & \text{if } n < d, \\ \tilde{\gamma}_{d'} & \text{if } n \geq d, \end{cases} \quad (3.8)$$

where k is given by (3.5).

The following seven directions are chosen for a Monte Carlo study on the type I error rates and power when using the ST_3 -plot for testing spherical symmetry:

$$\tilde{\mathbf{d}}_1, \tilde{\mathbf{d}}_2, \tilde{\mathbf{d}}_3, \tilde{\mathbf{d}}_4, \tilde{\mathbf{d}}_5, \tilde{\mathbf{d}}_6, \tilde{\mathbf{d}}_7. \quad (3.9)$$

The Monte Carlo study on type I error rates of the local acceptance region (3.3) is carried out by generating spherical samples from the following six spherical distributions by MATLAB code. These null distributions are discussed in detail in Chapter 3 of Fang et al. [12]. Here we only point out the corresponding parameters for the chosen spherical distributions without explaining their meaning. The six chosen spherical (null) distributions are: (1) the standard normal distribution $N_d(\mathbf{0}, \mathbf{I}_d)$; (2) the multivariate t -distribution with degrees of freedom $m = 5$; (3) the Kotz type distribution with $N = 2$, $s = 1$, and $r = 0.5$; (4) the Pearson

type VII distribution (PVII) with $m = 2$ and $N = 20$; (5) the Pearson type II distribution (PII) with $m = 3/2$; (6) the Cauchy distribution. The TFWW algorithm [22] pages 166–170 [23], is employed to generate empirical samples from these spherical distributions except the normal distribution whose samples can be generated from the MATLAB internal function. Table 1 gives the empirical type I error rates ($\alpha = 5\%$) of the ST_3 -plot in testing spherical symmetry for dimensions $d = 20$ and $d = 30$, where the seven directions are given by (3.9) and the type I error rates were calculated by

$$\text{type I error rate} = \frac{\text{number of rejections}}{\text{number of replications}}. \quad (3.10)$$

The local acceptance region given in (3.3) for $\alpha = 5\%$ is used to count the number of rejections for the selected null distributions. The simulation was done with 2,000 replications.

Based on Table 1, we can summarize the following empirical conclusions:

- (1) the ST_3 -plot seems to have better control of the type I error rates by using the five directions $\tilde{\mathbf{d}}_1, \dots, \tilde{\mathbf{d}}_5$ than by using the two directions $\tilde{\mathbf{d}}_6$ and $\tilde{\mathbf{d}}_7$, which tend to have lower type I error rates than the significance level $\alpha = 5\%$. The ST_3 -plot based on these two directions may over-accept the null hypothesis of spherical symmetry;
- (2) the performance of the ST_3 -plot on controlling the type I error rates tends to be slightly affected by the sample size. This may be due to the arrangement of the observation matrix $\tilde{\mathbf{X}}$ in Theorem 2.2. So we can expect the ST_3 -plot to have good control on type I error rates in the case of high dimension with a small sample size. This is a good indication for high-dimensional data analysis.

For $\alpha = 1\%$ and 10% , we obtained similar results to those in Table 1 on the type I error rates of the ST_3 -plot by using the acceptance region (3.3) and the same directions in (3.9). These are not presented to save space.

3.4. Power Study

The power of the ST_3 -plot in testing spherical symmetry is computed by using the acceptance regions (3.3) and the formula given by (3.10). The KS-type statistic (2.17) is computed in the same way as in Section 3.2. The following six nonspherical alternative distributions are selected:

- (1) the multivariate χ^2 -distribution comprises i.i.d. $\chi^2(1)$ -variables;
- (2) the multivariate exponential distribution comprises i.i.d. univariate exponential variables with a density function $\exp(-x)$, $x > 0$;
- (3) the multivariate gamma distribution that comprises of i.i.d. gamma variables, each has a gamma distribution with a density $f(x) = (x/4) \exp(-x/2)$ ($x > 0$);
- (4) the distribution $\text{nor} + \chi^2$ comprises i.i.d. marginals, $[d/2]$ marginals have a standard normal distribution $N(0, 1)$ and $d - [d/2]$ marginals have a chi-squared distribution $\chi^2(2)$. Here d is the sample dimension;
- (5) the distribution $\text{nor} + \text{Cauchy}$ comprises two independent marginals, one is the $[d/2]$ -dimensional standard normal distribution and the other has a $d_1 (= d - [d/2])$ -dimensional Cauchy distribution;

Table 1: Type I error rates of the ST_3 -plot for testing spherical symmetry.

