Generalizing Univariate Signed Rank Statistics For Testing and Estimating a Multivariate Location Parameter

Ola Hössjer[*]
Lund Institute of Technology

Christophe Croux[†]
University of Antwerp

Abstract. We generalize signed rank statistics to dimensions higher than one. This results in a class of orthogonally invariant and distribution free tests that can be used for testing spherical symmetry/location parameter. The corresponding estimator is orthogonally equivariant. Both the test and estimator can be chosen with asymptotic efficiency 1. The breakdown point of the estimator depends only on the scores, not on the dimension of the data. For elliptical distributions, we obtain an affine invariant test with the same asymptotic properties, if the signed rank statistic is applied to standardized data. We also present a method for computing the estimator numerically, and consider a real data example and some simulations. Finally, an application to detection of time-varying signals in spherically symmetric noise is given.

Key words and phrases. Affine invariant tests, asymptotic normality, breakdown point, distribution free tests, elliptical symmetry, orthogonal invariance, spherical symmetry.

1   Introduction

There exist several notions of symmetry for multivariate data. In this paper we will only be concerned with the concepts of *spherical* and *elliptical* symmetry. A random vector[1] $\mathbf{Y} \in I\!\!R^p$ is said to be spherically (elliptically) symmetric around $^-$ provided $\mathbf{Y} - ^-$ and $\mathbf{L}(\mathbf{Y} - ^-)$ have the same distribution for all orthogonal (bijective) linear operators on $I\!\!R^p$. Other more general notions of symmetry include central and angular symmetry. We refer to Small (1990) and Chaudhuri and Sengupta (1992) for more details. It is well known that signed rank statistics can be used for testing symmetry ($^- = \mathbf{0}$) of a univariate distribution, or alternatively, for estimating the center of symmetry (cf. e.g. Hettmansperger, 1984). In this paper we generalize signed rank statistics to the multivariate case. Let $a_n(1), \ldots, a_n(n)$ be a sequence of non-negative scores. For any vector $\mathbf{v} = (v_1, \ldots, v_n)$ in $I\!\!R^n$, define the distance

$$D_n(v_1, \ldots, v_n) = \sum_{i=1}^{n} a_n(R(|v_i|))|v_i|, \tag{1.1}$$

where $R(|v_i|)$ is the rank of $|v_i|$ among $|v_1|, \ldots, |v_n|$. Suppose that we have an i.i.d. sample $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, with each $\mathbf{X}_i$ having a spherically symmetric density around $^-$. Define

$$\hat{}_n(\mathbf{X}) = \arg\min D_n(\|\mathbf{X}_1 - ^-\|, \ldots, \|\mathbf{X}_n - ^-\|), \tag{1.2}$$

as an estimate of $^-$, with $\| \cdot \|$ denoting the Euclidean norm in $I\!\!R^p$. It is easy to see that $\hat{}_n$ is both location and orthogonally equivariant, that is for any $\mathbf{b} \in I\!\!R^p$ and orthogonal matrix $\mathbf{L}$,

$$\hat{}_n(\mathbf{LX} + \mathbf{b}) = \mathbf{L}\hat{}_n(\mathbf{X}) + \mathbf{b}, \tag{1.3}$$

with $\mathbf{LX} + \mathbf{b} = (\mathbf{LX}_1 + \mathbf{b}, \ldots, \mathbf{LX}_n + \mathbf{b})$. On the other hand, (1.3) need not hold if $\mathbf{L}$ is not orthogonal. We will also show that $\hat{}_n$ is asymptoically normally distributed, and calculate its breakdown point. The breakdown point depends on the scores only, not on p. Quite surprisingly, the asymptotic efficiency when using Wilcoxon scores tends to $3/4$ as p increases for multivariate gaussian data. On the other hand, it is possible to choose the scores so that the efficiency is 1 and the breakdown point tends to 0.5 as p increases.

The corresponding test for testing $H_0 : ^- = \mathbf{0}$ against $H_1 : ^- \neq \mathbf{0}$ is

$$\phi_1(\mathbf{X}) = \begin{cases} 1 \text{ if } \|\mathbf{T}_n(\mathbf{X})\|^2 > t \\ 0 \text{ if } \|\mathbf{T}_n(\mathbf{X})\|^2 < t, \end{cases} \tag{1.4}$$

where $\phi_1 = 1$ means "reject H_0",

$$\mathbf{T}_n(\mathbf{X}) = -[\frac{d}{d^-}D_n((\|\mathbf{X}_1 - ^-\|, \ldots, \|\mathbf{X}_n - ^-\|)]_{-=\mathbf{0}} = \sum_{i=1}^{n} a_n(R(\|\mathbf{X}_i\|)\mathbf{U}(\mathbf{X}_i) \tag{1.5}$$

and

$$\mathbf{U}(\mathbf{x}) = \begin{cases} \mathbf{0}, \text{if } \mathbf{x} = \mathbf{0} \\ \mathbf{x}/\|\mathbf{x}\|\text{if } \mathbf{x} \neq \mathbf{0} \end{cases}$$

---

[1] All vectors in $I\!\!R^p$ are tacitly assumed to be column vectors.

is the projection of $\mathbf{x}$ onto S^{p-1}, the unit sphere in $I\!\!R^p$. An important fact is that $\mathbf{T}_n(\mathbf{X})$ is distribution free under H_0, which makes it possible to calculate exact critical regions. The reason is that $\mathbf{T}_n(\mathbf{X}) = \sum_{i=1}^n a_n(i)\mathbf{U}(\mathbf{X}_{(i)})$, where $R(\|\mathbf{X}_{(i)}\|) = i$ and $\{\mathbf{U}(\mathbf{X}_{(i)})\}$ are i.i.d. and uniformly distributed on S^{p-1} (since the data has a spherically symmetric distribution).[2]

Note that $\|\mathbf{T}_n(\mathbf{L}\mathbf{X})\| = \|\mathbf{T}_n(\mathbf{X})\|$ for any orthogonal transformation. Hence, ϕ_1 is orthogonally invariant. Asymptotically, $\|\mathbf{T}_n(\mathbf{X})\|^2/n$ converges to a χ^2-distribution with p degrees of freedom, and to a non-central χ^2-distribution under contiguous alternatives. This makes it possible to construct approximate level α tests.

A drawback of ϕ_1 is that it is only invariant w.r.t. orthogonal transformations of the data, not all linear bijective transformations. Suppose $\mathbf{X}$ is a sample from an elliptical distribution, and that $\hat{\Sigma}_n$ is an estimate of the scatter matrix of the data, that is equivariant in the sense that

$$\hat{\Sigma}_n(\mathbf{L}\mathbf{X}) = \mathbf{L}\hat{\Sigma}_n(\mathbf{X})\mathbf{L}^T \tag{1.6}$$

for any non-singular matrix $\mathbf{L}$. We will show that the test

$$\phi_2(\mathbf{X}) = \begin{cases} 1, & \text{if } \|\mathbf{V}_n(\mathbf{X})\|^2 > t \\ 0, & \text{if } \|\mathbf{V}_n(\mathbf{X})\|^2 < t, \end{cases} \tag{1.7}$$

with

$$\mathbf{V}_n(\mathbf{X}) = \mathbf{T}_n(\hat{\Sigma}_n^{-1/2}(\mathbf{X})\mathbf{X}), \tag{1.8}$$

is invariant w.r.t. all linear bijective transformations of the data. Also, it has the same asymptotic behaviour as ϕ_1, if $\hat{\Sigma}_n$ is a $\sqrt{n}$-consistent estimator of the scatter matrix.

