

ON A TRANSFORMATION AND RE-TRANSFORMATION  
TECHNIQUE FOR CONSTRUCTING AN AFFINE  
EQUIVARIANT MULTIVARIATE MEDIAN

BIMAN CHAKRABORTY AND PROBAL CHAUDHURI

(Communicated by Wei Y. Loh)

ABSTRACT. An affine equivariant version of multivariate median is introduced. The proposed median is easy to compute and has some appealing geometric features that are related to the configuration of a multivariate data cloud. The transformation and re-transformation approach used in the construction of the median has some fundamental connection with the data driven co-ordinate system considered by Chaudhuri and Sengupta (1993, *Journal of the American Statistical Association*). Large sample statistical properties of the median are discussed and finite sample performance is investigated using Monte Carlo simulations.

1. INTRODUCTION

There are several versions of multivariate median that have been proposed and studied in the existing literature (see e.g. Small [16] and Chaudhuri [8] for detailed reviews). Some authors (e.g. Bickel [6], Barnett [5], Babu and Rao [3], Abdous and Theodorescu [1]) have considered the median vector formed by the univariate medians corresponding to the co-ordinate variables of a multivariate data set. Others (e.g. Haldane [11], Gower [10], Brown [7]) have discussed a multivariate extension of median defined as the minimizer (w.r.t  $\theta$ ) of the sum  $\sum_{i=1}^n \|X_i - \theta\|$ , where the  $X_i$ 's are  $d$ -dimensional multivariate observations and  $\|\cdot\|$  is the usual Euclidean  $L_2$ -norm. Asymptotic properties of this median (popularly known as spatial median) have been extensively studied by several people. It is obvious that the multivariate median, which consists of univariate medians of the real valued components of the data, is not equivariant under rotations or other affine transformations of the observations. However, given a set of multivariate observations, such a median is quite easy to compute being based on the usual medians of the univariate components of the data. The spatial median is equivariant under location transformations as well as rotations or orthogonal transformations of the data. But, for general configurations of multivariate data in dimensions two or higher (i.e. when  $d \geq 2$ ), spatial median cannot be expressed as a function of the data points in a closed form, and

*Key words and phrases.* Affine transformation, asymptotic distribution, equivariance, generalized variance.

it can be computed only by means of some iterative schemes. Further, it is not equivariant under arbitrary affine transformations.

Different affine equivariant versions of multivariate median have been proposed and investigated by Tukey [17], Oja [14] and Liu [13]. These medians however are not easy to compute for high dimensional data especially when  $d \geq 3$ , and the computational difficulty increases enormously as  $d$  increases. Like the spatial median, each of these medians is defined as the solution of an optimization problem, and none of them can be expressed as a function of the data points in a simple and convenient form if  $d \geq 2$ . The results on the asymptotic behavior of Oja's median and Liu's median have been worked out in the literature (see e.g. Arcones, Chen and Gine [2]). However, not much is known about asymptotic properties (e.g. rate of convergence, limiting distribution, asymptotic variability, etc.) of Tukey's median. In this note, we propose an affine equivariant version of multivariate median and study its statistical properties. The proposed median is quite easy to compute in any dimension, and its computation does not require any iterative optimization. As a matter of fact, this multivariate median can be expressed as a function of the data in a nice and elegant form, and it has some interesting connections with the fundamental geometry of a multivariate data cloud.

In Section 2, we introduce a way of constructing an affine equivariant multivariate median using a transformation and re-transformation technique, where the transformation used is closely related to the data-driven co-ordinate system introduced in Chaudhuri and Sengupta [9]. In Section 3, we investigate the asymptotic behavior of the proposed median. In particular, some interesting and useful results related to its limiting distribution are derived when the probability distribution underlying the data is elliptically symmetric. Based on these results, we suggest an adaptive procedure for choosing the appropriate transformation so that the asymptotic variation of the sample median is minimized in some appropriate sense leading to an efficient estimate of multivariate location.

## 2. AFFINE EQUIVARIANT MULTIVARIATE MEDIAN: THE TRANSFORMATION AND RE-TRANSFORMATION APPROACH

We begin by introducing some notation. Consider data points  $X_1, X_2, \dots, X_n$  in  $\mathbb{R}^d$ . Unless specified otherwise, all vectors in this paper will be column vectors, and the superscript  $T$  will be used to denote the transpose of a vector or a matrix. Define

$$S_n = \{\alpha | \alpha \subseteq \{1, 2, \dots, n\} \text{ and } |\alpha| = d + 1\}$$

