Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. **53**(2), 2011, 197–215 doi: 10.1111/j.1467-842X.2011.00624.x

AFFINE-EQUIVARIANT SPATIAL MEDIAN AND ITS USE IN THE MULTIVARIATE MULTI-SAMPLE LOCATION PROBLEM

FRANTIŠEK RUBLÍK[1] and JÁN SOMORČÍK[2]*

Slovak Academy of Sciences Bratislava and Comenius University Bratislava

Summary

The classical spatial median is not affine-equivariant, which often turns out to be an unfavourable property. In this paper, the asymptotic properties of an affine-equivariant modification of the spatial median are investigated. It is shown that under some weak regularity conditions, the modified spatial median computed by means of the sample norming matrix is asymptotically equivalent to the one computed by means of the population norming matrix, which yields its asymptotic normality. A consistent estimate of the asymptotic covariance matrix of the modified spatial median is also presented. These results are implemented in a scheme, where the sample norm is determined by means of the sample Dümbgen scatter matrix. The results are utilized in the construction of affine-invariant test statistics for testing the multi-sample hypothesis of equality of location parameters. The performance of the proposed tests is demonstrated through a simulation study.

Key words: affine-equivariant spatial median; asymptotic tests of location hypothesis; Dümbgen matrix.

1. Introduction

The spatial median of a random sample of points in $\mathbb{R}^d$ is defined as the point in $\mathbb{R}^d$ with the minimal sum of Euclidean distances to the sample points. It is used as an estimator of location. It is popular because of its robustness and ease of computation, for example by means of the improved version of the simple iterative algorithm of Weiszfeld (1937) proposed in Vardi & Zhang (2000).

However, there is an important deficiency of the spatial median: it is translation-, rotationally-, but not affine-equivariant. This means that if we perform a linear transformation of the coordinate system in which the d-dimensional data are measured, the spatial median of the data does not have to change accordingly. Let us illustrate the issue by an example. Suppose we have the measured body weight in kilograms and body height in centimetres of a sample of people. We compute the sample spatial median (79.4 kg, 180.2 cm)$^\top$, say. If we now express the original data in pounds and inches respectively, which is a linear change of scales, the lack of affine equivariance of the spatial median means that the sample spatial median computed from the 'new' data sample could be, for example, (168.9 lb, 72.3 in)$^\top$,

*Author to whom correspondence should be addressed.
[1]Institute of Measurement Science of the Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovakia.
[2]Faculty of Mathematics, Physics and Informatics of the Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia.
e-mail: somorcik@fmph.uniba.sk
Acknowledgments. The authors would like to thank the journal's editors and anonymous referees for their helpful comments and suggestions during the preparation of this paper. This work was supported by VEGA grant No. 1/0077/09 of the Science Grant Agency of the Slovak Republic.

which does not correspond to the previous (79.4 kg, 180.2 cm)$^\top$. Thus, the location of the sample expressed by the spatial median may depend on the coordinate system.

In this paper we investigate the properties of a modification of the spatial median that can be used to generate an affine-equivariant estimator of location. The idea is based on the transformation–re-transformation approach suggested by Rao (1988). Suppose that a positive-definite symmetric matrix W associated with the underlying random vector X is such that an affine transformation AX by a regular matrix A changes the matrix W into $(A^\top)^{-1}WA^{-1}$. For example, the inverse of the covariance matrix of the distribution (if it exists) can be taken for W. In general, W^{-1} having the above property is called a scatter matrix – see chapter 3.1 of Oja (2010). Therefore, let us call the matrix W the *inverse scatter matrix* of X or of the distribution of X. We denote by $W^{1/2}$ the unique positive-definite symmetric matrix satisfying $W^{1/2}W^{1/2} = W$.

Now, it is easy to see that if we perform an affine transformation of the data points $X_1, \dots, X_n$ using $W^{1/2}$, compute the spatial median of these new points and re-transform it using $W^{-1/2}$, we obtain an affine-equivariant estimate, which means that if we apply the above process to the data points $AX_1, \dots, AX_n$, where A can be any regular matrix, the result will be A times the result for $X_1, \dots, X_n$. Note that in the latter transformation–re-transformation we use $(A^\top)^{-1}WA^{-1}$ in place of W.

In practice, however, the value of the matrix W is often unknown and has to be estimated by an estimator W_n based on $X_1, \dots, X_n$. It will be subject to some regularity conditions, under which we will show in Section 2 that asymptotically it does not matter whether we use the modified spatial medians based on the true inverse scatter matrix W or on its estimate W_n.

We note that our approach is similar to the affine-equivariant version of the spatial median presented by Hettmansperger & Randles (2002). However, in Hettmansperger & Randles (2002) the location parameter and the scatter matrix are estimated simultaneously, which requires an iterative procedure whose convergence is not ensured. Moreover, the scatter matrix and its estimate are prescribed to be the Tyler matrices, in contrast to the freedom of choice provided in our approach. Lopuhaä (1992) treats a similar problem to that of our paper, but for general M-estimators of location. The spatial median can also be seen as an M-estimator of location, but, unfortunately, the theory of Lopuhaä (1992) is not applicable to it owing to some too restrictive assumptions about Lopuhaä's class of M-estimators. A version of affine-equivariant spatial median is discussed in Nevalainen, Larocque & Oja (2007) but under more restrictive assumptions about the underlying distribution, such as elliptical symmetry and continuous density. Finally, an affine-equivariant spatial median is developed by Chakraborty, Chaudhuri & Oja (1998). As in our paper, they use the transformation–re-transformation approach, but the transformation matrix is obtained in a completely different and computationally intensive way based on the so-called data-driven coordinate system. Moreover, elliptical symmetry is required.

We apply our modified spatial median to the d-dimensional q-sample location problem. Its setting is based on q independent random samples of sizes $n_1, \dots, n_q$ from d-variate probability distributions with densities of the form $f(x - \mu_1), \dots, f(x - \mu_q)$, respectively, where $x \in \mathbb{R}^d$ and f is an arbitrary non-negative function satisfying $\int_{\mathbb{R}^d} f(x)\mathrm{d}x = 1$. The aim of the inference is to test the hypothesis

$$H_0 : \mu_1 = \cdots = \mu_q. \tag{1}$$

Many tests not requiring normality assumptions have been developed for this problem. A test based on interdirections was constructed by Um & Randles (1998), a paper in which various other test procedures are also mentioned. Another test for this problem based on multivariate centred ranks is given by Hettmansperger, Möttönen & Oja (1998). The test of the general linear hypothesis in the multivariate linear model in Bai *et al.* (1990), based on the L_1-approach, is also applicable. However, as the authors note, its critical values are hard to obtain when the underlying distribution is not spherically symmetric. Finally, we mention the multivariate generalizations of the univariate sign and rank tests given by Oja (2010).

Two test statistics M_1 and M_2 based on spatial medians are presented by Somorčík (2006); details are in Section 3. Their structure is the same as that of the well-known Lawley–Hotelling test statistic, but they use sample spatial medians instead of sample means to estimate the location parameters. It was shown that this makes the test more robust in the case of heavy-tailed underlying distributions. However, because the spatial median is not an affine-equivariant estimate, the spatial median tests are not affine-invariant either. This means that the value of the test statistics and, thus, the result of the testing, may change after a linear transformation of the coordinate system in which the measurements were taken. Moreover, the lack of affine-invariance negatively affects the power of the spatial median tests, as has been shown by Somorčík (2007) in a simulation study.

For these reasons we present in Section 3 modified versions of test statistics M_1 and M_2 that use the affine-equivariant modified spatial median. Its use makes the tests affine-invariant, and their very good performance in the case of spherically symmetric underlying distributions remains in the case of elliptical symmetry. Some simulation results comparing the finite-sample performance of the improved spatial median tests with their competitors are included in Section 4, and a real-data example is given in Section 5. The proofs and related remarks are postponed to the Appendix or to the Supplementary Material published online at the ANZJS website. The Supplementary Material also contains a computer code in R that implements our proposed methods.

2. Asymptotic properties of the modified spatial median

As is well known, the spatial median of the d-dimensional random vector X is defined as

$$\mu := \arg\min_{M\in\mathbb{R}^d} \mathrm{E}(\|X - M\| - \|X\|),$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. Its sample counterpart, the sample spatial median of the random sample $X_1, \ldots, X_n$ from $\mathbb{R}^d$, is defined as

$$\hat{\mu} := \arg\min_{M\in\mathbb{R}^d} \sum_{i=1}^{n} \|X_i - M\|. \tag{2}$$

For any symmetric positive-definite $d \times d$ matrix V define a vector norm by the formula $|||x|||_V := (x^\top Vx)^{1/2}$. Let W be an inverse scatter matrix of the random vector X. By the spatial median $\theta(W)$ of X given by the norm $|||\cdot|||_W$ we understand the vector

$$\theta(W) := \arg\min_{M\in\mathbb{R}^d} \mathrm{E}(|||X - M|||_W - |||X|||_W).$$

Its finite-sample version is

$$\hat{\theta}_n(W) := \arg\min_{M\in\mathbb{R}^d} \sum_{i=1}^{n} |||X_i - M|||_W. \tag{3}$$

Let θ^W be the spatial median of the distribution of the vector $W^{1/2}X$ given by the standard Euclidean norm $\|\cdot\|$ and let $\hat{\theta}_n^W$ be its finite-sample version:

$$\hat{\theta}_n^W := \arg\min_{M\in\mathbb{R}^d} \sum_{i=1}^{n} \|W^{1/2}X_i - M\|. \tag{4}$$

It is easy to see that

$$\theta(W) = W^{-1/2}\theta^W, \quad \hat{\theta}_n(W) = W^{-1/2}\hat{\theta}_n^W. \tag{5}$$

In order to establish our asymptotic results it is natural to require that the estimated matrix W_n is 'close enough' to the theoretical W:

Assumption 1. The matrices $\{W_n\}_{n=1}^{\infty}$ and W are positive-definite, $W_n \to W$ a.s. and $W_n = W + O_P(n^{-1/2})$.

We will also require boundedness of the underlying density as in Chaudhuri (1992) and also directional symmetry. This kind of symmetry also involves spherical or elliptical symmetry as special cases.

Assumption 2. The random vector X has a density f with respect to the Lebesgue measure on $\mathbb{R}^d$, which is bounded on every bounded subset of $\mathbb{R}^d$.

Assumption 3. The distribution of X is directionally symmetric around a point $\mu \in \mathbb{R}^d$; in other words, $(X-\mu)/\|X-\mu\|$ has the same distribution as $-(X-\mu)/\|-(X-\mu)\|$.

Note that under Assumption 3 the centre of the directional symmetry and the spatial median of the distribution coincide. Moreover, it is easy to see that the distribution of $W^{1/2}X$ is also directionally symmetric, this time around $W^{1/2}\mu$, and therefore $\theta^W = W^{1/2}\mu$ and $\theta(W) \equiv \mu$. However, the sample counterparts $\hat{\theta}_n^W$ and $\hat{\theta}_n(W)$ in general do not equal $W^{1/2}\hat{\mu}$ and $\hat{\mu}$ respectively.

Define

$$U(x) := \begin{cases} \dfrac{x}{\|x\|} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \qquad P(x) := \begin{cases} \dfrac{1}{\|x\|}\left(I_d - \dfrac{xx^\top}{\|x\|^2}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

$$D_1 := \mathrm{E}(P(W^{1/2}X - \theta^W)), \quad D_2 := \mathrm{E}(U(W^{1/2}X - \theta^W) \cdot U^\top(W^{1/2}X - \theta^W)). \tag{6}$$

Because the distribution of $W^{1/2}X$ also satisfies Assumption 2, the results of Chaudhuri (1992) are applicable to $\hat{\theta}_n^W$, and one obtains

$$n^{1/2}\big(\hat{\theta}_n^W - \theta^W\big) \to \mathrm{N}_d\big(0, D_1^{-1}D_2D_1^{-1}\big) \quad \text{in distribution.} \tag{7}$$

This together with (5) gives

$$n^{1/2}\{\hat{\theta}_n(W) - \theta(W)\} \to \mathrm{N}_d(0, B), \text{ where } B := W^{-1/2} D_1^{-1} D_2 D_1^{-1} W^{-1/2}. \tag{8}$$

Our main theoretical result is the following assertion.

Theorem 1. *Under Assumptions 1, 2 and 3:* $\hat{\theta}_n(W_n) = \hat{\theta}_n(W) + o_P(n^{-1/2})$. *Consequently,*

$$n^{1/2}(\hat{\theta}_n(W_n) - \theta(W)) \to \mathrm{N}_d(0, B) \quad \text{in distribution.} \tag{9}$$

In statistical practice one needs a consistent estimate of B. The following theorem states that a strategy analogous to that of Somorčík (2006) and Bai *et al.* (1990) in the case of the classical spatial median (2) can be used.

Theorem 2. *Define*

$$\begin{aligned} \hat{D}_1 &:= \frac{1}{n} \sum_{i=1}^{n} P\big(W_n^{1/2} X_i - W_n^{1/2} \hat{\theta}_n(W_n)\big), \\ \hat{D}_2 &:= \frac{1}{n} \sum_{i=1}^{n} U\big(W_n^{1/2} X_i - W_n^{1/2} \hat{\theta}_n(W_n)\big) \cdot U^{\top}\big(W_n^{1/2} X_i - W_n^{1/2} \hat{\theta}_n(W_n)\big), \\ \hat{B}_n &:= W_n^{-1/2} \hat{D}_1^{-1} \hat{D}_2 \hat{D}_1^{-1} W_n^{-1/2}. \end{aligned} \tag{10}$$

Suppose that Assumptions 1, 2 and 3 hold. If $d \geq 3$, *then* $\hat{B}_n = B + o_P(1)$.

If the dimension $d = 2$ then the asymptotic properties of the estimator (10) turn out to be a more complicated problem, which occurs in a similar context also in Bose (1995) and Bose & Chaudhuri (1993). Consistency of this estimate can be established under the following conditions.

Assumption 4. Let W_{nk} denote the estimate of W based on all n data points except of the kth. Then $\max_{k=1,\ldots,n} \| W_{nk}^{1/2} - W^{1/2} \| \to 0$ a.s.