Directions	n	Normal	Multivar. t	Kotz	PVII	PII	Cauchy
$(d = 20 \text{ and } \alpha = 0.05)$							
\tilde{d}_1	20	0.0620	0.0580	0.0525	0.0445	0.0560	0.0440
	50	0.0480	0.0505	0.0490	0.0490	0.0520	0.0465
	100	0.0505	0.0485	0.0570	0.0485	0.0475	0.0540
	200	0.0470	0.0560	0.0545	0.0495	0.0495	0.0500
\tilde{d}_2	20	0.0535	0.0425	0.0465	0.0480	0.0485	0.0505
	50	0.0495	0.0510	0.0585	0.0550	0.0450	0.0455
	100	0.0395	0.0445	0.0490	0.0530	0.0480	0.0500
	200	0.0505	0.0560	0.0430	0.0495	0.0525	0.0500
\tilde{d}_3	20	0.0400	0.0365	0.0360	0.0485	0.0490	0.0550
	50	0.0495	0.0525	0.0465	0.0590	0.0460	0.0480
	100	0.0425	0.0540	0.0405	0.0535	0.0475	0.0540
	200	0.0455	0.0495	0.0515	0.0530	0.0490	0.0525
\tilde{d}_4	20	0.0465	0.0485	0.0480	0.0520	0.0530	0.0535
	50	0.0520	0.0530	0.0425	0.0445	0.0495	0.0475
	100	0.0555	0.0405	0.0340	0.0470	0.0435	0.0510
	200	0.0600	0.0480	0.0440	0.0520	0.0445	0.0425
\tilde{d}_5	20	0.0520	0.0405	0.0480	0.0505	0.0560	0.0445
	50	0.0420	0.0380	0.0355	0.0400	0.0425	0.0305
	100	0.0365	0.0375	0.0370	0.0395	0.0370	0.0375
	200	0.0405	0.0260	0.0330	0.0460	0.0375	0.0350
\tilde{d}_6	20	0.0525	0.0470	0.0515	0.0480	0.0475	0.0470
	50	0.0035	0.0030	0.0030	0.0030	0.0025	0.0020
	100	0.0035	0.0020	0.0010	0.0020	0.0040	0.0065
	200	0.0010	0.0015	0.0035	0.0020	0.0025	0.0025
\tilde{d}_7	20	0.0520	0.0405	0.0480	0.0505	0.0560	0.0445
	50	0.0065	0.0005	0.0050	0.0035	0.0040	0.0065
	100	0.0030	0.0005	0.0030	0.0000	0.0030	0.0075
	200	0.0015	0.0020	0.0035	0.0010	0.0025	0.0050
$(d = 30 \text{ and } \alpha = 0.05)$							
\tilde{d}_1	20	0.0485	0.0500	0.0455	0.0640	0.0625	0.0600
	50	0.0475	0.0465	0.0485	0.0510	0.0545	0.0560
	100	0.0505	0.0440	0.0540	0.0515	0.0390	0.0535
	200	0.0610	0.0485	0.0580	0.0580	0.0570	0.0575
\tilde{d}_2	20	0.0465	0.0485	0.0595	0.0615	0.0410	0.0490
	50	0.0480	0.0455	0.0515	0.0605	0.0465	0.0465
	100	0.0445	0.0675	0.0490	0.0515	0.0455	0.0440
	200	0.0460	0.0435	0.0500	0.0480	0.0455	0.0460

Table 1: Continued.