which is the collection of all subsets of size  $d + 1$  of  $\{1, 2, \dots, n\}$ . For a fixed  $\alpha = \{i_0, i_1, \dots, i_d\} \in S_n$ , let  $\mathbf{X}(\alpha, i_0)$  be the  $d \times d$  matrix whose columns are the random vectors  $(X_i - X_{i_0})$  with  $i \in \alpha$  and  $i \neq i_0$ . We assume that elements of  $\alpha$  are naturally ordered, and if the  $X_i$ 's are i.i.d. observations with a common probability distribution that happens to be absolutely continuous w.r.t. the Lebesgue measure on  $\mathbb{R}^d$ ,  $\mathbf{X}(\alpha, i_0)$  must be an invertible matrix with probability one. We will treat  $\mathbf{X}(\alpha, i_0)$  as a data based transformation matrix, and for each  $i \notin \alpha$ , write  $Y_i^{(\alpha, i_0)} = \{\mathbf{X}(\alpha, i_0)\}^{-1} X_i$  (cf. data-driven co-ordinate system discussed in Chaudhuri and Sengupta [9]). These transformed observations are a way of viewing the multivariate data cloud from a data-centric reference frame. Consider

$$Z_i^{(\alpha, i_0)} = \{\mathbf{X}(\alpha, i_0)\}^{-1} (X_i - X_{i_0}) = Y_i^{(\alpha, i_0)} - \{\mathbf{X}(\alpha, i_0)\}^{-1} X_{i_0} .$$

It is easy to verify (using straight-forward algebra and some minor modifications of the arguments used in proving Proposition 2.1 in Chaudhuri and Sengupta [9]) that for fixed  $\alpha \in S_n$  and  $i_0 \in \alpha$ , the  $Z_i^{(\alpha, i_0)}$ 's with  $1 \leq i \leq n$  and  $i \notin \alpha$  form a maximal invariant for the group of invertible affine transformations on  $\mathbb{R}^d$ .

Let  $\hat{\phi}_n^{(\alpha, i_0)}$  be the vector of medians of the co-ordinates of the  $d$ -dimensional transformed observations  $Y_i^{(\alpha, i_0)}$  such that  $1 \leq i \leq n$  and  $i \notin \alpha$ . Then define the multivariate median  $\hat{\theta}_n^{(\alpha, i_0)}$  for the original data by re-transforming  $\hat{\phi}_n^{(\alpha, i_0)}$  as  $\hat{\theta}_n^{(\alpha, i_0)} = \{\mathbf{X}(\alpha, i_0)\} \hat{\phi}_n^{(\alpha, i_0)}$ . The following Proposition asserts affine equivariance of  $\hat{\theta}_n^{(\alpha, i_0)}$  and we will prove it in the Appendix.

**Proposition 2.1.** *Let  $\hat{\theta}_n^{(\alpha, i_0)}$  be the multivariate median based on the data points  $X_1, X_2, \dots, X_n$  as described above. Suppose that  $\mathbf{A}$  is a fixed  $d \times d$  nonsingular matrix and  $\mathbf{b}$  is a fixed vector in  $\mathbb{R}^d$ . Then the multivariate median computed from  $\mathbf{A}X_1 + \mathbf{b}, \mathbf{A}X_2 + \mathbf{b}, \dots, \mathbf{A}X_n + \mathbf{b}$  in the same way as above (using the same index set  $\alpha$  and the same index  $i_0 \in \alpha$ ) will be  $\mathbf{A}\hat{\theta}_n^{(\alpha, i_0)} + \mathbf{b}$ .*

In the case of data arising from an elliptically symmetric distribution,  $\hat{\theta}_n^{(\alpha, i_0)}$  estimates the centre of elliptic symmetry of that distribution (see also Section 3.1). In general,  $\hat{\theta}_n^{(\alpha, i_0)}$  can be viewed as a descriptive statistic yielding a new concept of location of a multivariate data cloud. Note at this point that for  $d = 1$  and a fixed  $\alpha = \{i_0, i_1\} \in S_n$ ,  $\hat{\theta}_n^{(\alpha, i_0)}$  reduces to the usual univariate median of the  $X_i$ 's excluding the observations  $X_{i_0}$  and  $X_{i_1}$ . Hence the difference between  $\hat{\theta}_n^{(\alpha, i_0)}$  and the median of all the  $X_i$ 's with  $1 \leq i \leq n$  will be insignificant especially when the sample size  $n$  is large, and their asymptotic behavior will be identical. Specifically, if  $X_1, X_2, \dots, X_n$  are i.i.d. univariate observations with a common density  $f$  that has a median at  $\theta$ , and  $f$  is continuous and positive at  $\theta$ , the asymptotic distribution of  $\sqrt{n}(\hat{\theta}_n^{(\alpha, i_0)} - \theta)$  will be Gaussian with mean 0 and variance  $\{2f(\theta)\}^{-2}$ , which is the same as the asymptotic distribution of the median of all the  $X_i$ 's (see Bahadur [4]). The situation however is very different in higher dimensions. For  $d \geq 2$ , the asymptotic behavior of  $\hat{\theta}_n^{(\alpha, i_0)}$  critically depends on the matrix  $\mathbf{X}(\alpha, i_0)$ , and as we will see in Section 3, the selection of  $\alpha \in S_n$  and  $i_0 \in \alpha$  has a crucial impact on the asymptotic performance of  $\hat{\theta}_n^{(\alpha, i_0)}$ .