Assumption 5. $\max_{k=1,\ldots,n} \| W_{nk}^{1/2} - W_n^{1/2} \| = O_P(n^{-1/2})$.

Theorem 3. *Suppose that Assumptions 1–5 are fulfilled. Then the result of Theorem 2 holds also for* $d = 2$.

Because the proof of the previous theorem requires a relatively large amount of space, it is not included in this paper. Details can be found in Somorčík (2009); see also the remark in the Appendix.

The next result states that in some cases the classical and the modified spatial median asymptotically coincide, a fact that we will use to study the efficiency of the latter.

Theorem 4. *Suppose that Assumptions 1, 2 and 3 are fulfilled. If* $W = c \cdot I_d$ *for a positive scalar* c *then* $\hat{\theta}_n(W_n) = \hat{\mu} + o_P(n^{-1/2})$, *where* $\hat{\mu}$ *is given by (2).*

The validity of the condition $W = c{\cdot}I_d$ in the above theorem depends on the choice of the matrix W. However, because W is the inverse of a scatter matrix, it always holds, for instance, under a broad class of distributions with exchangeable and symmetric marginals. See theorem 3.1 of Oja (2010).

We note that the asymptotic relative efficiency (ARE) of an estimator $\hat{\xi}_1$ relative to another $\hat{\xi}_2$ is defined as $ARE(\hat{\xi}_1, \hat{\xi}_2) := (\det(V_2)/\det(V_1))^{1/d}$, where V_1 and V_2 are the asymptotic covariance matrices of $\hat{\xi}_1$ and $\hat{\xi}_2$ respectively. Hence, under the assumptions of Theorem 4,

$$ARE(\hat{\theta}_n(W_n), \hat{\mu}) = 1. \tag{11}$$

Suppose that the underlying distribution is elliptically symmetric; hence, its density is of the form $g(x^\top \Sigma^{-1} x))$. If the affine-equivariant $\hat{\theta}_n(W_n)$ and an affine-equivariant location estimator $\hat{\xi}$ have asymptotic covariance matrices B and V under the density $g(x^\top x)$ then the affine equivariance implies that their asymptotic covariance matrices under the aforementioned elliptical symmetry will be $\Sigma^{1/2} B \Sigma^{1/2}$ and $\Sigma^{1/2} V \Sigma^{1/2}$. This means that

$$ARE(\hat{\theta}_n(W_n), \hat{\xi}) = \left(\frac{\det(\Sigma^{1/2} V \Sigma^{1/2})}{\det(\Sigma^{1/2} B \Sigma^{1/2})}\right)^{1/d} = \left(\frac{\det(V)}{\det(B)}\right)^{1/d} = \frac{d_V}{d_B}, \tag{12}$$

where d_V and d_B are the diagonal elements of V and B respectively, as affine equivariance ensures that under spherical symmetry these matrices are of the form $V = d_V{\cdot}I_d$ and $B = d_B{\cdot}I_d$. Hence, what we obtain is the ARE of $\hat{\theta}_n(W_n)$ relative to $\hat{\xi}$ under the spherically symmetric density $g(x^\top x)$. $ARE(\hat{\mu}, \hat{\xi})$ has been evaluated in many papers under spherically symmetric distributions for various affine-equivariant estimators $\hat{\xi}$; see, for example, Bai *et al.* (1990), Oja & Niinimaa (1985) or Möttönen, Oja & Tienari (1997). Under spherical symmetry the condition $W = c{\cdot}I_d$ is guaranteed, (11) and (12) imply that $ARE(\hat{\theta}_n(W_n), \hat{\xi}) = ARE(\hat{\mu}, \hat{\xi})$, and this equality holds also under elliptically symmetric versions of these spherically symmetric distributions.

3. Modified spatial median in multi-sample location tests

In this section we use the modified affine-equivariant spatial medians for testing the hypothesis (1). Classical spatial medians were used by Somorčík (2006) in the statistics

$$M_1 := \sum_{a=1}^{q} n_a(\hat{\mu}_a - \bar{\mu})^\top \hat{V}^{-1}(\hat{\mu}_a - \bar{\mu}), \quad M_2 := \sum_{a=1}^{q} n_a(\hat{\mu}_a - \hat{\mu})^\top \hat{V}^{-1}(\hat{\mu}_a - \hat{\mu}),$$

where $\hat{\mu}_1, \ldots, \hat{\mu}_q$ are the classical spatial medians of the particular samples (see (2)), $\bar{\mu} := (1/n)\sum_{a=1}^{q} n_a \hat{\mu}_a$ is their weighted mean, $\hat{\mu}$ is the sample spatial median of the pooled sample and $\hat{V}$ is a consistent estimate of the asymptotic covariance matrix V of the sample spatial medians. The asymptotic distribution of M_1 and M_2 was proved to be $\chi^2_{(q-1)d}$ provided that H_0 holds and Assumption 2 is fulfilled; in this case, no symmetry assumptions about the underlying distribution are needed. In the test based on M_i, the null hypothesis (1) is rejected for large values of M_i; more precisely, it is rejected if M_i exceeds the αth critical value of the $\chi^2_{(q-1)d}$ distribution.

The modified spatial median can be used for testing the q-sample hypothesis of the equality of location parameters too. Let $X_{a1}, \ldots, X_{an_a}$ be a random sample from the d-

dimensional distribution with density of the form $f(x - \mu_a)$, ($x \in \mathbb{R}^d$ and f is an arbitrary non-negative function satisfying $\int_{\mathbb{R}^d} f(x)\mathrm{d}x = 1$), $a = 1, \ldots, q$, with the samples being independent. Let $\hat{\eta}_a := \hat{\theta}_{n_a}(W_n)$ denote the modified spatial median of the ath sample defined by (5) with W_n based on the pooled sample of all $n := n_1 + \cdots + n_q$ data points, and $\bar{\eta} := (1/n)\sum_{a=1}^{q} n_a \hat{\eta}_a$. Finally, $\hat{B} := \hat{B}_n$ denotes the estimate (10) of the asymptotic covariance matrix of the affine-equivariant spatial median computed from the pooled sample. Define new test statistics by the formulas

$$A_1 := \sum_{a=1}^{q} n_a(\hat{\eta}_a - \bar{\eta})^\top \hat{B}^{-1}(\hat{\eta}_a - \bar{\eta}), \quad A_2 := \sum_{a=1}^{q} n_a(\hat{\eta}_a - \hat{\eta})^\top \hat{B}^{-1}(\hat{\eta}_a - \hat{\eta}).$$

Because the modified spatial medians are affine-equivariant, these statistics are invariant under affine transformation of the data.

In the following theorem we consider also the Pitman alternatives, which means that in accordance with Assumption 3 the centre of directional symmetry of the ath sample is $\mu_a = \mu + h_a/n^{1/2}$. The constant vectors $h_a \in \mathbb{R}^d$ satisfy the natural condition $\sum_{a=1}^{q} p_a h_a = 0$, where $p_a = \lim(n_a/n) > 0$.

Theorem 5. *Suppose that Assumptions 1, 2 and 3 are fulfilled.*

(I) *A_1 and A_2 are asymptotically equal under the null hypothesis (1): $A_1 = A_2 + o_P(1)$. Moreover, both converge to $\chi^2_{(q-1)d}$ in distribution as $n_1 \to \infty, \ldots, n_q \to \infty$.*

(II) *Under the Pitman alternatives, the asymptotic distribution of A_1 is non-central chi-squared $\chi^2_{(q-1)d}(\delta_{A_1})$ with non-centrality parameter $\delta_{A_1} := \sum_{a=1}^{q} p_a h_a^\top B^{-1} h_a$, where B is defined in (8).*

(III) *If $W = c{\cdot}I_d$ for a positive scalar c, then $A_1 = M_1 + o_P(1)$ under H_0 or the Pitman alternatives and $A_2 = M_2 + o_P(1)$ under H_0.*

If the underlying distribution is such that the Pitman alternatives are contiguous to H_0 then Theorem 5(I) implies that the asymptotic distribution of A_2 under the Pitman alternatives is the same as that of A_1, with the same non-centrality parameter $\delta_{A_2} = \delta_{A_1}$. Also note that Theorem 5(I) in this paper and theorem 3 in Somorčík (2006) imply that A_1 and A_2 are asymptotically equal to the well-known several-sample spatial sign test statistic. Moreover, the affine invariance of A_1 and A_2 provides a similar relationship also under elliptical symmetry: A_1 and A_2 are asymptotically equal to the affine-invariant generalizations of the several-sample spatial sign test statistic: to the one based on interdirections and presented in Um & Randles (1998) and to the other one presented in Oja (2010) on p. 156 and considered also in the simulation study below.

To carry out the modification described in the general setting, we have chosen the Dümbgen matrix in the role of the scatter matrix W^{-1}. The Dümbgen matrix Υ of the distribution of X is defined by Dümbgen (1998) and Dümbgen & Tyler (2005) as the symmetric positive-definite matrix satisfying the equalities

$$\Upsilon = d \cdot \mathrm{E}\left(\frac{(X-Y)(X-Y)^\top}{(X-Y)^\top \Upsilon^{-1}(X-Y)} \middle| X - Y \neq 0 \right), \quad \mathrm{tr}(\Upsilon) = d,$$

where Y is an independent copy of X. Its sample version $\Upsilon_n = \Upsilon_n(X_1, \ldots, X_n)$ based on the data points $X_1, \ldots, X_n$ is defined as the positive-definite symmetric matrix satisfying

$$\Upsilon_n = \frac{d}{m} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \tau(X_i - X_j, \Upsilon_n), \quad \operatorname{tr}(\Upsilon_n) = d,$$

where m is the number of the indices $i < j$ such that $X_i \neq X_j$, $\tau(x, \Upsilon_n) = xx^\top / x^\top \Upsilon_n^{-1} x$ if $x \neq 0$ and $\tau(x, \Upsilon_n) = 0$ otherwise. The following theorem sets the minimal number of data points to ensure the almost sure existence of the modified spatial median based on the sample Dümbgen matrix.

Theorem 6. *Let $X_1, \ldots, X_n$ be a random sample from the distribution of the random vector X, having a density with respect to the d-dimensional Lebesque measure, and $\hat{\theta}_n(W_n)$ be the modified spatial median defined by (5) with $W_n := \Upsilon_n^{-1}$. Suppose that*

$$n > \max\left\{d, 2d\frac{(n-1)}{(n+1)}\right\}. \tag{13}$$

Then with probability 1 this median $\hat{\theta}_n(W_n)$ exists, is uniquely defined and affine-equivariant.

We shall employ the previous modified spatial median for testing (1). Thus in the rest of this section $W_n := \Upsilon_n^{-1}$. Because according to corollary 4.1 of Dümbgen (1998) Assumption 1 holds for this choice of W_n, according to Theorem 5 under validity of Assumptions 2 and 3 the test based on A_i rejects the null hypothesis (1) if A_i exceeds the αth critical value of the $\chi^2_{(q-1)d}$ distribution.

The choice of the Dümbgen matrix enables us to use a simpler form of $\hat{B}$. Dümbgen (1998) specifies certain classes of symmetric underlying distributions under which the Dümbgen matrix Υ is equal to the Tyler matrix Δ defined in Tyler (1987) by the implicit formula

$$\Delta = d \cdot \mathrm{E}\left(\frac{(X-\mu)(X-\mu)^\top}{(X-\mu)^\top \Delta^{-1} (X-\mu)} \middle| X \neq 0\right), \quad \operatorname{tr}(\Delta) = d.$$

Then the definition (6) immediately implies that $D_2 = (1/d) \cdot I_d$, as $W = \Upsilon^{-1} = \Delta^{-1}$ and $\theta^W = W^{1/2}\mu$. Thus in these cases it is not necessary to estimate D_2, and one may estimate B by the estimator $\hat{B} := W_n^{-1/2} \hat{D}_1^{-1} D_2 \hat{D}_1^{-1} W_n^{-1/2}$. This form of $\hat{B}$ is used in the forthcoming simulation study, as the symmetric classes mentioned in Dümbgen (1998) cover also elliptically symmetric distributions and linear transformations of distributions with independent, identically distributed components symmetric around a location parameter. In a general situation, however, $\hat{B} := \hat{B}_n$ given by (10) has to be used.

4. Simulation study

The performance of the obtained tests in the finite-sample case is illustrated by simulation estimates in Table 1 in the case of $q = 3$ samples of $n_1 = n_2 = n_3 = 30$ data points from 3-variate ($d = 3$) distributions, (possibly) differing only by the shifts. We consider multivariate spherically symmetric normal (normal distribution), multivariate spherically symmetric Cauchy (Cauchy distribution I) with density of the form *const* $\times$ $1/(1 + \|x\|^2)^{(d+1)/2}$, and multivariate but not spherically symmetric Cauchy with independent components

TABLE 1

Simulated probabilities of type I error (H_0 true = yes) and powers (H_0 true = no) of our affine-invariant A_1, A_2 and affine-non-invariant M_1, M_2; of the Lawley-Hotelling's T^2 and the Oja's Q^2_{sign} and Q^2_{rank}. A_1^{perm}, A_2^{perm}: permutation approach applied to A_1, A_2 (1000 random data rearrangements).

		Test statistic								
H_0 true	Distribution	A_1	A_2	A_1^{perm}	A_2^{perm}	M_1	M_2	T^2	Q^2_{sign}	Q^2_{rank}
Yes	Cauchy I	.070	.077	.050	.049	.073	.079	.023	.045	.046
	Normal	.047	.050	.053	.053	.049	.052	.068	.049	.049
	Cauchy II	.069	.077	.046	.046	.073	.080	.023	.052	.050
No	Cauchy I	.44	.453	.411	.407	.430	.449	.049	.497	.322
	Normal	.740	.749	.771	.770	.738	.749	.876	.762	.832
	Cauchy II	.230	.243	.197	.196	.228	.243	.033	.240	.157

Setup: $q = 3$ samples of $n_1 = n_2 = n_3 = 30$ data points from some $d = 3$-variate distributions. Each entry based on 5000 Monte Carlo simulations.