When $a_n(i) \equiv 1$, $\hat{\ }_n$ becomes the L_1-estimator, and ϕ_1 becomes a multidimensional extension of the univariate sign test. The asymptotic properties of the L_1-median were studied by Brown (1983), and its breakdown point by Lopuhaä and Rousseeuw (1991). Tests that are asymptotically equivalent to ϕ_1 when $a_n(i) \equiv 1$ and equivariant w.r.t. linear bijective transformations have been considered by Randles (1989), Brown and Hettmansperger (1989), Brown et al. (1992) and Hettmansperger et al. (1994). For Wilcoxon scores, that is $a_n(i) = i$, Peters and Randles (1990) have proposed a multivariate signed rank test that is similar to and asymptotically equivalent with ϕ_2, but based on the concept of interdirections (Randles, 1989) and the standard sample covariance matrix. For the two-sample problem, Randles and Peters (1990) consider a multivariate rank test with general score function, and compute its efficiency. Other multivariate generalizations of univariate signed rank tests are considered by Brown and Hettmansperger (1987), Chaudhuri (1992), Randles (1992), Jan and Randles (1994) and Hettmansperger et al. (1992).

The paper is organized as follows: In Section 2 we investigate the asymptotic properties of $\hat{\ }_n$ and ϕ_1, and the breakdown point of $\hat{\ }_n$ is considered in Section 3. In Section 4 we consider the invariance and asymptotic properties of ϕ_2. Asymptotic efficiencies and breakdown points are computed in Section 5 for multivariate gaussian data and various choices of score functions.

---

[2]If ties occur among $\{\|\mathbf{X}_i\|\}$, we may for instance choose the corresponding ranks randomly. Since we will assume that $\mathbf{X}_i$ has an absolutely continuous distribution, the probability that a tie occurs i 0. In the sequel, we will tacitly assume $\|\mathbf{X}_{(1)}\| < \ldots < \|\mathbf{X}_{(n)}\|$.

Computational issues, simulations and a real data example are discussed in Section 6. Finally, an analogue of ϕ_1, which has applications to communication theory, is considered in Section 7.

2   Asymptotics for orthogonally equivariant tests and estimators

We need the following regularity conditions.

(i) Let

$$\mathbf{X}_i = \bar{} + \mathbf{e}_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are i.i.d. random vectors with a spherically symmetric density $f(\mathbf{x}) = k(\|\mathbf{x}\|)$.

Let $G(r) = P(\|\mathbf{e}_1\| \leq r)$ be the distribution function of $\|\mathbf{e}_1\|$, so that

$$G'(r) = g(r) = p\,\omega_{p-1}k(r)r^{p-1}, \quad r > 0, \tag{2.2}$$

with ω_{p-1} the area of S^{p-1}. Given a vector $\mathbf{x} = (x_1, \ldots, x_p)$ in $I\!\!R^p$, let $\bigtriangledown f(\mathbf{x}) = (\partial f(\mathbf{x})/\partial x_1, \ldots, \partial f(\mathbf{x})/\partial x_p)^T$.

(ii) The Fisher information matrix $\int_{I\!\!R^p} \bigtriangledown f(\mathbf{x}) \bigtriangledown f(\mathbf{x})^T f(\mathbf{x})^{-1}d\mathbf{x} = I_1(G)\mathbf{I}_p$ of f w.r.t. the location parameter $\bar{}$ satisfies

$$0 < I_1(G) = \frac{1}{p}\int_0^\infty (\frac{k'(r)}{k(r)})^2 g(r)dr < \infty, \tag{2.3}$$

with $\mathbf{I}_p$ the $p \times p$ identity matrix.

(iii) The scores are generated from a function $h : [0,1] \rightarrow [0, \infty)$ in one of the following three ways: $a_n(i) = h(i/(n+1))$, $a_n(i) = \int_{(i-1)/n}^{i/n} h(u)du$ or $a_n(i) = Eh(U_{i:n})$, where $U_{i:n}$ is the i:th order statistic from a random sample of size n with a uniform distribution on $(0,1)$.

(iv) The function $h(u) = h_1(u) - h_2(u)$, where h_1 and h_2 are non-decreasing and square integrable on $(0,1)$.

We will consider the following sequence of contiguous alternatives to H_0:

$$H_{1n} : \quad \bar{} = \Delta/\sqrt{n}, \tag{2.4}$$

where $\Delta \in I\!\!R^p$ is arbitrary. Let also N_p denote a p-dimensional normal distribution.

Theorem 2.1 *Suppose (i)–(iv) hold. Then*

$$\frac{1}{\sqrt{n}}\mathbf{T}_n(\mathbf{X}) \xrightarrow{d} N_p(B(h,G)\Delta, A(h)\mathbf{I}_p) \text{ as } n \rightarrow \infty, \tag{2.5}$$

under H_{1n} in (2.4), where

$$B(h,G) = \frac{1}{p}\int_0^\infty h(G(r))(-\frac{k'(r)}{k(r)})g(r)dr \tag{2.6}$$

and

$$A(h) = \frac{1}{p}\int_0^1 h^2(u)du. \tag{2.7}$$

3

Corollary 2.2 *Suppose the threshold in (1.4) is chosen as $t = n\chi_\alpha^2(p)A(h)$, where $\chi_\alpha^2(p)$ is the $(1 - \alpha)$-quantile of a χ^2-distribution with p degrees of freedom. Then the power function*

$$\beta_n(\Delta) = P(\phi_1(\mathbf{X}) = 1|^- = \frac{\Delta}{\sqrt{n}}) \to P(\chi^2(p, \frac{B(h,G)^2}{A(h)}\|\Delta\|^2) > \chi_\alpha^2(p)) \text{ as } n \to \infty,$$

where $\chi^2(p, \delta)$ is a non-central χ^2-distribution with p degrees of freedom and non-centrality parameter δ.

We need the following additional regularity condition for Theorem 2.3:

(v) The score generating function h is non-negative and non-decreasing and not identically equal to 0.