In an attempt to make the spatial median affine equivariant, Rao [15] proposed normalization of data points by multiplying them with  $S^{-1/2}$ , where  $S$  is the usual sample dispersion matrix. In other words, his proposal was to minimize  $\sum_{i=1}^n \{(X_i - \theta)^T S^{-1} (X_i - \theta)\}^{1/2}$  w.r.t.  $\theta$ . While such a modified version of the spatial median is indeed affine equivariant, normalization of the observations using the square-root of the sample dispersion matrix cannot produce any affine equivariant modification of the vector of medians of real valued components of a multivariate data set. For a fixed  $\alpha \in S_n$  and  $i_0 \in \alpha$ , multiplication of the data points with the matrix  $\{\mathbf{X}(\alpha, i_0)\}^{-1}$  can be viewed as a different (and somewhat unconventional) way of normalizing the observations. Clearly, once we select  $\alpha \in S_n$  and  $i_0 \in \alpha$ , the computation of  $\hat{\phi}_n^{(\alpha, i_0)}$  and  $\hat{\theta}_n^{(\alpha, i_0)}$  is extremely simple in any dimension. One only needs to compute the usual univariate median for each co-ordinate of the transformed observations  $Y_i^{(\alpha, i_0)}$ , and then re-transform the resulting vector of medians

(i.e.  $\hat{\phi}_n^{(\alpha, i_0)}$ ) by multiplying it with  $\mathbf{X}(\alpha, i_0)$ . We conclude this Section with the following Theorem, which exposes an interesting geometric feature of  $\hat{\theta}_n^{(\alpha, i_0)}$ .

**Theorem 2.2.** *Fix  $\alpha \in S_n$  and  $i_0 \in \alpha$ , and let  $\theta \in \mathbb{R}^d$ . For each  $i \in \alpha$ , replace  $X_i$  by  $Z_i = X_i + \theta$ , and for each  $i \notin \alpha$ , replace  $X_i$  by  $Z_i = X_i + X_{i_0}$ . So, each data point is transformed by a location shift, where the shifting vector is either  $\theta$  or  $X_{i_0}$  depending on whether the data point to be shifted is used in the formation of the transformation matrix  $\mathbf{X}(\alpha, i_0)$  or not, respectively. Consider those simplices in  $\mathbb{R}^d$  each of which is formed by a collection of  $d+1$  points  $\{Z_{j_1}, Z_{j_2}, \dots, Z_{j_{d-1}}, Z_{i_0}, Z_i\}$  such that  $\{j_1, j_2, \dots, j_{d-1}\} \subset \alpha$  and  $i \notin \alpha$ . Then  $\theta = \hat{\theta}_n^{(\alpha, i_0)}$  minimizes the sum of the volumes of all such simplices.*

The proof of this Theorem will be given in the Appendix. Random simplices formed by data points play a very crucial role in the construction of Oja's median [14] as well as Liu's median [13]. The above Theorem indicates that they have a fundamental role in the construction of  $\hat{\theta}_n^{(\alpha, i_0)}$  too.

### 3. ASYMPTOTIC PROPERTIES OF PROPOSED MEDIAN

From now on we will assume that the  $X_i$ 's are i.i.d. observations with a common probability distribution that is absolutely continuous w.r.t. the Lebesgue measure on  $\mathbb{R}^d$ . Then, for a fixed  $\alpha \in S_n$  and  $i_0 \in \alpha$ , the transformed observations  $Y_i^{(\alpha, i_0)}$  with  $1 \leq i \leq n$  and  $i \notin \alpha$  are conditionally independently distributed with a common absolutely continuous distribution if we condition on the  $X_i$ 's for which  $i \in \alpha$ . It is now obvious that the limiting conditional distributions (conditioned on the  $X_i$ 's with  $i \in \alpha$ ) of both  $\hat{\phi}_n^{(\alpha, i_0)}$  and  $\hat{\theta}_n^{(\alpha, i_0)}$  will be normal in view of well-known asymptotic results about the univariate median that are applicable to a vector of univariate medians (see e.g. Babu and Rao [3]). When the common distribution of the  $X_i$ 's happens to be elliptically symmetric, it is possible to describe that limiting normal distribution explicitly by deriving a useful expression for the limiting dispersion matrix in terms of  $\mathbf{X}(\alpha, i_0)$ . This leads to valuable insights into the asymptotic performance of  $\hat{\theta}_n^{(\alpha, i_0)}$  as an estimate of the center of elliptic symmetry, and provides us with a way of adaptively selecting  $\alpha \in S_n$  and  $i_0 \in \alpha$  based on asymptotic variation of the resulting location estimate.