TABLE 2

Simulated probabilities of type I error (H_0 true = yes) and powers (H_0 true = no) of the affine-non-invariant M_1 (first lines) and M_2 (latter lines) under elliptical deformations.

		ρ										
H_0 true	Distribution	−0.4	−0.3	−0.2	−0.1	0	0.5	0.6	0.7	0.8	0.9	0.95
Yes	Cauchy I	.080	.074	.073	.072	.073	.080	.081	.085	.097	.114	.151
		.088	.082	.078	.079	.079	.086	.089	.094	.107	.132	.174
	Normal	.051	.049	.049	.049	.049	.050	.050	.050	.053	.066	.087
		.057	.053	.052	.052	.052	.054	.055	.055	.062	.077	.105
	Cauchy II	.075	.073	.071	.073	.073	.081	.088	.093	.112	.145	.212
		.085	.081	.078	.080	.080	.092	.098	.106	.124	.169	.246
No	Cauchy I	.421	.430	.430	.431	.430	.438	.436	.433	.425	.414	.404
		.445	.452	.449	.447	.449	.454	.454	.451	.446	.440	.441
	Normal	.726	.733	.735	.738	.738	.722	.714	.702	.682	.637	.585
		.738	.744	.744	.748	.749	.730	.721	.710	.694	.653	.612
	Cauchy II	.221	.218	.222	.226	.228	.255	.264	.267	.276	.303	.341
		.240	.234	.236	.240	.243	.272	.280	.286	.303	.331	.376

Setup: $q = 3$ samples of $n_1 = n_2 = n_3 = 30$ data points from some elliptically deformed $d = 3$-variate distributions. Elliptical deformations given by the square root of the 3×3 matrix $\Sigma = (\Sigma_{ij})$ with $\Sigma_{ij} = 1$ if $i = j$ and $\Sigma_{ij} = \rho$ otherwise. Each entry based on 5000 Monte Carlo simulations.

(Cauchy distribution II) where each of the d independent components has the classical univariate Cauchy distribution with location 0 and scale 1. The simulation results include other affine-invariant test statistics: the Lawley–Hotelling test statistic T^2 from Um & Randles (1998), and the sign and rank test statistics Q^2_{sign} and Q^2_{rank} presented in Oja (2010) on p. 156 and pp. 159–160. In computing the Dümbgen matrix we used the function duembgen.shape() from the package ICSNP in the statistical software R provided by Nordhausen *et al.* (2010). Q^2_{sign} and Q^2_{rank} were evaluated using the package MNM by Nordhausen, Möttönen & Oja (2009).

In Table 2, the effect of the lack of affine invariance of the statistics M_1 and M_2 is illustrated by multiplying the sample data with the square root of the 3×3 matrix $\Sigma = (\Sigma_{ij})$, where $\Sigma_{ij} = 1$ if $i = j$ and $\Sigma_{ij} = \rho$ otherwise. Thus $\rho \in (-1/2, 1)$ determines the 'shape' of the elliptical deformation of the data clouds. The condition $\rho > -1/2$ is necessary for Σ to be positive-definite.

Based on 5000 Monte Carlo replications, the values in both tables report the proportions of times H_0 was rejected while testing at the nominal level of 5%. Except for the columns A_1^{perm} and A_2^{perm}, the asymptotic approach was adopted, which means that the 5% critical value of the $\chi^2_{(q-1)d}$ distribution was used for rejection of H_0. To obtain columns A_1^{perm} and A_2^{perm}, the permutation instead of the asymptotic approach was applied to the statistics A_1 and A_2: for each original sample of data points with labels specifying their belonging to one of the $q = 3$ samples, 1000 random rearrangements of the labels were performed to obtain the permutation p-value corresponding to the original arrangement of the labels. Thus, the proportion of times the permutation p-value was found under 5% is reported in columns A_1^{perm} and A_2^{perm}. We note that because the inverse W_n of the Dümbgen matrix, the matrix $\hat{B}$, and the modified spatial median $\hat{\eta}$ are all given by the *pooled* sample, for each of the 1000 permutations of the labels only $\hat{\eta}_1$, $\hat{\eta}_2$, $\hat{\eta}_3$ and consequently $\bar{\eta}$ had to be recomputed.

From the point of view of the significance level, lines with 'H_0 true = yes' show that the statistics A_1, A_2 are more suitable for testing (1) than M_1, M_2. Indeed, for $\rho = 0$ (Table 1) the simulated probabilities of type I error of A_1 and A_2 are similar to those of M_1 and M_2. With ρ moving away from 0 (Table 2), however, the probabilities of type I error of M_1 and M_2 seem to increase, up to 0.246 in the case of Cauchy distribution II, whereas the probabilities for A_1 and A_2 stay constant because of the affine invariance. The probabilities of type I error of A_1 and A_2 are slightly over 0.05, but, as indicated by columns A_1^{perm} and A_2^{perm}, the permutation approach to the testing cures this deficiency. The simulated probabilities of type I error of Q^2_{sign} and Q^2_{rank} appear to be close enough to the nominal level, without the need for the permutation approach.

Lines with 'H_0 true = no' present results of simulations where we used the location parameters $\mu_1 := (0, 0.4, 0)^\top$, $\mu_2 := (-0.4, 0.4, 0.4)^\top$ and $\mu_3 := (0.4, 0, 0)^\top$ to violate the null hypothesis (1). For $\rho = 0$ (Table 1), the powers of A_1 and A_2 are similar to those of M_1 and M_2. With ρ moving away from 0 (Table 2), however, we see a decrease of the powers of M_1 and M_2 under Cauchy distribution I and in particular under multivariate normality, whereas the powers of A_1 and A_2 remain constant. Under Cauchy distribution II we observe an increase of simulated powers of M_1 and M_2, but, as already noted, under this distribution the simulated probabilities of type I error are unacceptably high.

In the case of 'H_0 true = no', when comparing A_1, A_2, A_1^{perm} and A_2^{perm} with their affine-invariant competitors T^2, Q^2_{sign} and Q^2_{rank}, one can see the superior performance when confronted with Q^2_{rank} and T^2 under both Cauchy distributions. This was previously observed for M_1 and M_2 vs. T^2 in the case of $\rho = 0$; see Somorčík (2006). The affine invariance of A_1, A_2, A_1^{perm} and A_2^{perm} preserves this favourable property also after an elliptical deformation of the data distribution. The performance of A_1 and A_2 when compared with Q^2_{sign} is similar, except for the domination of Q^2_{sign} under Cauchy distribution I and the slightly higher simulated probabilities of type I error of A_1 and A_2 under both Cauchy distributions, owing to small sample sizes. A_1^{perm} and A_2^{perm} with ideal probabilities of type I error result in slightly lower powers than those of A_1 and A_2 under both Cauchy distributions. The superior performance of T^2 under multivariate normality is not surprising, as it is associated with the well-known optimality of the sample mean under this distribution. However, the A_1, A_2, A_1^{perm} and A_2^{perm} also compete quite well in this case.

Under Cauchy distribution II, Hettmansperger's H from Hettmansperger *et al.* (1998) was included in the simulations, resulting in a good probability of type I error (0.044) and in power (0.240) similar to the powers of A_1 and A_2. However, the advantage of A_1 and A_2

compared with H lies in the fact that H is computationally much more intensive. Therefore, H is difficult to use for large data sets, and cross-validation would be very slow.

5. An example

Let us illustrate the usage of our tests based on A_1 and A_2 by an application to the famous Egyptian skull data from $q = 3$ epochs: around 4000, around 3000, and around 1850 BC. For each of the $n_1 = n_2 = n_3 = 30$ skulls from each epoch we have $d = 4$ measurements: maximum breadth, basibregmatic height, basialiveolar length, and nasal height of the skull. The hypothesis (1) means that the measurements do not change over time; the opposite would indicate interbreeding with immigrant populations. See, for example, the R package HSAUR for details about the data.

First we compute by means of the function duembgen.shape() the inverse of the sample Düembgen matrix W_n given by the pooled sample of all $n = n_1 + n_2 + n_3 = 90$ data points. Then we multiply all the data points by the matrix $W_n^{1/2}$ and calculate the classical spatial medians of the $q = 3$ transformed samples; one can use the function spatial.median() from the R package ICSNP, for example. In accordance with (5), multiplying these spatial medians by $W_n^{-1/2}$ yields the modified spatial medians $\hat{\eta}_1 = \hat{\theta}_{n_1}(W_n)$, $\hat{\eta}_2 = \hat{\theta}_{n_2}(W_n)$, $\hat{\eta}_3 = \hat{\theta}_{n_3}(W_n)$ and their weighted average $\bar{\eta}$. In the same way, we also compute the modified spatial median $\hat{\eta}$ of the pooled sample. It remains to calculate the estimate $\hat{B} = \hat{B}_n$ of the asymptotic covariance matrix given by (10); we again use the pooled sample and the modified spatial median $\hat{\eta}$ of the pooled sample to centre the data. Now everything is ready to enumerate the test statistics A_1 and A_2 defined in Section 3. Here we summarize the numerical results:

$$W_n = \begin{pmatrix} 0.90 & -0.11 & 0.04 & -0.26 \\ -0.11 & 0.88 & -0.03 & -0.24 \\ 0.04 & -0.03 & 0.67 & -0.01 \\ -0.26 & -0.24 & -0.01 & 2.08 \end{pmatrix}, \quad \begin{aligned} \hat{\eta}_1 &= (131.40, 133.67, 99.61, 50.46)^\top \\ \hat{\eta}_2 &= (132.29, 133.05, 98.93, 50.34)^\top \\ \hat{\eta}_3 &= (135.03, 133.87, 95.79, 50.34)^\top \\ \bar{\eta} &= (132.98, 133.55, 98.03, 50.38)^\top \\ \hat{\eta} &= (132.85, 133.56, 98.23, 50.46)^\top \end{aligned}$$

$$\hat{B} = \begin{pmatrix} 24.43 & 7.83 & -2.45 & 5.02 \\ 7.83 & 22.88 & -1.44 & 5.28 \\ -2.45 & -1.44 & 29.67 & -0.25 \\ 5.02 & 5.28 & -0.25 & 9.72 \end{pmatrix}, \quad \begin{aligned} A_1 &= 17.63 \\ A_2 &= 17.77. \end{aligned}$$

The p-values based on the asymptotic $\chi^2_{(q-1)d} = \chi^2_8$ distribution are 2.4% and 2.3% respectively. The p-values of the corresponding permutation test based on 1000 random rearrangements of the 'epoch' labels of the skulls are both 1.0%; see Section 4 for details on the permutation approach. Therefore, we reject H_0 at the usual nominal level of 5% and conclude that the Egyptian skulls seem to have changed over time. We note that to obtain the above numerical results the reader can use the computer code in R that can be found in the Supplementary Material published online at the ANZJS website.

6. Discussion

From a theoretical point of view, the first two theorems presented in our paper provide a wide spectrum of possibilities for making statistical inferences based on an affine-equivariant modification of the classical spatial median. These possibilities are given by the general choice of the inverse scatter matrix and by the fact that the assumptions about its estimate are rather weak.

We utilized the modified spatial median to improve our previously published multivariate multi-sample location tests M_1 and M_2. The simulation study has shown that their lack of affine invariance can be eliminated in this way, because the finite-sample performance of the affine-invariant procedures A_1 and A_2 did not become worse despite the necessity of estimating the unknown inverse scatter matrix. Moreover, the good performance in the case of heavy-tailed underlying distributions remained, also owing to good robust properties of the sample Dümbgen matrix employed in estimating the scatter matrix. However, the simulation study has also shown that under smaller sample sizes the probability of type I error is in general slightly above the nominal level, a phenomenon that did not appear when considering, for instance, the affine-invariant sign and rank competitors Q^2_{sign} and Q^2_{rank}. Fortunately, the presented simulation study has shown that the easy-to-apply permutation approach instead of the asymptotic testing overcomes this difficulty, with no substantial decrease of power observed.

Appendix: The proofs and related remarks

Frequently we will use the fact noted by Chaudhuri (1992) that $1/\|x\|^\alpha$ is finitely integrable in any bounded neighbourhood of 0 for $0 \leq \alpha < d$. An immediate consequence is that for such an α under Assumption 2 the integral $\mathrm{E}(1/\|X\|^\alpha)$ is finite.

We will suppose that Assumptions 1, 2 and 3 hold, without stating this fact explicitly in the lemmas below. Without loss of generality, we assume that μ, the centre of the directional symmetry of the underlying density, is a zero vector. Note that the directional symmetry from Assumption 3 implies that the distribution of $W^{1/2}X$ is also directionally symmetric and therefore $\theta^W = 0$, by lemma 5.3 of Chaudhuri (1992).

For a $d \times d$ matrix A we will use the notation $\|A\|$ to denote its Frobenius norm. Finally, for the sake of brevity let $\hat{\theta}_n$ stand for $\hat{\theta}_n^{W_n}$.

Proof of Theorem 1

The strategy of the proof is inspired by the proof of the Bahadur-type representation of the classical spatial median in Chaudhuri (1992). This approach relies on step-wise approximations that use mainly the Bernstein inequality and similar results of Serfling (1980). The methodology of our proof differs substantially from the proof of the asymptotic normality of a general class of affine-equivariant M-estimators of location presented by Lopuhaä (1992), as we do not utilize concepts such as permissible classes of functions, polynomial discrimination, or envelopes.

First we introduce Lemmas 1–10. Their proofs can be found in the Supplementary Material published online at the ANZJS website.

Lemma 1. *For all* $x, y \in \mathbb{R}^d$ *such that* $x \neq 0$:

(I) $\|U(x + y) - U(x)\| \leq 2\|y\|/\|x\|$.

(II) *If* $x + y \neq 0$: $\|P(x+y) - P(x)\| \le ((d^{1/2}+1)/\|x+y\|^2 + (d^{1/2}+5)/\|x\|^2)\cdot\|y\|$.
(III) *If* $x + y \neq 0$: $\|P(x+y) - P(x)\| \le (d^{1/2}+5)\cdot\|y\|/(\|x\|\cdot\|x+y\|)$.