Directions	n	Normal	Multivar. t	Kotz	PVII	PII	Cauchy
$\tilde{\mathbf{d}}_3$	20	0.0455	0.0630	0.0465	0.0530	0.0465	0.0565
	50	0.0510	0.0530	0.0405	0.0515	0.0555	0.0555
	100	0.0540	0.0475	0.0465	0.0535	0.0535	0.0515
	200	0.0535	0.0535	0.0585	0.0470	0.0470	0.0435
$\tilde{\mathbf{d}}_4$	20	0.0530	0.0490	0.0565	0.0455	0.0430	0.0470
	50	0.0600	0.0490	0.0540	0.0490	0.0525	0.0490
	100	0.0435	0.0490	0.0410	0.0505	0.0510	0.0495
	200	0.0510	0.0500	0.0515	0.0600	0.0465	0.0525
$\tilde{\mathbf{d}}_5$	20	0.0535	0.0470	0.0390	0.0540	0.0570	0.0410
	50	0.0420	0.0335	0.0365	0.0325	0.0455	0.0310
	100	0.0415	0.0350	0.0400	0.0370	0.0305	0.0310
	200	0.0420	0.0380	0.0360	0.0395	0.0440	0.0325
$\tilde{\mathbf{d}}_6$	20	0.0520	0.0600	0.0520	0.0605	0.0550	0.0405
	50	0.0040	0.0055	0.0030	0.0030	0.0060	0.0030
	100	0.0015	0.0015	0.0005	0.0010	0.0025	0.0020
	200	0.0005	0.0010	0.0005	0.0010	0.0005	0.0020
$\tilde{\mathbf{d}}_7$	20	0.0475	0.0500	0.0485	0.0540	0.0435	0.0360
	50	0.0325	0.0370	0.0375	0.0320	0.0325	0.0280
	100	0.0010	0.0005	0.0000	0.0000	0.0005	0.0035
	200	0.0005	0.0010	0.0015	0.0005	0.0015	0.0030

- (6) the distribution $t + \text{Kotz}$ has a similar meaning to that for the distribution given by (5), one marginal has a $[d/2]$ -dimensional t -distribution with parameter $m = 5$ and the other has a d_1 -dimensional Kotz type distribution with parameters $N = 2$, $s = 1$, and $r = 0.5$.

By a standard Monte Carlo technique, a $d \times n$ sample matrix $\tilde{\mathbf{X}}_0 = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ can be generated from the above six nonspherical alternative distributions according to their marginal distributions. Then $\tilde{\mathbf{X}}_0$ is centered by

$$\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_0 - E(\tilde{\mathbf{X}}_0), \quad (3.11)$$

where the mean value $E(\tilde{\mathbf{X}}_0)$ is taken for each element of $\tilde{\mathbf{X}}_0$. By this way, we obtain nonspherical samples distributed in the space R^d like those of spherical samples. Table 2 presents the power ($\alpha = 5\%$) of the T_3 -plot in testing spherical symmetry for the six nonspherical distributions by using the seven directions in (3.9).

Based on Table 2, we can summarize the following empirical conclusions:

- (1) the ST_3 -plot based on the direction $\tilde{\mathbf{d}}_1$ remarkably outperforms (has much higher power) the other six directions under the four choices of the sample sizes and

Table 2: Power of the ST_3 -plot in testing spherical symmetry for six nonspherical distributions.