**3.1. Behavior in the elliptically symmetric case.** Suppose that the  $X_i$ 's have an elliptically symmetric probability distribution with density  $\{\det(\Sigma)\}^{-1/2} \cdot f\{(x - \theta)^T \Sigma^{-1}(x - \theta)\}$ . Here  $\theta \in \mathbb{R}^d$  is the location of symmetry, and  $\Sigma$  is a  $d \times d$  positive definite matrix. Let us write  $\{\Sigma^{-1/2} \mathbf{X}(\alpha, i_0)\}^{-1} = R(\alpha, i_0) J(\alpha, i_0)$ , where  $R(\alpha, i_0)$  is a diagonal matrix with positive diagonal entries and  $J(\alpha, i_0)$  is a matrix whose rows are of unit length. Clearly, the rows of  $J(\alpha, i_0)$  are obtained by normalizing the rows of  $\{\Sigma^{-1/2} \mathbf{X}(\alpha, i_0)\}^{-1}$ , and the diagonal elements of  $R(\alpha, i_0)$  are the norms of those rows.

**Theorem 3.1.** *Fix  $\alpha \in S_n$  and  $i_0 \in \alpha$  as before. Assume that the density function  $f$  is such that any univariate marginal  $g$  of the spherically symmetric density  $f(x^T x)$  is differentiable and positive at zero. Then as  $n$  tends to infinity, the conditional distribution of  $\sqrt{n}(\hat{\theta}_n^{(\alpha, i_0)} - \theta)$  given the  $X_i$ 's with  $i \in \alpha$  converges weakly to a  $d$ -dimensional normal distribution with zero mean and with dispersion matrix  $c \Sigma^{1/2} \{J(\alpha, i_0)\}^{-1} \{D(\alpha, i_0)\} \{[J(\alpha, i_0)]^T\}^{-1} \Sigma^{1/2}$ . Here  $c = \{2g(0)\}^{-2}$ , and  $D(\alpha, i_0)$*

is the  $d \times d$  matrix whose diagonal elements are all equal to 1, and for  $i \neq j$ , its  $(i, j)$ -th element is  $(2/\pi) \sin^{-1} \gamma_{ij}$ ,  $\gamma_{ij}$  being the inner product of the  $i$ -th and the  $j$ -th row of  $J(\alpha, i_0)$ .

It follows from the preceding Theorem that  $\hat{\theta}_n^{(\alpha, i_0)}$  is an  $n^{1/2}$ -consistent estimate of  $\theta$ , and its conditional asymptotic generalized variance is

$$(c/n)^d \{\det(\Sigma)\} [\det\{D(\alpha, i_0)\}] [\det\{J(\alpha, i_0)\}]^{-2} .$$

Consider now the symmetric positive definite matrix

$$V(\alpha, i_0) = \{J(\alpha, i_0)\}^{-1} \{D(\alpha, i_0)\} \{[J(\alpha, i_0)]^T\}^{-1} .$$

Note that it depends only on the directions of the rows of  $\{\Sigma^{-1/2} \mathbf{X}(\alpha, i_0)\}^{-1}$  and not on their magnitudes. The following Theorem establishes a lower bound for  $\det\{V(\alpha, i_0)\}$ , and this yields a lower bound for conditional asymptotic generalized variance of  $\hat{\theta}_n^{(\alpha, i_0)}$ .