Lemma 2. *There exists a non-increasing sequence* $\{\delta_n\}_{n=1}^{\infty}$ *of real numbers such that* $\delta_n \to 0$ *and almost surely* $\|W_n^{1/2} - W^{1/2}\| < \delta_n$ *for all n sufficiently large.*

Lemma 3. *There exists a constant* K_1 *such that almost surely* $\|\hat{\theta}_n\| \le K_1$ *for all n sufficiently large.*

Lemma 4. *Let* K_1 *be the constant from Lemma 3 and* $\{\delta_n\}_{n=1}^{\infty}$ *be the sequence from Lemma 2. Put* $B_n := \{\theta; \|\theta\| < K_1, \theta = (v_1/n^4, \dots, v_d/n^4)^\top, v_j \text{ are integers}\}$ *and* $\mathcal{U}_n := \{U; U \text{ is a symmetric positive-definite matrix}, \|U - W^{1/2}\| \le \delta_n\}$. *Let* $\tilde{\mathcal{U}}_n$ *denote the set of all symmetric matrices* $V = (v_{ij}/n^4)_{i,j=1}^{d}$ *such that* v_{ij} *are integers and there is a matrix* $U = U(V) \in \mathcal{U}_n$ *for which* $\|V - U\| \le d/n^4$.

(I) *For every* $U \in \mathcal{U}_n$ *there exists* $V \in \tilde{\mathcal{U}}_n$ *such that* $\|V - U\| \le d/n^4$. *Moreover, for all n large enough if* $V \in \tilde{\mathcal{U}}_n$ *then V is positive-definite and*

$$\|V\| \le \|W^{1/2}\| + 1, \quad \|V^{-1}\| \le \|W^{-1/2}\| + 1, \quad \det(V) \le \det(W^{1/2}) + 1,$$
$$\det(V^{-1}) \le \det(W^{-1/2}) + 1, \quad \lambda_d(V) > \lambda_d(W^{1/2})/2,$$

where $\lambda_d(\cdot)$ *denotes the smallest eigenvalue.*

(II) *There exists a constant* $K_2 > 0$ *such that almost surely for all n sufficiently large*

$$\max_{\theta\in B_n}\max_{V\in\tilde{\mathcal{U}}_n}\left\|\frac{1}{n}\sum_{i=1}^{n} U(VX_i - \theta) - \mathrm{E}\big(U(VX - \theta)\big)\right\| \le K_2\left(\frac{\ln n}{n}\right)^{1/2}. \tag{14}$$

(III) *There exists a constant* $K_3 > 0$ *such that almost surely for all n sufficiently large*

$$\max_{\theta\in B_n}\max_{V\in\tilde{\mathcal{U}}_n}\frac{1}{n}\sum_{i=1}^{n}\Psi\left(\|VX_i - \theta\| \le \frac{\max\{1, \|X_i\|\}}{n^2}\right) \le K_3\frac{\ln n}{n}.$$

Lemma 5. *Throughout this lemma assume that V is a symmetric positive-definite* $d \times d$ *matrix. Let* $G(V, y) := \int_{\mathbb{R}^d} U(Vx + y) f(x)\mathrm{d}x = \int_{\mathbb{R}^d} U(x + y) f_V(x)\mathrm{d}x$ *and* $H(V, y) := \int_{\mathbb{R}^d}(U(x+y) - U(x)) \cdot (U(x+y) - U(x))^\top f_V(x)\mathrm{d}x$, *where* $f_V(\cdot)$ *is the density of the random vector VX.*

(I) *The function* $G(V, \cdot)$ *has continuous partial derivatives* $J_V(y) := \partial G/\partial y^\top = \int_{\mathbb{R}^d} P(x+y) f_V(x)\mathrm{d}x$ *and* $J_V(y)$ *is always positive-definite.*
(II) *If* $y_n \to y$ *and* $V_n \to V$ *then* $J_{V_n}(y_n) \to J_V(y)$.
(III) $y = 0$ *is the unique root of the equation* $G(V, y) = 0$.

Assume further, that a and M are constants,

$$\|V\| \le a, \|V^{-1}\| \le a \tag{15}$$

and $\|y\| \le M$. *Then there exist constants* $g > 0$, m_1, m_2, m_3 *not depending on y or V such that*

(IV) $\|G(V, y)\| \ge g\|y\|$,

(V) $\|G(V, y)\| \le m_1\|y\|$,
(VI) *for* $d \ge 3$: $\|H(V, y)\| \le m_2\|y\|^2$, *for* $d = 2$ *and* $\alpha \in (0, 1)$: $\|H(V, y)\| \le m_2\|y\|^{2\alpha}$,
(VII) *for* $d \ge 3$: $\|J_V(y) - J_V(0)\| \le m_3\|y\|$, *for* $d = 2$ *and* $\alpha \in (0, 1)$: $\|J_V(y) - J_V(0)\| \le m_3\|y\|^{\alpha}$.

Lemma 6. *Let $\mathcal{V}$ be a compact set of symmetric positive-definite matrices. Then there exists a constant $\tilde{M}$ such that $\|J_V(0) - J_W(0)\| \le \tilde{M}\|V - W\|$ for all $V, W \in \mathcal{V}$.*

Lemma 7. *There exists a constant K_4 such that almost surely for all n sufficiently large $\|\hat{\theta}_n(W_n)\| \le K_4\big((\ln n)/n\big)^{1/2}$.*

Lemma 8. *Let $M > 0$. Then*

$$\sup\left\{\|G(V, -\theta) + J_V(0)\theta\|; \|\theta\| \le M\left(\frac{\ln n}{n}\right)^{1/2}, V \text{ satisfies (15)}\right\} = Z_n,$$

where Z_n stands for $O((\ln n)/n)$ for $d \ge 3$. For $d = 2$ it stands for $o(((\ln n)/n)^{(1+\alpha)/2})$ where $\alpha \in (0, 1)$ is arbitrary.

Lemma 9. *Denote $\Gamma_n(V, \theta) := (1/n)\sum_{i=1}^{n}(U(VX_i) - U(VX_i - \theta) + G(V, -\theta))$, and $\tilde{B}_n := \{\theta; \|\theta\| \le 2K_4((\ln n)/n)^{1/2}, \theta = (v_1/n^4, \dots, v_d/n^4)^\top, v_j \text{ are integers}\}$, where the constant K_4 comes from Lemma 7. Then almost surely for all n sufficiently large*

$$\max_{\theta\in\tilde{B}_n, V\in\tilde{\mathcal{U}}_n} \|\Gamma_n(V, \theta)\| \le \mathcal{K}(n),$$

where $\mathcal{K}(n) = K_5(\ln n)/n$ if $d \ge 3$, and $\mathcal{K}(n) = o(((\ln n)/n)^{\alpha})$ if $d = 2$. Here K_5 is a suitably chosen constant and $\alpha \in (0, 1)$ is arbitrary.

Lemma 10. *Let $\{W_n\}_{n=1}^{\infty}$ be the sequence from Assumption 1. Then almost surely*

$$\frac{1}{n}\sum_{i=1}^{n} U(W_n^{1/2}X_i) = J_{W_n^{1/2}}(0) \cdot W_n^{1/2} \cdot \hat{\theta}_n(W_n) + R_n,$$

where $R_n = O((\ln n)/n)$ if $d \ge 3$. If $d = 2$, then $R_n = o(((\ln n)/n)^{\alpha})$ where $\alpha \in (0, 1)$ is arbitrary.

Main part of the proof of Theorem 1. For all non-zero vectors x and symmetric positive-definite matrices W define the vector-valued function $g(W, x) := W^{1/2}{\cdot}U(W^{1/2}x) = Wx/(x^\top Wx)^{1/2}$. Its partial derivatives are

$$\begin{aligned}\frac{\partial g(W, x)}{\partial w_{st}} &= \left(\left(e_s e_t^\top + e_t e_s^\top\right)W^{-1} - \frac{x_s x_t}{x^\top W x}\right)g(W, x) \text{ for } s < t,\\ &= \left(e_s e_s{}^\top W^{-1} - \frac{x_s^2}{2x^\top W x}\right)g(W, x) \text{ for } s = t,\end{aligned}$$

where $e_s (s = 1, \dots, d)$ are the unit vectors with 1 in the sth coordinate. Note that $x_s x_t/(x^\top Wx) = (W^{-1/2}U(W^{1/2}x)U^\top(W^{1/2}x)W^{-1/2})_{st}$ for $s, t = 1, \dots, d$ which means that $\partial g(W, x)/\partial w_{st} =: \zeta(W, U(W^{1/2}x))$ is a function of W and $U(W^{1/2}x)$ and its norm can be bounded by a constant not depending on x because $\|U(W^{1/2}x)\| = 1$. Since Assumption 3 implies that $U(W^{1/2}X)$ has the same distribution as $U(-W^{1/2}X)$, we obtain that $\partial g(W, X)/\partial w_{st} = \zeta(W, U(W^{1/2}X))$ has the same distribution as $\zeta(W, U(-W^{1/2}X)) = -\partial g(W, X)/\partial w_{st}$, and therefore $\mathrm{E}(\partial g(W, X)/\partial w_{st}) = 0$. Hence by the central limit theorem

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(W, X_i)}{\partial w_{st}} = O_P(n^{-1/2}). \tag{16}$$

After computing the formula for the second partial derivative $\partial^2 g(W^*, x)/(\partial w_{st}\partial w_{uv})$ one easily finds that there exist constants a and K such that

$$\left\|\frac{\partial^2 g(W^*, x)}{\partial w_{st}\partial w_{uv}}\right\| \le K \tag{17}$$

for all $x \neq 0$ and W^* such that W^* satisfies (15). Consider the second-order Taylor series expansion

$$\frac{1}{n}\sum_{i=1}^{n} g(W_n, X_i) = \frac{1}{n}\sum_{i=1}^{n} g(W, X_i) + \sum_{\substack{s,t=1,\ldots,d\\ s\le t}} \frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(W, X_i)}{\partial w_{st}}(W_n(s,t) - W(s,t)) + R_n, \tag{18}$$

where

$$R_n = \sum_{\substack{s,t,u,v=1,\ldots,d\\ s\le t;u\le v}} \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 g(W+\beta(W_n - W), X_i)}{\partial w_{st}\partial w_{uv}}\cdot(W_n(s,t) - W(s,t))(W_n(u,v) - W(u,v)),$$

$\beta \in (0, 1)$ and $W(s, t)$ denotes the (s, t)th element of W. According to Assumption 1: $W_n - W = O_P(n^{-1/2})$. This together with (16) and (17) applied to (18) gives that

$$\frac{1}{n}\sum_{i=1}^{n} g(W_n, X_i) = \frac{1}{n}\sum_{i=1}^{n} g(W, X_i) + o_P(n^{-1/2}). \tag{19}$$

Further, (5) and theorem 3.2 on p. 901 of Chaudhuri (1992) applied to the transformed data $W^{1/2}X_1, \ldots, W^{1/2}X_n$ ensure that

$$W^{1/2}\cdot\hat{\theta}_n(W) = \hat{\theta}_n^W = J_{W^{1/2}}^{-1}(0)\cdot W^{-1/2}\cdot\frac{1}{n}\sum_{i=1}^{n} g(W, X_i) + o_P(n^{-1/2}). \tag{20}$$

Now, Lemma 10, (19) and (20) imply that almost surely

$$\begin{aligned} W_n^{1/2}\cdot J_{W_n^{1/2}}(0)\cdot W_n^{1/2}\cdot\hat{\theta}_n(W_n) &= \frac{1}{n}\sum_{i=1}^{n} g(W_n, X_i) - W_n^{1/2}R_n \\ &= \frac{1}{n}\sum_{i=1}^{n} g(W, X_i) + o_P(n^{-1/2}) \\ &= W^{1/2}\cdot J_{W^{1/2}}(0)\cdot W^{1/2}\cdot\hat{\theta}_n(W) + o_P(n^{-1/2}) \end{aligned}$$

because $W_n^{1/2}R_n = o_P(n^{-1/2})$. This means that

$$\begin{aligned} \hat{\theta}_n(W_n) = {}& \left[\left(W_n^{1/2}\cdot J_{W_n^{1/2}}(0)\cdot W_n^{1/2}\right)^{-1}\cdot W^{1/2}\cdot J_{W^{1/2}}(0)\cdot W^{1/2} - I_d\right]\cdot\hat{\theta}_n(W) \\ & + \hat{\theta}_n(W) + o_P(n^{-1/2}). \end{aligned}$$

Note that the expression in square brackets is $o_P(1)$ and $\hat{\theta}_n(W)$ is $O_P(n^{-1/2})$ because of (8). The rest of the proof is obvious. □

Proof of Theorem 2

Lemma 11. *If the dimension* $d \geq 3$, *then* $(1/n)\sum_{i=1}^{n} 1/\|X_i - \hat{\theta}_n(W_n)\|^2 = O_P(1)$.