n	Direction	$\chi^2(1)$	Exp	Gamma	Nor + χ^2	Nor + Cauchy	$t + Kotz$
$(d = 20, \alpha = 5\%)$							
20	\tilde{d}_1	0.7700	0.5750	0.4000	0.7945	0.4085	0.3820
	\tilde{d}_2	0.0830	0.0635	0.0460	0.0660	0.0965	0.0940
	\tilde{d}_3	0.1310	0.0860	0.0640	0.0840	0.1250	0.2925
	\tilde{d}_4	0.0875	0.0580	0.0585	0.0585	0.0735	0.2405
	\tilde{d}_5	0.3755	0.2420	0.1725	0.2740	0.1815	0.3760
	\tilde{d}_6	0.0520	0.0450	0.0470	0.0490	0.0450	0.1220
	\tilde{d}_7	0.3755	0.2420	0.1725	0.2740	0.1815	0.3760
50	\tilde{d}_1	0.7625	0.5355	0.3435	0.8495	0.4235	0.4040
	\tilde{d}_2	0.0925	0.0685	0.0455	0.0825	0.1405	0.2740
	\tilde{d}_3	0.2965	0.1490	0.0830	0.1225	0.3200	0.3840
	\tilde{d}_4	0.0665	0.0570	0.0470	0.0480	0.0680	0.2315
	\tilde{d}_5	0.2540	0.1640	0.1130	0.2340	0.1795	0.2930
	\tilde{d}_6	0.0045	0.0065	0.0035	0.0040	0.0135	0.0045
	\tilde{d}_7	0.0050	0.0015	0.0055	0.0035	0.0115	0.0055
100	\tilde{d}_1	0.7310	0.5115	0.3205	0.8715	0.4250	0.4235
	\tilde{d}_2	0.0960	0.0670	0.0605	0.1020	0.3490	0.3670
	\tilde{d}_3	0.3995	0.2170	0.1055	0.1800	0.3980	0.4070
	\tilde{d}_4	0.0565	0.0395	0.0470	0.0415	0.0765	0.2605
	\tilde{d}_5	0.2310	0.1595	0.1070	0.2595	0.2195	0.3045
	\tilde{d}_6	0.0025	0.0030	0.0020	0.0040	0.0150	0.0030
	\tilde{d}_7	0.0030	0.0015	0.0040	0.0015	0.0165	0.0040
200	\tilde{d}_1	0.7200	0.4845	0.3070	0.8825	0.4080	0.4130
	\tilde{d}_2	0.0920	0.0605	0.0465	0.2155	0.4005	0.4065
	\tilde{d}_3	0.4740	0.2790	0.1545	0.2655	0.4065	0.4265
	\tilde{d}_4	0.0535	0.0550	0.0485	0.0320	0.0655	0.2380
	\tilde{d}_5	0.1755	0.1450	0.0995	0.2620	0.2315	0.2970
	\tilde{d}_6	0.0020	0.0035	0.0020	0.0045	0.0335	0.0020
	\tilde{d}_7	0.0010	0.0010	0.0025	0.0030	0.0385	0.0030
$(d = 30, \alpha = 5\%)$							
20	\tilde{d}_1	0.8440	0.6660	0.4775	0.8785	0.5000	0.5025
	\tilde{d}_2	0.1100	0.0805	0.0470	0.0705	0.0875	0.2755
	\tilde{d}_3	0.1010	0.0650	0.0515	0.0705	0.0915	0.2120
	\tilde{d}_4	0.1255	0.0980	0.0740	0.0715	0.0875	0.3445
	\tilde{d}_5	0.4740	0.3185	0.2040	0.3900	0.2420	0.5330
	\tilde{d}_6	0.0695	0.0565	0.0635	0.0545	0.0670	0.2090
	\tilde{d}_7	0.0840	0.0580	0.0540	0.0695	0.0580	0.2175
50	\tilde{d}_1	0.8360	0.5935	0.3875	0.8980	0.5060	0.6585
	\tilde{d}_2	0.0720	0.0610	0.0550	0.0710	0.1770	0.2005
	\tilde{d}_3	0.2345	0.1205	0.0930	0.1290	0.2145	0.8500
	\tilde{d}_4	0.0860	0.0680	0.0505	0.0535	0.0925	0.3610
	\tilde{d}_5	0.3450	0.2355	0.1410	0.3560	0.3920	0.5820
	\tilde{d}_6	0.0025	0.0030	0.0045	0.0035	0.0055	0.0300
	\tilde{d}_7	0.0585	0.0645	0.0415	0.0570	0.0740	0.3320

Table 2: Continued.