**Theorem 3.2.** *For the matrices  $D(\alpha, i_0)$  and  $J(\alpha, i_0)$  defined above, we have  $\det\{D(\alpha, i_0)\} \geq [\det\{J(\alpha, i_0)\}]^2$  so that  $\det\{V(\alpha, i_0)\} \geq 1$ . This lower bound is sharp in the sense that an exact equality in place of the inequality will hold if  $J(\alpha, i_0)$  happens to be an orthogonal matrix.*

**3.2. Adaptive choice of  $\alpha$ .** Theorem 3.2 implies that whatever  $f$  and  $\Sigma$  may be, the conditional asymptotic generalized variance of  $\hat{\theta}_n^{(\alpha, i_0)}$  cannot be smaller than  $(c/n)^d \det(\Sigma)$  for any choice of  $\alpha \in S_n$  and  $i_0 \in \alpha$ , and one should preferably choose  $\mathbf{X}(\alpha, i_0)$  in such a way that the columns of  $\Sigma^{-1/2} \mathbf{X}(\alpha, i_0)$  (i.e. the vectors  $\Sigma^{-1/2}(X_i - X_{i_0})$ , where  $i \in \alpha$  and  $i \neq i_0$ ) are as orthogonal as possible so that  $\det\{V(\alpha, i_0)\}$  becomes very close to one. Here we propose an adaptive way to select the best subset  $\alpha$  and  $i_0 \in \alpha$ . First, obtain some consistent estimate of the scale matrix  $\Sigma$ , say  $\hat{\Sigma}$ , that is equivariant under nonsingular linear transformation of the data. Then, normalize each data point  $X_i$  by  $\hat{\Sigma}^{-1/2}$ . Define  $Y_i = \hat{\Sigma}^{-1/2} X_i$  for  $1 \leq i \leq n$ . Choose  $\alpha \in S_n$  and  $i_0 \in \alpha$  and compute  $\det\{V(\alpha, i_0)\}$  based on  $Y_i$ 's as described before. Then minimize  $\det\{V(\alpha, i_0)\}$  over all choices of  $\alpha \in S_n$  and  $i_0 \in \alpha$ . Suppose that  $\alpha^*$  and  $i_0^*$  are some minimizers. Form  $\mathbf{X}(\alpha^*, i_0^*)$  and compute  $\hat{\theta}_n^{(\alpha^*, i_0^*)}$  from the original observations  $X_i$ .

In this adaptive selection of  $\alpha \in S_n$  and  $i_0 \in \alpha$ , we have to search from a total number of  $(d+1) \binom{n}{d+1}$  possible cases, and for each such case, we have to invert a  $d \times d$  matrix. However, once the matrix  $\mathbf{X}(\alpha^*, i_0^*)$  is formed, the computation of  $\hat{\theta}_n^{(\alpha^*, i_0^*)}$  is straightforward as it does not require any further optimization or iterative computation. As a result, the adaptive version of the transformation and re-transformation estimate continues to remain computationally advantageous compared to several other equivariant versions of multivariate median considered in the existing literature.

In this paper, we are not presenting any analytic study of the robustness and the efficiency of the adaptive estimate, and we intend to pursue it in a future paper. However, we are presenting some finite sample simulation results on the efficiency of the adaptive estimate compared with the vector of co-ordinate wise medians, when data points are generated from (1) a bivariate standard normal distribution and (2) a bivariate Laplace distribution with density  $(2\pi)^{-1} \exp\{-(x^2 + y^2)^{1/2}\}$ . For calculating efficiencies, we have estimated the determinant of the  $2 \times 2$  mean

square and product error matrix based on 2000 Monte Carlo replications for our adaptive estimate as well as the vector of co-ordinate wise medians. Results are summarized in Table 1. The adaptive estimate is based on  $n$  observations and we have tried three different sample sizes.  $Eff_1$  gives the efficiency of the adaptive estimate compared to the co-ordinate wise median when the latter is based on  $(n - d - 1)$  observations, while  $Eff_2$  gives the efficiency when the co-ordinate wise median is based on  $n$  observations.

TABLE 1. Finite sample efficiency

GAUSSIAN	$n = 20$	$n = 30$	$n = 40$
$Eff_1$	1.01162	1.03836	1.04814
$Eff_2$	0.82544	0.90015	0.91812
LAPLACE	$n = 20$	$n = 30$	$n = 40$
$Eff_1$	0.98563	0.97195	0.92566
$Eff_2$	0.74914	0.82285	0.86649

We observe that for both the distributions considered here,  $Eff_1$  is larger than  $Eff_2$  as one would expect. The reason behind comparing with the co-ordinate wise median based on  $(n - d - 1)$  observations is that we are constructing the matrix  $\mathbf{X}(\alpha^*, i_0^*)$  using  $(d + 1)$  data points, and then  $(n - d - 1)$  transformed data points are being used to construct the location estimate. The adaptive construction of  $\mathbf{X}(\alpha^*, i_0^*)$  is comparable with estimation of the unknown scale matrix  $\Sigma$  associated with the elliptically symmetric distribution from which the observations are generated. So, in the efficiencies reported in Table 1, we are actually comparing the case when  $\Sigma$  is unknown and the adaptive version of the transformation re-transformation estimate is used with the case when  $\Sigma$  is known (without loss of generality  $\Sigma$  can be taken to be the identity matrix) and the vector of the co-ordinate wise median is used.