Proof of Lemma 11. Let the random events

$$E_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\|X_i - \hat{\theta}_n(W_n)\|^2} > M \right\}, \quad F_n = \left\{ \|\hat{\theta}_n(W_n)\| > K_4 \left(\frac{\ln n}{n} \right)^{1/2} \right\}.$$

Then

$$\Pr(E_n) \leq \Pr(E_n \cap F_n^C) + \Pr(F_n) = \Pr(E_n \cap F_n^C) + o(1), \tag{21}$$

where the last equality follows from Lemma 7. Further,

$$\Pr\bigl(E_n \cap F_n^C\bigr) \leq \Pr\bigl(E_n \cap F_n^C \cap G_n^C\bigr) + \Pr(G_n), \tag{22}$$

where G_n denotes the random event $\min\{\|X_1\|, \|X_2\|, \ldots, \|X_n\|\} \leq 2K_4((\ln n)/n)^{1/2}$. Note that for $d \geq 3$

$$\Pr(G_n) \leq \sum_{i=1}^{n} \Pr\left(\|X_i\| \leq 2K_4 \left(\frac{\ln n}{n} \right)^{1/2} \right) \leq n M_{10} M_{11} \left(2K_4 \left(\frac{\ln n}{n} \right)^{1/2} \right)^d \overset{n\to\infty}{\longrightarrow} 0, \tag{23}$$

where the constant M_{10} bounds the density of X in the ball with centre 0 and radius K_4, cf. Assumption 2. The constant M_{11} comes from the formula for the volume of a d-dimensional ball. On the set $F_n^C \cap G_n^C$, however, the inequality $\|X_i - \hat{\theta}_n(W_n)\| > K_4((\ln n)/n)^{1/2}$ holds for $i = 1, \ldots, n$ and therefore

$$\left| \frac{1}{\|X_i - \hat{\theta}_n(W_n)\|} - \frac{1}{\|X_i\|} \right| \leq \frac{\|\hat{\theta}_n(W_n)\|}{\|X_i - \hat{\theta}_n(W_n)\| \cdot \|X_i\|} < \frac{1}{\|X_i\|}.$$

This means that

$$\Pr\left(E_n \cap F_n^C \cap G_n^C\right) \leq \Pr\left(\frac{1}{n} \sum_{i=1}^{n} \frac{4}{\|X_i\|^2} > M \right). \tag{24}$$

Since $\mathrm{E}(1/\|X\|^2) < \infty$ because $d \geq 3$, employing the law of large numbers and taking into account (21)–(24) one obtains that the lemma is true. □

Main part of the proof of Theorem 2. In view of Assumption 1 it is sufficient to show that $\hat{D}_2 = D_2 + o_P(1)$ and $\hat{D}_1 = D_1 + o_P(1)$.

Define $D_2^* := (1/n)\sum_{i=1}^{n} U(W^{1/2}X_i - W^{1/2}\hat{\theta}_n(W_n)) \cdot U^\top(W^{1/2}X_i - W^{1/2}\hat{\theta}_n(W_n))$. It is easy to derive by Lemma 1 (I) that $\|U(x)U^\top(x) - U(y)U^\top(y)\| \leq 4\|x - y\|/\|y\|$. This and Assumption 1 imply that

$$\|\hat{D}_2 - D_2^*\| \leq 4\frac{\|W_n^{1/2} - W^{1/2}\|}{\lambda_d(W^{1/2})} = o_P(1).$$

Similarly, for $D_2^{**} := (1/n)\sum_{i=1}^{n} U(W^{1/2}X_i)\cdot U^{\top}(W^{1/2}X_i)$ the inequality

$$\|D_2^* - D_2^{**}\| \le 4\|W^{1/2}\| \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|W^{1/2}X_i\|}\right)\|\hat\theta_n(W_n)\|$$

holds. But Assumption 2 and the law of large numbers imply that $\mathrm{E}(1/\|X\|) < \infty$ and $(1/n)\sum_{i=1}^{n} 1/\|X_i\| = O_P(1)$, which together with (9) ensures that $\|D_2^* - D_2{}^{**}\| = o_P(1)$. Finally, note that by the law of large numbers $D_2^{**} = D_2 + o_P(1)$. Hence, we conclude that $\hat D_2 = D_2 + o_P(1)$.

Now, put $D_1^* := (1/n)\sum_{i=1}^{n} P(W_n^{1/2}X_i)$ and $D_1^{**} := (1/n)\sum_{i=1}^{n} P(W^{1/2}X_i)$. Similar arguments to above and Lemma 1 (III) ensure that almost surely for all n large enough

$$\begin{aligned}\|D_1^* - D_1^{**}\| &\le (d^{1/2}+5)\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{\|W_n^{1/2} - W^{1/2}\|\cdot\|X_i\|}{\|W_n^{1/2}X_i\|\cdot\|W^{1/2}X_i\|}\\ &\le \frac{(d^{1/2}+5)\cdot\|W_n^{1/2} - W^{1/2}\|}{\lambda_d^2(W^{1/2})/2}\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|X_i\|} = o_P(1).\end{aligned} \tag{25}$$

Further, by Lemma 1 (II) we obtain that

$$\begin{aligned}\|\hat D_1 - D_1^*\| &\le \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d^{1/2}+1}{\|X_i - \hat\theta_n(W_n)\|^2} + \frac{d^{1/2}+5}{\|X_i\|^2}\right)\cdot\frac{1}{\lambda_d^2(W_n^{1/2})}\cdot\|W_n^{1/2}\|\cdot\|\hat\theta_n(W_n)\|\\ &\le \left(\frac{1}{n}\sum_{i=1}^{n}\frac{d^{1/2}+1}{\|X_i - \hat\theta_n(W_n)\|^2} + \frac{1}{n}\sum_{i=1}^{n}\frac{d^{1/2}+5}{\|X_i\|^2}\right)\\ &\quad\cdot\frac{1}{\lambda_d^2(W^{1/2})/2}\cdot(\|W^{1/2}\|+1)\cdot\|\hat\theta_n(W_n)\|,\end{aligned} \tag{26}$$

where the second inequality holds almost surely for all n large enough. $\|\hat\theta_n(W_n)\|$ is $o_P(1)$ by (9) and the second sum is $O_P(1)$, cf. proof of Lemma 11. Finally, let us apply Lemma 11 to the first sum so that we obtain that $\|\hat D_1 - D_1^*\| = o_P(1)$. Moreover, the law of large numbers ensures that $D_1^{**} = D_1 + o_P(1)$. Hence, with (25) in mind we obtain that $\hat D_1 = D_1 + o_P(1)$ and the proof is complete. □

Remark on the proof of Theorem 3. It is natural to ask why the rather simple proof of Theorem 2 cannot be applied also in the case of $d = 2$. Problems arise when proving $\|\hat D_1 - D_1^*\| = o_P(1)$, because the expansion (26) is useless. Note that one cannot rely on $\mathrm{E}(1/\|X\|^2) < \infty$ any more, because it may not hold under the weak Assumption 2 when $d = 2$. Therefore, the frequently used $(1/n)\sum_{i=1}^{n} 1/\|X_i\|^2 = O_P(1)$ becomes questionable. This ruins also the proof of Lemma 11. Attempts to prove at least $(1/n)\sum_{i=1}^{n} 1/\|X_i - \hat\theta_n(W_n)\| = O_P(1)$, a statement similar to Lemma 11, reveal that also the final convergence in (23) depends strongly on the condition $d \ge 3$. This is given by the fact that the convergence to zero of the radius $K_4((\ln n)/n)^{1/2}$ of the ball, which according to Lemma 7 encapsulates $\hat\theta_n(W_n)$, is insufficiently slow for $d = 2$.

The above complications forced us to utilize the methodology of Bai *et al.* (1990), developed to prove the consistency of the asymptotic covariance matrix estimate of the

L_1-estimator in general linear regression models. However, this approach required us to establish some asymptotic properties of $\hat{\theta}_{nk}(W_{nk})$, which is defined in the same way as $\hat{\theta}_n(W_n)$, but with the kth observation omitted. For this reason it was necessary to impose additional Assumptions 4 and 5. See Somorčík (2009) for a detailed proof of Theorem 3.

Proof of Theorem 4. If W is a scalar multiple of the identity matrix, then the estimate $\hat{\theta}_n(W)$ does not depend on W and $\hat{\theta}_n(W) = \hat{\mu}$. Theorem 1 completes the proof. □

Proofs of Theorems 5 and 6. See the Supplementary Material published online at the ANZJS website.

Supporting Information

Additional Supporting Information may be found in the online version of this article at http://wileyonlinelibrary.com/journal/anzs

Data S1. Supplementary materials

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

BAI, Z.D., CHEN, X.R., MIAO, B.Q. & RAO, C.R. (1990). Asymptotic theory of least distances estimate in multivariate linear models. *Statistics* **21**, 503–519.

BOSE, A. (1995). Estimating the asymptotic dispersion of the L1 median. *Ann. Inst. Statist. Math.* **47**, 267–271.

BOSE, A. & CHAUDHURI, P. (1993). On the dispersion of multivariate median. *Ann. Inst. Statist. Math.* **45**, 541–550.

CHAKRABORTY, B., CHAUDHURI, P. & OJA, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica* **8**, 767–784.

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

DÜMBGEN, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50**, 471–491.

DÜMBGEN, L. & TYLER, D. (2005). On the breakdown properties of some multivariate M-functionals. *Scand. J. Statist.* **32**, 247–264.

HETTMANSPERGER, T.P., MÖTTÖNEN, J. & OJA, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica* **8**, 785–800.

HETTMANSPERGER, T.P. & RANDLES, R.H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89**, 851–860.

LOPUHAÄ, H.P. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Annals Statist.* **20**, 398–413.

MILASEVIC, P. & DUCHARME, G.R. (1987). Uniqueness of the spatial median. *Annals Statist.* **15**, 1332–1333.

MÖTTÖNEN, J., OJA, H. & TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Annals Statist.* **25**, 542–552.

NEVALAINEN, J., LAROCQUE, D. & OJA, H. (2007). On the multivariate spatial median for clustered data. *Canad. J. Statist.* **35**, 215–231.

NORDHAUSEN, K., MÖTTÖNEN, J. & OJA, H. (2009). MNM: Multivariate Nonparametric Methods. An approach based on spatial signs and ranks [Computer software manual]. Available from http://CRAN.R-project.org/package=MNM (R package version 0.95-1)

NORDHAUSEN, K., SIRKIA, S., OJA, H. & TYLER, D.E. (2010). ICSNP: Tools for multivariate nonparametrics [Computer software manual]. Available from http://CRAN.R-project.org/package=ICSNP (R package version 1.0-7)

OJA, H. (2010). *Multivariate Nonparametric Methods with R*. New York: Springer.

OJA, H. & NIINIMAA, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **47**, 372–377.

RAO, C.R. (1988). Methodology based on the L_1-norm, in statistical inference. *Sankhya Series A* **50**, 289–313.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. London: Chapman & Hall.

SOMORČÍK, J. (2006). Tests using spatial median. *Austrian Journal of Statistics* **35**, 331–338.

SOMORČÍK, J. (2007). Performance of some spatial median tests under elliptical symmetry. *Measurement Science Review* **7**, 43–50.

SOMORČÍK, J. (2009). Inference based on spatial median (PhD Thesis). Comenius Univiversity, Bratislava.

TYLER, D. (1987). A distribution-free M-estimator of multivariate scatter. *Annals Statist.* **15**, 234–251.

UM, Y. & RANDLES, R.H. (1998). Nonparametric tests for the multivariate multisample location problem. *Statist. Sinica* **8**, 801–812.

VARDI, Y. & ZHANG, C.H. (2000). The multivariate L1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97**, 1423–1426.

WEISZFELD, E. (1937). Sur le point pour lequel la somme des distances de n points donne's est minimum. *Tohoku Math. J. (2)* **43**, 355–386.

Supplementary material

Proof of Lemma 1

The inequality (I) can be found in Chaudhuri (1992) and (III) is just a simple corollary. Finally, making use of (I) and an other inequality $\big|1/\|x+y\| - 1/\|x\|\big| \leq (1/\|x+y\|^2 + 1/\|x\|^2)\|y\|$ from Chaudhuri (1992), one obtains (II).

Proof of Lemma 2

An application of the Egoroff theorem yields that there exist positive integers n_r^* and sets A_r, $r = 1, 2, \ldots$ such that $1 - \Pr(A_r) < 1/2^r$ and for all $n \geq n_r^*$ the inequality $\|W_n^{1/2} - W^{1/2}\| < 1/2^r$ holds on A_r. Let $B_m = \cap_{r=m}^{\infty} A_r$. Then $1 - \Pr(B_m) \leq \sum_{r=m}^{\infty} 1/2^r \overset{m\to\infty}{\longrightarrow} 0$ and the probability of the set $B = \cup_{m=1}^{\infty} B_m$ equals 1. Put $n_r = \sum_{j=1}^{r} n_j^*$ and $\delta_n = 1$ or $1/2^r$, if $n < n_1$ or $n_r \leq n < n_{r+1}$, respectively. Since $\{B_m\}_{m=1}^{\infty}$ is an increasing sequence of sets and on B_m the inequality $\|W_n^{1/2} - W^{1/2}\| < \delta_n$ holds if $n \geq n_m$, the lemma is proved. $\square$

Proof of Lemma 3

The lemma is almost the same as Lemma 5.2 on p. 906 in Chaudhuri (1992). The only distinction is that instead of the random sample $X_1, \ldots, X_n$ our data points are $W_n^{1/2}X_1, \ldots, W_n^{1/2}X_n$, where W_n is the random matrix. It is shown on p. 907 in Chaudhuri (1992) that for K_1 large enough almost surely for all n sufficiently large

$$\frac{1}{n}\sum_{i=1}^{n} \Psi\left(\|X_i\| > \frac{K_1}{4}\right) \leq 2\delta, \tag{S.1}$$

where $\Psi(\cdot)$ is the indicator function of the set. Further, Assumption 1 ensures that almost surely for all n large enough $\|W_n^{1/2} - W^{1/2}\| \leq 1$, which means that

$$\|W_n^{1/2}X_i\| \leq \|W_n^{1/2}\| \cdot \|X_i\| \leq (\|W_n^{1/2} - W^{1/2}\| + \|W^{1/2}\|) \cdot \|X_i\| \leq (1 + \|W^{1/2}\|) \cdot \|X_i\|.$$

Hence $\|W_n^{1/2}X_i\| > K_1'/4$ implies that $\|X_i\| > K_1/4$, where $K_1' := (1 + \|W^{1/2}\|) \cdot K_1$, and by (S.1) almost surely $(1/n)\sum_{i=1}^n \Psi(\|W_n^{1/2}X_i\| > K_1'/4) \leq 2\delta$ for all n large enough. Finally, making use of this inequality one can prove similarly as in Chaudhuri (1992) that if $\|\theta\| > K_1'$ then $(1/n)\sum_{i=1}^n(\|W_n^{1/2}X_i - \theta\| - \|W_n^{1/2}X_i\|) > 0$, which completes the proof. □

Proof of Lemma 4

(I) Let $U = \{u_{ij}\}_{i,j=1}^d \in \mathcal{U}_n$ and $v_{ij} = \text{sign}(u_{ij}) \cdot \max\{v; v \text{ is integer}, v \leq |u_{ij}|n^4\}$. Then $|v_{ij}/n^4 - u_{ij}| \leq 1/n^4$, the matrix $V = (v_{ij}/n^4)_{i,j=1}^d$ belongs to $\tilde{\mathcal{U}}_n$ and $\|V - U\| \leq d/n^4$. Further, for all $V \in \tilde{\mathcal{U}}_n$ and corresponding $U \in \mathcal{U}_n$

$$\|V - W^{1/2}\| \leq \|V - U\| + \|U - W^{1/2}\| \leq \frac{d}{n^4} + \delta_n \to 0,$$

and since the matrix $W^{1/2}$ is positive-definite, for all n large enough also V is positive-definite. The remaining statements are immediate consequences of continuity of the functions $\|\cdot\|$, $\det(\cdot)$ and $\lambda_d(\cdot)$.