Theorem 2.3 *Suppose (i)–(v) holds. The estimate $\hat{}_n$ defined in (1.2) then satisfies*

$$\sqrt{n}(\hat{}_n - {}^-) \xrightarrow{d} N_p(\mathbf{0}, \frac{A(h)}{B(h,G)^2}\mathbf{I}_p) \text{ as } n \to \infty. \tag{2.8}$$

3 Breakdown point of $\hat{}_n$

We will use a version of the breakdown point introduced by Donoho and Huber (1983). Let $\mathbf{X}'$ be a corrupted sample obtained by replacing at most m of the original points. The maximum bias caused by the contamination is

$$\text{bias}(m; \hat{}_n, \mathbf{X}) = \sup_{\mathbf{X}'} \|\hat{}_n(\mathbf{X}') - \hat{}_n(\mathbf{X})\|,$$

and the breakdown point

$$\varepsilon_n^*(\hat{}_n; \mathbf{X}) = \min\{\frac{m}{n}; \text{ bias}(m; \hat{}_n, \mathbf{X}) = \infty\}. \tag{3.1}$$

Let us make the following assumption:

(vi) The scores $a_n(1), \ldots, a_n(n)$ are non-negative and non-decreasing and $a_n(n) > 0$.

Put

$$n^* = \min\{j; \sum_{i=n-j+1}^{n} a_n(i) \geq \sum_{i=1}^{n-j} a_n(i)\}. \tag{3.2}$$

Theorem 3.1 *Given (vi), the breakdown point for the estimator $\hat{}_n$ defined in (1.2) is given by*

$$\varepsilon_n^*(\hat{}_n; \mathbf{X}) = n^*/n, \tag{3.3}$$

with n^* defined in (3.2).

Corollary 3.2 *Suppose (vi) holds and that $\{a_n(i)\}$ are generated from a function h in any of the three ways described in (iii). If h has at most a finite number of discontinuities, $\varepsilon_n^*(\hat{}_n; \mathbf{X}) \to \varepsilon^*(h)$ as $n \to \infty$, where $\varepsilon^*(h)$ is defined by*

$$\int\limits_0^{1-\varepsilon^*(h)} h(u)du = \int\limits_{1-\varepsilon^*(h)}^1 h(u)du. \tag{3.4}$$

Hence, we see that the breakdown point is the same for all p, in particular the same as for the univariate case (cf. Hettmansperger, 1984, p. 90).

4 Properties of ϕ_2

In this section we consider the test ϕ_2 introduced in (1.7). We first have:

Proposition 4.1 *Suppose the scatter matrix estimate $\hat{\Sigma}_n$ satisfies (1.6). Then*

$$\|\mathbf{V}_n(\mathbf{L}\mathbf{X})\| = \|\mathbf{V}_n(\mathbf{X})\|$$

for any non-singular matrix $\mathbf{L}$. In particular, the test ϕ_2 defined in (1.7) is invariant w.r.t. such transformations of the data.

We will need the following extra conditions for the asymptotic properties of $\mathbf{V}_n$:

(vii) The data are given by $\mathbf{X}_i = \mathbf{L}(\bar{} + \mathbf{e}_i)$, $\quad i = 1, \ldots, n$, where the errors $\mathbf{e}_i$ are i.i.d. random vectors with density $f(\mathbf{x}) = k(\|\mathbf{x}\|)$, and $\bar{}$ and $\mathbf{L}$ are unknown parameters.

(viii) It holds that $0 < I_1(G) < \infty$, $0 < I_2(G) = \int_0^\infty (\frac{k'(r)}{k(r)})^2 r^2 g(r)dr < \infty$ and $g(r) \le C_1/r$ for some constant $C_1 > 0$.

(ix) The score function h defined in (iii) is bounded.

(x) The statistic $\hat{\Sigma}_n$ is a $\sqrt{n}$-consistent estimator of $\Sigma = \mathbf{L}\mathbf{L}^T$.

As an example of a $\sqrt{n}$-consistent and robust estimator $\hat{\Sigma}_n$ we have the S-estimator (cf. Rousseeuw and Leroy 1987). The asymptotic behaviour of $\mathbf{V}_n(\mathbf{X})$ is given by the following theorem:

Theorem 4.2 *Suppose (iii)–(iv) and (vii)–(x) are satisfied. Then, if H_{1n} (cf. (2.4)) holds for the model in (vii),*

$$\frac{1}{\sqrt{n}}\mathbf{V}_n(\mathbf{X}) \xrightarrow{d} N_p(B(h, G)\Delta', A(h)\mathbf{I}_p),$$

where Δ' is a vector (depending on $\mathbf{L}$) with $\|\Delta'\| = \|\Delta\|$, and $B(h, G)$ and $A(h)$ are defined as in (2.6) and (2.7) respectively.

Corollary 4.3 *The asymptotic power function of ϕ_2 is the same as for ϕ_1, see Corollary 2.2.*

5 Asymptotic efficiencies

From Corollary 2.2, Theorem 2.3 and Corollary 4.3 we see that the asymptotic prestanda of ϕ_1, $\hat{\tilde{}}_n$ and ϕ_2 all depend on making the quantity

$$\frac{B(h,G)^2}{A(h)} = \frac{\left(\int_0^\infty \psi(r)\frac{k'(r)}{k(r)}g(r)dr\right)^2}{p\int_0^\infty \psi(r)^2 g(r)dr} \tag{5.1}$$

as large as possible, where $\psi = h \circ G$. The ratio in (5.1) is maximized for

$$h_G(u) = -\frac{k'(G^{-1}(u))}{k(G^{-1}(u))}, \tag{5.2}$$

giving the maximal value $I_1(G)$ (cf. (2.3)). The asymptotic efficiency becomes

$$e(h,G) = \frac{B(h,G)^2}{A(h)I_1(G)}. \tag{5.3}$$

In Table 1 we have computed the efficiency $e(h,G)$ for multivariate gaussian data ($k(r) = \exp(-r^2/2)/((2\pi\sigma^2)^{p/2})$) and two choices of h: sign scores ($h(u) \equiv 1$) and Wilcoxon scores ($h(u) = u$). We see that the efficiency of the sign scores approach 1 as p increases whereas the efficiency for the Wilcoxon scores decrease towards 3/4. The optimal score function $h = h_G$ defined in (5.2) has efficiency 1. See also Brown (1983), Hettmansperger et al. (1994), Peters and Randles (1990) and Randles and Peters (1990).

Table 1: Asymptotic efficiencies for gaussian data

p	1	2	3	4	5	6	7	8	9	10	∞
sign	0.637	0.785	0.849	0.884	0.905	0.920	0.931	0.940	0.946	0.951	1.000
Wilcoxon	0.955	0.985	0.975	0.961	0.949	0.938	0.928	0.920	0.913	0.907	0.750

In Table 2 we have computed the asymptotic breakdown point for the optimal score function (still assuming gaussian data). We see that $\varepsilon^*(h_G) \to 0.5$ as $p \to \infty$. As a comparison, $\varepsilon^*(h)$ equals 0.5 for sign scores and 0.293 for Wilcoxon scores, regardless of the value of p. The reason that $\varepsilon^*(h_G)$ depends on p is that G depends on p. When $p = 1$, h_G reduces to the normal scores function $\Phi^{-1}((u+1)/2)$, with Φ the c.d.f. of the standard normal distribution.