#### 4. APPENDIX: THE PROOFS

*Proof of Proposition 2.1.* First observe that in view of the way the matrix  $\mathbf{X}(\alpha, i_0)$  has been constructed, if the  $X_i$ 's are transformed to  $(\mathbf{A}X_i + \mathbf{b})$ 's,  $\mathbf{X}(\alpha, i_0)$  will be transformed to  $\mathbf{A}\mathbf{X}(\alpha, i_0)$ . Also, note that the  $Y_i^{(\alpha, i_0)}$ 's remain invariant under a linear transformation of the  $X_i$ 's, and the vector of medians of the real valued co-ordinates of a multivariate data set is equivariant under a location shift of the data points. Hence, for the transformed data points  $(\mathbf{A}X_i + \mathbf{b})$ ,  $\hat{\phi}_n^{(\alpha, i_0)}$  will be transformed to  $\hat{\phi}_n^{(\alpha, i_0)} + \{\mathbf{A}\mathbf{X}(\alpha, i_0)\}^{-1}\mathbf{b}$ . Consequently  $\hat{\theta}_n^{(\alpha, i_0)}$ , which was originally defined as  $\{\mathbf{X}(\alpha, i_0)\}\hat{\phi}_n^{(\alpha, i_0)}$ , will be transformed to  $\mathbf{A}\hat{\theta}_n^{(\alpha, i_0)} + \mathbf{b}$ .  $\square$

*Proof of Theorem 2.2.* For  $Z = (z_1, z_2, \dots, z_d) \in \mathbb{R}^d$ , let  $|Z|$  denote the  $L_1$ -norm of  $Z$  defined as  $|Z| = \sum_{i=1}^d |z_i|$ . Then it is easy to see that

$$\hat{\theta}_n^{(\alpha, i_0)} = \arg \min_{\theta \in \mathbb{R}^d} \sum_{1 \leq i \leq n; i \neq \alpha} |\{\mathbf{X}(\alpha, i_0)\}^{-1}(X_i - \theta)|.$$

Now,  $\{\mathbf{X}(\alpha, i_0)\}^{-1}(X_i - \theta)$  can be viewed as a solution (in  $Z$ ) of the system of linear equations  $\{\mathbf{X}(\alpha, i_0)\}Z = (X_i - \theta)$ . So, if one applies the well-known Cramer's rule

for solving a system of linear equations, the absolute value of any component of the  $d$ -dimensional vector  $\{\mathbf{X}(\alpha, i_0)\}^{-1}(X_i - \theta)$  will be of the form

$$|\det\{\mathbf{X}(\alpha, i_0)\}^{-1}| \det\{(X_{j_1} - X_{i_0}, X_{j_2} - X_{i_0}, \dots, X_{j_{d-1}} - X_{i_0}, X_i - \theta)\}.$$

The proof of the Theorem is now complete in view of the fact that

$$|\det\{(X_{j_1} - X_{i_0}, X_{j_2} - X_{i_0}, \dots, X_{j_{d-1}} - X_{i_0}, X_i - \theta)\}|$$

is the volume of the simplex in  $\mathbb{R}^d$ , which is formed by the set of  $d + 1$  points  $\{Z_{j_1}, Z_{j_2}, \dots, Z_{j_{d-1}}, Z_{i_0}, Z_i\}$  as described in the statement of the Theorem.  $\square$