(II) By the definition of $\mathcal{U}_n, \tilde{\mathcal{U}}_n$ there exists $\tilde{K}_1$ such that $\|V\| \leq \tilde{K}_1$ whenever $V \in \tilde{\mathcal{U}}_n$. For such a $V = (v_{ij}/n^4)_{i,j=1}^d$ the only possible values of v_{ij} are $0, \pm 1, \pm 2, \ldots, \pm\lfloor \tilde{K}_1 n^4 \rfloor$ and $\tilde{\mathcal{U}}_n$ has at most

$$\#\tilde{\mathcal{U}}_n \leq (2\tilde{K}_1 n^4 + 1)^{d^2} \leq (2\tilde{K}_1 + 1)^{d^2} n^{4d^2} \tag{S.2}$$

elements. Similarly it can be shown that

$$\#B_n \leq (2K_1 + 1)^d n^{4d}. \tag{S.3}$$

Let F_n be the event defined in (14) and F_n^C denote the complementary event. An application of the inequalities $\sqrt{a_1^2 + \cdots + a_d^2} \leq |a_1| + \cdots + |a_d|$, $\max_x \sum_j a_j(x) \leq \sum_j \max_x a_j(x)$

and sub-additivity of probability ensure that

$$\Pr(F_n^C) \le \sum_{j=1}^{d} \sum_{\theta \in B_n} \sum_{V \in \tilde{\mathcal{U}}_n} \Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} U_j(VX_i - \theta) - \mathrm{E}(U_j(VX - \theta)) \right| > \frac{K_2}{d} \sqrt{\frac{\ln n}{n}} \right), \tag{S.4}$$

where $U_j(x)$ denotes the j-th coordinate of $U(x)$. Since the function $U_j(\cdot)$ is bounded by -1 and 1, the Lemma on p. 75 in Serfling (1980) implies

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} U_j(VX_i - \theta) - \mathrm{E}(U_j(VX - \theta)) \right| > \frac{K_2}{d} \sqrt{\frac{\ln n}{n}} \right) \le 2 \exp\left(\frac{-2n \left(\frac{K_2}{d} \sqrt{\frac{\ln n}{n}} \right)^2}{(1 - (-1))^2} \right). \tag{S.5}$$

Combining (S.2)–(S.5) we get that for K_2 sufficiently large $\sum_{n=1}^{\infty} \Pr(F_n^C) < \infty$. This together with the Borel-Cantelli lemma completes the proof.

(III) Let $V \in \tilde{\mathcal{U}}_n$. The random vector $Y = VX$ has the density of the form $f_V(y) := f(V^{-1}y)|\det(V^{-1})|$. Using Assumption 2 and Lemma 4 (I) we obtain that

$$\sup_{n \ge n_0} \sup\{f_V(y); \|y\| \le K_1 + 1, V \in \tilde{\mathcal{U}}_n\} = M_1 < +\infty.$$

Let $\theta \in B_n$. Then the set $A_n = \{y; \|y - \theta\| \le 1/n^2\} \subseteq \{y; \|y\| \le K_1 + 1\}$ and for a suitable constant M_2

$$\Pr\left(\|VX - \theta\| \le \frac{1}{n^2} \right) = \int_{A_n} f_V(y) \mathrm{d}y \le M_2 \left(\frac{1}{n^2} \right)^d. \tag{S.6}$$

By Lemma 4 (I) for n large enough $\lambda_d(V) > \lambda_d(W^{1/2})/2 =: K_0 > 0$ and $1/(K_0 n^2) < 1/2$. Then

$$\left\{ x; \|Vx - \theta\| \le \frac{\|x\|}{n^2}, \|x\| \ge 1 \right\} \subseteq \left\{ x; \|x - V^{-1}\theta\| \le \frac{\|x\|}{K_0 n^2}, \|x\| \ge 1 \right\} \subseteq$$
$$\subseteq C_n := \{x; r_1 \le \|x\| \le r_2, \|x\| \ge 1\}, \quad r_1 = \frac{\|V^{-1}\theta\|}{1 + \frac{1}{K_0 n^2}}, r_2 = \frac{\|V^{-1}\theta\|}{1 - \frac{1}{K_0 n^2}}.$$

It can be shown that there exist M_3, M_4 not depending on n such that $r_2 < M_3$ and the Lebesgue measure of C_n is dominated by M_4/n^2. By Assumption 2 there exists a constant M_5 such that $\sup_{\|x\|\le M_3} f(x) = M_5$. Combining these facts we obtain

$$\Pr\left(\|VX-\theta\| \le \frac{\|X\|}{n^2}, \|X\| \ge 1\right) \le \Pr(C_n) \le M_5 \frac{M_4}{n^2}. \tag{S.7}$$

The inequalities (S.6) and (S.7) enable us to estimate from above

$$\begin{aligned}\mathrm{E}(\Psi) &:= \mathrm{E}\left(\Psi\left(\|VX-\theta\| \le \frac{\max\{1,\|X\|\}}{n^2}\right)\right) \le \\ &\le \Pr\left(\|VX-\theta\| \le \frac{1}{n^2}\right) + \Pr\left(\|VX-\theta\| \le \frac{\|X\|}{n^2}, \|X\| \ge 1\right) \le \frac{M_6}{n^2}\end{aligned}$$

for a suitable constant M_6. Then also $Var(\Psi) \le E(\Psi^2) = E(\Psi) \le M_6/n^2$. Put

$$F_n^C := \left\{\omega; \max_{\theta\in B_n} \max_{V\in\tilde{\mathcal{U}}_n} \frac{1}{n}\sum_{i=1}^n \Psi_i > K_3 \frac{\ln n}{n}\right\},$$

where $\Psi_i := \Psi(\|VX_i-\theta\| \le \max\{1,\|X_i\|\}/n^2)$. Then for all n sufficiently large

$$\begin{aligned}\Pr(F_n^C) &\le \Pr\left(\max_{\theta\in B_n} \max_{V\in\tilde{\mathcal{U}}_n} \left|\frac{1}{n}\sum_{i=1}^n \Psi_i - \mathrm{E}(\Psi)\right| > K_3\frac{\ln n}{n} - \frac{M_6}{n^2}\right) \le \\ &\le \Pr\left(\max_{\theta\in B_n} \max_{V\in\tilde{\mathcal{U}}_n} \left|\frac{1}{n}\sum_{i=1}^n \Psi_i - \mathrm{E}(\Psi)\right| > \frac{K_3}{2}\cdot\frac{\ln n}{n}\right)\end{aligned}$$

Now, proceeding similarly as in the proof of part (II) of this lemma and using the Bernstein inequality on p. 95 in Serfling (1980) we obtain that for K_3 large enough $\Pr(F_n^C) \le M_7/n^2$ for a suitable constant M_7 and all n large enough. Then the proof proceeds as the proof of part (II). □

Proof of Lemma 5

(I) & (III) Note that for any given regular matrix V the density $f_V(\cdot)$ satisfies Assumption 2. Assumption 3 ensures that the distribution of VX is also directionally symmetric and according to Chaudhuri (1992) the spatial median of the distribution of the vector VX is 0, irrespective of V. Therefore (I) and (III) are now consequences of Lemma 5.3 of Chaudhuri (1992).

(II) It can be shown that

$$\|P(V_n x + y_n)\|^{3/2} \leq \left(\frac{\sqrt{d}+1}{\lambda_d(V_n)}\right)^{3/2} \frac{1}{\|x + V_n^{-1} y_n\|^{3/2}}. \tag{S.8}$$

For n large enough $1/\lambda_d(V_n) \leq 2/\lambda_d(V)$ and as $V_n^{-1} y_n$ is bounded,

$$\sup_n \int_{\mathbb{R}^d} \frac{1}{\|x + V_n^{-1} y_n\|^{3/2}} f(x) \mathrm{d}x = \sup_n \int_{\mathbb{R}^d} \frac{1}{\|x\|^{3/2}} f(x - V_n^{-1} y_n) \mathrm{d}x < \infty.$$

This together with (S.8) means that $\{P(V_n X + y_n)\}_{n=1}^{\infty}$ is uniformly integrable. Since $J_{V_n}(y_n) = \int_{\mathbb{R}^d} P(V_n x + y_n) f(x) \mathrm{d}x$, an application of Vitali convergence theorem yields (II).

(IV) Fix $\epsilon > 0$. Let $Q > 0$ be a constant. If $\|z\| > M + Q$, then $\|z + y\| > Q$ and

$$\|P(z+y) - P(z)\| \leq \left(\frac{1}{Q} + \frac{1}{Q+M}\right) \cdot (\sqrt{d}+1).$$

This means that for Q large enough

$$\int_{\|z\| > M+Q} \|P(z+y) - P(z)\| f_V(z) \mathrm{d}z < \frac{\epsilon}{2} \tag{S.9}$$

for all regular V and $\|y\| \leq M$. Let $\{y_n\}_{n=1}^{\infty}$ be an arbitrary sequence such that $\|y_n\| \leq M$ and $y_n \to 0$. Since the inequalities

$$\sup\{f_V(z); \|z\| \leq M + Q, V \text{ satisfies (15)}\} =: C < \infty,$$
$$\|P(z+y_n) - P(z)\|^{3/2} \leq (\sqrt{d}+1)^{3/2} \cdot \left(\frac{1}{\|z+y_n\|} + \frac{1}{\|z\|}\right)^{3/2}$$

hold, uniform integrability and the Vitali convergence theorem can be utilized in a similar way as in the proof of part (II) to show that

$$\int_{\|z\| \leq M+Q} \|P(z+y_n) - P(z)\| f_V(z) \mathrm{d}z \leq \int_{\|z\| \leq M+Q} \|P(z+y_n) - P(z)\| C \mathrm{d}z \to 0$$

as $n \to \infty$. Combining this result with (S.9) one obtains that there exists $\delta > 0$ such that if $\|y\| < \delta$ then

$$\|J_V(y) - J_V(0)\| < \epsilon, \tag{S.10}$$

irrespective of V satisfying (15). Let G_j and ${}_jJ_V$ denote the j-th coordinate and j-th row of G and J respectively. Using the mean value theorem one obtains that

$$G_j(V, y) = G_j(V, y) - G_j(V, 0) = {}_jJ_V(\xi_j y)y$$

for some $\xi_j \in (0, 1)$. Applying Euclidean norm we obtain

$$\|G(V, y)\| \geq \|J_V(0)y\| - \sum_{j=1}^{d} \|{}_jJ_V(\xi_j y) - {}_jJ_V(0)\| \cdot \|y\|. \tag{S.11}$$

Let $g_1 := \inf\{\lambda; \lambda \text{ is the smallest eigenvalue of } J_V(0), V \text{ satisfies (15)}\}$. Since part (II) of this lemma holds, g_1 is a positive number and in accordance with (S.10) there exists $\delta > 0$ such that $\|J_V(y) - J_V(0)\| < g_1/(2d)$ for all $\|y\| \leq \delta$ and V satisfying (15). This together with (S.11) means that $\|G(V, y)\| \geq (g_1/2)\|y\|$ for all $\|y\| \leq \delta$. If $M \leq \delta$ it is enough to put $g := g_1/2$. If not, the function $G(V, \cdot)/\| \cdot \|$ is continuous and positive (cf. (III)) on the compact set $\{y; \delta \leq y \leq M\}$, and obviously is here bounded from below by a $g_2 > 0$ for all V satisfying (15). Thus putting $g := \min\{g_1/2, g_2\}$ one obtains that (IV) holds.

(V) By Lemma 1 (I) we obtain

$$\begin{aligned} \|G(V, y)\| &= \|G(V, y) - G(V, 0)\| \leq 2\|y\| \int_{\mathbb{R}^d} \frac{1}{\|x\|} f_V(x) \mathrm{d}x \leq \\ &\leq 2\|y\| \left(\int_{\|x\| \leq 1} \frac{1}{\|x\|} f_V(x) \mathrm{d}x + \int_{\|x\| > 1} f_V(x) \mathrm{d}x \right). \end{aligned}$$

The second integral is bounded by 1. Since $\sup\{f_V(x); \|x\| \leq 1, V \text{ satisfies (15)}\}$ is a real number, the existence of m_1 is proved.

(VI) If $d \geq 3$, it is easy to derive by means of Lemma 1 (I) that $\|H(V, y)\| \leq 4\|y\|^2 \int_{\mathbb{R}^d} (1/\|x\|^2) f_V(x) \mathrm{d}x$. Now the proof continues similarly as in part (V).

Let $d = 2$. We shall use the inequality

$$\frac{\|y\|^\beta}{\|x\| \cdot \|x+y\|} \le 2^\beta (\|x\|^{\beta-2} + \|x+y\|^{\beta-2}) \tag{S.12}$$

from Chaudhuri (1992) which is true for $\beta \in (0,1)$. Employing Lemma 1 (I) and (S.12) we obtain that for $\alpha \in (1/2, 1)$ and $\beta := 2 - 2\alpha \in (0,1)$

$$\begin{aligned}\|y\|^{\beta-2} \cdot \|H(V,y)\| &\le 4 \int_{\mathbb{R}^d} \frac{\|y\|^\beta}{\|x\| \cdot \|x+y\|} f_V(x)\mathrm{d}x \le \\ &\le 4 \cdot 2^\beta \left(\int_{\mathbb{R}^d} \frac{1}{\|x\|^{2-\beta}} f_V(x)\mathrm{d}x + \int_{\mathbb{R}^d} \frac{1}{\|x+y\|^{2-\beta}} f_V(x)\mathrm{d}x \right).\end{aligned}$$

Note that $2 - \beta \in (1,2)$, hence, the integrals are finite. We can continue similarly as in the proof of part (V) to prove that the first integral is bounded, irrespective of y and V. The second integral can be written in the form $\int_{\mathbb{R}^d} (1/\|x\|^{2-\beta}) f_V(x-y)\mathrm{d}x$ and treated similarly. Hence, we have shown that for $\alpha \in (1/2, 1)$ there exists m_2 such that

$$\|H(V,y)\| \le m_2 \|y\|^{2-\beta} = m_2 \|y\|^{2\alpha}.$$

For $\alpha \in (0, 1/2)$ put $\alpha^* = \alpha + 1/2$ and use the proved inequality.