n	Direction	$\chi^2(1)$	Exp	Gamma	Nor + χ^2	Nor + Cauchy	$t + \text{Kotz}$
100	$\tilde{\mathbf{d}}_1$	0.7925	0.5760	0.3505	0.9250	0.5065	0.8675
	$\tilde{\mathbf{d}}_2$	0.0810	0.0585	0.0555	0.0850	0.2760	0.4265
	$\tilde{\mathbf{d}}_3$	0.3580	0.1910	0.1025	0.2055	0.3845	0.9765
	$\tilde{\mathbf{d}}_4$	0.0575	0.0605	0.0630	0.0555	0.0910	0.3695
	$\tilde{\mathbf{d}}_5$	0.3070	0.1905	0.1255	0.3915	0.5625	0.7005
	$\tilde{\mathbf{d}}_6$	0.0005	0.0020	0.0020	0.0000	0.0130	0.0095
	$\tilde{\mathbf{d}}_7$	0.0010	0.0015	0.0005	0.0020	0.0165	0.0060
200	$\tilde{\mathbf{d}}_1$	0.7950	0.5335	0.3220	0.9310	0.5075	0.9605
	$\tilde{\mathbf{d}}_2$	0.0905	0.0615	0.0510	0.1405	0.2920	0.6455
	$\tilde{\mathbf{d}}_3$	0.4615	0.2595	0.1330	0.2900	0.3910	0.9995
	$\tilde{\mathbf{d}}_4$	0.0555	0.0625	0.0595	0.0485	0.0112	0.3780
	$\tilde{\mathbf{d}}_5$	0.2865	0.1920	0.1255	0.4315	0.5718	0.7440
	$\tilde{\mathbf{d}}_6$	0.0015	0.0000	0.0015	0.0025	0.0140	0.0095
	$\tilde{\mathbf{d}}_7$	0.0005	0.0025	0.0000	0.0020	0.0175	0.0095

almost all of the selected nonspherical distributions. So $\tilde{\mathbf{d}}_1$ can be considered as a general choice for the T_3 -plot in testing high-dimensional spherical symmetry;

- (2) the ST_3 -plot is slightly affected by an increase of the sample size. Because of the rearrangement of the observation matrix $\tilde{\mathbf{X}} = (\mathbf{x}_1, \dots, \mathbf{x}_n) : d \times n$, instead of $n \times d$, the sample size n is implicitly taken as the sample dimension d , and the sample dimension d is taken as the sample size n . This is a rotation of the regular observation matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)' : n \times d$. This rotation results in a power decrease of the ST_3 -plot when the sample size n increases, and it results in a power increase of the T_3 -plot when the sample dimension d increases. This can be observed from Table 2. Therefore, the ST_3 -plot is especially suitable for testing very high-dimensional spherical symmetry with a relatively small sample size.

3.5. Practical Illustration

To illustrate how to apply the ST_3 -plot in practice, we employ a subset of a real data set. The data set was used in Walker and Wright [24] and was called the VDP (vertical density profile) data set. As described by Walker and Wright [24], manufacturers of engineered wood boards, which include particle board and medium density fiberboard, are very concerned about the density properties of the board produced. The density is measured using a profilometer which uses a laser device to take a series of measurements across the thickness of the board. A profilometer takes multiple measurements on a sample (usually a 2×2 inch piece) to form the vertical density profile of the board. The VDP data subset that we illustrate here consists of 45 measurements taken 0.014 inch apart, and comprises 2 groups:

- (A) group A consists of 9 subjects A_1, \dots, A_9 ;
- (B) group B consists of 11 subjects B_1, \dots, B_{11} .

We can consider each subject as an observation with 45 measurements. Then each observation has a dimension of 45. Based on the structure of the complete VDP data set,

groups A and B have a sample size 9 and 11, respectively. Because a spherical distribution always has a zero mean vector, the selected two groups of VDP data should be shifted to have the origin as its central location. This is realized by subtracting the group sample mean from each observation in groups A and B, respectively. That is,

CA. centerized subgroup A: the sample mean from the 9 subjects is subtracted from each observation in subgroup A;

CB. centerized subgroup B: the sample mean from the 11 subjects is subtracted from each observation in subgroup B.

After the centerization, each of the above subgroups is comparable to a sample from a spherical distribution, which has a zero mean.