*Proof of Theorem 3.1.* In view of affine equivariance of the location estimate  $\hat{\theta}_n^{(\alpha, i_0)}$ , it is sufficient to prove the Theorem in the special case when  $\theta =$  the zero vector in  $\mathbb{R}^d$  and  $\Sigma =$  the  $d \times d$  identity matrix. Then, given the  $X_i$ 's for which  $i \in \alpha$ , the transformed observations  $Y_i^{(\alpha, i_0)}$  with  $i \notin \alpha$  are conditionally i.i.d. random vectors with common density  $|\det\{\mathbf{X}(\alpha, i_0)\}| f\{y^T[\mathbf{X}(\alpha, i_0)]^T[\mathbf{X}(\alpha, i_0)]y\}$ . Let  $r_1, \dots, r_d$  be the diagonal entries of  $R(\alpha, i_0)$ . In view of the main result in Babu and Rao [3] on asymptotic distribution of the vector of univariate quantiles of a multivariate data, the conditional distribution of  $n^{1/2}\hat{\phi}_n^{(\alpha, i_0)}$  will converge weakly to a  $d$ -variate normal distribution with zero mean, and the limiting dispersion matrix will be such that its  $k$ -th diagonal entry will be  $cr_k^2$ , and for  $k \neq l$ , its  $(k, l)$ -th element will be  $4cr_k r_l \{Pr(U_{ik}^{(\alpha, i_0)} > 0 \text{ and } U_{il}^{(\alpha, i_0)} > 0) - 1/4\}$ . Here  $U_{ik}^{(\alpha, i_0)}$  and  $U_{il}^{(\alpha, i_0)}$  are the  $k$ -th and the  $l$ -th components of  $Y_i^{(\alpha, i_0)}$  respectively. Note that we are using the fact that for a  $d$ -dimensional random vector  $Z$  with a spherically symmetric distribution, the distribution of the random variable  $\mathbf{a}^T Z$  is the same for any  $\mathbf{a} \in \mathbb{R}^d$  such that  $\mathbf{a}^T \mathbf{a} = 1$ . Also, since the conditional distribution of  $Y_i^{(\alpha, i_0)}$  is elliptically symmetric around the origin in  $\mathbb{R}^d$ ,  $Pr\{U_{ik}^{(\alpha, i_0)} > 0 \text{ and } U_{il}^{(\alpha, i_0)} > 0\}$  does not depend on the density  $f$ . Recall that the rows of  $J(\alpha, i_0)$  are of unit length obtained by normalizing the rows of  $\{\mathbf{X}(\alpha, i_0)\}^{-1}$ . We now have the following by some routine analytic computation:

$$Pr(U_{ik}^{(\alpha, i_0)} > 0 \text{ and } U_{il}^{(\alpha, i_0)} > 0) = 1/4 + (1/2\pi)\sin^{-1}\gamma_{kl}.$$

So, the dispersion matrix of the conditional asymptotic distribution of  $n^{1/2}\hat{\phi}_n^{(\alpha, i_0)}$  is  $c\{R(\alpha, i_0)\}\{D(\alpha, i_0)\}\{R(\alpha, i_0)\}$ . Next recall that

$$\hat{\theta}_n^{(\alpha, i_0)} = \mathbf{X}(\alpha, i_0)\hat{\phi}_n^{(\alpha, i_0)} = \{J(\alpha, i_0)\}^{-1}\{R(\alpha, i_0)\}^{-1}\hat{\phi}_n^{(\alpha, i_0)}.$$

The proof of the Theorem is now complete by straightforward algebra.  $\square$

The following well-known Fact will be used in the proof of Theorem 3.2. A proof of this Fact has been discussed in Lancaster [12].

**Fact 4.1.** Let  $X$  and  $Y$  be  $p$ - and  $q$ -dimensional normal random vectors. Then

$$\max_{\mathbf{a} \in \mathbb{R}^p, \mathbf{b} \in \mathbb{R}^q} Corr(\mathbf{a}^T X, \mathbf{b}^T Y) = \max_{\eta, \psi} Corr(\eta(X), \psi(Y)),$$

where ‘‘Corr’’ stands for the usual correlation coefficient, and  $\eta : \mathbb{R}^p \rightarrow \mathbb{R}$ ,  $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$  are measurable functions such that  $\eta(X)$  and  $\psi(Y)$  have finite second moments.

*Proof of Theorem 3.2.* Denote  $\{J(\alpha, i_0)\}\{J(\alpha, i_0)\}^T$  by  $P$ , and for notational convenience, we will write simply  $D$  for  $D(\alpha, i_0)$ . We will prove the result by induction on the dimension  $d$ . For  $d = 2$ , the matrices  $D$  and  $P$  can be written as,

$$D = \begin{bmatrix} 1 & \frac{2}{\pi} \sin^{-1}(\cos\delta) \\ \frac{2}{\pi} \sin^{-1}(\cos\delta) & 1 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & \cos\delta \\ \cos\delta & 1 \end{bmatrix},$$

where  $0 < \delta < \pi$  is the angle between the two rows of  $J(\alpha, i_0)$ . So,  $\det(D) = 1 - \{\frac{2}{\pi} \sin^{-1}(\cos\delta)\}^2 = 1 - (1 - \frac{2}{\pi}\delta)^2$  and  $\det(P) = 1 - \cos^2\delta$ . Now, for  $0 < \delta < \pi$ ,  $\cos^2\delta \geq (1 - \frac{2}{\pi}\delta)^2$ , and the equality holds if and only if  $\delta = \pi/2$ . This proves the result for  $d = 2$ .