(VII) For $d \ge 3$ Lemma 1 (II) ensures that

$$\|J_V(y) - J_V(0)\| \le \left(\int_{\mathbb{R}^d} \frac{\sqrt{d}+1}{\|x+y\|^2} f_V(x)\mathrm{d}x + \int_{\mathbb{R}^d} \frac{\sqrt{d}+5}{\|x\|^2} f_V(x)\mathrm{d}x \right) \|y\|.$$

The boundedness of the second integral was treated in the proof of part (VI). The first integral can be written in the form $\int_{\mathbb{R}^d} \left((\sqrt{d}+1)/\|x\|^2\right) f_V(x-y)\mathrm{d}x$ and its boundedness is now obvious.

In the case of $d = 2$ we apply Lemma 1 (III) and (S.12) to obtain that for $\beta := 1 - \alpha \in (0,1)$

$$\begin{aligned}\|y\|^{\beta-1} \cdot \|J_V(y) - J_V(0)\| &\le \int_{\mathbb{R}^d} (\sqrt{d}+5) \frac{\|y\|^\beta}{\|x\| \cdot \|x+y\|} f_V(x)\mathrm{d}x \le \\ &\le (\sqrt{d}+5) 2^\beta \left(\int_{\mathbb{R}^d} \frac{1}{\|x\|^{2-\beta}} f_V(x)\mathrm{d}x + \int_{\mathbb{R}^d} \frac{1}{\|x+y\|^{2-\beta}} f_V(x)\mathrm{d}x \right).\end{aligned}$$

The proof can be completed in the same way as the proof of (VI) for $d = 2$. $\square$

Proof of Lemma 6

Since the set $\mathcal{V}$ is compact and the smallest eigenvalue is a continuous function of its symmetric matrix argument, there exists a positive real number λ such that $\|Ax\| \geq \lambda\|x\|$ for all $A \in \mathcal{V}$. Using this inequality and Lemma 1 (III) one obtains

$$\|P(Vx) - P(Wx)\| \leq \frac{\sqrt{d}+5}{\lambda^2} \cdot \frac{1}{\|x\|} \cdot \|V - W\|.$$

This implies that $\|J_V(0) - J_W(0)\| \leq \left((\sqrt{d}+5)/\lambda^2\right) \int_{\mathbb{R}^d} (1/\|x\|) f(x) \mathrm{d}x \cdot \|V - W\|$ and the proof is complete. $\square$

Proof of Lemma 7

Since $\hat{\theta}_n(W_n) = W_n^{-1/2}\hat{\theta}_n$ and Lemma 2 holds, it is sufficient to prove the lemma for $\hat{\theta}_n$ instead of $\hat{\theta}_n(W_n)$. Let $\theta_n^* \in B_n$ and $V_n^* \in \tilde{\mathcal{U}}_n$ be the elements of B_n and $\tilde{\mathcal{U}}_n$, which are the nearest to $\hat{\theta}_n$ and $W_n^{1/2}$, respectively. Lemma 2 ensures that almost surely for all n large enough $W_n^{1/2} \in \mathcal{U}_n$ which together with Lemma 4 (I) means that

$$\|V_n^* - W_n^{1/2}\| \leq \frac{d}{n^4}. \tag{S.13}$$

Lemma 3 implies that almost surely for all n large enough

$$\|\hat{\theta}_n - \theta_n^*\| \leq \frac{d}{n^4}. \tag{S.14}$$

If $\|V_n^* X_i - \theta_n^*\| \geq \max\{1, \|X_i\|\}/n^2$ then using Lemma 1 (I), (S.13) and (S.14) one obtains

$$\|U(W_n^{1/2}X_i - \hat{\theta}_n) - U(V_n^* X_i - \theta_n^*)\| \leq \frac{4d}{n^2}. \tag{S.15}$$

Denote $\Psi_i^* := \Psi\left(\|V_n^* X_i - \theta_n^*\| \leq \max\{1/n^2, \|X_i\|/n^2\}\right)$. Then

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^n U(V_n^* X_i - \theta_n^*)\right\| &\leq \frac{1}{n}\sum_{i=1}^n \|U(W_n^{1/2} X_i - \hat{\theta}_n) - U(V_n^* X_i - \theta_n^*)\| \cdot \Psi_i^* + \\ &+ \frac{1}{n}\sum_{i=1}^n \|U(W_n^{1/2} X_i - \hat{\theta}_n) - U(V_n^* X_i - \theta_n^*)\| \cdot (1 - \Psi_i^*) \\ &+ \frac{1}{n}\left\|\sum_{i=1}^n U(W_n^{1/2} X_i - \hat{\theta}_n)\right\|.\end{aligned}$$

We apply Lemma 4 (III) to the first sum. The second sum can be bounded from above using (S.15) and the norm of the third one by 1, cf. Fact 5.5 on p. 909 in Chaudhuri (1992). Thus we obtain that almost surely for all n large enough

$$\left\|\frac{1}{n}\sum_{i=1}^n U(V_n^* X_i - \theta_n^*)\right\| \leq 2K_3 \frac{\ln n}{n} + \frac{4d}{n^2} + \frac{1}{n}. \tag{S.16}$$

Similarly as (S.15) it can be shown that if $\|V_n^* X_i - \theta\| \geq \max\{1/n^2, \|X_i\|/n^2\}$ then

$$\|U(W_n^{1/2} X_i - \theta) - U(V_n^* X_i - \theta)\| \leq \frac{2d}{n^2}$$

and analogously as in (S.16)

$$\max_{\theta \in B_n} \left\|\frac{1}{n}\sum_{i=1}^n (U(W_n^{1/2} X_i - \theta) - U(V_n^* X_i - \theta))\right\| \leq 2K_3 \frac{\ln n}{n} + \frac{2d}{n^2} \tag{S.17}$$

almost surely for all n large enough. Taking into account both (S.16) and (S.17) with $\theta_n^* \in B_n$ in place of θ we obtain that almost surely for all n large enough

$$\left\|\frac{1}{n}\sum_{i=1}^n U(W_n^{1/2} X_i - \theta_n^*)\right\| \leq (4K_3 + 6d + 1)\frac{\ln n}{n}. \tag{S.18}$$

Let K_4' be a constant, $\theta \in B_n$ and $\|\theta\| > K_4'\sqrt{(\ln n)/n}$. Obviously,

$$\left\| \frac{1}{n}\sum_{i=1}^{n} U(W_n^{1/2}X_i - \theta) \right\| \geq \|G(V_n^*, -\theta)\| - \left\| \frac{1}{n}\sum_{i=1}^{n} U(V_n^* X_i - \theta) - G(V_n^*, -\theta) \right\| - \\ - \left\| \frac{1}{n}\sum_{i=1}^{n} \left(U(W_n^{1/2}X_i - \theta) - U(V_n^* X_i - \theta) \right) \right\|.$$

Lemma 5 (IV) can be applied to the first term because the existence of the constant a required by this lemma is ensured almost surely for all n large enough by Lemma 4 (I). Lemma 4 (II) may be used for the second term and (S.17) for the third term. We obtain

$$\left\| \frac{1}{n}\sum_{i=1}^{n} U(W_n^{1/2}X_i - \theta) \right\| \geq g\|\theta\| - K_2\sqrt{\frac{\ln n}{n}} - \left(2K_3\frac{\ln n}{n} + \frac{2d}{n^2} \right) \geq \\ \geq (gK_4' - K_2 - 2K_3 - 2d)\sqrt{\frac{\ln n}{n}}$$

almost surely for all n large enough. Hence if K_4' is large enough then $\|\theta_n^*\| \leq K_4'\sqrt{(\ln n)/n}$ by (S.18) and the rest of the proof follows from (S.14). $\square$

Proof of Lemma 8

We apply the mean value theorem to the j-th component of the function $G(V, \cdot)$ and by (III) and (I) of Lemma 5 we obtain

$$(G(V, -\theta) + J_V(0)\theta)_j = G_j(V, 0) + \frac{\partial G_j}{\partial \theta^\top}(\alpha_j \theta) \cdot (-\theta) + {}_j J_V(0)\theta = {}_j \left(-J_V(\alpha_j\theta) + J_V(0) \right)\theta,$$

for a suitable $\alpha_j \in (0,1)$ depending on θ. Now Lemma 5 (VII) can be utilized and one obtains the inequalities

$$Z_n \leq \sup\left\{ m_3\|\theta\|^2; \|\theta\| \leq M\sqrt{\frac{\ln n}{n}} \right\} = O\left(\frac{\ln n}{n} \right),$$
$$Z_n \leq \sup\left\{ m_3\|\theta\|^{\alpha+1}; \|\theta\| \leq M\sqrt{\frac{\ln n}{n}} \right\} = o\left(\left(\frac{\ln n}{n} \right)^{(1+\alpha)/2} \right),$$

where the upper line holds for $d \geq 3$ and the lower one for $d = 2$. $\square$

Proof of Lemma 9

From Lemma 5 (III) we have $G(V,0)=0$. It implies that

$$Var\big(U(VX)-U(VX-\theta)+G(V,-\theta)\big)=H(V,-\theta)-G(V,-\theta)\cdot G^{\top}(V,-\theta).$$

If $V\in\tilde{\mathcal{U}}_n$ then according to Lemma 4 (I) the matrix V satisfies condition (15) for n large enough. Hence, if also $\theta\in\tilde{B}_n$ then Lemma 5 (V), (VI) can be used and one obtains that

$$\max_{j=1,\dots,d} Var\big(U_j(VX)-U_j(VX-\theta)+G_j(V,-\theta)\big)\le\mathcal{K}_0(n),$$

where $\mathcal{K}_0(n)=M_8(\ln n)/n$ if $d\ge 3$ and $\mathcal{K}_0(n)=o\Big(\big((\ln n)/n\big)^{\alpha}\Big)$ if $d=2$, M_8 is a suitably chosen constant and $\alpha\in(0,1)$ is arbitrary. Let

$$F_n^C=\left\{\max_{\theta\in\tilde{B}_n}\max_{V\in\tilde{\mathcal{U}}_n}\|\Gamma_n(V,\theta)\|>\mathcal{K}_1(n)\right\},$$

where $\mathcal{K}_1(n)=K_5(\ln n)/n$ if $d\ge 3$ and $\mathcal{K}_1(n)=K_5\big((\ln n)/n\big)^{\alpha}$ if $d=2$. Following the steps in the proof of Lemma 4 (II), using the Bernstein inequality on p. 95 in Serfling (1980) and choosing K_5 sufficiently large one obtains that for a suitable constant M_9 and all n large enough $\Pr(F_n^C)\le M_9/n^2$. Now the proof can be completed as that of Lemma 4 (II). $\square$

Proof of Lemma 10

Recall $V_n^*\in\tilde{\mathcal{U}}_n$ and $\theta_n^*\in B_n$ defined in the proof of Lemma 7. It was shown therein that almost surely for all n large enough $\hat{\theta}_n\le K_4\sqrt{(\ln n)/n}$. Therefore taking into account (S.14) one obtains that $\theta_n^*\in\tilde{B}_n$ almost surely for all n large enough. Then according to Assumption 1 almost surely for all n large enough $\lambda_d(W_n^{1/2})>\lambda_d(W^{1/2})/2$ which together with Lemma 1 (I) and (S.13) means that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \left(U(W_n^{1/2} X_i) - U(V_n^* X_i) \right) \right\| \leq \frac{1}{n} \sum_{i=1}^{n} 2 \frac{\|(W_n^{1/2} - V_n^*) X_i\|}{\|W_n^{1/2} X_i\|} \leq$$
$$\leq \frac{1}{n} \sum_{i=1}^{n} 2 \frac{\frac{d}{n^4} \|X_i\|}{\frac{1}{2} \lambda_d(W^{1/2}) \|X_i\|} = \frac{4d}{\lambda_d(W^{1/2}) n^4} \quad \text{(S.19)}$$

almost surely for all n large enough. Consider the decomposition

$$\frac{1}{n} \sum_{i=1}^{n} U(V_n^* X_i) = \frac{1}{n} \sum_{i=1}^{n} \left(U(V_n^* X_i) - U(V_n^* X_i - \theta_n^*) + G(V_n^*, -\theta_n^*) \right) +$$
$$+ \frac{1}{n} \sum_{i=1}^{n} U(V_n^* X_i - \theta_n^*) - \left(G(V_n^*, -\theta_n^*) + J_{V_n^*}(0) \theta_n^* \right) + J_{V_n^*}(0) \theta_n^*.$$

Application of Lemma 9, (S.16) and Lemma 8 to the first, second and third term respectively gives that almost surely

$$\begin{gathered} \frac{1}{n} \sum_{i=1}^{n} U(V_n^* X_i) = \delta(n) + J_{V_n^*}(0) \theta_n^*, \\ \text{where } \delta(n) = O\left(\frac{\ln n}{n}\right) \text{ if } d \geq 3, \text{ and } \delta(n) = o\left(\left(\frac{\ln n}{n}\right)^{\alpha}\right) \text{ if } d = 2. \end{gathered} \quad \text{(S.20)}$$

Further, (5) implies the equality

$$J_{V_n^*}(0) \cdot \theta_n^* = J_{V_n^*}(0) \cdot (\theta_n^* - \hat{\theta}_n) + \left(J_{V_n^*}(0) - J_{W_n^{1/2}}(0) \right) \cdot \hat{\theta}_n + J_{W_n^{1/2}}(0) \cdot W_n^{1/2} \cdot \hat{\theta}_n(W_n).$$