We see that the sign scores and the optimal scores behave similarly when p is large. This may be explained as follows: First note that $h_G(u) = G^{-1}(u)$ for gaussian data. Now G is the distribution of $\sqrt{X_{11}^2 + \ldots + X_{1p}^2}$ if $\mathbf{X}_1 = (X_{11}, \ldots, X_{1p})$. Hence, by the law of large numbers $G(\sqrt{p} \cdot)$ is close

Table 2: Breakdown points for optimal scores, gaussian data

p	1	2	3	4	5	6	7	8	9	10	∞
$\varepsilon^*(h_G)$	0.239	0.306	0.340	0.360	0.375	0.386	0.394	0.400	0.406	0.411	0.500

to a one point distribution at 1 when p is large, and so $h_G/\sqrt{p}$ is almost constant $= 1$. See also Chaudhuri (1992) for a discussion of this phenomenon. In fact, the same argument applies as soon as the marginals of $\mathbf{X}_1$ have finite variance. This includes the elliptically symmetric power family of Randles (1989), but not the multivariate t-distributions treated in Hettmansperger et al. (1994). In the latter case, $\sum_{i=1}^{p} X_{1i}^2/p$ has limiting χ^2-distribution.

6   Computation, simulation, and a real data example

In order to compute the estimate $\hat{\theta}_n$ in (1.2) we have to minimize the function $S_n(\theta) = D_n(\|\mathbf{X}_1 - \theta\|, \ldots, \|\mathbf{X}_n - \theta\|)$. Note that this is a convex function since

$$S_n(\theta) = \sup_{\pi \in \mathcal{P}_n} \sum_{i=1}^{n} a(\pi(i))\|\mathbf{X}_i - \theta\|, \tag{6.1}$$

where $\mathcal{P}_n$ is the group of n-permutations. If not all the points $\mathbf{X}_i$ lie on a line (in which case we essentially have a one-dimensional problem), then $\hat{\theta}_n = \text{argmin-}\theta\, S_n(\theta)$ is unique and can be characterized by the solution of the inequality

$$\sum_{\mathbf{X}_i = \theta} a_n(R(\|\mathbf{X}_i - \theta\|)) \geq \|\sum_{\mathbf{X}_i \neq \theta} a_n(R(\|\mathbf{X}_i - \theta\|)) \frac{\mathbf{X}_i - \theta}{\|\mathbf{X}_i - \theta\|}\|. \tag{6.2}$$

This follows from the results of Kemperman (1987) combined with (6.1). When the right hand side of (6.2) is not equal to zero, then we have a degenerate solution which equals one of the observations.

Define now the weights $w_i(\theta)$ by

$$w_i(\theta) = \begin{cases} \frac{a_n(R(\|\mathbf{X}_i - \theta\|))}{\|\mathbf{X}_i - \theta\|} & \mathbf{X}_i \neq \theta \\ 0 & \mathbf{X}_i = \theta \end{cases} \tag{6.3}$$

and a vector

$$\Delta(\theta) = \frac{\sum_{i=1}^{n} w_i(\theta)(\mathbf{X}_i - \theta)}{\sum_{i=1}^{n} w_i(\theta)}, \tag{6.4}$$

which points in the opposite direction as the gradient. The solution of the minimization problem satisfies $\Delta(\theta) = 0$, except when we have a degenerate solution.

In order to compute $\hat{\theta}_n$ we propose the following iterative algorithm, which is a steepest descent algorithm combined with stephalving. From a provisional solution $\hat{\theta}_{n,k}$ we compute $\Delta(\hat{\theta}_{nk})$. We first take a step $\Delta(\hat{\theta}_{n,k})$ from $\hat{\theta}_{n,k}$. However, we only do this if the objective function $S_n(\theta)$ gets smaller. If not, we only take a step $2^{-j}\Delta(\hat{\theta}_k)$, where j is the smallest integer i such that $S(\hat{\theta}_{n,k} + 2^{-i}\Delta(\hat{\theta}_{n,k})) < S(\hat{\theta}_{n,k})$. Then we take $\hat{\theta}_{n,k+1} = \hat{\theta}_{n,k} + 2^{-j}\Delta(\hat{\theta}_{n,k})$. In this way, the value of the objective function decreases with every step. We propose to stop the iteration process when the objective function does not decrease when we take a step smaller (in euclidean norm) than *tol* in the direction of $\Delta(\theta)$. A second control variable is *maxstep*, which gives a limit on the maximum number of steps. An algorithm in pseudo-code looks like:

- Initialize *tol* and *maxstep* to control the precision.

- $k = 0$ and $\hat{\theta}_{n,0}$ is an initial estimator, for example the coordinatewise median

- do while $k \geq maxstep$

  - if $\|\Delta(\hat{\theta}_{n,k})\| < tol$ then *maxhalf* =0
    else *maxhalf*$=(\ln(\|\Delta(\hat{\theta}_{n,k})\|) - \ln(tol))/\ln(2)$.

  - $\hat{\theta}_{n,k+1} = \hat{\theta}_{n,k} + \Delta(\hat{\theta}_{n,k})$

  - $j = 0$

  - do while $S(\hat{\theta}_{n,k+1}) > S(\hat{\theta}_{n,k})$ and $j \leq maxhalf$

    * $j = j + 1$
    * $\hat{\theta}_{n,k+1} = \hat{\theta}_{nk} + \Delta(\hat{\theta}_{n,k})/2^j$

    endo

  - if $j > maxhalf$ then return$(\hat{\theta}_{n,k})$

  - k=k+1

  endo

- If $k > maxstep$ then 'iteration failed'

Extensive experiments have shown that this algorithm is quite fast and that it always converges , even in the case of degenerate solutions (we were however not able to prove this rigorously).

In the one-dimensional case an algorithm was already given by Hettmansperger and Utts (1977). The generalization of their algorithm to higher dimensions is immediate, and given by

$$\hat{\theta}_{n,k+1} = \frac{\sum_{i=1}^{n} w_i(\hat{\theta}_{n,k})\mathbf{X}_i}{\sum_{i=1}^{n} w_i(\hat{\theta}_{n,k})}. \tag{6.5}$$

This algorithm omits the stephalving and then convergence is no longer guaranteed, as was already mentioned by Cheng and Hettmansperger (1983). The estimates $\hat{\theta}_{n,k}$ generated by (6.5) can be considered as k-step estimators. However, if we start from the coordinatewise median the orthogonal equivariance is not longer guaranteed, and if we start from the mean the robustness is not longer guaranteed. Therefore full iteration is advised, hence convergence is important.

An alternative approach is a generalization of Gower (1974). Gower's algorithm uses in each step a bisection method to find the minimum of the objective function in the direction of the steepest descent. This algorithm is much slower than ours. One could also use a Newton-step instead of (6.4). This corresponds essentially to the approach of Bedall and Zimmermann (1978) who gave an algorithm for the L_1-median. Instead of (6.4) we now take a step $\Delta_2(\theta)$ where $H(\theta).\Delta_2(\theta) = \Delta(\theta)$ and

$$H(\theta) = \frac{\sum_{i=1}^{n} w_i(\theta)\left(\mathbf{I}_p - \mathbf{U}(\mathbf{X}_i - \theta)\mathbf{U}(\mathbf{X}_i - \theta)'\right)}{\sum_{i=1}^{n} w_i(\theta)}.$$

It gives a better algorithm for the sign scores, but with general scores this method has no advantages.