For illustration purpose, in choosing the projection directions for the ST_3 -plots for the data of group A and group B as defined above, we consider four directions:

- (1) $\mathbf{d}_1 = \tilde{\mathbf{d}}_1$ determined by (3.5);
- (2) $\mathbf{d}_2 = \tilde{\mathbf{d}}_2$ determined by (3.5);
- (3) $\mathbf{d}_5 = \tilde{\mathbf{d}}_5$ determined by (3.8);
- (4) $\mathbf{d}_6 = \tilde{\mathbf{d}}_6$ determined by (3.8).

On each of these four directions, the T_3 -function (2.15) (i.e., the ST_3 -plot) and the acceptance region given by (3.3) are plotted in Figure 2 for the centerized subsets CA and CB.

The following facts can be observed:

- (1) the ST_3 -plot for the centerized data in subgroup A (the CA plots) at the projection direction \mathbf{d}_5 goes beyond the 90%-acceptance region, showing evidence of nonspherical symmetry of the data at the significance level $\alpha = 10\%$. As a result, it can be concluded that the null hypothesis of spherical symmetry for the centerized data in subgroup A is rejected at $\alpha = 10\%$;
- (2) the ST_3 -plots for the centerized data in subgroup B (the CB plots) at the projection directions \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_6 go beyond the 95%-acceptance regions, showing evidence of nonspherical symmetry of the data at the significance level $\alpha = 5\%$. As a result, it can be concluded that the null hypothesis of spherical symmetry for the centerized data in subgroup B is rejected at $\alpha = 5\%$.

The evidence of nonspherical symmetry of the data in subgroups A and B above implies that it may be inappropriate to set up a random effects model $\mathbf{x}_i - \bar{\mathbf{x}} = \sigma \mathbf{e}_i$ with spherically distributed random effects \mathbf{e}_i for the centerized data $\{\mathbf{x}_i - \bar{\mathbf{x}} : i = 1, \dots, n\}$ ($n = 9$ for subgroup A and $n = 11$ for subgroup B), where σ stands for a scale parameter. The illustration of detecting nonspherical symmetry in Figure 2 could provide a way to regression diagnostics with the assumption of spherically distributed error terms. One of the regression diagnostic techniques is to check if the residual random vectors like $\{\mathbf{y}_i - \hat{\mathbf{y}}_i : i = 1, \dots, n\}$ are approximately i.i.d. d -dimensional normal deviates $N_d(\mathbf{0}, \sigma^2 \mathbf{I}_d)$ by using the probability-plot method, where \mathbf{I}_d stands for the identity matrix and σ denotes an unknown standard deviation. If it shows a lack of fit for the normal assumption, one could consider testing the spherical symmetry for the residual random vectors $\{\mathbf{y}_i - \hat{\mathbf{y}}_i : i = 1, \dots, n\}$ by providing the ST_3 -plots as in Figure 2. If no evidence of nonspherical symmetry can be detected from