Thus Lemma 5 (II), (S.14), Lemma 6, (S.13) and Lemma 3 ensure that almost surely

$$J_{V_n^*}(0) \cdot \theta_n^* = O(n^{-4}) + J_{W_n^{1/2}}(0) \cdot W_n^{1/2} \cdot \hat{\theta}_n(W_n). \quad \text{(S.21)}$$

Combine (S.19), (S.20) and (S.21) and the proof is complete. □

Proof of Theorem 5

(I) By *multiplied samples* we will understand the original q samples with each data point multiplied by $W^{1/2}$. Let the $\hat{\mu}_a^W$'s denote the classical spatial medians of the multiplied samples given by (4) and let $\bar{\mu}^W := (1/n)\sum_{a=1}^q n_a \hat{\mu}_a^W$ be their weighted average. Since according to (5) the equality $\mu_i^W = W^{1/2}\mu_i$ holds, the difference $\hat{\mu}_a^W - \bar{\mu}^W = T_1 + T_2 + T_3$, where $T_1 = \hat{\mu}_a^W - W^{1/2}\mu_a$, $T_2 = W^{1/2}\mu_a - \sum_{b=1}^q (n_b/n)\cdot W^{1/2}\mu_b$, and $T_3 = -\sum_{b=1}^q (n_b/n)\cdot(\hat{\mu}_b^W - W^{1/2}\mu_b)$. T_1 and T_3 are $O_P(n^{-1/2})$ by (7), both under H_0 or the Pitman alternatives. Thus obviously

$$\hat{\mu}_a^W - \bar{\mu}^W = O_P(n^{-1/2}) \tag{S.22}$$

under H_0 or the Pitman alternatives. Then by Theorem 1, (5), Theorem 2 and (S.22) we obtain

$$\begin{aligned} A_1 &= \sum_{a=1}^q n_a \big(W^{-1/2}\hat{\mu}_a^W - W^{-1/2}\bar{\mu}^W + o_P(n^{-1/2})\big)^\top (B^{-1} + o_P(1)) \cdot \\ &\quad \cdot \big(W^{-1/2}\hat{\mu}_a^W - W^{-1/2}\bar{\mu}^W + o_P(n^{-1/2})\big) = \\ &= \sum_{a=1}^q n_a (\hat{\mu}_a^W - \bar{\mu}^W)^\top D_1 D_2^{-1} D_1 (\hat{\mu}_a^W - \bar{\mu}^W) + o_P(1) = \\ &= \sum_{a=1}^q n_a (\hat{\mu}_a^W - \bar{\mu}^W)^\top \big((\hat{V}^W)^{-1} + o_P(1)\big)(\hat{\mu}_a^W - \bar{\mu}^W) + o_P(1) = M_1^W + o_P(1) \end{aligned} \tag{S.23}$$

under H_0 or Pitman alternatives, where $\hat{V}^W = V^W + o_P(1)$ is the estimate of $V^W := D_1^{-1} D_2 D_1^{-1}$ in the case of multiplied samples given in Somorčík (2006) and M_1^W is the test statistics M_1 computed from multiplied samples. Similar reasoning as above, making also use of the fact that $\hat{\mu}_a^W - \hat{\mu}^W = O_P(n^{-1/2})$ under H_0, where $\hat{\mu}^W$ denotes the spatial median of the multiplied pooled sample, implies that

$$A_2 = M_2^W + o_P(1) \tag{S.24}$$

under H_0. Since we know from Theorem 1 of Somorčík (2006) that $M_1^W = M_2^W + o_P(1)$ under H_0 and M_1^W converges to $\chi^2_{(q-1)d}$ in distribution under H_0, the result (I) follows.

(II) Since in accordance with the assumptions the centre of directional symmetry of the a-th multiplied sample is $W^{1/2}\mu + W^{1/2}h_a/\sqrt{n}$, according to Theorem 2 of Somorčík (2006) the asymptotic distribution of M_1^W under Pitman alternatives is noncentral chi-squared $\chi^2_{(q-1)d}(\delta_{M_1^W})$ with noncentrality parameter

$$\delta_{M_1^W} = \sum_{a=1}^{q} p_a (W^{1/2} h_a)^\top (V^W)^{-1} W^{1/2} h_a = \sum_{a=1}^{q} p_a h_a^\top B^{-1} h_a = \delta_{A_1},$$

which together with (S.23) completes the proof.

(III) If $W = c \cdot I_d$ then M_1^W and M_2^W do not depend on W and $M_1^W = M_1$, $M_2^W = M_2$. (S.23) and (S.24) complete the proof. $\square$

Proof of Theorem 6

Let $P_n(A) = \big(2/(n(n+1))\big)\#\{(i,j); 1 \le i < j \le n, X_i - X_j \in A\}$ denote the empirical probability generated by $\{X_i - X_j; 1 \le i < j \le n\}$. The continuity of the distribution implies that we may assume without the loss of generality that for $i = 1, \ldots, n-d$

$$i_1 < i_2 < \ldots < i_d \in \{i+1, i+2, \ldots, n\} \Longrightarrow rank(X_i - X_{i_1}, \ldots, X_i - X_{i_d}) = d. \quad \text{(S.25)}$$

Suppose that $V \subset \mathbb{R}^d$ is a linear space and $1 \le v = \dim(V) < d$. Putting $A = \{X_i - X_j; 1 \le i < j \le n\}$, $A_i = \{X_i - X_j; i < j \le n\}$ one obtains by means of (S.25) that

$$P_n(V) = P_n(V \cap A) = \sum_{i=1}^{n-1} P_n(V \cap A_i) \le \frac{2}{n(n+1)} \Big(\sum_{i=1}^{n-v} v + \sum_{i=1}^{v-1} i \Big) = \frac{\delta(v)}{n(n+1)},$$

where $\delta(v) = 2(n-v)v + (v-1)v$. But

$$\frac{\delta(v)}{n(n+1)} - \frac{v}{d} = \frac{v\big(2d(n-1) - n(n+1)\big)}{n(n+1)d}$$

which together with (13) means that $P_n(V) < v/d$ and the existence result follows from Theorem 2.1 of Dümbgen (1998) and the Theorem of Milasevic & Ducharme (1987). The affine equivariance follows from formula (4.1) of Dümbgen (1998) and from (3) of this paper. $\square$

Supporting Information 2

```
# Below we provide the commented code of the R function "RublikSomorcik(...)" which computes the values
# of our test statistics A_1 and A_2.
# It also computes the p-values of the corresponding tests based on the asymptotic or permutation approach.
#
# In our implementation, W_n is the inverse of be the sample Duembgen matrix of the pooled sample.
# To evaluate sample spatial medians and the sample Duembgen matrix, RublikSomorcik(...) relies on the functions
# spatial.median(...) and duembgen.shape(...) from the R package ICSNP. Therefore, make sure that ICSNP is installed.
#
# The matrix D_2 is estimated. In some cases of symmetry (see the main paper), D_2 can be set to diag(d)/d.
# This requires manual change of the code.
#
# RublikSomorcik(...) was built under R 2.11.1.
#
#
#___________
# ARGUMENTS:
#
# X: Matrix or data.frame with "n_1+...+n_q" rows and "1+d" columns.
#    The first column of X is used to indicate the samples the corresponding rows belong to.
#
# Stat: Which test statistic should be computed? Possible values are "A1" (default) and "A2".
#
# Permute: Use the permutation approach to compute the p-value? Possible values are TRUE and FALSE (default).
#
# N.of.perm: Number of random permutations to obtain the permutation p-value (default is 1000). Used only if Permute=TRUE.
#
#
#___________
# VALUE:
#
# A list is returned containing the following components:
#
# value: Value of the test statistic.
#
# Stat: Name of the evaluated test statistic ("A1", or "A2").
#
# p.value: The corresponding p-value based on the asymptotic (if Permute=FALSE) or permutation (if Permute=TRUE) approach.
#
# Permute: Was the permutation approach used to compute the p-value?
#
# N.of.perm: Number of random permutations to obtain the permutation p-value. NA is returned if Permute=FALSE.
#
#
#___________
# EXAMPLES OF USAGE:
#
#  library(HSAUR)
#  X <- skulls[1:90,] #Egyptian skulls from three epochs
#
# # p-values based on the asymptotic distribution:
#  RublikSomorcik(X)
#  RublikSomorcik(X,Stat="A2")
# # p-values based on the permutation approach:
#  set.seed(1)
#  RublikSomorcik(X,Permute=TRUE,N.of.perm=1000)
#  RublikSomorcik(X,Stat="A2",Permute=TRUE,N.of.perm=1000)




RublikSomorcik <- function(X, Stat="A1", Permute=FALSE, N.of.perm=1000)
{
 # the package ICSNP provides functions to compute the spatial median [function spatial.median(...)]
 # and the Duembgen matrix [function duembgen.shape(...)]:
  require(ICSNP)

 # preparation of the input data X:
 # - the constants n_1,...,n_q [stored in vector n], q, and d are obtained
 # - rows are possibly reordered to collate the rows of each of the "q" samples
  X[,1] <- as.factor(X[,1])
  XX <- NULL
  n <- NULL
  for(X.level in levels(X[,1]))
   if( sum(X[,1]==X.level)>0 )
         {
          n <- c(n,sum(X[,1]==X.level))
          XX <- rbind( XX , X[X[,1]==X.level,-1] )
         }
  X <- XX
  q <- length(n)
  d <- ncol(X)

 # Wn.Sqr.Rt is the square root of W_n and inv.Wn.Sqr.Rt is its inverse:
  Duembgen.eigen <- eigen( duembgen.shape(X) )
  inv.Wn.Sqr.Rt <- Duembgen.eigen$vectors %*% diag( sqrt(Duembgen.eigen$values) ) %*% t(Duembgen.eigen$vectors)
  Wn.Sqr.Rt <- solve(inv.Wn.Sqr.Rt)

 # Y = the original data transformed by the square root of W_n:
  Y <- as.matrix(X) %*% Wn.Sqr.Rt

 # columns of the matrix MEDIANS are the modified spatial medians \hat\eta_1,...,\hat\eta_q
  MEDIANS <- matrix(ncol=q,nrow=d)
  for(j in 1:q) MEDIANS[,j] <- inv.Wn.Sqr.Rt %*% spatial.median( Y[(sum(n[1:j])-n[j]+1):sum(n[1:j]),] )

 # Median.transformed is the classical spatial median of the pooled sample of the transformed data:
```

```
  Median.transformed <- spatial.median(Y)

 # inv.B is the inverse of the sample covariance matrix \hat{B}:
  D1 <- matrix(0,ncol=d,nrow=d)
  D2 <- matrix(0,ncol=d,nrow=d)
  for(i in 1:sum(n))
        {
         Distance <- dist( rbind(Y[i,],Median.transformed) )
         D1 <- D1 + (diag(d) - (Y[i,]-Median.transformed)%*%t(Y[i,]-Median.transformed) / Distance^2 ) / Distance
         D2 <- D2 + (Y[i,]-Median.transformed)%*%t(Y[i,]-Median.transformed) / Distance^2
        }
  D1 <- D1/sum(n)
  D2 <- D2/sum(n)
  inv.B <- Wn.Sqr.Rt %*% D1 %*% solve(D2) %*% D1 %*% Wn.Sqr.Rt

 # location parameter used in the test statistic:
 # A1: weighted average \bar\eta of the the modified spatial medians \hat\eta_1,...,\hat\eta_q
 # A2: modified spatial median \hat\eta of the pooled sample
  Location <- switch(Stat,
                     A1 = MEDIANS %*% n / sum(n),
                     A2 = inv.Wn.Sqr.Rt %*% Median.transformed
                    )

 # final evaluation of the test statistic:
  A <- 0
  for (j in 1:q) A <- A+n[j]*t(MEDIANS[,j]-Location) %*% inv.B %*% (MEDIANS[,j]-Location)
  A <- as.numeric(A)

  if(Permute)
   {#_________beginning of permutations__________

    # save current .Random.seed, if it currently exists:
     if(exists(".Random.seed")) {Existed.Random.seed <- TRUE; old.seed <- .Random.seed} else Existed.Random.seed <- FALSE

    # values of the test statistic based on the N.of.perm permutations of the data will be stored in permA:
     permA <- rep(0,times=N.of.perm)

     for(Counter in 1:N.of.perm)
     {
      # generate random permutation of the rows:
       Permutation <- sample(1:sum(n),size=sum(n),replace=FALSE)

      # columns of the matrix MEDIANS are the modified spatial medians \hat\eta_1,...,\hat\eta_q of the permuted data:
       for(j in 1:q) MEDIANS[,j] <- inv.Wn.Sqr.Rt %*% spatial.median( Y[Permutation[(sum(n[1:j])-n[j]+1):sum(n[1:j])],] )

      # in the case of A1 new location parameter used in the test statistic has to be recomputed,
      # in the case of A2 the \hat\eta does not depend on the permutations
       if(Stat=="A1") Location <- MEDIANS %*% n / sum(n)

      # final evaluation of the test statistic based on the permuted data:
       for (j in 1:q) permA[Counter] <- permA[Counter]+n[j]*t(MEDIANS[,j]-Location) %*% inv.B %*% (MEDIANS[,j]-Location)
     }

    # restore the old .Random.seed [or remove .Ramdom.seed if it previously did not exist]:
     if(Existed.Random.seed) assign(".Random.seed", old.seed, envir=.GlobalEnv) else rm(.Random.seed,envir=.GlobalEnv)

    # the value of the test statistic and the permutation p-value is returned:
     return( list(value=A, Stat=Stat, p.value=sum(permA>A)/N.of.perm, Permute=Permute, N.of.perm=N.of.perm) )

   }#_________end of permutations__________


 # if no permutations, the value of the test statistic and the asymptotic p-value are returned:
  return( list(value=A, Stat=Stat, p.value=1-pchisq(A,df=(q-1)*d), Permute=Permute, N.of.perm=NA))
}
```