Extensive experiments with sign, Wilcoxon and optimal scores have shown that the proposed algorithm yields better solutions (in the sense of a smaller value of the objective function than

Table 3: Location Estimates for the Hawkins-Bradu-Kass data-set

	$\hat{\Sigma}_n$ = S-estimator			$\hat{\Sigma}_n$ = Covariance		
	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
Sign	1.690	2.156	2.137	2.280	3.341	5.557
Wilcoxon	1.742	2.226	2.417	3.672	6.592	8.819
Optimal	1.743	2.253	2.137	3.218	5.557	7.280
mean(1-75)	3.207	5.597	7.230	3.207	5.597	7.230
mean(15-75)	1.538	1.780	1.687	1.538	1.780	1.687

the other discussed methods in generally less time.) The algorithm is very fast. On a 486 PC it takes less than a second to compute $\hat{\bar{}}_n$ with $tol= 10^{-6}$ for a data set with 50 observations in 4 dimensions.

As an example, we computed the location estimate $\hat{\bar{}}_n$ for the exploratory variables of the well-known data set of Hawkins, Bradu, and Kass (1984, Table 4). This data set consists of 75 observations in 3 dimensions, of which it is known that the first 14 observations are outliers. In Table 3 we give the location estimates for the rank estimates based on sign scores, Wilcoxon scores and optimal scores. We also give the mean of the whole data set and the mean of the last 61 observations. We used both a 50% breakdown S-estimator with biweight ρ-function and the classical empirical covariance-matrix for the estimate of the scatter of the data used in (1.8). The S-estimator was computed using the algorithm of Ruppert (1992). We see that the mean is the most sensitive to the outliers, while the sign scores give the most robust estimates. Note that it is really necessary to use a robust estimate of the scatter matrix when prescaling the data, otherwise the location estimator breaks down as was illustrated in Rousseeuw and Leroy (1987, page 271-273).

We performed a simulation study to compare the performance of the tests at finite samples. We compared the rank-based test ϕ_1 with sign scores, Wilcoxon scores and optimal normal scores with the Hotelling T^2 test. The Hotelling T^2 test is defined as $H = n\overline{X}'S^{-1}\overline{X}$, where $\overline{X}$ is the average and S is the classical covariance of the observations. We generated 10,000 samples and for each sample we looked whether the H_0 hypothesis $\bar{} = 0$ was rejected or not at a level of $\alpha = 0.05$. For the rank tests the value of the treshhold was taken as in Corollary 2.2, and for the Hotelling test it was chosen as $F(p, n-p, \alpha)\frac{p(n-1)}{n-p}$, where $F(p, n-p, \alpha)$ is the α-upper quantile of an F-distribution with p and $n-p$ degrees of freedom.

In the experiment each sample consists of $n = 20$ observations drawn from the distribution $(1 - \varepsilon)N(0, \mathbf{I}_p) + \varepsilon N(0, 10\mathbf{I}_p)$ for various values of ε and p. In Table 4 we reported the observed relative frequencies (with standard errors smaller than 0.005) Note that the values for the rank-based method are more or less stable under contamination, which is not the case for the Hotelling test. The Hotelling test is much too conservative under contamination: it is not capable of detecting deviations from the model, which can be considered as a kind of masking effect. These conclusions stay true in higher dimensions.

This can also be illustrated by a real data example. Consider the perspiration data (Johnson

9

Table 4: Observed Relative Frequency of Rejecting H_0 with $\alpha = 0.05$

		Contamination ε			
		0	0.1	0.2	0.3
$p = 1$	Sign	0.042	0.043	0.037	0.040
	Wilcoxon	0.043	0.048	0.042	0.046
	Optimal	0.032	0.036	0.030	0.034
	Hotelling	0.049	0.017	0.024	0.035
$p = 2$	Sign	0.051	0.045	0.048	0.047
	Wilcoxon	0.044	0.039	0.039	0.041
	Optimal	0.037	0.031	0.034	0.035
	Hotelling	0.051	0.015	0.006	0.011
$p = 3$	Sign	0.048	0.045	0.045	0.049
	Wilcoxon	0.043	0.042	0.039	0.045
	Optimal	0.039	0.037	0.035	0.043
	Hotelling	0.051	0.023	0.007	0.006
$p = 4$	Sign	0.048	0.044	0.047	0.051
	Wilcoxon	0.041	0.037	0.041	0.044
	Optimal	0.041	0.036	0.039	0.041
	Hotelling	0.050	0.032	0.011	0.004

and Wichern 1988, page 174; Peters and Randles 1990), which consists of 20 observations in 3 dimensions. Johnson and Wichern showed that the assumption of a multivariate normal distribution is reasonable. We test whether $\bar{} = (4, 50, 10)$ using the rank tests and the Hotelling test. We first compute P-values for the clean data set (I) and then for the contaminated data set (II), where we changed the value of the first and the second variable of the first observation. We compute ϕ_2 where $\hat{\Sigma}_n$ in formula (1.8) is either a 50% breakdown S-estimator or the empirical covariance matrix. We see in Table 5 that the P-value of the Hotelling test increases under contamination, although it is clear that the contamination makes the H_0 hypothesis less acceptable. This is a consequence of the masking effect. We see that the P-values of the rank test change in the right direction. Again,

Table 5: P-values for the clean and contaminated Perspiration data

	$\hat{\Sigma}_n = $ S-estimator		$\hat{\Sigma}_n = $ Covariance	
	I	II	I	II
Sign	0.1024	0.0163	0.0922	0.6811
Wilcoxon	0.0675	0.0147	0.0352	0.3407
Optimal	0.0756	0.0146	0.0455	0.4133
Hotelling	0.0649	0.3928	0.0649	0.3928

we see that is really necessary to use a robust estimator of scatter for prescaling the data.

7 Time-varying signals

Consider the following model:
$$\mathbf{X}_i = \mu \mathbf{s}_i + \mathbf{e}_i, \quad i = 1, \ldots, n, \tag{7.1}$$

where $\mathbf{s}_1, \ldots, \mathbf{s}_n \in I\!\!R^p$ are known, μ is an unknown scalar and $\{\mathbf{e}_i\}$ are i.i.d. random vectors with a spherically symmetric density. The statistic

$$W_n(\mathbf{X}) = \sum_{i=1}^{n} a_n(R(\|\mathbf{X}_i\|)) \mathbf{s}_i^T \mathbf{U}(\mathbf{X}_i) \tag{7.2}$$

may be used for testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ as follows:

$$\phi_3(\mathbf{X}) = \begin{cases} 1 \text{ if } |W_n(\mathbf{X})| > t \\ 0 \text{ if } |W_n(\mathbf{X})| < t. \end{cases}$$

Note that W_n is distribution free under H_0. An important communication application of (7.1) when $p = 2$ is the detection of bandpass signals in bandpass noise. The two components of $\mathbf{X}_i$ are then the in-phase and quadrature components of the signal. For instance, in radar applications it is common to assume spherically symmetric noise. (cf. Kassam and Poor, 1985). For the asymptotic investigation, assume that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{s}_i^T \mathbf{s}_i = s^2 > 0 \text{ as } n \to \infty. \tag{7.3}$$