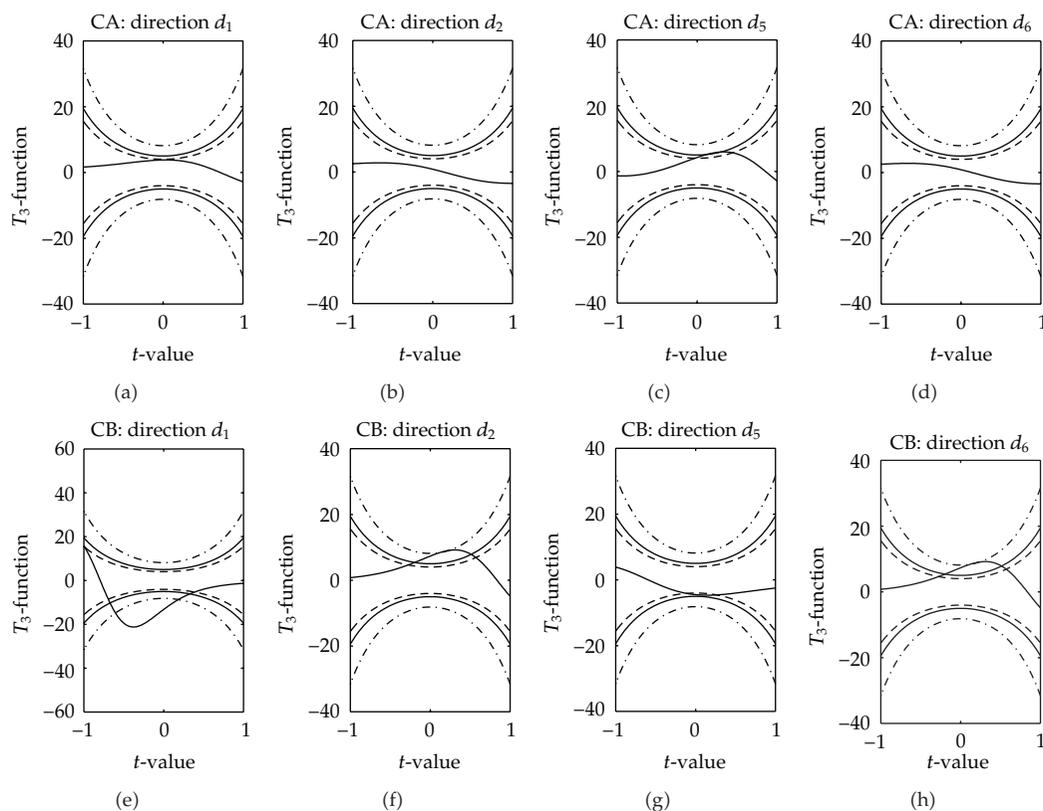


Figure 2: ST_3 -Plots for the two centered subsets of data in Groups A and B from the VDP data on four selected projection directions. d_1 , d_2 , d_5 , and d_6 have the same meaning as that for \tilde{d}_1 , \tilde{d}_2 , \tilde{d}_5 , and \tilde{d}_6 as in Tables 1 and 2, respectively. For each plot, the region between the two dashdotted curves stands for the 99% acceptance region, the region between the two real-line curves for the 95% acceptance region, and the region between the two dashed curves for the 90% acceptance region.

the ST_3 -plots, the regression model for the observed data $\{y_i : i = 1, \dots, n\}$ could be extended to have spherically distributed error terms. More discussion on regression models with spherically distributed error terms and statistical inference under elliptical distributions (which contains spherical distributions as a special case) can be referred to Fraser and Ng [25] and Fang and Zhang [18]. The illustration of detecting nonspherical symmetry for high-dimensional data by the ST_3 -plot provides a graphical tool for goodness-of-fit problems in generalized multivariate analysis.

4. Concluding Remarks

Ghosh's [16] original T_3 -plot is an effective graphical method for detecting nonnormality of univariate data. The T_3 -plot was extended to detecting nonmultinormality of high-dimensional data by Fang et al. [17]. In this paper we found another application of the T_3 -plot in testing high-dimensional spherical symmetry by providing approximate local acceptance regions. The simulation results in Section 3 show that the local acceptance regions given by (3.3) have feasible performance. Although we have not been able to find an optimal

projection direction in applying the ST_3 -plot to real high-dimensional data analysis, those directions used in the Monte Carlo study in Section 3 can provide potential users with a good reference. Theoretically, any projection direction subject to the condition in Theorem 2.2 can be applied to the ST_3 -plot for testing high-dimensional spherical symmetry. Some directions may perform better than others, as demonstrated in Section 3. For general purpose, the idea in analysis of principal components can provide a guideline for choosing projection directions when applying the ST_3 -plot to real high-dimensional data analysis. This is illustrated by (3.4), (3.7) and the VDP data set in Section 3.

In this paper we emphasize the ST_3 -plot for testing spherical symmetry for the case of high dimension with a small sample size. For regular cases of testing spherical symmetry, the Q-Q plots proposed by Li et al. [11], and those analytical methods summarized in Fang and Liang [13], or the methods mentioned in the relatively new references in Section 1, should be used. There has been a lack of new effective methods for analysis of high-dimensional data with a small sample size since the past few years. So the ST_3 -plot in this paper sheds some additional light on the area of high-dimensional data analysis.