Now assume that the result is true for dimension  $d - 1 \geq 2$ . Partition the  $d \times d$  matrix  $D$  as  $\begin{bmatrix} 1 & \mathbf{d}^T \\ \mathbf{d} & D_* \end{bmatrix}$  and the  $d \times d$  matrix  $P$  as  $\begin{bmatrix} 1 & \mathbf{p}^T \\ \mathbf{p} & P_* \end{bmatrix}$ . Note that  $P$  can be viewed as the correlation matrix of a  $d$ -dimensional normal random vector  $(U, W_1, \dots, W_{d-1})$ , and  $D$  can be viewed as the correlation matrix of the random vector  $(I(U > 0), I(W_1 > 0), \dots, I(W_{d-1} > 0)) = (V, Z_1, \dots, Z_{d-1})$  (say), where  $I$  is the usual 0–1 valued indicator function. Write  $W = (W_1, \dots, W_{d-1})$  and  $Z = (Z_1, \dots, Z_{d-1})$ . Then using Fact 4.1 stated above, we get

$$\max_{\mathbf{b} \in \mathbb{R}^{d-1}} \text{Corr}(U, \mathbf{b}^T W) \geq \max_{\mathbf{b} \in \mathbb{R}^{d-1}} \text{Corr}(V, \mathbf{b}^T Z).$$

But on the LHS above, we have the multiple correlation coefficient between  $U$  and  $W$ , and on the RHS, we have the multiple correlation coefficient between  $V$  and  $Z$ . Therefore, we must have  $\mathbf{p}^T P_*^{-1} \mathbf{p} \geq \mathbf{d}^T D_*^{-1} \mathbf{d}$ . The induction hypothesis implies that  $\det(D_*) \geq \det(P_*)$ . The proof of the Theorem is now complete by observing that  $\det(D) = \{\det(D_*)\}(1 - \mathbf{d}^T D_*^{-1} \mathbf{d})$  and  $\det(P) = \{\det(P_*)\}(1 - \mathbf{p}^T P_*^{-1} \mathbf{p})$ .  $\square$

#### ACKNOWLEDGEMENTS

Research presented here has been partially supported by a grant from the Indian Statistical Institute. The authors are thankful to an anonymous referee for several helpful comments.

#### REFERENCES

1. Abdous, B. and Theodorescu, R. (1992), Note on the spatial quantile of a random vector. *Statistics & Probability Letters*, **13**, 333–336. MR **93d**:62094
2. Arcones, M.A., Chen, Z. and Giné, E. (1994), Estimators related to U-processes with applications to multivariate medians : Asymptotic normality. *The Annals of Statistics*, **22**, 1460–1477. CMP 95:06
3. Babu, G. J. and Rao, C. R. (1988), Joint asymptotic distribution of marginal quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis*, **27**, 15–23. MR **90e**:62078
4. Bahadur, R.R. (1966), A note on quantiles in large samples. *The Annals of Mathematical Statistics*, **37**, 577–580. MR **32**:6522
5. Barnett, V. (1976), The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society, Series A*, **139**, 318–354. MR **56**:4060
6. Bickel, P. J. (1964), On some alternative estimates of shift in the  $p$ -variate one sample problem. *Annals of Mathematical Statistics*, **35**, 1079–1090. MR **29**:2904
7. Brown, B. M. (1983), Statistical use of spatial median. *Journal of the Royal Statistical Society, Series B*, **45**, 25–30. MR **85a**:62073
8. Chaudhuri, P. (1992), Multivariate location estimation using extension of  $R$ -estimates through  $U$ -statistics type approach. *The Annals of Statistics*, **20**, 897–916. MR **93h**:62072

9. 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. MR **94j**:62099
10. Gower, J. C. (1974), The mediancenter. *Journal of the Royal Statistical Society, Series C*, **23**, 466–470.
11. Haldane, J. B. S. (1948), Note on the median of a multivariate distribution. *Biometrika*, **35**, 414–415.
12. Lancaster, H. O. (1969), *The Chi-Squared Distribution*. Wiley, New York. MR **40**:6667
13. Liu, R. Y. (1990), On a notion of data depth based on random simplices. *The Annals of Statistics*, **18**, 405–414. MR **91d**:62068
14. Oja, H. (1983), Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, **1**, 327–332. MR **85a**:62091
15. Rao, C. R. (1988), Methodology based on the  $L_1$ -norm in statistical inference. *Sankhyā, Series A*, **50**, 289–313. MR **91e**:62181
16. Small, C. G. (1990), A survey of multidimensional medians. *International Statistical Review*, **58**, 263–277.
17. Tukey, J. W. (1975), Mathematics and picturing data. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974* (Ed. R.D. James), vol. 2, pp. 523–531. MR **55**:26