Consider a sequence of contiguous alternatives $H_{1n} : \mu = \Delta/\sqrt{n}$, where Δ is a scalar. Suppose the threshold $t = \sqrt{nA(h)}s\lambda_{\alpha/2}$ in (7.2), with $\lambda_{\alpha/2}$ the $(1 - \alpha/2)$-quantile of the standard normal distribution. Then it is shown in Hössjer and Croux (1993) that the resulting test ϕ_3 is asymptotically level α with power function

$$\beta_n(\Delta) = P(\phi_3(\mathbf{X}) = 1 | \mu = \frac{\Delta}{\sqrt{n}}) \to 2 - \Phi(\lambda_{\alpha/2} - \frac{B(h, G)s\Delta}{\sqrt{A(h)}}) - \Phi(\lambda_{\alpha/2} + \frac{B(h, G)s\Delta}{\sqrt{A(h)}}),$$

provided (i)-(iv) and (7.3) hold, and Φ is the c.d.f. of the standard normal distribution.

A Appendix

Proof of Theorem 2.1. Put $\mathbf{T}_{n\Delta} = \sum_{i=1}^{n} a_n(R(\|\mathbf{e}_i + \Delta/\sqrt{n}\|))\mathbf{U}(\mathbf{e}_i + \Delta/\sqrt{n})$. Then clearly, $\mathbf{T}_n(\mathbf{X}) = \mathbf{T}_{n\Delta}$ when $^- = \Delta/\sqrt{n}$ in (2.1). Introduce also $\mathbf{S}_n(\mathbf{X}) = \sum_{i=1}^{n} h(G(\|\mathbf{X}_i\|))\mathbf{U}(\mathbf{X}_i)$, and define $\mathbf{S}_{n\Delta}$ as the value of $\mathbf{S}_n(\mathbf{X})$ when $^- = \Delta/\sqrt{n}$ in (2.1). It follows from Hàjek and Sĭdàk (1967, Theorem VI.7) that

$$E\|\frac{1}{\sqrt{n}}(\mathbf{T}_{n0} - \mathbf{S}_{n0})\|^2 \to 0. \tag{A.1}$$

Also, it may be shown as in Hàjek and Sĭdàk (1967, Section VI.2.1) that the density $q_n(\mathbf{x}) = \prod_1^n f(\mathbf{x}_i - \Delta/\sqrt{n})$ is contiguous w.r.t. $p_n(\mathbf{x}) = \prod_1^n f(\mathbf{x}_i)$ in $I\!R^{pn}$ for any $\Delta \in I\!R^p$. This and (A.1) imply that

$$\frac{1}{\sqrt{n}}(\mathbf{T}_{n\Delta} - \mathbf{S}_{n\Delta}) \xrightarrow{p} \mathbf{0}, \quad \forall \Delta \in I\!R^p. \tag{A.2}$$

In view of (A.2), it suffices to show that

$$\frac{1}{\sqrt{n}}\mathbf{S}_{n\Delta} \xrightarrow{d} N_p(B(h,G)\Delta, A(h)\mathbf{I}_p) \quad \forall \Delta \in I\!R^p. \tag{A.3}$$

When $\Delta = \mathbf{0}$, (A.3) follows immediately from the (multivariate) Central Limit Theorem. For general Δ, (A.3) then follows from LeCam's third Lemma (cf. Hàjek and Sĭdàk, 1967, Lemma V.I.4). The proof makes use of the fact that $(\mathbf{S}_{n0}, \log(q_n(\mathbf{X})/p_n(\mathbf{X})))^T/\sqrt{n}$ is asymptotically normal

$$N_{p+1}\left((\mathbf{0}, -\frac{1}{2}I_1(G)\|\Delta\|^2)^T, \begin{pmatrix} A(h)\mathbf{I}_p & B(h,G)\Delta \\ B(h,G)\Delta^T & I_1(G)\|\Delta\|^2 \end{pmatrix}\right),$$

which is also a consequence of the CLT. □

Proof of Theorem 2.3. Assume w.l.o.g. that $\bar{} = \mathbf{0}$. Then $\sqrt{n}\hat{}_n = \arg\min_\Delta D_{n\Delta} := -\hat{\Delta}_n$, where $D_{n\Delta} = D_n(\|\mathbf{e}_1 + \Delta/\sqrt{n}\|, \ldots, \|\mathbf{e}_n + \Delta/\sqrt{n}\|)$. Define $Q_{n\Delta} = D_{n\mathbf{0}} + \mathbf{T}_{n\mathbf{0}} \cdot \Delta/\sqrt{n} + B(h,G)\|\Delta\|^2/2$ as a quadratic approximation to $D_{n\Delta}$. We will first prove that

$$\bigtriangledown(D_{n\Delta} - Q_{n\Delta}) = \frac{1}{\sqrt{n}}(\mathbf{T}_{n\Delta} - \mathbf{T}_{n\mathbf{0}}) - B(h,G)\Delta \xrightarrow{p} \mathbf{0}, \quad \forall \Delta \in I\!R^p. \tag{A.4}$$

In view of (A.2), (A.4) will follow if we show that

$$\frac{1}{\sqrt{n}}(\mathbf{S}_{n\Delta} - \mathbf{S}_{n\mathbf{0}}) - B(h,G)\Delta \xrightarrow{p} \mathbf{0}. \tag{A.5}$$

Suppose first that h is bounded and let $\psi = h \circ G$. Then

$$\text{tr}\left(\text{Var}(\frac{1}{\sqrt{n}}(\mathbf{S}_{n\Delta} - \mathbf{S}_{n\mathbf{0}}))\right) \leq E\|\psi(\mathbf{e}_1 + \frac{\Delta}{\sqrt{n}})\mathbf{U}(\mathbf{e}_1 + \frac{\Delta}{\sqrt{n}}) - \psi(\mathbf{e}_1)\mathbf{U}(\mathbf{e}_1)\|^2. \tag{A.6}$$

Since ψ is monotone and $\mathbf{e}_1$ has an absolutely continuous density f,

$$\psi(\mathbf{e}_1 + \frac{\Delta}{\sqrt{n}})\mathbf{U}(\mathbf{e}_1 + \frac{\Delta}{\sqrt{n}}) \xrightarrow{a.s.} \psi(\mathbf{e}_1)\mathbf{U}(\mathbf{e}_1). \tag{A.7}$$

Hence, (A.6)–(A.7) and dominated convergence imply that

$$\text{tr}\left(\text{Var}(\frac{1}{\sqrt{n}}(\mathbf{S}_{n\Delta} - \mathbf{S}_{n\mathbf{0}}))\right) \to 0,$$

and together with (A.3) this proves (A.5). When ψ is unbounded, (A.4) may be established by approximating $\mathbf{T}_{n\Delta}$ with a statistic having scores generated from a truncated version of h, see the proof of Theorem 3.1 in Jurečkovà (1969).