References

- [1] K.-C. Li, "Sliced inverse regression for dimension reduction," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 316–342, 1991.
- [2] K.-C. Li, "On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 1025–1039, 1992.
- [3] R. D. Cook, "Principal Hessian directions revisited," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 84–94, 1998.
- [4] R. D. Cook and B. Li, "Dimension reduction for conditional mean in regression," *The Annals of Statistics*, vol. 30, no. 2, pp. 455–474, 2002.
- [5] M. J. R. Healy, "Multivariate normal plotting," *Applied Statistics*, vol. 17, pp. 157–161, 1968.
- [6] N. J. H. Small, "Plotting squared radii," *Biometrika*, vol. 65, pp. 657–658, 1978.
- [7] S. K. Ahn, " F -probability plot and its application to multivariate normality," *Communications in Statistics*, vol. 21, pp. 997–1023, 1992.
- [8] J. A. Koziol, "Probability plots for assessing multivariate normality," *The Statistician*, vol. 42, pp. 161–173, 1993.
- [9] B. M. Brown and T. P. Hettmansperger, "Normal scores, normal plots, and tests for normality," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1668–1675, 1996.
- [10] J.-J. Liang and P. M. Bentler, "A t -distribution plot to detect non-multinormality," *Computational Statistics & Data Analysis*, vol. 30, no. 1, pp. 31–44, 1999.
- [11] R. Li, K. T. Fang, and L. X. Zhu, "Some Q-Q probability plots to test Spherical and Elliptical symmetry," *Journal of Computational and Graphical Statistics*, vol. 6, pp. 435–450, 1997.
- [12] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London, UK, 1990.
- [13] K. T. Fang and J. Liang, "Testing Spherical and Elliptical symmetry," in *Encyclopedia of Statistical Sciences*, S. Kotz, C. B. Read, and D. L. Banks, Eds., vol. 3, pp. 686–691, John Wiley & Sons, New York, NY, USA, 1999.
- [14] V. I. Koltchinskii and L. Li, "Testing for Spherical symmetry of a multivariate distribution," *Journal of Multivariate Analysis*, vol. 65, no. 2, pp. 228–244, 1998.
- [15] C. Diks and H. Tong, "A test for symmetries of multivariate probability distributions," *Biometrika*, vol. 86, no. 3, pp. 605–614, 1999.
- [16] S. Ghosh, "A new graphical tool to detect non-normality," *Journal of the Royal Statistical Society B*, vol. 58, no. 4, pp. 691–702, 1996.
- [17] K.-T. Fang, R.-Z. Li, and J.-J. Liang, "A multivariate version of Ghosh's T_3 -plot to detect non-multinormality," *Computational Statistics & Data Analysis*, vol. 28, no. 4, pp. 371–386, 1998.
- [18] K. T. Fang and Y. T. Zhang, *Generalized Multivariate Analysis*, Springer, Berlin, Germany, 1990.

- [19] J. Läuter, "Exact t and F tests for analyzing studies with multiple endpoints," *Biometrics*, vol. 52, no. 3, pp. 964–970, 1996.
- [20] K. T. Fang, L. X. Zhu, and P. M. Bentler, "A necessary test of goodness of fit for sphericity," *Journal of Multivariate Analysis*, vol. 45, no. 1, pp. 34–55, 1993.
- [21] J. Läuter, E. Glimm, and S. Kropf, "New multivariate tests for data with an inherent structure," *Biometrical Journal*, vol. 38, no. 1, pp. 5–23, 1996.
- [22] Y. Tashiro, "On methods for generating uniform random points on the surface of a sphere," *Annals of the Institute of Statistical Mathematics*, vol. 29, no. 2, pp. 295–300, 1977.
- [23] K.-T. Fang and Y. Wang, *Number-Theoretic Methods in Statistics*, Chapman and Hall, London, UK, 1994.
- [24] E. Walker and S. P. Wright, "Comparing curves using additive models," *Journal of Quality Technology*, vol. 34, pp. 118–129, 2002.
- [25] D. A. S. Fraser and K. W. Ng, "Multivariate regression analysis with Spherical error," in *Multivariate Analysis—V*, P. R. Krishnaiah, Ed., pp. 369–386, North-Holland, 1980.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