Since the scores $a_n(i)$ are non-negative and non-decreasing, it follows from Theorem 2.1 in McKean and Schrader (1980) that $D_{n\Delta}$ is convex function of Δ. Since also $D_{n\mathbf{0}} = Q_{n\mathbf{0}}$, Lemma 3.1 in Heiler and Willers (1988) and (A.4) imply that

$$\sup_{\Delta \in K} |D_{n\Delta} - Q_{n\Delta}| \xrightarrow{p} \mathbf{0} \tag{A.8}$$

for any compact subset K of $I\!\!R^p$. Let $\hat{\Delta}_{nQ}$ be the unique minimizer of $Q_{n\Delta}$. Then (A.8) and the convexity of $D_{n\Delta}$ imply that

$$\hat{\Delta}_n + \hat{\Delta}_{nQ} = o_p(1). \tag{A.9}$$

Since

$$\hat{\Delta}_{nQ} = -\frac{1}{\sqrt{n}} B(h, G)^{-1} \mathbf{T}_{n\mathbf{0}}, \tag{A.10}$$

the theorem follows from (A.9), (A.10) and Theorem 2.1, with $\Delta = \mathbf{0}$. $\qquad\square$

Proof of Theorem 3.1.

$\underline{\varepsilon_n^* \leq n^*/n}$

Suppose n^* points, say $\mathbf{X}_1, \ldots, \mathbf{X}_{n^*}$, are replaced to form a new sample $\mathbf{X}' = (\mathbf{X}'_1, \ldots, \mathbf{X}'_n)$. For ease of notation, put $D'_n(\bar{\theta}) = D_n(\|\mathbf{X}'_1 - \bar{\theta}\|, \ldots, \|\mathbf{X}'_n - \bar{\theta}\|)$. Pick L so large that $\|\mathbf{X}_i\| < L$, $i = n^*+1, \ldots, n$ and put $\mathbf{X}'_i = M\mathbf{e}_1$, $i = 1, \ldots, n^*$, with $M > 3L$. Given any $\bar{\theta} \neq M\mathbf{e}_1$, let $\mathbf{t}(\bar{\theta}) = \mathbf{U}(M\mathbf{e}_1 - \bar{\theta}) \in S^{p-1}$ be the direction from $\bar{\theta}$ towards $M\mathbf{e}_1$. Since

$$\begin{cases} \mathbf{t}(\bar{\theta}) \cdot \mathbf{U}(\mathbf{X}'_i - \bar{\theta}) = 1, \\ R(\|\mathbf{X}'_i - \bar{\theta}\|) > n - n^*, \end{cases} \quad \forall \|\bar{\theta}\| \leq L, \; i \leq n^*,$$

it follows from (3.2) that

$$\left[\frac{dD'_n(\bar{\theta} + r\mathbf{t}(\bar{\theta}))}{dr} \right]_{r=0} = -\mathbf{t}(\bar{\theta}) \cdot \sum_{i=1}^{n} a_n(R(\|\mathbf{X}'_i - \bar{\theta}\|)) \mathbf{U}(\mathbf{X}'_i - \bar{\theta}) \leq$$

$$\leq \sum_{i=1}^{n-n^*} a_n(i) - \sum_{i=n-n^*+1}^{n} a_n(i) \leq 0, \quad \forall \|\bar{\theta}\| \leq L. \tag{A.11}$$

Assume now that $D'_n(\tilde{\theta}_n) = \inf_{\bar{\theta}; \|\bar{\theta}\| \leq L/2} D'_n(\bar{\theta})$. According to (A.11) we may then find an $r > 0$ such that $L/2 < \|\tilde{\theta}_n + r\mathbf{t}(\tilde{\theta}_n)\| \leq L$ and and $D'_n(\tilde{\theta}_n + r\mathbf{t}(\tilde{\theta}_n)) \leq D'_n(\tilde{\theta}_n)$. Since L can be chosen as large as we please, we have a breakdown.

$\underline{\varepsilon_n^* \geq n^*/n}$

Suppose now that $\mathbf{X}_1, \ldots, \mathbf{X}_{n^*-1}$ are replaced. Note that

$$\sum_{i=1}^{n-n^*+1} a_n(i) - \sum_{i=n-n^*+2}^{n} a_n(i) = \delta > 0, \tag{A.12}$$

in view of (3.2). Choose L so large that $\|\mathbf{X}_i\| < L$, $i = 1, \ldots, n$. Then, for any $\mathbf{t} \in S^{p-1}$,

$$\left[\frac{dD'_n(r\mathbf{t})}{dr} \right]_{r=L} = -\mathbf{t} \cdot \mathbf{T}_n(\mathbf{X}'_1 - L\mathbf{t}, \ldots, \mathbf{X}'_n - L\mathbf{t}) =$$

$$= -\mathbf{t} \cdot \sum_{i=1}^{n} a_n(R(\|\mathbf{X}'_i - L\mathbf{t}\|)) \mathbf{U}(\mathbf{X}'_i - L\mathbf{t}). \tag{A.13}$$

13

It is clear that $\lim_{L\to\infty}\inf_{\substack{\mathbf{t}\in S^{p-1}\\i\geq n^*}}-\mathbf{t}\cdot\mathbf{U}(\mathbf{X}_i'-L\mathbf{t})=1$. Given $\varepsilon>0$, choose $L_0=L_0(\varepsilon)$ so large that $-\mathbf{t}\cdot\mathbf{U}(\mathbf{X}_i'-L\mathbf{t})\geq1-\varepsilon$ for all $L\geq L_0$, $i\geq n^*$ and $\mathbf{t}\in S^{p-1}$. It follows from (vi) and (A.12)–(A.13) that

$$
\begin{aligned}
\left[\frac{dD_n'(r\mathbf{t})}{dr}\right]_{r=L} &> (1-\varepsilon)\sum_{i\geq n^*}a_n(R(\|\mathbf{X}_i'-L\mathbf{t}\|))-\sum_{i<n^*}a_n(R(\|\mathbf{X}_i'-L\mathbf{t}\|))\geq\\
&\geq (1-\varepsilon)\sum_{i=1}^{n-n^*+1}a_n(i)-\sum_{i=n-n^*+2}^{n}a_n(i)=\delta-\varepsilon\sum_{i=1}^{n-n^*+1}a_n(i)>0,
\end{aligned}
$$

for all $L\geq L_0$ and $\mathbf{t}\in S^{p-1}$, provided ε is small enough. Therefore, $\hat{}_n(\mathbf{X}')$ must be located inside $\{\mathbf{x};\ \|\mathbf{x}\|\leq L\}$. $\square$

Proof of Proposition 4.1. See Hössjer and Croux (1993). $\square$

Proof of Theorem 4.2. According to (vii) and (x), $\mathbf{T}_n(\Sigma^{-1/2}\mathbf{X})=\mathbf{T}_n(\mathbf{R}^-+\mathbf{Re}_1,\dots,\mathbf{R}^-+\mathbf{Re}_n)$, where $\mathbf{R}=\Sigma^{-1/2}\mathbf{L}$ is orthogonal (as is easily seen). Putting $\Delta'=\mathbf{R}\Delta$, it follows from Theorem 2.1 that

$$\frac{1}{\sqrt{n}}\mathbf{T}_n(\Sigma^{-1/2}\mathbf{X})\xrightarrow{d}N_p(B(h,G)\Delta',A(h)\mathbf{I}_p). \tag{A.14}$$

Put $\hat{\Sigma}_n=\hat{\Sigma}_n(\mathbf{X})$. Because of (A.14), it suffices to show that

$$\frac{1}{\sqrt{n}}(\mathbf{T}_n(\hat{\Sigma}_n^{-1/2}\mathbf{X})-\mathbf{T}_n(\Sigma^{-1/2}\mathbf{X}))\xrightarrow{p}\mathbf{0}. \tag{A.15}$$

Because of contiguity, it suffices to establish (A.15) when $\Delta=\mathbf{0}$. Assumption (x) implies that

$$\hat{\Gamma}_n=\hat{\Sigma}_n^{-1/2}\Sigma^{1/2}=\mathbf{I}_p+O_p(1/\sqrt{n}). \tag{A.16}$$

W.l.o.g. we may assume $\mathbf{R}=\mathbf{I}_p$ (this holds when $\mathbf{L}=\Sigma^{1/2}$). Formula (A.15) is then equivalent to

$$\frac{1}{\sqrt{n}}(\mathbf{T}_n(\hat{\Gamma}_n\mathbf{e})-\mathbf{T}_n(\mathbf{e}))\xrightarrow{p}\mathbf{0}, \tag{A.17}$$

with $\mathbf{e}=(\mathbf{e}_1,\dots,\mathbf{e}_n)$. Given any number $M>0$, introduce $C_n(M)=\{fl=(\gamma_{ij})_{i,j=1}^p;\ |\gamma_{ij}|\leq M/\sqrt{n}\}$, put $\Gamma_n(fl)=\mathbf{I}_p+fl/\sqrt{n}$, and $\mathbf{Z}_n(fl)=(\mathbf{T}_n(\Gamma_n(fl)\mathbf{e})-\mathbf{T}_n(\mathbf{e}))/\sqrt{n}$. It remains to show that

$$\sup_{fl\in C_n(M)}\|\mathbf{Z}_n(fl)\|=o_p(1),\quad\forall M>0. \tag{A.18}$$

Formula (A.17) will then follow from (A.16) and (A.18). For a proof of (A.18) we refer to Hössjer and Croux (1993). $\square$

References

Bedall, F.K., and Zimmermann, H. (1979). Algorithm AS 143, The Mediancentre, *Applied Statistics*, **28**, 325-328.

Brown, B.M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society*, Ser. B, **45**, 25-30.

Brown, B.M. and Hettmansperger, T.P. (1987). Affine invariant rank methods in the bivariate location model. *Journal of the Royal Statistical Society*, Ser. B, **49**, 301-310.

Brown, B.M. and Hettmansperger, T.P. (1989). An affine invariant bivariate version of the sign test. *Journal of the Royal Statistical Society*, Ser. B, **51**, 117-125.

Brown, B.M., Hettmansperger, T.P, Nyblom, J. and Oja, H. (1992). On certain bivariate sign tests and medians. *Journal of the American Statistical Association,* **87**, 127-135.

Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *The Annals of Statistics*, **20**, 897-916.

Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension: Inference based on the geometry of the data cloud, *Journal of the American Statistical Association*, **88**, 1363-1370.

Cheng, K-S. and Hettmansperger, T.P. (1983). Weighted least squares rank estimates, *Communications in Statistics, Theory and Methods*, **12**, 1069-1086.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point, in *A festschrift for Erich Lehmann*, eds. P. Bickel, K. Doksum, and J.L. Hodges, Jr. Wadsworth, Belmont, CA.

Gower, J.C. (1974). Algorithm AS 78, The Mediancentre, *Applied Statistics,* **23**, 466-470.

Hàjek, J., and Šidàk, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hawkins, D.M., Bradu, D., and Kass, G.V. (1984). Location of Several Outliers in Multiple-Regression Data Using Elemental Subsets, *Technometrics,* **26**, 197-208.

Heiler, S. and Willers, R. (1988). Asymptotic normality of R-estimates in the linear model, *Statistics*, **19**, 173-184.

Hettmansperger, T.P. (1984). *Statistical Inference Based on Ranks*, Wiley, New York.

Hettmansperger, T.P., Nyblom, J. and Oja, J. (1992). In L_1-*statistical analysis and related methods*, ed. Y. Dodge, Amsterdam: North-Holland, 267-278.

Hettmansperger, T.P., Nyblom, J. and Oja, J. (1994). Affine invariant multivariate one-sample sign tests. *Journal of the Royal Statistical Society*, Ser. B, **56**, 221-234.

Hettmansperger, T.P. and Utts, J.M. (1977). Robustness properties for a simple class of rank estimates, *Communications in Statistics, Theory and Methods*, **6**, 855-868.

Hössjer, O. and Croux, C. (1993). Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter. University of Antwerpen, Dept. of Mathematics & Computer Science, report 93-19.

Jan and Randles, R.H. (1994). A multivariate signed sum test for the one-sample location problem. To appear in *Journal of Nonparametric Statistics*.

Johnson, R.A., Wichern, D.W. (1988). *Applied Multivariate Statistical Analysis, Second Edition,* Prentice Hall, New Yersey.

Jurečkovà, J. (1969). Asymptotic linearity of a rank statistic in regression parameter, *The Annals of Mathematical Statistics*, **40**, 1889-1900.

Kassam, S.A. and Poor H.V. (1985). Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, **73**, 433-481.

Kemperman J.H.M (1987). The Median of a Finite Measure on a Banach Space, In *Statistical Data Analysis*, ed. Y. Dodge, North-Holland, Amsterdam, 217–231.

Lopuhaä, H.P. and Rousseeuw, P. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices, *The Annals of Statistics*, **19**, 229-248.

McKean, J.W. and Schrader, R.M. (1980). The geometry of robust procedures in linear models, *Journal of the Royal Statistical Society*, Ser. B, **42**, 366-371.

Peters, D., and Randles, R.H. (1990). A Multivariate Signed-Rank Test for the One-Sample Location Problem, *Journal of the American Statistical Association,* **85**, 552-557.

Randles, R.H. (1989). A Distribution-Free Multivariate Sign Test Based on Interdirections, *Journal of the American Statistical Association,* **84**, 1045–1050.

Randles, R.H. (1992). In L_1-*statistical analysis and related methods,* ed. Y. Dodge, Amsterdam: North-Holland, 295-304.

Randles, R.H. and Peters, D. (1990). Multivariate rank tests for the two-sample location problem. *Commun. Statist.-Theory Meth.,* **19**(11), 4225–4238.

Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection,* Wiley, New York.

Ruppert, D. (1992). Computing S-estimators for regression and multivariate location/ dispersion, to appear in *Journal of Computational and Graphical Statistics*.

Small, C.G. (1990). A survey of multidimensional medians, *International Statistical Review*, **58**, 263-277.